IDENTIFYING CANCER GENE SUBTYPES FROM GENE EXPRESSION BY CO-CLUSTERING ALGORITHM AND SUPPORT VECTOR MACHINE

LOGENTHIRAN A/L MACHAP

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

FEBRUARY 2021

DEDICATION

This thesis is dedicated to my lovely grandmother,
Mrs S. SOORANAM MARIAPPAN,
who taught me the best and useful knowledge from her life experience.
Without you, none of my success would be possible. Thank you maa!

ACKNOWLEDGEMENT

"Dear GOD Thank you for giving me the strength and encouragement especially during all the challenging moments in completing this thesis. I am truly grateful for your exceptional love and grace during this research journey."

Firstly, I would like to express my sincere gratitude to my advisor/supervisor Dr Afnizanfaizal Abdullah for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study. I am also very thankful to my co-supervisor Dr Zuraini Ali Shah for her guidance, advice, and motivation. I extend my gratitude to Dr Zalmiyah Zakaria, been my supervisor for a short period where I have learned some good knowledge from her advice and experience.

I am also genuinely thankful to the University of Technology Malaysia (UTM) for making it possible for me to study here. Besides this, I am grateful to the Ministry of Education Malaysia since they offered me with MyBrain15 scholarship during my research studies.

I have great pleasure in acknowledging my gratitude to my colleagues and fellow research scholars at Synthetic Biology Research Group (SynBio) and Artificial Intelligence and Bioinformatics Research Group (AIBIG) in UTM. Especially my dear friends Kohbalan, Chan Weng Howe, Kauthar, Nies Hui Yen being there at times when I required motivation and propelling me on the course of this thesis and for assisting me in the collection of data for my research. Their support, encouragement, and credible ideas have been great contributors to the completion of the thesis.

My acknowledgement would be imperfect without thanking the major source of my strength, my family. The blessings and sacrifices of my dearest parents Mr. Machap & Mrs. Jayanthi and the love and care of my brother Mr. Muralee and sister-in-law Mrs Baby; my sister Ms Karpagambal. I thank them for putting up with me in difficult moments where I felt stumped and for stimulating me on to follow my dream of getting this Ph.D. This would not have been possible without their thoughtful and concerned love and support is always given to me.

Last but not the least; I would like to express my wholehearted gratitude to my wife Mrs. Vaahini Kalaivanan, understanding me to the core, has been my best friend and great companion, loved, supported, encouraged, entertained, sacrificed, and helped me get through this agonizing period in the most positive way.

ABSTRACT

Cancer subtype information is significant to understand tumour heterogeneity. Present methods to find cancer subtypes have focused on utilizing traditional clustering algorithms such as hierarchical clustering. Since most of these methods depend on high dimensional data, the drawback is to divide the genes into different clusters, where a gene or a condition only belongs to one cluster. A gene may contribute to more than one biological process, so a gene may belong to multiple clusters. Besides, the centroid in the objective function of network-assisted co-clustering for the identification of cancer subtypes (NCIS) dragged with outliers. So, these outliers get their cluster instead of being ignored. Hence, this research is focusing on improving the NCIS method. Enhanced NCIS (iNCIS) is basically assigned weights to genes base on a gene interaction network, and it imperatively optimizes the sum-squared residue to get co-clusters. Next, supervised infinite feature selection with multiple support vector machine (SinfFS-mSVM) is proposed to obtain significant genes from a high dimensional data by using the classes obtained from iNCIS and improve the accuracy of classification. The effectiveness of iNCIS and SinfFS-mSVM is being evaluated on a large-scale Breast Cancer (BRCA) and Glioblastoma Multiforme (GBM) from The Cancer Genome Atlas (TCGA) project. From the implementation, there are five breast cancer gene subtypes and four glioblastoma multiforme cancer gene subtypes that have been successfully identified. The weighted co-clustering approach in iNCIS provides a unique solution to integrate gene network interaction into the clustering process. The improvement of the co-clustering Rand Index and F1-measure is 54.5% and 33.9% for BRCA and 34.2% and 31.5% for GBM. Meanwhile, a significant gene subset with higher classification accuracy was selected from SinfFS-mSVM. The classification accuracy for the selected gene subset improved by 3.00% and 2.99% for BRCA and GBM, correspondingly. Furthermore, biological validation conducted on the selected genes from each subtype is to justify the validity of the results. In conclusion, the empirical study on large-scale cancer datasets using iNCIS and SinfFS-mSVM comprehensively find cancer gene subtypes and genes by achieving higher clustering and classification accuracy. Future works are needed to integrate more comprehensive gene network information and to select optimal parameters.

ABSTRAK

Maklumat subjenis barah adalah penting untuk pemahaman keheterogenan ketumbuhan. Kaedah yang sedia ada hanya akan mengenal pasti subjenis kanser dimana sebahagian besarnya ditumpukan dalam penggunaan algoritma kluster tradisional seperti kluster berhierarki. Kebanyakan kaedah ini hanya bergantung kepada data berdimensi tinggi di mana kelemahannya adalah pembahagian gen ke dalam kluster-kluster yang berbeza, di mana satu gen atau suatu syarat hanya menjadi kepunyaan satu kluster sahaja. Walaubagaimanapun, jika satu gen boleh menyumbang kepada lebih daripada satu proses biologi, maka satu gen boleh dimiliki oleh lebih daripada satu kluster. Selain itu, sentroid dalam fungsi objektif kumpulan rangkaian membantu kelompok bersama untuk pengenalpastian subjenis kanser (NCIS) diseret dengan unsur dengan nilai tersisih. Jadi, nilai tersisih ini mendapat kluster mereka yang sepatutnya diabaikan. Maka, penyelidikan ini menumpukan kepada penambahbaikkan kaedah NCIS. Penambahbaikkan NCIS (iNCIS), pada asasnya menetapkan pemberat kepada gen berdasarkan rangkaian gen dan ia berulang untuk mengoptimumkan jumlah kuasa dua lebihan bagi perolehan kluster. Seterusnya, penyeliaan pilihan rencana tidak terhingga dengan mesin vektor sokongan berbilang (SinfFS-mSVM) dicadangkan untuk memperoleh gen-gen penting dari data berdimensi tinggi dengan menggunakan kelas yang diperolehi daripada iNCIS dan seterusnya memperbaiki ketepatan pengkelas gen. Keberkesanan iNCIS and SinfFS-mSVM dinilai menerusi data barah payudara dan barah otak (GBM) daripada projek “Atlas Genom Kanser”. Melalui pelaksanaan ini, terdapat lima subjenis gen barah bagi barah payudara dan empat subjenis gen barah bagi barah otak berjaya dikenal pasti. Pendekatan klusteran bersama pemberat dalam iNCIS memberikan satu penyelesaian unik interaksi rangkaian gen ke dalam proses klusteran. Jumlah peningkatan Indeks Rand dan F1-mengukur dalam klusteran bersama masing-masing adalah 54.5% dan 33.9% bagi BRCA dan 34.2% dan 31.5% bagi GBM. Sementara itu, subset gen yang penting dipilih dengan ketepatan pengklasifikasian yang lebih tinggi. Kejituan pengklasifikasian untuk subset gen terpilih masing-masing meningkat sebanyak 3.00% dan 2.99% untuk BRCA dan GBM. Pengesahan biologi dijalankan ke atas gen-gen terpilih daripada setiap subjenis untuk menjustifikasi kesahihan keputusan. Sebagai kesimpulan, kajian empirikal ke atas kedua-dua data set barah menggunakan iNCIS dan SinfFS-mSVM secara menyeluruh berjaya mengenal pasti subjenis kanser dan gen dengan mencapai kualiti kluster bersama dan ketepatan pengklasifikasian yang lebih baik. Kerja penambahbaikkan diperlukan pada masa hadapan untuk gabungan maklumat rangkaian gen yang lebih komprehensif dan juga memilih parameter optimum.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BLAST	-	Basic alignment search tool
BRCA	-	Breast cancer
cDNA	-	complementary Deoxyribonucleic acid
DAVID	-	Database for Annotation, Visualization, and Integrated Discovery
DNA	-	Deoxyribonucleic acid
GBM	-	Glioblastoma Multiforme
GFC	-	Gene Functional Classification
GNI	-	Gene Network Information
GO	-	Gene Ontology
HMM	-	Hidden Markov Model
iNCIS	-	improved Network assisted Co-clustering for the Identification of Cancer Subtypes
infFS	-	infinite feature selection
KEGG	-	Kyoto Encyclopedia of Genes and Genomes
MAD	-	Median Absolute Deviation
MatLab	-	Matrix Laboratory
miRNA	-	Micro Ribonucleic Acid
mRNA	-	Messenger Ribonucleic Acid
MSA	-	Multiple Sequence Alignment
MSSRCC	-	Minimize Sum-Square Residue Co-Clustering
NCI	-	National Cancer Institute
NCIS	-	Network assisted Co-clustering algorithm for the Identification of Cancer Subtypes
NetBC	-	Network Bi-Clustering
NMAD	-	Normalized median absolute deviation
PID	-	Pathway Interaction Database
RF	-	Random Forest
RI	-	Rand Index
RNA		Ribonucleic Acid
rRNA		ribosomal Ribonucleic Acid

SinfFS	-	Supervised Infinite Feature Selection
SNMTF	-	Semi-Nonnegative Matrix tri-Factorization
SVM	-	Support Vector Machine
TCGA	-	The Cancer Genome Atlas
TF	-	Transcription factors
tRNA		Transfer Ribonucleic Acid

CHAPTER 1

INTRODUCTION

1.1 Background

Cells are the basic structure of living organisms. The human body comprised trillions of cells (Lodish *et al.*, 2008). Generally, cells undergo a development process called mitosis. During mitosis, a cell replicates all its genetic information. Deoxyribonucleic acid (DNA) present in cancerous cells multiplies and grows uncontrollably. Hence, this inconsistent cell development produces cancerous cells. The newly developed cells contain damaged DNA. Therefore, cancer diseases are known as complex and heterogeneity. The complexity of cancer disease is caused by many genes and pathways (Győrffy *et al.*, 2016; Shukla *et al.*, 2013). Generally, the same types of cancer patients hardly have regularly mutated genes. Therefore, 10%-30% of patients with the same type of cancer are only able to react for a given drug. (Kaiser, 2006; Snyder *et al.*, 2014). Various types of research are ongoing to detect the right diagnosis and prognosis for cancer. Therefore, DNA microarray gene expression technology data are used in the research. Gene expression defines the translation of genetic information from genes through messenger Ribonucleic Acid (mRNA) to proteins. Messenger RNA is also known as mRNA is complementary to one of the DNA strands to form single-stranded RNA (Gilbert *et al.*, 2016).

Over the last decades, the DNA microarray gene expression profiling has made evolution in the biological field (Liang *et al.*, 2004). Thousands of genes expression level can be measured through this technology. Whereas most of the old methods capable of measuring one or a few genes at a time (Lockhart *et al.*, 1996; Schena *et al.*, 1995). Normally, gene expression can be structured in a data matrix. The rows denote genes and columns denote conditions. The impact of microarray technology essentially in cancer and genetic field is very high. It has been shown through various research studies. On that note, scientists are actively working on the

identification of valuable biological information from gene expression data. Numerous applications are developed for disease diagnosis (Saiki *et al.*, 2008), gene function recognition (Eisen *et al.*, 1998), pharmaceutical target discovery (Corn and El-Deiry, 2007), and pathway analysis (Werner, 2008).

Biological systems are complex due to the presence of a huge number of genes. Thus, the computational method is required to analyse the genes from the expression dataset. Generated gene expression dataset is usually without any known features. The primary objective is to group the gene according to their expression under different conditions. Generally, the genes are acts as regulators that underlie phenotypic differences under the conditions such as disease stage, tissue types and developmental stages, (van Dam *et al.*, 2017). Genes present in the same groups are considered to have similarity features. However, genes in different groups are considered to have dissimilar features. This grouping method is essential in the process of cancer gene subtype identification. Biological information has been integrated as guidance in the grouping process. But only limited research studies have applied biological knowledge. For example, Gene Ontology (GO) and Gene Network Information (GNI) are used as assistance (Ashburner *et al.*, 2000).

The nature of GNI offers a rich source of evidence for implying functional associations of genes and proteins. The GNI also plays a significant role in every biochemical process. A strong analysis of this network helps to disclose the disease progression (Hwang *et al.*, 2018). The requirement of network information has been identified in recent research works. However, existing computational methods rarely use this gene network information (Grimes *et al.*, 2019; Yan *et al.*, 2017). Integrating gene network information with the clustering process assists in identifying similarities among gene expressions. This identification can describe the function of an unknown cancer gene subtype by predicting similar genes.

Clustering is one of the methods for exploration and interpretation of high dimensional microarray gene expression data (also known as one-way clustering). In this clustering method usually, genes, samples, or both, are clustered together (Andreopoulos *et al.*, 2009). Traditional clustering, for instance, k-means clustering

(Tothill *et al.*, 2008), hierarchical clustering with heat map visualization (Eisen *et al.*, 1998), and self-organizing maps (Nikkilä *et al.*, 2002) identify biologically significant groups of genes or samples successfully. Yet, these methods do not utilize the data entirely as clustering is done for genes first and then for samples or vice versa. Hence, genes that are co-expressed in another subset of samples are undetected. In addition, other limitations in the traditional clustering algorithms are struggling in determining the number of clusters, random initialization of genes, and engagements in gene subtypes (Oyelade *et al.*, 2016). The cluster number in the traditional clustering determines through numerous trial and errors. Moreover, the random initialization of genes provides uneven results. Additionally, the experiments take a long time because of multiple runs. Usually, this is because of the presence of non-significant genes (Zhou *et al.*, 2018), missing values (Vazifehdan *et al.*, 2018) in the dataset. The presence of significant and non-significant genes will increase the conflicts among genes. The reason is a gene consists of multiple functions and pathways (Andreopoulos *et al.*, 2009). When there is no prior pre-processing is done, then the genes cause different types of disease properties. This leads to uneven results in prediction and classification accuracy.

Thus, Cheng and Church (2000) have introduced a promising method to analyse gene expression which is co-clustering. In the co-clustering method, the clustering is done simultaneously for genes and samples. Co-clustering methods generally called unsupervised methods. Thus, they do not have any guidance or information to be used in exploring gene expression. Hence, semi-supervised methods are introduced. These methods integrate some critical information to guide the clustering process effectively. However, there is still a drawback in this clustering method. This is due to the applications of non-comprehensive biological knowledge with gene expression datasets. For example, incomplete gene network information, transcription factors (TF), microRNA, and metabolic network cause inaccurate results.

Besides, classification is also a well-known approach in microarray analysis. This approach is used to acquire biological information from the gene expression data especially in cancer diagnosis. In classification, the microarray data set is

divided into training data and testing data. The training data set is used to train a classifier to classify the test data to identify sample class. In recent years, many classification algorithms have been developed. For example, support vector machine (SVM) (Kang *et al.*, 2019) and random forest (Nguyen *et al.*, 2013; Nguyen and Bui, 2019). SVM uses a hyperplane with a maximized margin to split the given sample classes. Random forest is a decision tree-based ensemble learning approach. Due to robustness to sparse and noisy data, SVM has been widely used in microarray data analysis (Kang *et al.*, 2019).

Microarray gene expression data is known as high dimensional data. This is because; it contains many genes and a small number of samples. It is essential to select significant genes from high dimensional microarray data. This is to construct a strong classifier to obtain better classification performance by selecting significant genes or features (Ma and Huang, 2008). This process is known as gene selection or feature selection. Commonly, there are three types of gene selection methods available. They are filter, wrapper, and embedded methods. These methods are distinguished based on their way of incorporation. Yet each has its advantages and disadvantages (Saeys *et al.*, 2007).

1.2 Challenges for Cancer Gene Subtypes Identification

Implementation of the co-clustering algorithm offers an excellent tool for studying genes to gene relationships in the biological process. However, the fundamental challenge of cancer gene subtype identification is the cluster prediction. Hence, the first challenge of this research is to handle mRNA changes during gene expression experiments. Gene expression result is represented by positive and negative values. A positive value denotes an increase in expression and negative values denote a decrease in expression. Naturally, a gene can have one or various functions. The function of a gene depends on the occurrence of other gene products. The second challenge is data quality issues in gene expression datasets. Thus, the clustering and classification will be affected by the data quality such as missing values, irregular expression, and non-significant genes. This is because; the

relationships among genes are identified by biological information and expressional features from the data. The third challenge is the presence of high and low-level expression values in gene expression datasets. When the expression value is high, then the gene is possible to be grouped in multiple clusters. The lower expression values may be missed out from the group, which is considered necessary in certain subtype identification. Thus, handling of lower expression values from gene expression is becoming a challenging task for the researchers. On that note, this research is focusing on identifying significant genes from the presence of many genes in high dimensional data.

1.3 Current Methods for Cancer Gene Subtype Identification

Generally, there are two categories of approaches for cancer gene subtypes identification. The first is experimental (*in vivo* and *in vitro*) and the second is computational analyses (*in silico*) (more details are presented in Chapter 2). Figure 1.1 shows an overview of these two categories of approaches.

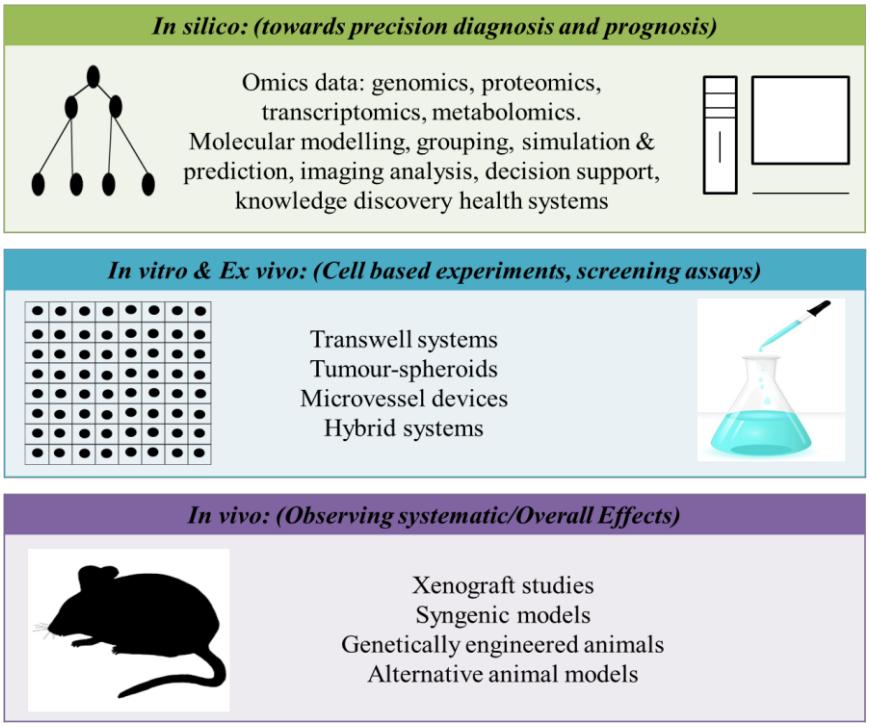


Figure 1.1 Cancer research preclinical techniques (Jean-Quartier *et al.*, 2018)

Animals are regularly used *in vivo* research studies. For instance, the rat is used for cancer diagnosis and prognosis (Yee *et al.*, 2015). This type of preclinical model can help in the identification of novel targets. However, there is no sufficient information provided on the target validation (Gould *et al.*, 2015). Furthermore, *in vitro* assists in the learning of cellular processes of tumour mechanisms (Katt *et al.*, 2016). Besides, *ex vivo* methods are utilized for drug reaction prediction in cancer patients (Vaira *et al.*, 2010). Therefore, researchers are planning to reduce the cost spent on wet laboratory experiments. This is because purchasing of biological and chemical materials for an experiment is costly. In specific, in silico approaches have been proposed for refining clinical and general biomedical studies involving laboratory work (Jean-Quartier *et al.*, 2018).

In-silico method executes through a computer or via computer simulation. Thus, *in silico* methods can be defined as computational approaches. This method incorporates biological analysis and simulation. Examples are computational validation, inference, prediction, classification, and modelling. Various cancer research is conducted by different peoples and organizations. For example, the National Cancer Institute has done the Cancer Genome Anatomy Project. The prime objective of this research is to identify cancer gene subtype from gene expression profiles and gene network information data of the patient to produce output for diagnosis and prognosis.

1.4 Problem Statement

Traditional clustering methods usually do not utilize the full benefit of the gene expression data. The clustering is done for genes first and then for samples or vice versa. Hence, groups of genes that are co-expressed or share similar characteristics only in a subset of samples are identified (Denitto *et al.*, 2017a; Huang *et al.*, 2014; Wang *et al.*, 2002). Figure 1.2 shows traditional clustering clusters a gene into a single cluster. This means a gene do not present in the other cluster. Thus, the problem in these clusters is considered low-quality clusters to be used in cancer diagnosis and prognosis. This is because a gene can involve in more

than one biological process that can be grouped into multiple clusters. In different conditions of the biological process, there are possibilities of a gene to be clustered in numerous clusters (Gasch and Eisen, 2002). The aim is to cluster genes that participate not only in single but more than one biological process into multiple clusters. For instance, in a recently published study, a network-assisted co-clustering algorithm (NCIS) for microarray analysis was proposed (Liu *et al.*, 2014). This NCIS method integrates gene network information to cluster genes and samples into biologically significant clusters concurrently. Despite the excellent performance shown in NCIS, the cancer gene subtype identification can still be improved on the objective function. The centroid in the objective function of (NCIS) dragged with outliers that get their cluster instead of being ignored. This cluster can be low quality cluster with the presence of outliers. Hence, the NCIS need to be improved to obtain better clusters to be used further in classification.

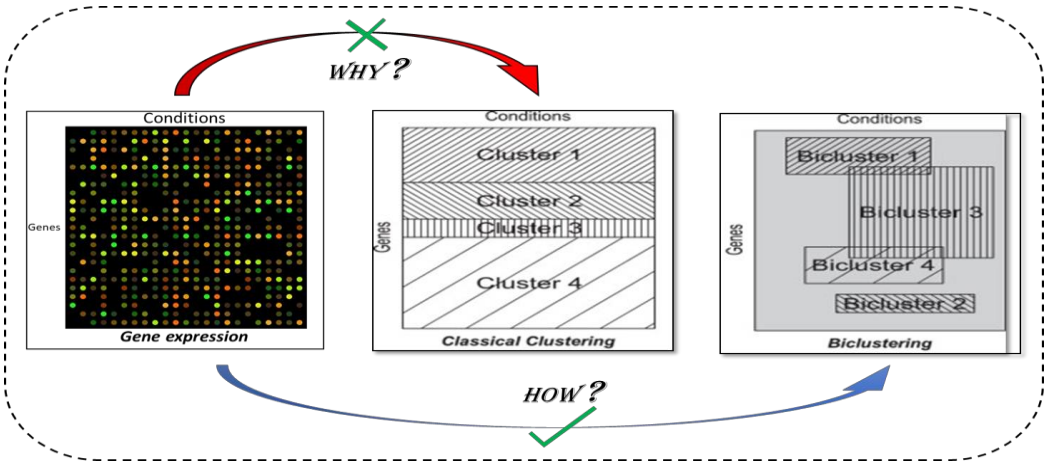
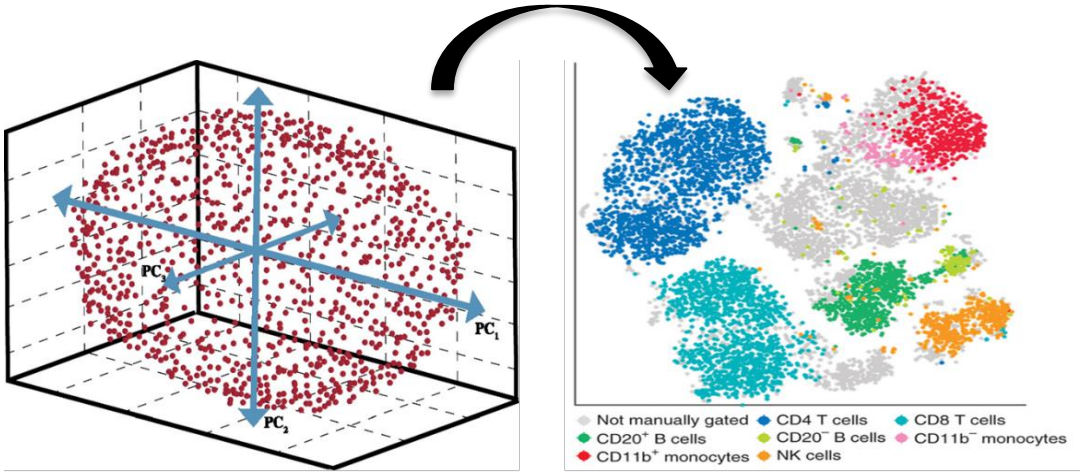


Figure 1.2 Traditional clustering and co-clustering method

The fact of high dimensional microarray gene expression data could cause the performance to drop in the analysis. High dimensional microarray data is which has hundreds to thousands of affected genes but limited number samples (Amir *et al.*, 2013; Tomasev *et al.*, 2014). The presence of non-significant genes in the classifier construction could lead to a classifier with weak discriminative power. Therefore, the cancer gene subtypes obtain will be huge and inaccurate for cancer diagnosis and prognosis. Figure 1.3 (a) shows the general view of high dimensional data for all

genes which will be classified according to suitable criteria as in Figure 1.3 (b). Therefore, the aim is to select significant genes and then classify them to help in diagnosis. The significant genes are selected by feature selection algorithm, Supervised infinite Feature Selection algorithm then classify those genes with multi-class support vector machine (SinFS-mSVM).



(a) *(b)*

Figure 1.3 High dimensional data (a) (Chester and Maecker, 2015) (b) (Amir *et al.*, 2013)

Thus, the main problem in this research is the ineffectiveness of the gene clustering due to the objective function of NCIS is dragged with outliers and selection of significant gene leads to the low performance of the results with the existence of non-significant genes in the microarray gene expression to study complex disease cancer. On that note, the hypothesis of this research is set as, if the co-clustering method is improved by minimizing sum-squared residue, then the accuracy for cancer gene classification increases. Therefore, this research proposes to deal with the above-mentioned problems based on the following research questions:

(i) How to efficiently identify cancer gene subtype co-clusters in the gene expression changes to improve clustering quality?

(ii) How to ensure to select significant gene for cancer gene subtype to improve the efficiency of identification of significant genes to improve the classification performance?

1.5 Research Goal and Objectives

The goal of this research is to propose and develop an effective improved objective function of the network-assisted co-clustering algorithm by minimizing sum-squared residue for the identification of cancer gene subtypes from high dimensional microarray data. The objectives are as follows:

(a) To propose an improved NCIS by minimizing sum-squared residue for efficient identification of cancer gene subtype in gene expression for better co-clustering performance.
(b) To propose a supervised infFS-mSVM method for multiple cancer gene subset selections besides to obtain better classification performance for selected significant gene subset.

1.6 Research Scope and Significance

This research work is primarily to address the problems in microarray gene expression data and traditional clustering. Presence of non-significant gens in gene expression data and grouping a gene into a single cluster are major issues. A single gene may fall into multiple clusters with the presence of significant genes. The existing NCIS algorithm improved due to objective function which cannot identify the outliers. Hence, this addressed by minimizing sum-squared residue to identify co-clusters of cancer gene subtype. Then, the significant genes are identified by infinite feature selection (infFS) method to be classified further. Two well-known cancer datasets are used to evaluate the effectiveness of the proposed methods. They are breast cancer (BRCA) (Network, 2012) and glioblastoma multiforme (GBM) (Verhaak *et al.*, 2010). These microarray datasets are used together with gene

network information from Reactome (Croft *et al.*, 2010), NCI (National Cancer Institute)-Nature curated Pathway Interaction Database (PID), Kyoto Encyclopaedia of Genes and Genomes (KEGG), GeneCards, and Database for Annotation, Visualization and Integrated Discovery (DAVID). The development of the proposed methods in this research is based on Network-assisted Co-clustering for the Identification of Cancer Subtypes (NCIS) in Matlab, and the performance is justified by comparing with NCIS (Liu *et al.*, 2014), MSSRCC (Cho and Dhillon, 2008), and NetBC (Yu *et al.*, 2017a). Identified genes are validated through biological literature in *Genecards* (www.genecards.com) to justify the relationship of these genes with the disease.

The significance of this research is addressed from co-clustering and classification accuracy of identified cancer gene subtypes using microarray data. The proposed method efficiently incorporates network information with gene expression across samples to produce co-clusters. Also, this method has the potential to handle high dimensional variations and integrating prior knowledge in analysing datasets. Besides this, it also expands the understanding of the cellular and molecular-level mechanisms of cancer. The obtained co-clusters assist to improve classification accuracy. Then, these genes are validated biologically for their contribution to a cancer type. This validation is to identify the drug target genes, biomarkers, tumour suppressor, etc. Additionally, the drug target of specific cancer can study from the validated genes. On top of that, it would be beneficial for the cancer genomics field, especially the medical field to enhance diagnosis and prognosis.

1.7 Thesis Organization

This thesis comprises of six chapters. The remaining chapters of the thesis are described as follows:

i. Chapter 2: This chapter provides a literature review of the research. There are fundamental concepts in molecular biology such as central dogma, genes, gene expression, microarray technology, machine learning, data mining,

clustering and classification, gene network and, review on previous research works in gene and subtypes identification using data mining approaches is also described. A comparative analysis of the selected algorithms is also presented. Finally, a description of current research trends and directions are explained.

ii. Chapter 3: This chapter presents the research methodology. It shows the main research framework described in this research. Then, microarray cancer datasets details are described. Besides, performance measurement for the verification and validation process of the research is presented.

iii. Chapter 4: This chapter illustrates details of standard NCIS and proposed improved NCIS (iNCIS) algorithms to identify cancer gene subtypes. A detailed comparison of the proposed method with the previous work is presented. Suitable parameter settings to experiment are also listed - results obtained from the research presented with previous research results. Finally, a comprehensive discussion about the results and analysis are described.

iv. Chapter 5: This chapter provides the Supervised Infinite Feature Selection (SinfFS) algorithm with a multiclass support vector machine (SVM) for further cancer gene ranking for classification. Classification accuracy and error rate for gene selection are presented. Furthermore, biological validation for selected genes and their pathways are illustrated.

v. Chapter 6: This chapter discusses the general conclusion and contribution of the research from the achieved results. Together with this, topics for future studies also have been suggested.

REFERENCES

Abba, M. C., Drake, J. A., Hawkins, K. A., Hu, Y., Sun, H., Notcovich, C., et al. (2004). Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression. *Breast Cancer Research, 6*(5), R499.

Abdel-Fatah, T., Balls, G., Miles, A., Moseley, P., Green, A., Rees, R., et al. (2012). Abstract P6-07-09: Identification of Trophinin associated protein (TROAP) as a novel biological marker in breast cancer (BC): Co-expression of TROAP and TOPO2A predicts response of anthracycline based chemotherapy (ATC-CT). *Cancer Research, 72*(24 Supplement), P6-07-09-P06-07-09.

Abdulazeez, S. (2019). Molecular simulation studies on B-cell lymphoma/leukaemia 11A (BCL11A). *American journal of translational research, 11*(6), 3689.

Abdullah, A., Deris, S., Hashim, S. Z. M., Mohamad, M. S., and Arjunan, S. N. V. (2011, 5-8 Dec. 2011). *An improved local best searching in Particle Swarm Optimization using Differential Evolution.* Paper presented at the 2011 11th International Conference on Hybrid Intelligent Systems (HIS), 115-120.

Aggarwal, C. C. (2014). Feature selection for classification: A review. In *Data Classification* (pp. 63-90): Chapman and Hall/CRC.

Agnihotri, S., Burrell, K., Buczkowicz, P., Remke, M., Golbourn, B., Chornenkyy, Y., et al. (2014). ATM regulates 3-Methylpurine-DNA glycosylase and promotes therapeutic resistance to alkylating agents. *Cancer Discovery*, CD-14-0157.

Agrawal, R., Gehrke, J. E., Gunopulos, D., and Raghavan, P. (1999). Automatic subspace clustering of high dimensional data for data mining applications: Google Patents.

Al-Batah, M. S., Zaqaibeh, B. M., Alomari, S. A., and Alzboon, M. S. (2019). Gene Microarray Cancer Classification using Correlation Based Feature Selection Algorithm and Rules Classifiers. *International Journal of Online and Biomedical Engineering (iJOE), 15*(08), 62-73.

Alavi Majd, H., Shahsavari, S., Baghestani, A. R., Tabatabaei, S. M., Khadem Bashi, N., Rezaei Tavirani, M., et al. (2016). Evaluation of Plaid Models in Biclustering of Gene Expression Data. *Scientifica, 2016*.

Alcantara Llaguno, Sheila R., Wang, Z., Sun, D., Chen, J., Xu, J., Kim, E., et al. (2015). Adult Lineage-Restricted CNS Progenitors Specify Distinct Glioblastoma Subtypes. *Cancer Cell, 28*(4), 429-440.

Algfoor, Z. A., Sunar, M. S., and Abdullah, A. (2017). A new weighted pathfinding algorithms to reduce the search time on grid maps. *Expert Systems with Applications, 71*, 319-331.

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature, 403*(6769), 503-511.

Alpaydin, E. (2014). *Introduction to machine learning*: MIT press.

Amir, E.-a. D., Davis, K. L., Tadmor, M. D., Simonds, E. F., Levine, J. H., Bendall, S. C., et al. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature Biotechnology, 31*, 545.

An, J., Liew, A. W.-C., and Nelson, C. C. (2012). Seed-Based Biclustering of Gene Expression Data. *PLOS ONE, 7*(8), e42431.

Anaissi, A., Kennedy, P. J., Goyal, M., and Catchpoole, D. R. (2013). A balanced iterative random forest for gene selection from microarray data. *BMC bioinformatics, 14*(1), 261.

Andreopoulos, B., An, A., Wang, X., and Schroeder, M. (2009). A roadmap of clustering algorithms: finding a match for a biomedical application. *Briefings in Bioinformatics, 10*(3), 297-314.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nature genetics, 25*(1), 25.

Ashford, P., Pang, C. S. M., Moya-Garcia, A. A., Adeyelu, T., and Orengo, C. (2018). A CATH domain functional family based approach to identify putative cancer driver genes and driver mutations. *bioRxiv*, 399014.

Aswad, L., Yenamandra, S. P., Ow, G. S., Grinchuk, O., Ivshina, A. V., and Kuznetsov, V. A. (2015). Genome and transcriptome delineation of two major oncogenic pathways governing invasive ductal breast cancer development. *Oncotarget, 6*(34), 36652-36674.

Ayadi, W., Elloumi, M., and Hao, J.-K. (2009). A biclustering algorithm based on a Bicluster Enumeration Tree: application to DNA microarray data. *BioData Mining, 2*(1), 9.

Azar, A. T., and El-Said, S. A. (2014). Performance analysis of support vector machines classifiers in breast cancer mammography recognition. *Neural Computing and Applications, 24*(5), 1163-1177.

Bagheri-Yarmand, R., Mandal, M., Taludker, A. H., Wang, R.-A., Vadlamudi, R. K., Kung, H.-J., et al. (2001). Etk/Bmx tyrosine kinase activates Pak1 and regulates tumorigenicity of breast cancer cells. *Journal of Biological Chemistry, 276*(31), 29403-29409.

Bai, Y., Qiao, L., Xie, N., Shi, Y., Liu, N., and Wang, J. (2017). Expression and prognosis analyses of the Tob/BTG antiproliferative (APRO) protein family in human cancers. *PLOS ONE, 12*(9), e0184902.

Bajikar, S. S., Wang, C.-C., Borten, M. A., Pereira, E. J., Atkins, K. A., and Janes, K. A. (2017). Tumor-Suppressor Inactivation of GDF11 Occurs by Precursor Sequestration in Triple-Negative Breast Cancer. *Developmental Cell, 43*(4), 418-435.e413.

Baldi, P., and Hatfield, G. W. (2002). *DNA microarrays and gene expression: from experiments to data analysis and modeling*: Cambridge university press.

Baldi, P., and Long, A. D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics, 17*(6), 509-519.

Bansal, M., Gatta, G. D., and di Bernardo, D. (2006). Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics, 22*(7), 815-822.

Bao, L., Guo, T., Wang, J., Zhang, K., and Bao, M. (2020). Prognostic genes of triple-negative breast cancer identified by weighted gene co-expression network analysis. *Oncology Letters, 19*(1), 127-138.

Bar, E. E., Lin, A., Mahairaki, V., Matsui, W., and Eberhart, C. G. (2010). Hypoxia Increases the Expression of Stem-Cell Markers and Promotes Clonogenicity in Glioblastoma Neurospheres. *The American Journal of Pathology, 177*(3), 1491-1502.

Barabasi, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet, 5*(2), 101-113.

Barnett, D. H., Sheng, S., Howe Charn, T., Waheed, A., Sly, W. S., Lin, C.-Y., et al. (2008). Estrogen Receptor Regulation of Carbonic Anhydrase XII through a Distal Enhancer in Breast Cancer. *Cancer Research, 68*(9), 3505-3515.

Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M., and Yakhini, Z. (2000). Tissue classification with gene expression profiles. *Journal of computational biology, 7*(3-4), 559-583.

Benabdeslem, K., and Allab, K. (2013). Bi-clustering continuous data with self-organizing map. *Neural Computing and Applications, 22*(7), 1551-1562.

Berenguer, J., Lagerweij, T., Zhao, X. W., Dusoswa, S., van der Stoop, P., Westerman, B., et al. (2018). Glycosylated extracellular vesicles released by glioblastoma cells are decorated by CCL18 allowing for cellular uptake via chemokine receptor CCR8. *Journal of extracellular vesicles, 7*(1), 1446660-1446660.

Bergmann, S., Ihmels, J., and Barkai, N. (2003). Iterative signature algorithm for the analysis of large-scale gene expression data. *Physical review E, 67*(3), 031902.

Bhakta, S., Crocker, L. M., Chen, Y., Hazen, M., Schutten, M. M., Li, D., et al. (2018). An Anti-GDNF Family Receptor Alpha 1 (GFRA1) Antibody–Drug Conjugate for the Treatment of Hormone Receptor–Positive Breast Cancer. *Molecular Cancer Therapeutics, 17*(3), 638-649.

Bhat, R., Vasaikar, S., Bae, L., Carmine, D. A., Cataldo, M. L., Nanda, S., et al. (2018). Abstract 1926: Npy1r as a prognostic marker and a novel drug target in estrogen receptor-positive breast cancer. *Cancer Research, 78*(13 Supplement), 1926-1926.

Binai, N. A., Carra, G., Löwer, J., Löwer, R., and Wessler, S. (2013). Differential gene expression in ERα-positive and ERα-negative breast cancer cells upon leptin stimulation. *Endocrine, 44*(2), 496-503.

Birinci, H., Şen, B., Sayğılı, S., Ölmez, E., İnan, S., and Özbilgin, K. (2017). Investigation of the Effects of Paclitaxel and Pycnogenol on Inflammatory Response (PTX3, BDNF, IGF2R) in Human Breast Cancer Cell Line. *Proceedings, 1*(10), 1024.

Bø, T. H., and Jonassen, I. (2002). New feature subset selection procedures for classification of expression profiles. *Genome biology, 3*(4), research0017. 0011.

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2012). A review of feature selection methods on synthetic data. *Knowledge and Information Systems, 34*.

Bolón-Canedo, V., Sánchez-Maroño, N., and Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems, 34*(3), 483-519.

Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., Benítez, J. M., and Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Information Sciences, 282*, 111-135.

Bozdağ, D., Parvin, J. D., and Catalyurek, U. V. (2009). A biclustering method to discover co-regulated genes using diverse gene expression datasets. In *Bioinformatics and Computational Biology* (pp. 151-163): Springer.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Buchanan, F. G., and DuBois, R. N. (2006). Emerging roles of β-arrestins. *Cell cycle, 5*(18), 2060-2063.

Budczies, J., Brockmöller, S. F., Müller, B. M., Barupal, D. K., Richter-Ehrenstein, C., Kleine-Tebbe, A., et al. (2013). Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism. *Journal of Proteomics, 94*, 279-288.

Busygin, S., Prokopyev, O., and Pardalos, P. M. (2008). Biclustering in data mining. *Computers & Operations Research, 35*(9), 2964-2987.

Cadenas, J. M., Garrido, M. C., and MartíNez, R. (2013). Feature subset selection filter–wrapper based on low quality data. *Expert systems with applications, 40*(16), 6241-6252.

Canul-Reich, J., Hall, L. O., Goldgof, D. B., Korecki, J. N., and Eschrich, S. (2012). Iterative feature perturbation as a gene selector for microarray data. *International Journal of Pattern Recognition and Artificial Intelligence, 26*(05), 1260003.

Carmona-Saez, P., Pascual-Marqui, R. D., Tirado, F., Carazo, J. M., and Pascual-Montano, A. (2006). Biclustering of gene expression data by non-smooth non-negative matrix factorization. *BMC Bioinformatics, 7*(1), 78.

Carter, G. W., Prinz, S., Neou, C., Shelby, J. P., Marzolf, B., Thorsson, V., et al. (2007). Prediction of phenotype and gene expression for combinations of mutations. *Molecular systems biology, 3*(1).

Casci, T. (2010). Gene networks: Meaningful connections. *Nature Reviews Genetics*.

Casey, T., Bond, J., Tighe, S., Hunter, T., Lintault, L., Patel, O., et al. (2009). Molecular signatures suggest a major role for stromal cells in development of invasive breast cancer. *Breast Cancer Research and Treatment, 114*(1), 47-62.

Cerami, E., Demir, E., Schultz, N., Taylor, B. S., and Sander, C. (2010). Automated network analysis identifies core pathways in glioblastoma. *PloS one, 5*(2), e8918.

Chan, W.-C., Ho, M.-R., Li, S.-C., Tsai, K.-W., Lai, C.-H., Hsu, C.-N., et al. (2012). MetaMirClust: Discovery of miRNA cluster patterns using a data-mining approach. *Genomics, 100*(3), 141-148.

Chandrashekar, G., and Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering, 40*(1), 16-28.

Chatterjee, N., Wang, W.-L. W., Conklin, T., Chittur, S., and Tenniswood, M. (2013). Histone deacetylase inhibitors modulate miRNA and mRNA expression, block metaphase, and induce apoptosis in inflammatory breast cancer cells. *Cancer biology & therapy, 14*(7), 658-671.

Chen, H.-L., Yang, B., Wang, G., Wang, S.-J., Liu, J., and Liu, D.-Y. (2012). Support vector machine based diagnostic system for breast cancer using swarm intelligence. *Journal of medical systems, 36*(4), 2505-2519.

Cheng, C.-J., Lin, Y.-C., Tsai, M.-T., Chen, C.-S., Hsieh, M.-C., Chen, C.-L., et al. (2009). SCUBE2 Suppresses Breast Tumor Cell Proliferation and Confers a Favorable Prognosis in Invasive Breast Cancer. *Cancer Research, 69*(8), 3634-3641.

Cheng, L., Schneider, B. P., and Li, L. (2016a). A bioinformatics approach for precision medicine off-label drug drug selection among triple negative breast cancer patients. *Journal of the American Medical Informatics Association, 23*(4), 741-749.

Cheng, W.-Y., Kandel, J. J., Yamashiro, D. J., Canoll, P., and Anastassiou, D. (2012). A Multi-Cancer Mesenchymal Transition Gene Expression Signature Is Associated with Prolonged Time to Recurrence in Glioblastoma. *PLOS ONE, 7*(4), e34705.

Cheng, W., Ren, X., Zhang, C., Cai, J., Liu, Y., Han, S., et al. (2016b). Bioinformatic profiling identifies an immune-related risk signature for glioblastoma. *Neurology, 86*(24), 2226-2234.

Cheng, Y., and Church, G. M. (2000). *Biclustering of Expression Data*. Paper presented at the Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology.

Chester, C., and Maecker, H. T. (2015). Algorithmic Tools for Mining High-Dimensional Cytometry Data. *The Journal of Immunology, 195*(3), 773-779.

Cheung, H. C., Hai, T., Zhu, W., Baggerly, K. A., Tsavachidis, S., Krahe, R., et al. (2009). Splicing factors PTBP1 and PTBP2 promote proliferation and migration of glioma cell lines. *Brain, 132*(8), 2277-2288.

Cho, H., and Dhillon, I. S. (2008). Coclustering of Human Cancer Microarrays Using Minimum Sum-Squared Residue Coclustering. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 5*(3), 385-400.

Cho, H., Dhillon, I. S., Guan, Y., and Sra, S. (2004). *Minimum sum-squared residue co-clustering of gene expression data.* Paper presented at the Proceedings of the 2004 SIAM International Conference on Data Mining, 114-125.

Chou, C.-H., Huang, M.-J., Chen, C.-H., Shyu, M.-K., Huang, J., Hung, J.-S., et al. (2015). Up-regulation of C1GALT1 promotes breast cancer cell growth through MUC1-C signaling pathway. *Oncotarget, 6*(8), 6123.

Chou, J.-S., Chiu, C.-K., Farfoura, M., and Al-Taharwa, I. (2010). Optimizing the prediction accuracy of concrete compressive strength based on a comparison of data-mining techniques. *Journal of Computing in Civil Engineering, 25*(3), 242-253.

Choudhury, Y., Tay, F. C., Lam, D. H., Sandanaraj, E., Tang, C., Ang, B.-T., et al. (2012). Attenuated adenosine-to-inosine editing of microRNA-376a* promotes invasiveness of glioblastoma cells. *The Journal of Clinical Investigation, 122*(11), 4059-4076.

Clark, T. G., Bradburn, M. J., Love, S. B., and Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. *British journal of cancer, 89*(2), 232-238.

Clausen, K. A., Blish, K. R., Birse, C. E., Triplette, M. A., Kute, T. E., Russell, G. B., et al. (2011). SOSTDC1 differentially modulates Smad and beta-catenin activation and is down-regulated in breast cancer. *Breast Cancer Research and Treatment, 129*(3), 737-746.

Clavreul, A., Etcheverry, A., Tétaud, C., Rousseau, A., Avril, T., Henry, C., et al. (2015). Identification of two glioblastoma-associated stromal cell subtypes with different carcinogenic properties in histologically normal surgical margins. *Journal of Neuro-Oncology, 122*(1), 1-10.

Clevert, D.-A., Unterthiner, T., Povysil, G., and Hochreiter, S. (2017). Rectified factor networks for biclustering of omics data. *Bioinformatics, 33*(14), i59-i66.

Colman, H., Zhang, L., Sulman, E. P., McDonald, J. M., Shooshtari, N. L., Rivera, A., et al. (2010). A multigene predictor of outcome in glioblastoma. *Neuro-Oncology, 12*(1), 49-57.

Conroy, S., Kruyt, F. A. E., Wagemakers, M., Bhat, K. P. L., and den Dunnen, W. F. A. (2018). IL-8 associates with a pro-angiogenic and mesenchymal subtype in glioblastoma. *Oncotarget, 9*(21), 15721-15731.

Corn, P. G., and El-Deiry, W. S. (2007). Microarray analysis of p53-dependent gene expression in response to hypoxia and DNA damage. *Cancer biology & therapy, 6*(12), 1858-1866.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature, 227*(5258), 561-563.

Croft, D., O'kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., et al. (2010). Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research, 39*(suppl_1), D691-D697.

Cui, J., Shan, M., Yan, R., and Wu, Y. (2014). Aero-engine fault diagnosis using improved local discriminant bases and support vector machine. *Mathematical Problems in Engineering, 2014*.

Dai, X., Chen, A., and Bai, Z. (2014). Integrative investigation on breast cancer in ER, PR and HER2-defined subgroups using mRNA and miRNA expression profiling. *Scientific Reports, 4*, 6566.

Dairou, J., Atmane, N., Rodrigues-Lima, F., and Dupret, J.-M. (2004). Peroxynitrite Irreversibly Inactivates the Human Xenobioticmetabolizing Enzyme Arylamine N-Acetyltransferase 1 (NAT1) in Human Breast Cancer Cells A CELLULAR AND MECHANISTIC STUDY. *Journal of Biological Chemistry, 279*(9), 7708-7714.

Daxin, J., Chun, T., and Aidong, Z. (2004). Cluster analysis for gene expression data: a survey. *IEEE Transactions on Knowledge and Data Engineering, 16*(11), 1370-1386.

de França, F. O., and Coelho, A. L. (2015). A biclustering approach for classification with mislabeled data. *Expert Systems with Applications, 42*(12), 5065-5075.

Deng, H., and Runger, G. (2013). Gene selection with guided regularized random forest. *Pattern Recognition, 46*(12), 3483-3489.

Deng, L., Pei, J., Ma, J., and Lee, D. L. (2004). *A rank sum test method for informative gene discovery.* Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 410-419.

Deng, S., Zhou, H., Xiong, R., Lu, Y., Yan, D., Xing, T., et al. (2007). Over-expression of genes and proteins of ubiquitin specific peptidases (USPs) and proteasome subunits (PSs) in breast cancer tissue observed by the methods of RFDD-PCR and proteomics. *Breast Cancer Research and Treatment, 104*(1), 21-30.

Deng, Y., He, T., Wu, Y., Vanka, P., Yang, G., Huang, Y., et al. (2005). Computationally analyzing the possible biological function of YJL103C-an ORF potentially involved in the regulation of energy process in yeast. *International journal of molecular medicine, 15*(1), 123-127.

Denisko, D., and Hoffman, M. M. (2018). Classification and interaction in random forests. *Proceedings of the National Academy of Sciences*, 201800256.

Denitto, M., Farinelli, A., Figueiredo, M. A., and Bicego, M. (2017a). A biclustering approach based on factor graphs and the max-sum algorithm. *Pattern Recognition, 62*, 114-124.

Denitto, M., Farinelli, A., Figueiredo, M. A. T., and Bicego, M. (2017b). A biclustering approach based on factor graphs and the max-sum algorithm. *Pattern Recognition, 62*(Supplement C), 114-124.

Dennis, G., Sherman, B., Hosack, D., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology, 4*(5), P3.

Dhillon, I. S., Mallela, S., and Modha, D. S. (2003). *Information-theoretic co-clustering*. Paper presented at the Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.

Díaz-Uriarte, R., and De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics, 7*(1), 3.

DiMeo, T. A., Anderson, K., Phadke, P., Feng, C., Perou, C. M., Naber, S., et al. (2009). A Novel Lung Metastasis Signature Links Wnt Signaling with Cancer Cell Self-Renewal and Epithelial-Mesenchymal Transition in Basal-like Breast Cancer. *Cancer Research, 69*(13), 5364-5373.

Ding, C., and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology, 3*(02), 185-205.

Ding, Y., Herman, J. A., Toledo, C. M., Lang, J. M., Corrin, P., Girard, E. J., et al. (2017). ZNF131 suppresses centrosome fragmentation in glioblastoma stem-like cells through regulation of HAUS5. *Oncotarget, 8*(30), 48545-48562.

Ding, Y., and Wilkins, D. (2006). *Improving the performance of SVM-RFE to select genes in microarray data.* Paper presented at the BMC bioinformatics, S12.

Divina, F., and Aguilar-Ruiz, J. S. (2006). Biclustering of expression data with evolutionary computation. *IEEE transactions on knowledge and data engineering, 18*(5), 590-602.

Dolezal, J. M., Dash, A. P., and Prochownik, E. V. (2018). Diagnostic and prognostic implications of ribosomal protein transcript expression patterns in human cancers. *BMC Cancer, 18*(1), 275.

Draghici, S., Kulaeva, O., Hoff, B., Petrov, A., Shams, S., and Tainsky, M. A. (2003). Noise sampling method: an ANOVA approach allowing robust selection of differentially regulated genes measured by DNA microarrays. *Bioinformatics, 19*(11), 1348-1359.

Duan, K.-B., Rajapakse, J. C., Wang, H., and Azuaje, F. (2005). Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE transactions on nanobioscience, 4*(3), 228-234.

Dunn, P. M. (2003). Gregor Mendel, OSA (1822–1884), founder of scientific genetics. *Archives of Disease in Childhood-Fetal and Neonatal Edition, 88*(6), F537-F539.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences, 95*(25), 14863-14868.

El Akadi, A., Amine, A., El Ouardighi, A., and Aboutajdine, D. (2011). A two-stage gene selection scheme utilizing MRMR filter and GA wrapper. *Knowledge and Information Systems, 26*(3), 487-500.

Ellis, M. J., Tao, Y., Luo, J., A'Hern, R., Evans, D. B., Bhatnagar, A. S., et al. (2008). Outcome Prediction for Estrogen Receptor–Positive Breast Cancer Based on Postneoadjuvant Endocrine Therapy Tumor Characteristics. *JNCI: Journal of the National Cancer Institute, 100*(19), 1380-1388.

Endogenous, H., Breast Cancer Collaborative, G., Key, T. J., Appleby, P. N., Reeves, G. K., and Roddam, A. W. (2010). Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. *The Lancet. Oncology, 11*(6), 530-542.

Epstein, R. J. (2004). The CXCL12–CXCR4 chemotactic pathway as a target of adjuvant breast cancer therapies. *Nature Reviews Cancer, 4*(11), 901.

Erciyes, K. (2015). *Distributed and sequential algorithms for bioinformatics* (Vol. 23): Springer.

Erdem-Eraslan, L., Heijsman, D., de Wit, M., Kremer, A., Sacchetti, A., van der Spek, P. J., et al. (2015). Tumor-specific mutations in low-frequency genes affect their functional properties. *Journal of Neuro-Oncology, 122*(3), 461-470.

Eren, K., Deveci, M., Kucuktunc, O., and Catalyurek, U. V. (2013). A comparative analysis of biclustering algorithms for gene expression data. *Brief Bioinform, 14*(3), 279-292.

Esseghir, S., Reis-Filho, J., Kennedy, A., James, M., O'hare, M., Jeffery, R., et al. (2006). Identification of transmembrane proteins as potential prognostic markers and therapeutic targets in breast cancer by a screen for signal sequence encoding transcripts. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland, 210*(4), 420-430.

Fallahpour, S., Navaneelan, T., De, P., and Borgo, A. (2017). Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data. *CMAJ open, 5*(3), E734-E739.

Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S., Nobel, A. B., et al. (2006). Concordance among gene-expression–based predictors for breast cancer. *New England Journal of Medicine, 355*(6), 560-569.

Fanale, D., Amodeo, V., Bazan, V., Insalaco, L., Incorvaia, L., Barraco, N., et al. (2016). Can the microRNA expression profile help to identify novel targets for zoledronic acid in breast cancer? *Oncotarget, 7*(20), 29321-29332.

Fang, X., Huang, Z., Zhou, W., Wu, Q., Sloan, A. E., Ouyang, G., et al. (2014). The Zinc Finger Transcription Factor ZFX Is Required for Maintaining the Tumorigenic Potential of Glioblastoma Stem Cells. *STEM CELLS, 32*(8), 2033-2047.

Farmer, P., Bonnefoi, H., Anderle, P., Cameron, D., Wirapati, P., Becette, V., et al. (2009). A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nature medicine, 15*(1), 68.

Farquad, M., and Bose, I. (2012). Preprocessing unbalanced data using support vector machine. *Decision Support Systems, 53*(1), 226-233.

Fiannaca, A., La Rosa, M., La Paglia, L., Rizzo, R., and Urso, A. (2015). Analysis of miRNA expression profiles in breast cancer using biclustering. *BMC Bioinformatics, 16*(4), S7.

Flanagan, J. M., Funes, J. M., Henderson, S., Wild, L., Carey, N., and Boshoff, C. (2009). Genomics screen in transformed stem cells reveals RNASEH2A, PPAP2C, and ADARB1 as putative anticancer drug targets. *Molecular Cancer Therapeutics, 8*(1), 249-260.

Flores, J. L., Inza, I., Larrañaga, P., and Calvo, B. (2013). A new measure for gene expression biclustering based on non-parametric correlation. *Computer methods and programs in biomedicine, 112*(3), 367-397.

Foley, J., Nickerson, N. K., Nam, S., Allen, K. T., Gilmore, J. L., Nephew, K. P., et al. (2010). EGFR signaling in breast cancer: Bad to the bone. *Seminars in Cell & Developmental Biology, 21*(9), 951-960.

Fontanillo, C., Aibar, S., Sanchez-Santos, J. M., and De Las Rivas, J. (2012). Combined analysis of genome-wide expression and copy number profiles to identify key altered genomic regions in cancer. *BMC genomics, 13*(5), 1.

Fu, Y., Zheng, S., An, N., Athanasopoulos, T., Popplewell, L., Liang, A., et al. (2011). β-catenin as a potential key target for tumor suppression. *International journal of cancer, 129*(7), 1541-1551.

Gabrusiewicz, K., Rodriguez, B., Wei, J., Hashimoto, Y., Healy, L. M., Maiti, S. N., et al. (2016). Glioblastoma-infiltrated innate immune cells resemble M0 macrophage phenotype. *JCI insight, 1*(2).

Gan, X., Liew, A. W.-C., and Yan, H. (2008). Discovering biclusters in gene expression data based on high-dimensional linear geometries. *BMC Bioinformatics, 9*(1), 209.

Gao, C., McDowell, I. C., Zhao, S., Brown, C. D., and Engelhardt, B. E. (2016). Context Specific and Differential Gene Co-expression Networks via Bayesian Biclustering. *PLOS Computational Biology, 12*(7), e1004791.

Gardner, T. S., and Faith, J. J. (2005). Reverse-engineering transcription control networks. *Physics of Life Reviews, 2*(1), 65-88.

Garner, J. M., Ellison, D. W., Finkelstein, D., Ganguly, D., Du, Z., Sims, M., et al. (2015). Molecular Heterogeneity in a Patient-Derived Glioblastoma Xenoline Is Regulated by Different Cancer Stem Cell Populations. *PLOS ONE, 10*(5), e0125838.

Gasch, A. P., and Eisen, M. B. (2002). Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome biology, 3*(11), research0059. 0051.

Gaul, W., and Schader, M. (1996). A new algorithm for two-mode clustering. In *Data analysis and information systems* (pp. 15-23): Springer.

Gee, J. M. W., Robertson, J. F. R., Ellis, I. O., Nicholson, R. I., and Hurst, H. C. (1999). Immunohistochemical analysis reveals a tumour suppressor-like role for the transcription factor AP-2 in invasive breast cancer. *The Journal of Pathology, 189*(4), 514-520.

Geisenberger, C., Mock, A., Warta, R., Rapp, C., Schwager, C., Korshunov, A., et al. (2015). Molecular profiling of long-term survivors identifies a subgroup of glioblastoma characterized by chromosome 19/20 co-gain. *Acta Neuropathologica, 130*(3), 419-434.

Gelman, I. H. (2010). Emerging Roles for SSeCKS/Gravin/AKAP12 in the Control of Cell Proliferation, Cancer Malignancy, and Barriergenesis. *Genes & Cancer, 1*(11), 1147-1156.

Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*(14), 2225-2236.

Getz, G., Levine, E., and Domany, E. (2000). Coupled two-way clustering analysis of gene microarray data. *Proceedings of the National Academy of Sciences, 97*(22), 12079-12084.

Ghosh, M. G., Thompson, D. A., and Weigel, R. J. (2000). PDZK1 and GREB1 Are Estrogen-regulated Genes Expressed in Hormone-responsive Breast Cancer 1, 2. *Cancer Research, 60*(22), 6367-6375.

Gianinna, A., Alhadi, B., and Devvi, S. (2017). Implementation of plaid model biclustering method on microarray of carcinoma and adenoma tumor gene expression data. *Journal of Physics: Conference Series, 893*(1), 012046.

Gianni, L., Zambetti, M., Clark, K., Baker, J., Cronin, M., Wu, J., et al. (2005). Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *Journal of clinical oncology, 23*(29), 7265-7277.

Gil Del Alcazar, C. R., Huh, S. J., Ekram, M. B., Trinh, A., Liu, L. L., Beca, F., et al. (2017). Immune Escape in Breast Cancer During In Situ to Invasive Carcinoma Transition. *Cancer Discovery*, CD-17-0222.

Gilbert, W. V., Bell, T. A., and Schaening, C. (2016). Messenger RNA modifications: Form, distribution, and function. *Science, 352*(6292), 1408-1412.

Goel, H. L., Pursell, B., Chang, C., Shaw, L. M., Mao, J., Simin, K., et al. (2013). GLI1 regulates a novel neuropilin-2/α6β1 integrin based autocrine pathway that contributes to breast cancer initiation. *EMBO Molecular Medicine, 5*(4), 488-508.

Gould, S. E., Junttila, M. R., and de Sauvage, F. J. (2015). Translational value of mouse models in oncology drug development. *Nature medicine, 21*(5), 431.

Grant, N. J., Coates, P. J., Woods, Y. L., Bray, S. E., Morrice, N. A., Hastie, C. J., et al. (2015). Phosphorylation of a splice variant of collapsin response mediator protein 2 in the nucleus of tumour cells links cyclin dependent kinase-5 to oncogenesis. *BMC Cancer, 15*(1), 885.

Grimes, T., Potter, S. S., and Datta, S. (2019). Integrating gene regulatory pathways into differential network analysis of gene expression data. *Scientific Reports, 9*(1), 5479.

Grinde, M. T., Skrbo, N., Moestue, S. A., Rødland, E. A., Borgan, E., Kristian, A., et al. (2014). Interplay of choline metabolites and genes in patient-derived breast cancer xenografts. *Breast Cancer Research, 16*(1), R5.

Gruber, A. D., and Pauli, B. U. (1999). Tumorigenicity of Human Breast Cancer Is Associated with Loss of the Ca²⁺-activated Chloride Channel CLCA2. *Cancer Research, 59*(21), 5488-5491.

Guan, X., Vengoechea, J., Zheng, S., Sloan, A. E., Chen, Y., Brat, D. J., et al. (2014). Molecular Subtypes of Glioblastoma Are Relevant to Lower Grade Glioma. *PLOS ONE, 9*(3), e91216.

Guo, J., Zhu, P., Wu, C., Yu, L., Zhao, S., and Gu, X. (2003). In silico analysis indicates a similar gene expression pattern between human brain and testis. *Cytogenetic and genome research, 103*(1-2), 58-62.

Gupta, M. K., Polisetty, R. V., Ramamoorthy, K., Tiwary, S., Kaur, N., Uppin, M. S., et al. (2013). Secretome analysis of Glioblastoma cell line-HNGC-2. *Molecular BioSystems, 9*(6), 1390-1400.

Győrffy, B., Bottai, G., Fleischer, T., Munkácsy, G., Budczies, J., Paladini, L., et al. (2016). Aberrant DNA methylation impacts gene expression and prognosis in breast cancer subtypes. *International journal of cancer, 138*(1), 87-97.

Györffy, B., Lanczky, A., Eklund, A. C., Denkert, C., Budczies, J., Li, Q., et al. (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Research and Treatment, 123*(3), 725-731.

Han, F., Sun, W., and Ling, Q.-H. (2014). A Novel Strategy for Gene Selection of Microarray Data Based on Gene-to-Class Sensitivity Information. *PLOS ONE, 9*(5), e97530.

Harrington, W. R., Sengupta, S., and Katzenellenbogen, B. S. (2006). Estrogen Regulation of the Glucuronidation Enzyme UGT2B15 in Estrogen Receptor-Positive Breast Cancer Cells. *Endocrinology, 147*(8), 3843-3850.

Harris, L. N., Ismaila, N., McShane, L. M., Andre, F., Collyar, D. E., Gonzalez-Angulo, A. M., et al. (2016). Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 34*(10), 1134-1150.

Hartigan, J. A. (1972). Direct clustering of a data matrix. *Journal of the american statistical association, 67*(337), 123-129.

Hartwell, L., Goldberg, M. L., Fischer, J. A., Hood, L., and Aquadro, C. F. (2008). *Genetics: from genes to genomes*: McGraw-Hill New York.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). Unsupervised learning. In *The elements of statistical learning* (pp. 485-585): Springer.

Hayes, J., Peruzzi, P. P., and Lawler, S. (2014). MicroRNAs in cancer: biomarkers, functions and therapy. *Trends in molecular medicine, 20*(8), 460-469.

He, R., Liu, P., Xie, X., Zhou, Y., Liao, Q., Xiong, W., et al. (2017). circGFRA1 and GFRA1 act as ceRNAs in triple negative breast cancer by regulating miR-34a. *Journal of Experimental & Clinical Cancer Research, 36*(1), 145.

Henriksen, M., Johnsen, K. B., Olesen, P., Pilgaard, L., and Duroux, M. (2014). MicroRNA Expression Signatures and Their Correlation with Clinicopathological Features in Glioblastoma Multiforme. *NeuroMolecular Medicine, 16*(3), 565-577.

Henriques, R., Ferreira, F. L., and Madeira, S. C. (2017). BicPAMS: software for biological data analysis with pattern-based biclustering. *BMC Bioinformatics, 18*, 82.

Henriques, R., and Madeira, S. C. (2014). BicSPAM: flexible biclustering using sequential patterns. *BMC Bioinformatics, 15*(1), 130.

Herbert, A., Gerry, N. P., McQueen, M. B., Heid, I. M., Pfeufer, A., Illig, T., et al. (2006). A common genetic variant is associated with adult and childhood obesity. *Science, 312*(5771), 279-283.

Hess, K. R., Anderson, K., Symmans, W. F., Valero, V., Ibrahim, N., Mejia, J. A., et al. (2006). Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *Journal of clinical oncology, 24*(26), 4236-4244.

Hessel, E. V., de Wit, M., Wolterink-Donselaar, I. G., Karst, H., de Graaff, E., van Lith, H. A., et al. (2014). Identification of Srp9 as a febrile seizure susceptibility gene. *Annals of clinical and translational neurology, 1*(4), 239-250.

Hino, O., and Kobayashi, T. (2017). Mourning Dr. Alfred G. Knudson: the two-hit hypothesis, tumor suppressor genes, and the tuberous sclerosis complex. *Cancer Science, 108*(1), 5-11.

Hlaváč, V., Brynychová, V., Václavíková, R., Ehrlichová, M., Vrána, D., Pecha, V., et al. (2013a). The expression profile of ATP-binding cassette transporter genes in breast carcinoma. *Pharmacogenomics, 14*(5), 515-529.

Hlaváč, V., Brynychová, V., Václavíková, R., Ehrlichová, M., Vrána, D., Pecha, V., et al. (2013b). The expression profile of ATP-binding cassette transporter genes in breast carcinoma. *Pharmacogenomics, 14*(5), 515-529.

Hoelzinger, D. B., Mariani, L., Weis, J., Woyke, T., Berens, T. J., McDonough, W., et al. (2005). Gene expression profile of glioblastoma multiforme invasive phenotype points to new therapeutic targets. *Neoplasia, 7*(1), 7-16.

Hofmann, T., and Puzicha, J. (1999). *Latent class models for collaborative filtering.* Paper presented at the IJCAI.

Hofree, M., Shen, J. P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. *Nature Methods, 10*, 1108.

Holst, F., Stahl, P. R., Ruiz, C., Hellwinkel, O., Jehan, Z., Wendland, M., et al. (2007). Estrogen receptor alpha (ESR1) gene amplification is frequent in breast cancer. *Nature Genetics, 39*, 655.

Hon, G. C., Hawkins, R. D., Caballero, O. L., Lo, C., Lister, R., Pelizzola, M., et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome research, 22*(2), 246-258.

Hopp, L., Willscher, E., Löffler-Wirth, H., and Binder, H. (2015). Function shapes content: DNA-methylation marker genes and their impact for molecular mechanisms of glioma. *Journal of Cancer Research Updates, 4*(4), 127-148.

Hossain, A., Kuo, M. T., and Saunders, G. F. (2006). Mir-17-5p regulates breast cancer cell proliferation by inhibiting translation of AIB1 mRNA. *Molecular and cellular biology, 26*(21), 8191-8201.

Houari, A., Ayadi, W., and Yahia, S. B. (2017). Mining Negative Correlation Biclusters from Gene Expression Data using Generic Association Rules. *Procedia Computer Science, 112*(Supplement C), 278-287.

Howell, D. C. (2014). Median absolute deviation. *Wiley statsRef: statistics reference online*.

Hsieh, T.-H., Hsu, C.-Y., Tsai, C.-F., Long, C.-Y., Wu, C.-H., Wu, D.-C., et al. (2015). HDAC Inhibitors Target HDAC5, Upregulate MicroRNA-125a-5p, and Induce Apoptosis in Breast Cancer Cells. *Molecular Therapy, 23*(4), 656-666.

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols, 4*(1), 44-57.

Huang, J., Luo, Q., Xiao, Y., Li, H., Kong, L., and Ren, G. (2017). The implication from RAS/RAF/ERK signaling pathway increased activation in epirubicin treated triple negative breast cancer. *Oncotarget, 8*(64), 108249-108260.

Huang, L., Zhang, H. H., Zeng, Z.-B., and Bushel, P. R. (2013). Improved sparse multi-class SVM and its application for gene selection in cancer classification. *Cancer informatics, 12*, CIN. S10212.

Huang, S.-Y., Sun, H.-J., Huang, C.-D., Chung, I.-F., and Su, C.-H. (2014). *A modified fuzzy co-clustering (MFCC) approach for microarray data analysis.* Paper presented at the Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on, 267-272.

Huang, X.-D., Wang, Z.-F., Dai, L.-M., and Li, Z.-Q. (2012). Microarray analysis of the hypoxia-induced gene expression profile in malignant C6 glioma cells. *Asian Pacific Journal of Cancer Prevention, 13*(9), 4793-4799.

Hudelist, G., Singer, C. F., Pischinger, K. I., Kaserer, K., Manavi, M., Kubista, E., et al. (2006). Proteomic analysis in human breast cancer: identification of a

characteristic protein expression profile of malignant breast epithelium. *Proteomics, 6*(6), 1989-2002.

Hussain, S. F., and Ramazan, M. (2016). Biclustering of human cancer microarray data using co-similarity based co-clustering. *Expert Systems with Applications, 55*(Supplement C), 520-531.

Hwang, S., Kim, C. Y., Yang, S., Kim, E., Hart, T., Marcotte, E. M., et al. (2018). HumanNet v2: human gene networks for disease research. *Nucleic Acids Research, 47*(D1), D573-D580.

Ishibashi, Y., Ohtsu, H., Ikemura, M., Kikuchi, Y., Niwa, T., Nishioka, K., et al. (2017). Serum TFF1 and TFF3 but not TFF2 are higher in women with breast cancer than in women without breast cancer. *Scientific Reports, 7*(1), 4846.

Ivanov, S., Liao, S.-Y., Ivanova, A., Danilkovitch-Miagkova, A., Tarasova, N., Weirich, G., et al. (2001). Expression of Hypoxia-Inducible Cell-Surface Transmembrane Carbonic Anhydrases in Human Cancer. *The American Journal of Pathology, 158*(3), 905-919.

Ivanov, S. V., Panaccione, A., Nonaka, D., Prasad, M. L., Boyd, K. L., Brown, B., et al. (2013). Diagnostic SOX10 gene signatures in salivary adenoid cystic and breast basal-like carcinomas. *British Journal Of Cancer, 109*, 444.

Jabbir, F., Irfan, M., Mustafa, G., and Ahmad, H. I. (2019). Bioinformatics approaches to explore the phylogeny and role of BRCA1 in breast cancer. *Critical Reviews™ in Eukaryotic Gene Expression, 29*(6).

Jackson, M., Hassiotou, F., and Nowak, A. (2015). Glioblastoma stem-like cells: at the root of tumor recurrence and a therapeutic target. *Carcinogenesis, 36*(2), 177-185.

Jansen, M. P. H. M., Sas, L., Sieuwerts, A. M., Van Cauwenberghe, C., Ramirez-Ardila, D., Look, M., et al. (2015). Decreased expression of ABAT and STC2 hallmarks ER-positive inflammatory breast cancer and endocrine therapy resistance in advanced disease. *Molecular Oncology, 9*(6), 1218-1233.

Jean-Quartier, C., Jeanquartier, F., Jurisica, I., and Holzinger, A. (2018). In silico cancer research towards 3R. *BMC Cancer, 18*(1), 408.

Jiang, L., Chen, S., Zhao, D., Yan, J., Chen, J., Yang, C., et al. (2018a). MNX1 reduces sensitivity to anoikis by activating TrkB in human glioma cells. *Molecular medicine reports, 18*(3), 3271-3279.

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018b). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nature Medicine, 24*(10), 1550-1558.

Jiang, T., Li, Z., Shang, X., Chen, B., Li, W., and Yin, Z. (2016). Constrained query of order-preserving submatrix in gene expression data. *Frontiers of Computer Science, 10*(6), 1052-1066.

Jin, D., and Lee, H. (2015). A Computational Approach to Identifying Gene-microRNA Modules in Cancer. *PLOS Computational Biology, 11*(1), e1004042.

Jin, L., Zhang, Y., Li, H., Yao, L., Fu, D., Yao, X., et al. (2012). Differential secretome analysis reveals CST6 as a suppressor of breast cancer bone metastasis. *Cell Research, 22*, 1356.

Jin, X., Jeon, H.-M., Jin, X., Kim, E.-J., Yin, J., Jeon, H.-Y., et al. (2016). The ID1-CULLIN3 Axis Regulates Intracellular SHH and WNT Signaling in Glioblastoma Stem Cells. *Cell Reports, 16*(6), 1629-1641.

Johansson, I., Nilsson, C., Berglund, P., Lauss, M., Ringnér, M., Olsson, H., et al. (2012). Gene expression profiling of primary male breast cancers reveals two

unique subgroups and identifies N-acetyltransferase-1 (NAT1) as a novel prognostic biomarker. *Breast Cancer Research, 14*(1), R31.

Joshi, A. D., Parsons, D. W., Velculescu, V. E., and Riggins, G. J. (2011). Sodium ion channel mutations in glioblastoma patients correlate with shorter survival. *Molecular cancer, 10*(1), 17.

Junttila, T. T., Sundvall, M., Lundin, M., Lundin, J., Tanner, M., Härkönen, P., et al. (2005). Cleavable ErbB4 Isoform in Estrogen Receptor–Regulated Growth of Breast Cancer Cells. *Cancer Research, 65*(4), 1384-1393.

Juratli, T. A., McCabe, D., Nayyar, N., Williams, E. A., Silverman, I. M., Tummala, S. S., et al. (2018). DMD genomic deletions characterize a subset of progressive/higher-grade meningiomas with poor outcome. *Acta Neuropathologica, 136*(5), 779-792.

Kaiser, J. (2006). First Pass at Cancer Genome Reveals Complex Landscape. *Science, 313*(5792), 1370-1370.

Kammerer, R., Buchner, A., Palluch, P., Pongratz, T., Oboukhovskij, K., Beyer, W., et al. (2011). Induction of Immune Mediators in Glioma and Prostate Cancer Cells by Non-Lethal Photodynamic Therapy. *PLOS ONE, 6*(6), e21834.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2011). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research, 40*(D1), D109-D114.

Kang, C., Huo, Y., Xin, L., Tian, B., and Yu, B. (2019). Feature selection and tumor classification for microarray data using relaxed Lasso and generalized multi-class support vector machine. *Journal of Theoretical Biology, 463*, 77-91.

Kaplan, E. L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association, 53*(282), 457-481.

Kappadakunnel, M., Eskin, A., Dong, J., Nelson, S. F., Mischel, P. S., Liau, L. M., et al. (2010). Stem cell associated gene expression in glioblastoma multiforme: relationship to survival and the subventricular zone. *Journal of neuro-oncology, 96*(3), 359-367.

Karagoz, K., Sinha, R., and Arga, K. Y. (2015). Triple negative breast cancer: a multi-omics network discovery strategy for candidate targets and driving pathways. *Omics: a journal of integrative biology, 19*(2), 115-130.

Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol, 9*(10), 770-780.

Karolchik, D., Hinrichs, A. S., and Kent, W. J. (2009). The UCSC Genome Browser. *Current protocols in bioinformatics, Chapter 1*, Unit1.4-Unit1.4.

Katt, M. E., Placone, A. L., Wong, A. D., Xu, Z. S., and Searson, P. C. (2016). In vitro tumor models: advantages, disadvantages, variables, and selecting the right platform. *Frontiers in bioengineering and biotechnology, 4*, 12.

Kenny, P. A., and Bissell, M. J. (2007). Targeting TACE-dependent EGFR ligand shedding in breast cancer. *The Journal of Clinical Investigation, 117*(2), 337-345.

Khaled, W. T., Lee, S. C., Stingl, J., Chen, X., Ali, H. R., Rueda, O. M., et al. (2015). BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. *Nature communications, 6*, ncomms6987.

Kim, C., Tang, G., Pogue-Geile, K. L., Costantino, J. P., Baehner, F. L., Baker, J., et al. (2011). Estrogen receptor (ESR1) mRNA expression and benefit from tamoxifen in the treatment and prevention of estrogen receptor-positive breast

cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 29*(31), 4160-4167.

King, S. I., Purdie, C. A., Bray, S. E., Quinlan, P. R., Jordan, L. B., Thompson, A. M., et al. (2012). Immunohistochemical detection of Polo-like kinase-1 (PLK1) in primary breast cancer is associated with TP53mutation and poor clinical outcome. *Breast Cancer Research, 14*(2), R40.

Király, A., Abonyi, J., Laiho, A., and Gyenesei, A. (2012, 10-10 Dec. 2012). *Biclustering of High-throughput Gene Expression Data with Bicluster Miner.* Paper presented at the 2012 IEEE 12th International Conference on Data Mining Workshops, 131-138.

Kleftogiannis, D., Korfiati, A., Theofilatos, K., Likothanassis, S., Tsakalidis, A., and Mavroudi, S. (2013). Where we stand, where we are moving: Surveying computational techniques for identifying miRNA genes and uncovering their regulatory role. *Journal of biomedical informatics, 46*(3), 563-573.

Klopocki, E., Kristiansen, G., Wild, P. J., Klaman, I., Castanos-Velez, E., Singer, G., et al. (2004). Loss of SFRP1 is associated with breast cancer progression and poor prognosis in early stage tumors. *International journal of oncology, 25*(3), 641-649.

Kluger, Y., Basri, R., Chang, J. T., and Gerstein, M. (2003). Spectral Biclustering of Microarray Data: Coclustering Genes and Conditions. *Genome Research, 13*(4), 703-716.

Kok, M., Koornstra, R. H., Margarido, T. C., Fles, R., Armstrong, N. J., Linn, S. C., et al. (2009). Mammosphere-derived gene set predicts outcome in patients with ER-positive breast cancer. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland, 218*(3), 316-326.

Kong, Y., and Yu, T. (2018). A Deep Neural Network Model using Random Forest to Extract Feature Representation for Gene Expression Data Classification. *Scientific reports, 8*(1), 16477.

Koren, A., Polak, P., Nemesh, J., Michaelson, Jacob J., Sebat, J., Sunyaev, Shamil R., et al. (2012). Differential Relationship of DNA Replication Timing to Different Forms of Human Mutation and Variation. *The American Journal of Human Genetics, 91*(6), 1033-1040.

Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering, 160*, 3-24.

Krijgsman, O., Roepman, P., Zwart, W., Carroll, J. S., Tian, S., de Snoo, F. A., et al. (2012). A diagnostic gene profile for molecular subtyping of breast cancer associated with treatment response. *Breast Cancer Research and Treatment, 133*(1), 37-47.

Kroes, R. A., Dawson, G., and Moskal, J. R. (2007). Focused microarray analysis of glyco-gene expression in human glioblastomas. *Journal of Neurochemistry, 103*(s1), 14-24.

Kruthika, B. S., Jain, R., Arivazhagan, A., Bharath, R. D., Yasha, T. C., Kondaiah, P., et al. (2018). Transcriptome profiling reveals PDZ binding kinase as a novel biomarker in peritumoral brain zone of glioblastoma. *Journal of Neuro-Oncology*.

Lacroix, M. (2006). Significance, detection and markers of disseminated breast cancer cells. *13*(4), 1033.

Laczko, R., Szauter, K. M., Jansen, M. K., Hollosi, P., Muranyi, M., Molnar, J., et al. (2007). Active lysyl oxidase (LOX) correlates with focal adhesion kinase

(FAK)/paxillin activation and migration in invasive astrocytes. *Neuropathology and Applied Neurobiology, 33*(6), 631-643.

Laé, M., La Rosa, P., Mandel, J., Reyal, F., Hupé, P., Terrier, P., et al. (2016). Whole-genome profiling helps to classify phyllodes tumours of the breast. *Journal of Clinical Pathology, 69*(12), 1081-1087.

Laible, M., Schlombs, K., Kaiser, K., Veltrup, E., Herlein, S., Lakis, S., et al. (2016). Technical validation of an RT-qPCR in vitro diagnostic test system for the determination of breast cancer molecular subtypes by quantification of ERBB2, ESR1, PGR and MKI67 mRNA levels from formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer, 16*(1), 398.

Lakizadeh, A., and Jalili, S. (2016). BiCAMWI: A Genetic-Based Biclustering Algorithm for Detecting Dynamic Protein Complexes. *PLOS ONE, 11*(7), e0159923.

LaMarca, H. L., and Rosen, J. M. (2007). Estrogen regulation of mammary gland development and breast cancer: amphiregulin takes center stage. *Breast Cancer Research, 9*(4), 304.

Lappano, R., Jacquot, Y., and Maggiolini, M. (2018). GPCR Modulation in Breast Cancer. *International journal of molecular sciences, 19*(12), 3840.

Lavrov, A. V., Chelysheva, E. Y., Smirnikhina, S. A., Shukhov, O. A., Turkina, A. G., Adilgereeva, E. P., et al. (2016). Frequent variations in cancer-related genes may play prognostic role in treatment of patients with chronic myeloid leukemia. *BMC Genetics, 17*(1), S14.

Lazar, C., Taminau, J., Meganck, S., Steenhoff, D., Coletta, A., Molter, C., et al. (2012). A survey on filter techniques for feature selection in gene expression microarray analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 9*(4), 1106-1119.

Lazzeroni, L., and Owen, A. (2002). Plaid models for gene expression data. *Statistica sinica*, 61-86.

Le, T. D., Liu, L., Zhang, J., Liu, B., and Li, J. (2015). From miRNA regulation to miRNA–TF co-regulation: computational approaches and challenges. *Briefings in Bioinformatics, 16*(3), 475-496.

Lea, R. W., Dawson, T., Martinez-Moreno, C. G., El-Abry, N., and Harvey, S. (2015). Growth hormone and cancer: GH production and action in glioma? *General and Comparative Endocrinology, 220*, 119-123.

Lee, C.-C., Chen, P.-H., Ho, K.-H., Shih, C.-M., Cheng, C.-H., Lin, C.-W., et al. (2017). The microRNA-302b-inhibited insulin-like growth factor-binding protein 2 signaling pathway induces glioma cell apoptosis by targeting nuclear factor IA. *PLOS ONE, 12*(3), e0173890.

Lee, E.-K., Han, G.-Y., Park, H. W., Song, Y.-J., and Kim, C.-W. (2010a). Transgelin promotes migration and invasion of cancer stem cells. *Journal of proteome research, 9*(10), 5108-5117.

Lee, H. J., Lee, J.-J., Song, I. H., Park, I. A., Kang, J., Yu, J. H., et al. (2015a). Prognostic and predictive value of NanoString-based immune-related gene signatures in a neoadjuvant setting of triple-negative breast cancer: relationship to tumor-infiltrating lymphocytes. *Breast Cancer Research and Treatment, 151*(3), 619-627.

Lee, J.-K., Chang, N., Yoon, Y., Yang, H., Cho, H., Kim, E., et al. (2015b). USP1 targeting impedes GBM growth by inhibiting stem cell maintenance and radioresistance. *Neuro-oncology, 18*(1), 37-47.

Lee, J. W., Lee, J. B., Park, M., and Song, S. H. (2005). An extensive comparison of recent classification tools applied to microarray data. *Computational Statistics & Data Analysis, 48*(4), 869-885.

Lee, K.-M., Park, S.-K., Kim, S.-U., Doll, M. A., Yoo, K.-Y., Ahn, S.-H., et al. (2003). N-acetyltransferase (NAT1, NAT2) and glutathione S-transferase (GSTM1, GSTT1) polymorphisms in breast cancer. *Cancer Letters, 196*(2), 179-186.

Lee, S. E., Simons, S. B., Heldt, S. A., Zhao, M., Schroeder, J. P., Vellano, C. P., et al. (2010b). RGS14 is a natural suppressor of both synaptic plasticity in CA2 neurons and hippocampal-based learning and memory. *Proceedings of the National Academy of Sciences, 107*(39), 16994-16998.

Lehmann, B. D., Bauer, J. A., Chen, X., Sanders, M. E., Chakravarthy, A. B., Shyr, Y., et al. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of Clinical Investigation, 121*(7), 2750-2767.

Lewis-Wambi, J. S., Cunliffe, H. E., Kim, H. R., Willis, A. L., and Jordan, V. C. (2008). Overexpression of CEACAM6 promotes migration and invasion of oestrogen-deprived breast cancer cells. *European journal of cancer (Oxford, England : 1990), 44*(12), 1770-1779.

Lewis, D. D., Yang, Y., Rose, T. G., and Li, F. (2004). Rcv1: A new benchmark collection for text categorization research. *Journal of machine learning research, 5*(Apr), 361-397.

Li, A., Liu, Z., Li, M., Zhou, S., Xu, Y., Xiao, Y., et al. (2016a). HDAC5, a potential therapeutic target and prognostic biomarker, promotes proliferation, invasion and migration in human breast cancer. *Oncotarget, 7*(25), 37966-37978.

Li, G., Ma, Q., Tang, H., Paterson, A. H., and Xu, Y. (2009). QUBIC: a qualitative biclustering algorithm for analyses of gene expression data. *Nucleic Acids Research, 37*(15), e101-e101.

Li, M., Xiao, A., Floyd, D., Olmez, I., Lee, J., Godlewski, J., et al. (2017). CDK4/6 inhibition is more active against the glioblastoma proneural subtype. *Oncotarget, 8*(33), 55319.

Li, X., Cowell, J. K., and Sossey-Alaoui, K. (2004). CLCA2 tumour suppressor gene in 1p31 is epigenetically regulated in breast cancer. *Oncogene, 23*, 1474.

Li, X., Feltus, F. A., Sun, X., Wang, Z., and Luo, F. (2010, 18-21 Dec. 2010). *A non-parameter Ising model for network-based identification of differentially expressed genes in recurrent breast cancer patients.* Paper presented at the 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 214-217.

Li, X., Liu, Y., Granberg, K. J., Wang, Q., Moore, L. M., Ji, P., et al. (2015). Two mature products of MIR-491 coordinate to suppress key cancer hallmarks in glioblastoma. *Oncogene, 34*(13), 1619.

Li, Y.-R., and Yang, W.-X. (2016). Myosins as fundamental components during tumorigenesis: diverse and indispensable. *Oncotarget, 7*(29), 46785-46812.

Li, Y., Guo, M., Fu, Z., Wang, P., Zhang, Y., Gao, Y., et al. (2016b). Immunoglobulin superfamily genes are novel prognostic biomarkers for breast cancer. *Oncotarget, 8*(2), 2444-2456.

Li, Y., Luo, Y., Wang, X., Shen, S., Yu, H., Yang, J., et al. (2012). Tumor suppressor gene NGX6 induces changes in protein expression profiles in colon cancer HT-29 cells. *Acta Biochimica et Biophysica Sinica, 44*(7), 584-590.

Li, Y., Min, W., Li, M., Han, G., Dai, D., Zhang, L., et al. (2016c). Identification of hub genes and regulatory factors of glioblastoma multiforme subgroups by RNA-seq data analysis. *International journal of molecular medicine, 38*(4), 1170-1178.

Liang, M., Cowley, A. W., and Greene, A. S. (2004). High throughput gene expression profiling: a molecular approach to integrative physiology. *The Journal of Physiology, 554*(Pt 1), 22-30.

Liang, Y., Liu, C., Luan, X.-Z., Leung, K.-S., Chan, T.-M., Xu, Z.-B., et al. (2013). Sparse logistic regression with a L 1/2 penalty for gene selection in cancer classification. *BMC bioinformatics, 14*(1), 198.

Libreros, S., Garcia-Areas, R., and Iragavarapu-Charyulu, V. (2013). CHI3L1 plays a role in cancer through enhanced production of pro-inflammatory/pro-tumorigenic and angiogenic factors. *Immunologic Research, 57*(1), 99-105.

Lim, S.-O., Li, C.-W., Xia, W., Lee, H.-H., Chang, S.-S., Shen, J., et al. (2016). EGFR Signaling Enhances Aerobic Glycolysis in Triple-Negative Breast Cancer Cells to Promote Tumor Growth and Immune Escape. *Cancer Research*.

Lin, B., Lee, H., Yoon, J.-G., Madan, A., Wayner, E., Tonning, S., et al. (2015). Global analysis of H3K4me3 and H3K27me3 profiles in glioblastoma stem cells and identification of SLC17A7 as a bivalent tumor suppressor gene. *Oncotarget, 6*(7), 5369-5381.

Lin, Y.-C., Lee, Y.-C., Li, L.-H., Cheng, C.-J., and Yang, R.-B. (2014). Tumor suppressor SCUBE2 inhibits breast-cancer cell migration and invasion through the reversal of epithelial–mesenchymal transition. *Journal of Cell Science, 127*(1), 85-100.

Lind, G. E., Danielsen, S. A., Ahlquist, T., Merok, M. A., Andresen, K., Skotheim, R. I., et al. (2011). Identification of an epigenetic biomarker panel with high sensitivity and specificity for colorectal cancer and adenomas. *Molecular cancer, 10*(1), 85.

Lison, P. (2015). An introduction to machine learning. *Language Technology Group (LTG), 1, 35*.

Liu, F., Hon, Gary C., Villa, Genaro R., Turner, Kristen M., Ikegami, S., Yang, H., et al. (2015a). EGFR Mutation Promotes Glioblastoma through Epigenome and Transcription Factor Network Remodeling. *Molecular Cell, 60*(2), 307-318.

Liu, L., Xu, Q., Cheng, L., Ma, C., Xiao, L., Xu, D., et al. (2015b). NPY1R is a novel peripheral blood marker predictive of metastasis and prognosis in breast cancer patients. *Oncology letters, 9*(2), 891-896.

Liu, Q., Nguyen, D. H., Dong, Q., Shitaku, P., Chung, K., Liu, O. Y., et al. (2009). Molecular properties of CD133+ glioblastoma stem cells derived from treatment-refractory recurrent brain tumors. *Journal of Neuro-Oncology, 94*(1), 1-19.

Liu, R.-Z., Li, S., Garcia, E., Glubrecht, D. D., Yin Poon, H., Easaw, J. C., et al. (2016). Association between cytoplasmic CRABP2, altered retinoic acid signaling, and poor prognosis in glioblastoma. *Glia, 64*(6), 963-976.

Liu, S., Xu, C., Zhang, Y., Liu, J., Yu, B., Liu, X., et al. (2018). Feature selection of gene expression data for Cancer classification using double RBF-kernels. *BMC Bioinformatics, 19*(1), 396.

Liu, X., Gao, Y., Zhao, B., Li, X., Lu, Y., Zhang, J., et al. (2015c). Discovery of microarray-identified genes associated with ovarian cancer progression. *International journal of oncology, 46*(6), 2467-2478.

Liu, X., and Wang, L. (2006). Computing the maximum similarity bi-clusters of gene expression data. *Bioinformatics, 23*(1), 50-56.

Liu, Y., Gu, Q., Hou, J., Han, J., and Ma, J. (2014). A network-assisted co-clustering algorithm to discover cancer subtypes based on gene expression. *BMC Bioinformatics, 15*(1), 37.

Liu, Y., Wu, S., Liu, Z., and Chao, H. (2017). A fuzzy co-clustering algorithm for biomedical data. *PLOS ONE, 12*(4), e0176536.

Llombart-Cussac, A., Cortés, J., Paré, L., Galván, P., Bermejo, B., Martínez, N., et al. (2017). HER2-enriched subtype as a predictor of pathological complete response following trastuzumab and lapatinib without chemotherapy in early-stage HER2-positive breast cancer (PAMELA): an open-label, single-group, multicentre, phase 2 trial. *The Lancet Oncology, 18*(4), 545-554.

Locatelli, M., Ferrero, S., Martinelli Boneschi, F., Boiocchi, L., Zavanone, M., Maria Gaini, S., et al. (2013). The long pentraxin PTX3 as a correlate of cancer-related inflammation and prognosis of malignancy in gliomas. *Journal of Neuroimmunology, 260*(1), 99-106.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., et al. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology, 14*, 1675.

Lodish, H., Darnell, J. E., Berk, A., Kaiser, C. A., Krieger, M., Scott, M. P., et al. (2008). *Molecular cell biology*: Macmillan.

Loussouarn, D., Campion, L., Leclair, F., Campone, M., Charbonnel, C., Ricolleau, G., et al. (2009). Validation of UBE2C protein as a prognostic marker in node-positive breast cancer. *British journal of cancer, 101*(1), 166.

Luker, K. E., and Luker, G. D. (2006). Functions of CXCL12 and CXCR4 in breast cancer. *Cancer letters, 238*(1), 30-41.

Lyons, M. J., Duron, R., Molinero, I., Sangiuolo, F., and Holden, K. R. (2010). Novel CLCN1 mutation in carbamazepine-responsive myotonia congenita. *Pediatric neurology, 42*(5), 365-368.

Ma, S., and Huang, J. (2008). Penalized feature selection and classification in bioinformatics. *Briefings in bioinformatics, 9*(5), 392-403.

Ma, S., Huang, Y., Huang, J., and Fang, K. (2012). Gene network-based cancer prognosis analysis with sparse boosting. *Genetics Research, 94*(04), 205-221.

Ma, X.-J., Dahiya, S., Richardson, E., Erlander, M., and Sgroi, D. C. (2009). Gene expression profiling of the tumor microenvironment during breast cancer progression. *Breast Cancer Research, 11*(1), R7.

Madeira, S. C., and Oliveira, A. L. (2004). Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 1*(1), 24-45.

Maldonado, S., Weber, R., and Basak, J. (2011). Simultaneous feature selection and classification using kernel-penalized support vector machines. *Information Sciences, 181*(1), 115-128.

Malorni, L., Piazza, S., Ciani, Y., Guarducci, C., Bonechi, M., Biagioni, C., et al. (2016). A gene expression signature of retinoblastoma loss-of-function is a predictive biomarker of resistance to palbociclib in breast cancer cell lines and is prognostic in patients with ER positive early breast cancer. *Oncotarget, 7*(42), 68012-68022.

Mao, Y., and Yang, Y. (2019). A Wrapper Feature Subset Selection Method Based on Randomized Search and Multilayer Structure. *BioMed Research International, 2019*.

Marques, E., Englund, J. I., Tervonen, T. A., Virkunen, E., Laakso, M., Myllynen, M., et al. (2015). Par6G suppresses cell proliferation and is targeted by loss-of-function mutations in multiple cancers. *Oncogene, 35*, 1386.

Martin, K. J., Patrick, D. R., Bissell, M. J., and Fournier, M. V. (2008). Prognostic Breast Cancer Signature Identified from 3D Culture Model Accurately Predicts Clinical Outcome across Independent Datasets. *PLOS ONE, 3*(8), e2994.

Masili-Oku, S. M., Almeida, B. G. L. d., Bacchi, C. E., Filassi, J. R., Baracat, E. C., and Carvalho, F. M. (2017). Lymphocyte-predominant triple-negative breast carcinomas in premenopausal patients: Lower expression of basal immunohistochemical markers. *The Breast, 31*, 34-39.

May, F. E., and Westley, B. R. (2015). TFF3 is a valuable predictive biomarker of endocrine response in metastatic breast cancer. *Endocrine-related cancer*, ERC-15-0129.

McGrath, J. E. (2015). *Genetic Screen Identifies Candidate Breast Cancer Tumor Dormancy Suppressor Genes Using Cellecta's Decipher Pooled shRNA Libraries.* State University of New York at Buffalo.

McGuire, A., Lowery, A. J., Kell, M. R., Kerin, M. J., and Sweeney, K. J. (2017). Locoregional recurrence following breast cancer surgery in the trastuzumab era: A systematic review by subtype. *Annals of surgical oncology, 24*(11), 3124-3132.

Mefford, D., and Mefford, J. (2012). Stromal Genes Add Prognostic Information to Proliferation and Histoclinical Markers: A Basis for the Next Generation of Breast Cancer Gene Signatures. *PLOS ONE, 7*(6), e37646.

Mercier, I., Casimiro, M. C., Wang, C., Rosenberg, A. L., Quong, J., Minkeu, A., et al. (2008). Human breast cancer-associated fibroblasts (CAFs) show caveolin-1 down-regulation and RB tumor suppressor functional inactivation: Implications for the response to hormonal therapy. *Cancer Biology & Therapy, 7*(8), 1212-1225.

Metzger-Filho, O., Sun, Z., Viale, G., Price, K. N., Crivellari, D., Snyder, R. D., et al. (2013). Patterns of recurrence and outcome according to breast cancer subtypes in lymph node–negative disease: results from International Breast Cancer Study Group Trials VIII and IX. *Journal of Clinical Oncology, 31*(25), 3083.

Michalak, E. M. (2017). The mammary stem cell field wakes up to hibernating cells. *Stem cell investigation, 4*, 45-45.

Michiels, S., Koscielny, S., and Hill, C. (2007). Interpretation of microarray data in cancer. *Br J Cancer, 96*(8), 1155-1158.

Milinkovic, V., Bankovic, J., Rakic, M., Stankovic, T., Skender-Gazibara, M., Ruzdijic, S., et al. (2013). Identification of Novel Genetic Alterations in Samples of Malignant Glioma Patients. *PLOS ONE, 8*(12), e82108.

Miselis, N. R., Linn, D., Restaino, C., Baral, T., Xia, J., Ueda, R., et al. (2017). Abstract 577: Antagonism of the co-inhibitory receptor BTLA enhances efficacy of anti-PD-1 treatment in murine syngeneic tumor models. *Cancer Research, 77*(13 Supplement), 577-577.

Mishra, D. (2012). Discovery of Overlapping Pattern Biclusters from Gene Expression Data using Hash based PSO. *Procedia Technology, 4*(Supplement C), 390-394.

Mohankumar, K. M., Currle, D. S., White, E., Boulos, N., Dapper, J., Eden, C., et al. (2015). An in vivo screen identifies ependymoma oncogenes and tumor-suppressor genes. *Nature Genetics, 47*, 878.

Momenzadeh, M., Sehhati, M., and Rabbani, H. (2019). A novel feature selection method for microarray data classification based on hidden Markov model. *Journal of biomedical informatics, 95*, 103213.

Morales, A. G., Pezuk, J. A., Brassesco, M. S., de Oliveira, J. C., de Paula Queiroz, R. G., Machado, H. R., et al. (2013). BUB1 and BUBR1 inhibition decreases proliferation and colony formation, and enhances radiation sensitivity in pediatric glioblastoma cells. *Child's Nervous System, 29*(12), 2241-2248.

Motley, A. M., Ward, G. P., and Hettema, E. H. (2008). Dnm1p-dependent peroxisome fission requires Caf4p, Mdv1p and Fis1p. *Journal of cell science, 121*(10), 1633-1640.

Mukhopadhyay, A., Maulik, U., and Bandyopadhyay, S. (2009). A novel coherence measure for discovering scaling biclusters from gene expression data. *Journal of bioinformatics and computational biology, 7*(05), 853-868.

Muntasell, A., Cabo, M., Servitja, S., Tusquets, I., Martínez-García, M., Rovira, A., et al. (2017). Interplay between Natural Killer Cells and Anti-HER2 Antibodies: Perspectives for Breast Cancer Immunotherapy. *Frontiers in Immunology, 8*(1544).

Murray, A., Madden, S. F., Synnott, N. C., Klinger, R., O'Connor, D., O'Donovan, N., et al. (2017). Vitamin D receptor as a target for breast cancer therapy. *24*(4), 181.

Naderi, A., Teschendorff, A. E., Barbosa-Morais, N. L., Pinder, S. E., Green, A. R., Powe, D. G., et al. (2006). A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene, 26*, 1507.

Nam, S., Chang, H. R., Jung, H. R., Gim, Y., Kim, N. Y., Grailhe, R., et al. (2015). A pathway-based approach for identifying biomarkers of tumor progression to trastuzumab-resistant breast cancer. *Cancer Letters, 356*(2, Part B), 880-890.

Naresh, A., Long, W., Vidal, G. A., Wimley, W. C., Marrero, L., Sartor, C. I., et al. (2006). The ERBB4/HER4 Intracellular Domain 4ICD Is a BH3-Only Protein Promoting Apoptosis of Breast Cancer Cells. *Cancer Research, 66*(12), 6412-6420.

Natarajan, J., and Ganapathy, J. (2007). Functional gene clustering via gene annotation sentences, MeSH and GO keywords from biomedical literature. *Bioinformation, 2*(5), 185.

Natsoulis, G., El Ghaoui, L., Lanckriet, G. R. G., Tolley, A. M., Leroy, F., Dunlea, S., et al. (2005). Classification of a large microarray data set: algorithm comparison and analysis of drug signatures. *Genome research, 15*(5), 724-736.

Nayak, S. B., Bhat, V. R., Upadhyay, D., and Udupa, S. L. (2003). Copper and ceruloplasmin status in serum of prostate and colon cancer patients. *Indian journal of physiology and pharmacology, 47*(1), 108-110.

Need, E. F., Selth, L. A., Harris, T. J., Birrell, S. N., Tilley, W. D., and Buchanan, G. (2012). Research Resource: Interplay between the Genomic and Transcriptional Networks of Androgen Receptor and Estrogen Receptor α in Luminal Breast Cancer Cells. *Molecular Endocrinology, 26*(11), 1941-1952.

Network, C. G. A. (2012). Comprehensive molecular portraits of human breast tumours. *Nature, 490*(7418), 61.

Network, C. G. A. R. (2015). Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *New England Journal of Medicine, 372*(26), 2481-2498.

Nevo, I., Woolard, K., Cam, M., Li, A., Webster, J. D., Kotliarov, Y., et al. (2014). Identification of molecular pathways facilitating glioma cell invasion in situ. *PloS one, 9*(11), e111783.

Nguyen, C., Wang, Y., and Nguyen, H. N. (2013). Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic.

Nguyen, H., and Bui, X.-N. (2019). Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. *Natural Resources Research, 28*(3), 893-907.

Ni, M., Chen, Y., Lim, E., Wimberly, H., Bailey, Shannon T., Imai, Y., et al. (2011). Targeting Androgen Receptor in Estrogen Receptor-Negative Breast Cancer. *Cancer Cell, 20*(1), 119-131.

Nikkilä, J., Törönen, P., Kaski, S., Venna, J., Castrén, E., and Wong, G. (2002). Analysis and visualization of gene expression data using self-organizing maps. *Neural networks, 15*(8-9), 953-966.

Nitze, I., Schulthess, U., and Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proc. of the 4th GEOBIA*, 7-9.

O'Connor, C. M., Adams, J. U., and Fairman, J. (2010). Essentials of cell biology. *Cambridge, MA: NPG Education, 1*.

Odibat, O., and Reddy, C. K. (2014). Efficient mining of discriminative co-clusters from gene expression data. *Knowledge and Information Systems, 41*(3), 667-696.

Oehler, V. G., Yeung, K. Y., Choi, Y. E., Bumgarner, R. E., Raftery, A. E., and Radich, J. P. (2009). *The derivation of diagnostic markers of chronic myeloid leukemia progression from microarray data* (Vol. 114).

Oghabian, A., Kilpinen, S., Hautaniemi, S., and Czeizler, E. (2014). Biclustering methods: biological relevance and application in gene expression analysis. *PloS one, 9*(3), e90801.

Olar, A., and Aldape, K. D. (2014). Using the molecular classification of glioblastoma to inform personalized treatment. *The Journal of pathology, 232*(2), 165-177.

Ono, Y., Chiba, S., Yano, H., Nakayama, N., Saio, M., Tsuruma, K., et al. (2016). Glycoprotein nonmetastatic melanoma protein B (GPNMB) promotes the progression of brain glioblastoma via Na+/K+-ATPase. *Biochemical and Biophysical Research Communications, 481*(1), 7-12.

Organization, W. H. (2019). World Health Organization-WHO.

Orr, B. A., Haffner, M. C., Nelson, W. G., Yegnasubramanian, S., and Eberhart, C. G. (2012). Decreased 5-Hydroxymethylcytosine Is Associated with Neural Progenitor Phenotype in Normal Brain and Shorter Survival in Malignant Glioma. *PLOS ONE, 7*(7), e41036.

Oyelade, J., Isewon, I., Oladipupo, F., Aromolaran, O., Uwoghiren, E., Ameh, F., et al. (2016). Clustering Algorithms: Their Application to Gene Expression Data. *Bioinformatics and biology insights, 10*, 237-253.

Özeş, A. R., Wang, Y., Zong, X., Fang, F., Pilrose, J., and Nephew, K. P. (2017). Therapeutic targeting using tumor specific peptides inhibits long non-coding

RNA HOTAIR activity in ovarian and breast cancer. *Scientific Reports, 7*(1), 894.
Padfield, E., Ellis, H. P., and Kurian, K. M. (2015). Current Therapeutic Advances Targeting EGFR and EGFRvIII in Glioblastoma. *Frontiers in Oncology, 5*(5).
Padilha, V. A., and Campello, R. J. G. B. (2017). A systematic comparative evaluation of biclustering techniques. *BMC Bioinformatics, 18*(1), 55.
Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*: Stanford InfoLabo. Document Number)
Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., et al. (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol, 24*(23), 3726-3734.
Palaniappan, A., Ramar, K., and Ramalingam, S. (2016). Computational Identification of Novel Stage-Specific Biomarkers in Colorectal Cancer Progression. *PLOS ONE, 11*(5), e0156665.
Pandey, P., Sliker, B., Peters, H. L., Tuli, A., Herskovitz, J., Smits, K., et al. (2016). Amyloid precursor protein and amyloid precursor-like protein 2 in cancer. *Oncotarget, 7*(15), 19430-19444.
Papadopoulos, K., Wattanaarsakit, P., Prasongchean, W., and Narain, R. (2016). Gene therapies in clinical trials. In *Polymers and Nanomaterials for Gene Therapy* (pp. 231-256): Elsevier.
Paplomata, E., and O'Regan, R. (2014). The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. *Therapeutic advances in medical oncology, 6*(4), 154-166.
Parry, J. J., Andrews, R., and Rogers, B. E. (2007). MicroPET imaging of breast cancer using radiolabeled bombesin analogs targeting the gastrin-releasing peptide receptor. *Breast cancer research and treatment, 101*(2), 175-183.
Patel, V. N., Gokulrangan, G., Chowdhury, S. A., Chen, Y., Sloan, A. E., Koyutürk, M., et al. (2013). Network Signatures of Survival in Glioblastoma Multiforme. *PLOS Computational Biology, 9*(9), e1003237.
Pati, S. K., Das, A. K., and Ghosh, A. (2013). *Gene selection using multi-objective genetic algorithm integrating cellular automata and rough set theory.* Paper presented at the International Conference on Swarm, Evolutionary, and Memetic Computing, 144-155.
Patil, S. S., Railkar, R., Swain, M., Atreya, H. S., Dighe, R. R., and Kondaiah, P. (2015). Novel anti IGFBP2 single chain variable fragment inhibits glioma cell migration and invasion. *Journal of Neuro-Oncology, 123*(2), 225-235.
Paul, D., Su, R., Romain, M., Sébastien, V., Pierre, V., and Isabelle, G. (2017). Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier. *Computerized Medical Imaging and Graphics, 60*, 42-49.
Pellikainen, J. M., and Kosma, V.-M. (2007). Activator protein-2 in carcinogenesis with a special reference to breast cancer—A mini review. *International Journal of Cancer, 120*(10), 2061-2067.
Pennisi, E. (2001). The Human Genome. *Science, 291*(5507), 1177-1180.
Perry, C., Agarwal, D., Abdel-Fatah, T. M. A., Lourdusamy, A., Grundy, R., Auer, D. T., et al. (2014). Dissecting DNA repair in adult high grade gliomas for patient stratification in the post-genomic era. *Oncotarget, 5*(14), 5764-5781.
Perry, R. J., Mast, F. D., and Rachubinski, R. A. (2009). Endoplasmic reticulum-associated secretory proteins Sec20p, Sec39p, and Dsl1p are involved in peroxisome biogenesis. *Eukaryotic cell, 8*(6), 830-843.

Perscheid, C. (2020). Integrative biomarker detection on high-dimensional gene expression data sets: a survey on prior knowledge approaches. *Briefings in bioinformatics*.

Pesce, E.-R., Blatch, G. L., and Edkins, A. L. (2015). Hsp40 Co-chaperones as Drug Targets: Towards the Development of Specific Inhibitors. In *Heat Shock Protein Inhibitors* (pp. 163-195): Springer.

Petermann, A., Stampnik, Y., Cui, Y., Morrison, H., Pachow, D., Kliese, N., et al. (2015). Deficiency of the protein-tyrosine phosphatase DEP-1/PTPRJ promotes matrix metalloproteinase-9 expression in meningioma cells. *Journal of Neuro-Oncology, 122*(3), 451-459.

Peterson, L. E., and Kovyrshina, T. (2018). DNA Repair Gene Expression Adjusted by the PCNA Metagene Predicts Survival in Multiple Cancers. *bioRxiv*, 446377.

Phillips, H. S., Hains, J. M., Laramee, G. R., Rosenthal, A., and Winslow, J. W. (1990). Widespread expression of BDNF but not NT3 by target areas of basal forebrain cholinergic neurons. *Science, 250*(4978), 290-294.

Phuong, T. M., and Nhung, N. P. (2013). Predicting gene function using similarity learning. *BMC genomics, 14*(4), S4.

Piatetsky-Shapiro, G., Khabaza, T., and Ramaswamy, S. (2003). *Capturing best practice for microarray gene expression data analysis*. Paper presented at the Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.

Pindah, W., Nordin, S., Seman, A., and Said, M. S. M. (2015, 21-23 April 2015). *Review of dimensionality reduction techniques using clustering algorithm in reconstruction of gene regulatory networks.* Paper presented at the 2015 International Conference on Computer, Communications, and Control Technology (I4CT), 172-176.

Pirooznia, M., Yang, J. Y., Yang, M. Q., and Deng, Y. (2008). A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics, 9*(1), S13.

Plourde, K. V., Labrie, Y., Desjardins, S., Belleau, P., Ouellette, G., Durocher, F., et al. (2013). Analysis of ZNF350/ZBRK1 promoter variants and breast cancer susceptibility in non-BRCA1/2 French Canadian breast cancer families. *Journal of Human Genetics, 58*(2), 59-66.

Pontes, B., Giráldez, R., and Aguilar-Ruiz, J. S. (2015). Biclustering on expression data: A review. *Journal of Biomedical Informatics, 57*(Supplement C), 163-180.

Popov, V. M., Zhou, J., Shirley, L. A., Quong, J., Yeow, W.-S., Wright, J. A., et al. (2009). The Cell Fate Determination Factor DACH1 Is Expressed in Estrogen Receptor-α–Positive Breast Cancer and Represses Estrogen Receptor-α Signaling. *Cancer Research, 69*(14), 5752-5760.

Prasad, A., Paruchuri, V., Preet, A., Latif, F., and Ganju, R. K. (2008). Slit-2 induces a tumor-suppressive effect by regulating β-catenin in breast cancer cells. *Journal of Biological Chemistry, 283*(39), 26624-26633.

Prieto-Dominguez, N., Shull, A. Y., and Teng, Y. (2018). Making way for suppressing the FGF19/FGFR4 axis in cancer. *Future Medicinal Chemistry, 10*(20), 2457-2469.

Pukkila, P. J. (2001). Molecular biology: the central dogma. *e LS*.

Qian, B., Deng, Y., Im, J. H., Muschel, R. J., Zou, Y., Li, J., et al. (2009). A Distinct Macrophage Population Mediates Metastatic Breast Cancer Cell Extravasation, Establishment and Growth. *PLOS ONE, 4*(8), e6562.

Quintero, M., Adamoski, D., Reis, L. M. d., Ascenção, C. F. R., Oliveira, K. R. S. d., Gonçalves, K. d. A., et al. (2017). Guanylate-binding protein-1 is a potential new therapeutic target for triple-negative breast cancer. *BMC Cancer, 17*(1), 727.

Ray, P. S., Wang, J., Qu, Y., Sim, M.-S., Shamonki, J., Bagaria, S. P., et al. (2010). FOXC1 Is a Potential Prognostic Biomarker with Functional Significance in Basal-like Breast Cancer. *Cancer Research*, 0008-5472.CAN-0009-4120.

Ribas, G., Milne, R. L., Gonzalez-Neira, A., and Benítez, J. (2008). Haplotype patterns in cancer-related genes with long-range linkage disequilibrium: no evidence of association with breast cancer or positive selection. *European Journal of Human Genetics, 16*(2), 252.

Rittschof, C. C., and Robinson, G. E. (2016). Chapter Five - Behavioral Genetic Toolkits: Toward the Evolutionary Origins of Complex Phenotypes. In V. Orgogozo (Ed.), *Current Topics in Developmental Biology* (Vol. 119, pp. 157-204): Academic Press.

Robinson, D. R., Wu, Y.-M., Vats, P., Su, F., Lonigro, R. J., Cao, X., et al. (2013). Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nature genetics, 45*(12), 1446.

Roffo, G., Melzi, S., and Cristani, M. (2015). *Infinite feature selection.* Paper presented at the Proceedings of the IEEE International Conference on Computer Vision, 4202-4210.

Roll, J. D., Rivenbark, A. G., Jones, W. D., and Coleman, W. B. (2008). DNMT3b overexpression contributes to a hypermethylator phenotype in human breast cancer cell lines. *Molecular Cancer, 7*(1), 15.

Ronca, R., Giacomini, A., Rusnati, M., and Presta, M. (2015). The potential of fibroblast growth factor/fibroblast growth factor receptor signaling as a therapeutic target in tumor angiogenesis. *Expert Opinion on Therapeutic Targets, 19*(10), 1361-1377.

Rose, A. A. N., Biondini, M., Curiel, R., and Siegel, P. M. (2017). Targeting GPNMB with glembatumumab vedotin: Current developments and future opportunities for the treatment of cancer. *Pharmacology & Therapeutics, 179*, 127-141.

Rossi, M., Magnoni, L., Miracco, C., Mori, E., Tosi, P., Pirtoli, L., et al. (2011). β-catenin and Gli1 are prognostic markers in glioblastoma. *Cancer biology & therapy, 11*(8), 753-761.

Rossi, M. R., Huntoon, K., and Cowell, J. K. (2005). Differential expression of the LGI and SLIT families of genes in human cancer cells. *Gene, 356*, 85-90.

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., and Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic acids research, 47*(W1), W5-W10.

Saeys, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23*(19), 2507-2517.

Sahoo, G., and Kumar, Y. (2012). Analysis of parametric & non parametric classifiers for classification technique using WEKA. *International Journal of Information Technology and Computer Science (IJITCS), 4*(7), 43-49.

Saidi, A., Javerzat, S., Bellahcène, A., De Vos, J., Bello, L., Castronovo, V., et al. (2008). Experimental anti-angiogenesis causes upregulation of genes

associated with poor survival in glioblastoma. *International Journal of Cancer, 122*(10), 2187-2198.
Saiki, T., Kawai, T., Morita, K., Ohta, M., Saito, T., Rokutan, K., et al. (2008). Identification of Marker Genes for Differential Diagnosis of Chronic Fatigue Syndrome. *Molecular Medicine, 14*(9-10), 599-607.
Saleh, S. M., Bertos, N. R., Gruosso, T., Gigoux, M., Souleimanova, M., Zhao, H., et al. (2017). Identification of interacting stromal axes in triple-negative breast cancer. *Cancer Research*, canres.3427.2016.
Sampat, M. P., Bovik, A. C., Aggarwal, J. K., and Castleman, K. R. (2005). Supervised parametric and non-parametric classification of chromosome images. *Pattern Recognition, 38*(8), 1209-1223.
Sano, A., Sangai, T., Maeda, H., Nakamura, M., Hasebe, T., and Ochiai, A. (2007). Kallikrein 11 expressed in human breast cancer cells releases insulin-like growth factor through degradation of IGFBP-3. *International journal of oncology, 30*(6), 1493-1498.
Sattiraju, A., Pandya, D., Sai, K. S., Wadas, T., Herpai, D., Debinski, W., et al. (2016). IL13RA2 targeted alpha particle therapy against glioblastoma. *Journal of Nuclear Medicine, 57*(supplement 2), 634-634.
Schaefer, C. F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., et al. (2008). PID: the pathway interaction database. *Nucleic acids research, 37*(suppl_1), D674-D679.
Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science, 270*(5235), 467-470.
Schill, R., Solbrig, S., Wettig, T., and Spang, R. (2018). Modelling cancer progression using Mutual Hazard Networks. *bioRxiv*, 450841.
Seymour, T., Nowak, A., and Kakulas, F. (2015). Targeting Aggressive Cancer Stem Cells in Glioblastoma. *Frontiers in Oncology, 5*(159).
Shah, M., Marchand, M., and Corbeil, J. (2012). Feature selection with conjunctions of decision stumps and learning from microarray data. *IEEE transactions on pattern analysis and machine intelligence, 34*(1), 174-186.
Sharan, R., Ulitsky, I., and Shamir, R. (2007). Network-based prediction of protein function. *Molecular Systems Biology, 3*(1).
Sharma, A., Imoto, S., and Miyano, S. (2012). A top-r feature selection algorithm for microarray gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 9*(3), 754-764.
Sharma, G. N., Dave, R., Sanadya, J., Sharma, P., and Sharma, K. K. (2010). Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology & research, 1*(2), 109-126.
Sheng, Q., Moreau, Y., and De Moor, B. (2003). Biclustering microarray data by Gibbs sampling. *Bioinformatics, 19*(suppl_2), ii196-ii205.
Shimizu, T., Sugiura, T., Wakayama, T., Kijima, A., Nakamichi, N., Iseki, S., et al. (2011). PDZK1 Regulates Breast Cancer Resistance Protein in Small Intestine. *Drug Metabolism and Disposition*, dmd.111.040295.
Shirahata, M., Iwao-Koizumi, K., Saito, S., Ueno, N., Oda, M., Hashimoto, N., et al. (2007). Gene expression-based molecular diagnostic system for malignant gliomas is superior to histological diagnosis. *Clinical Cancer Research, 13*(24), 7341-7356.

Shoemaker, B. A., and Panchenko, A. R. (2007). Deciphering protein–protein interactions. Part I. Experimental techniques and databases. *PLoS computational biology, 3*(3), e42.

Shreem, S. S., Abdullah, S., and Nazri, M. Z. A. (2014). Hybridising harmony search with a Markov blanket for gene selection problems. *Information Sciences, 258*, 108-121.

Shukla, S., Patric, I. R. P., Patil, V., Shwetha, S. D., Hegde, A. S., Chandramouli, B. A., et al. (2014). Methylation silencing of ULK2, an autophagy gene, is essential for astrocyte transformation and tumor growth. *Journal of Biological Chemistry, 289*(32), 22306-22318.

Shukla, S., Pia Patric, I. R., Thinagararjan, S., Srinivasan, S., Mondal, B., Hegde, A. S., et al. (2013). A DNA Methylation Prognostic Signature of Glioblastoma: Identification of NPTX2-PTEN-NF-κB Nexus. *Cancer Research, 73*(22), 6563-6573.

Shur, I., Zemer-Tov, E., Socher, R., and Benayahu, D. (2007). SVEP1 expression is regulated in estrogen-dependent manner. *Journal of Cellular Physiology, 210*(3), 732-739.

Siletz, A., Kniazeva, E., Jeruss, J., and Shea, L. (2013). Transcription factor networks in invasion-promoting breast carcinoma-associated fibroblasts. *Cancer Microenvironment, 6*(1), 91-107.

Singh, A., Bhanot, G., and Khiabanian, H. (2018a). TuBA: Tunable Biclustering Algorithm Reveals Clinically Relevant Tumor Transcriptional Profiles in Breast Cancer. *bioRxiv*, 245712.

Singh, A. J., Ramsey, S. A., Filtz, T. M., and Kioussi, C. (2018b). Differential gene regulatory networks in development and disease. *Cellular and Molecular Life Sciences, 75*(6), 1013-1025.

Singh, D. K., Kollipara, R. K., Vemireddy, V., Yang, X.-L., Sun, Y., Regmi, N., et al. (2017). Oncogenes Activate an Autonomous Transcriptional Regulatory Circuit That Drives Glioblastoma. *Cell Reports, 18*(4), 961-976.

Singh, P., Jenkins, L. M., Horst, B., Alers, V., Pradhan, S., Kaur, P., et al. (2018c). Inhibin Is a Novel Paracrine Factor for Tumor Angiogenesis and Metastasis. *Cancer Research, 78*(11), 2978-2989.

Smid, M., Wang, Y., Zhang, Y., Sieuwerts, A. M., Yu, J., Klijn, J. G. M., et al. (2008). Subtypes of Breast Cancer Show Preferential Site of Relapse. *Cancer Research, 68*(9), 3108-3114.

Smyth, G. K. (2005). Limma: linear models for microarray data. In *Bioinformatics and computational biology solutions using R and Bioconductor* (pp. 397-420): Springer.

Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J. M., Desrichard, A., et al. (2014). Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *New England Journal of Medicine, 371*(23), 2189-2199.

Soinov, L. A., Krestyaninova, M. A., and Brazma, A. (2002). Towards reconstruction of gene networks from expression data by supervised learning. *Genome Biology*.

Sokhansanj, B., Fitch, J. P., Quong, J., and Quong, A. (2004). Linear fuzzy gene network models obtained from microarray data by exhaustive search. *BMC Bioinformatics, 5*(1), 108.

Srivastava, S., and Joshi, N. (2014). Clustering Techniques Analysis for Microarray Data.

Stanton, S. E., and Disis, M. L. (2016). Clinical significance of tumor-infiltrating lymphocytes in breast cancer. *Journal for ImmunoTherapy of Cancer, 4*(1), 59.

Statnikov, A., Wang, L., and Aliferis, C. F. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC bioinformatics, 9*(1), 319.

Stirzaker, C., Zotenko, E., Song, J. Z., Qu, W., Nair, S. S., Locke, W. J., et al. (2015). Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nature Communications, 6*, 5899.

Stoddard, F. R., 2nd, Brooks, A. D., Eskin, B. A., and Johannes, G. J. (2008). Iodine alters gene expression in the MCF7 breast cancer cell line: evidence for an anti-estrogen effect of iodine. *International journal of medical sciences, 5*(4), 189-196.

Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *science, 302*(5643), 249-255.

Sun, P., Guo, J., and Baumbach, J. (2013a). BiCluE - Exact and heuristic algorithms for weighted bi-cluster editing of biomedical data. *BMC Proceedings, 7*(Suppl 7), S9-S9.

Sun, T., Wang, J., Li, X., Lv, P., Liu, F., Luo, Y., et al. (2013b). Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. *Computer methods and programs in biomedicine, 111*(2), 519-524.

Sun, Y., Wang, Y., Fan, C., Gao, P., Wang, X., Wei, G., et al. (2014). Estrogen promotes stemness and invasiveness of ER-positive breast cancer cells through Gli1 activation. *Molecular Cancer, 13*(1), 137.

Sundvall, M., Iljin, K., Kilpinen, S., Sara, H., Kallioniemi, O.-P., and Elenius, K. (2008). Role of ErbB4 in Breast Cancer. *Journal of Mammary Gland Biology and Neoplasia, 13*(2), 259-268.

Suzuki, M., Shigematsu, H., Shames, D. S., Takahashi, T., Sunaga, N., Shivapurkar, N., et al. (2005). Methylation and gene silencing of RRAD gene in lung and breast cancers. *Cancer Research, 65*(9 Supplement), 218-218.

Szulzewsky, F., Pelz, A., Feng, X., Synowitz, M., Markovic, D., Langmann, T., et al. (2015). Glioma-Associated Microglia/Macrophages Display an Expression Profile Different from M1 and M2 Polarization and Highly Express Gpnmb and Spp1. *PLOS ONE, 10*(2), e0116644.

Tanay, A., Sharan, R., and Shamir, R. (2002). Discovering statistically significant biclusters in gene expression data. *Bioinformatics, 18*(suppl_1), S136-S144.

Tanay, A., Sharan, R., and Shamir, R. (2005). Biclustering algorithms: A survey. *Handbook of computational molecular biology, 9*(1-20), 122-124.

Tang, J., Alelyani, S., and Liu, H. (2014). Feature selection for classification: A review. *Data classification: Algorithms and applications*, 37.

Tang, Y., Zhang, Y.-Q., and Huang, Z. (2007). Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 4*(3), 365-381.

Taverna, S. D., Ilin, S., Rogers, R. S., Tanny, J. C., Lavender, H., Li, H., et al. (2006). Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Molecular cell, 24*(5), 785-796.

Taylor, K. M., Morgan, H. E., Smart, K., Zahari, N. M., Pumford, S., Ellis, I. O., et al. (2007). The emerging role of the LIV-1 subfamily of zinc transporters in breast cancer. *Molecular medicine (Cambridge, Mass.), 13*(7-8), 396-406.

ten Hoorn, S., Trinh, A., de Jong, J., Koens, L., and Vermeulen, L. (2018). Classification of Colorectal Cancer in Molecular Subtypes by Immunohistochemistry. In *Colorectal Cancer* (pp. 179-191): Springer.

Teng, L., and Chan, L. (2008). Discovering biclusters by iteratively sorting with weighted correlation coefficient in gene expression data. *Journal of Signal Processing Systems, 50*(3), 267-280.

Teplyuk, N. M., Uhlmann, E. J., Wong, A. H.-K., Karmali, P., Basu, M., Gabriely, G., et al. (2015). MicroRNA-10b inhibition reduces E2F1-mediated transcription and miR-15/16 activity in glioblastoma. *Oncotarget, 6*(6), 3770.

The Cancer Genome Atlas, N., Koboldt, D. C., Fulton, R. S., McLellan, M. D., Schmidt, H., Kalicki-Veizer, J., et al. (2012). Comprehensive molecular portraits of human breast tumours. *Nature, 490*, 61.

Thomas, J. G., Olson, J. M., Tapscott, S. J., and Zhao, L. P. (2001). An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Research, 11*(7), 1227-1236.

Tie, X., Han, S., Meng, L., Wang, Y., and Wu, A. (2013). NFAT1 is highly expressed in, and regulates the invasion of, glioblastoma multiforme cells. *PLoS One, 8*(6), e66008.

Toillon, R.-A., Magné, N., Laïos, I., Castadot, P., Kinnaert, E., Van Houtte, P., et al. (2007). Estrogens decrease γ-ray–induced senescence and maintain cell cycle progression in breast cancer cells independently of p53. *International Journal of Radiation Oncology*Biology*Physics, 67*(4), 1187-1200.

Tomasev, N., Radovanovic, M., Mladenic, D., and Ivanovic, M. (2014). The Role of Hubness in Clustering High-Dimensional Data. *IEEE Transactions on Knowledge and Data Engineering, 26*(3), 739-751.

Tong, Y., Hao, J., Tu, Q., Yu, H., Yan, L., Li, Y., et al. (2017). A tree shrew glioblastoma model recapitulates features of human glioblastoma. *Oncotarget, 8*(11), 17897-17907.

Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., et al. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical cancer research, 14*(16), 5198-5208.

Toy, W., Shen, Y., Won, H., Green, B., Sakr, R. A., Will, M., et al. (2013). ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nature Genetics, 45*, 1439.

Treiber, J. M., Steed, T. C., Brandel, M. G., Patel, K. S., Dale, A. M., Carter, B. S., et al. (2018). Molecular physiology of contrast enhancement in glioblastomas: An analysis of The Cancer Imaging Archive (TCIA). *Journal of Clinical Neuroscience, 55*, 86-92.

Trevino, V., Falciani, F., and Barrera-Saldaña, H. A. (2007). DNA Microarrays: a Powerful Genomic Tool for Biomedical and Clinical Research. *Molecular Medicine, 13*(9), 527-541.

Tripathi, A., King, C., de la Morenas, A., Perry, V. K., Burke, B., Antoine, G. A., et al. (2008). Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. *International Journal of Cancer, 122*(7), 1557-1566.

Tripathi, A., Ren, Y., Jeffrey, P. D., and Hughson, F. M. (2009). Structural characterization of Tip20p and Dsl1p, subunits of the Dsl1p vesicle tethering complex. *Nature structural & molecular biology, 16*(2), 114.

Tseng, Y.-H., Vicent, D., Zhu, J., Niu, Y., Adeyinka, A., Moyers, J. S., et al. (2001). Regulation of growth and tumorigenicity of breast cancer cells by the low molecular weight GTPase Rad and nm23. *Cancer Research, 61*(5), 2071-2079.

Tsigelny, I. F., Kouznetsova, V. L., Lian, N., and Kesari, S. (2016). Molecular mechanisms of OLIG2 transcription factor in brain cancer. *Oncotarget, 7*(33), 53074-53101.

Tung, J.-N., Ko, C.-P., Yang, S.-F., Cheng, C.-W., Chen, P.-N., Chang, C.-Y., et al. (2016). Inhibition of pentraxin 3 in glioma cells impairs proliferation and invasion in vitro and in vivo. *Journal of Neuro-Oncology, 129*(2), 201-209.

Tyanova, S., Albrechtsen, R., Kronqvist, P., Cox, J., Mann, M., and Geiger, T. (2016). Proteomic maps of breast cancer subtypes. *Nature Communications, 7*, 10259.

Uppalapati, P., Xiang, Y., and Huang, K. (2010). *Predicting prognostic markers for glioma using gene co-expression network analysis*. Paper presented at the Proceedings of the first ACM international conference on bioinformatics and computational biology, 546-551.

Vaira, V., Fedele, G., Pyne, S., Fasoli, E., Zadra, G., Bailey, D., et al. (2010). Preclinical model of organotypic culture for pharmacodynamic profiling of human tumors. *Proceedings of the National Academy of Sciences, 107*(18), 8352-8356.

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in Bioinformatics, 19*(4), 575-592.

van Roosmalen, W., Le Dévédec, S. E., Golani, O., Smid, M., Pulyakhina, I., Timmermans, A. M., et al. (2015). Tumor cell migration screen identifies SRPK1 as breast cancer metastasis determinant. *The Journal of Clinical Investigation, 125*(4), 1648-1664.

Vastrad, B., Vastrad, C., Godavarthi, A., and Chandrashekar, R. (2017). Molecular mechanisms underlying gliomas and glioblastoma pathogenesis revealed by bioinformatics analysis of microarray data. *Medical Oncology, 34*(11), 182.

Vauléon, E., Tony, A., Hamlat, A., Etcheverry, A., Chiforeanu, D. C., Menei, P., et al. (2012). Immune genes are associated with human glioblastoma pathology and patient survival. *BMC medical genomics, 5*, 41-41.

Vazifehdan, M., Moattar, M. H., and Jalali, M. (2018). A hybrid Bayesian network and tensor factorization approach for missing value imputation to improve breast cancer recurrence prediction. *Journal of King Saud University - Computer and Information Sciences*.

Verbovšek, U., Motaln, H., Rotter, A., Atai, N. A., Gruden, K., Van Noorden, C. J. F., et al. (2014). Expression analysis of all protease genes reveals cathepsin K to be overexpressed in glioblastoma. *PloS one, 9*(10), e111819-e111819.

Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., et al. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell, 17*(1), 98-110.

Vladimir, F. (2005). Identifying Gene Regulatory Networks from Gene Expression Data. In *Handbook of Computational Molecular Biology* (pp. 27-21-27-29): Chapman and Hall/CRC.

Voduc, K. D., Cheang, M. C. U., Tyldesley, S., Gelmon, K., Nielsen, T. O., and Kennecke, H. (2010). Breast Cancer Subtypes and the Risk of Local and Regional Relapse. *Journal of Clinical Oncology, 28*(10), 1684-1691.

Wanat, J. J., Kim, K. P., Koszul, R., Zanders, S., Weiner, B., Kleckner, N., et al. (2008). Csm4, in collaboration with Ndj1, mediates telomere-led chromosome dynamics and recombination during yeast meiosis. *PLoS genetics, 4*(9), e1000188.

Wanderley, M. F. B., Gardeux, V., Natowicz, R., and de Pádua Braga, A. (2013). *GA-KDE-Bayes: an evolutionary wrapper method based on non-parametric density estimation applied to bioinformatics problems.* Paper presented at the ESANN.

Wang, F., Wang, A. Y., Chesnelong, C., Yang, Y., Nabbi, A., Thalappilly, S., et al. (2017a). ING5 activity in self-renewal of glioblastoma stem cells via calcium and follicle stimulating hormone pathways. *Oncogene, 37*, 286.

Wang, G., Platt-Higgins, A., Carroll, J., de Silva Rudland, S., Winstanley, J., Barraclough, R., et al. (2006). Induction of Metastasis by S100P in a Rat Mammary Model and Its Association with Poor Survival of Breast Cancer Patients. *Cancer Research, 66*(2), 1199-1207.

Wang, G., Song, Q., Xu, B., and Zhou, Y. (2013a). Selecting feature subset for high dimensional data via the propositional FOIL rules. *Pattern Recognition, 46*(1), 199-214.

Wang, H., Wang, W., Yang, J., and Yu, P. S. (2002). *Clustering by pattern similarity in large data sets.* Paper presented at the Proceedings of the 2002 ACM SIGMOD international conference on Management of data, 394-405.

Wang, J., Ray, P. S., Sim, M. S., Zhou, X. Z., Lu, K. P., Lee, A. V., et al. (2012). FOXC1 regulates the functions of human basal-like breast cancer cells by activating NF-κB signaling. *Oncogene, 31*, 4798.

Wang, M., Shang, X., Li, X., Liu, W., and Li, Z. (2013b). Efficient mining differential co-expression biclusters in microarray datasets. *Gene, 518*(1), 59-69.

Wang, M., Zhang, L., Xiong, Z., and Yang, T. (2017b). An efficient usage rate biclustering algorithm in real-valued function-resource dataset. *Journal of Algorithms & Computational Technology, 11*(3), 281-291.

Wang, Q., Lu, X., Zhao, S., Pang, M., Wu, X., Wu, H., et al. (2017c). Moesin Expression Is Associated with Glioblastoma Cell Proliferation and Invasion. *Anticancer research, 37*(5), 2211-2218.

Wang, X., Liu, J., Cheng, Y., Liu, A., and Chen, E. (2018). Dual Hypergraph Regularized PCA for Biclustering of Tumor Gene Expression Data. *IEEE Transactions on Knowledge and Data Engineering*, 1-1.

Wang, Y., Liu, J., Ying, X., Lin, P. C., and Zhou, B. P. (2016a). Twist-mediated Epithelial-mesenchymal Transition Promotes Breast Tumor Cell Invasion via Inhibition of Hippo Pathway. *Scientific Reports, 6*, 24606.

Wang, Y., Moncayo, G., Morin Jr, P., Xue, G., Grzmil, M., Lino, M., et al. (2013c). Mer receptor tyrosine kinase promotes invasion and survival in glioblastoma multiforme. *Oncogene, 32*(7), 872.

Wang, Y., Tetko, I. V., Hall, M. A., Frank, E., Facius, A., Mayer, K. F., et al. (2005). Gene selection from microarray data for cancer classification—a machine learning approach. *Computational biology and chemistry, 29*(1), 37-46.

Wang, Y. K., Print, C. G., and Crampin, E. J. (2013d). Biclustering reveals breast cancer tumour subgroups with common clinical features and improves prediction of disease recurrence. *BMC Genomics, 14*(1), 102.

Wang, Z., Dahiya, S., Provencher, H., Muir, B., Carney, E., Coser, K., et al. (2007). The prognostic biomarkers HOXB13, IL17BR, and CHDH are regulated by estrogen in breast cancer. *Clinical Cancer Research, 13*(21), 6327-6334.

Wang, Z., Li, G., W. Robinson, R., and Huang, X. (2016b). *UniBic: Sequential row-based biclustering algorithm for analysis of gene expression data* (Vol. 6).

Wang, Z., Yang, J., Xu, G., Wang, W., Liu, C., Yang, H., et al. (2014). Targeting miR-381-NEFL axis sensitizes glioblastoma cells to temozolomide by regulating stemness factors and multidrug resistance factors. *Oncotarget, 6*(5), 3147-3164.

Watson, J. D., and Crick, F. (2003). A structure for deoxyribose nucleic acid. 1953. *Nature, 421*(6921), 397.

Watson, P. H., Chia, S. K., Wykoff, C. C., Han, C., Leek, R. D., Sly, W. S., et al. (2003). Carbonic anhydrase XII is a marker of good prognosis in invasive breast carcinoma. *British Journal Of Cancer, 88*, 1065.

Weber, B. (2015). *Modeling Cancer Metabolism.* University for Health Sciences, Medical Informatics and Technology.

Wei, Z., and Li, H. (2007). A Markov random field model for network-based analysis of genomic data. *Bioinformatics, 23*(12), 1537-1544.

Werner, T. (2008). Bioinformatics applications for pathway analysis of microarray data. *Current Opinion in Biotechnology, 19*(1), 50-54.

Wesley, U. V., Clark, P., Kuo, J. S., and Dempsey, R. J. (2017). Abstract 4773: Dysregulated expression of pentraxin-3 in glioblastoma (GBM) stem cells: Implications for GBM invasion and progression. *Cancer Research, 77*(13 Supplement), 4773-4773.

Williams, C., Edvardsson, K., Lewandowski, S. A., Ström, A., and Gustafsson, J. Å. (2007). A genome-wide study of the repressive effects of estrogen receptor beta on estrogen receptor alpha signaling in breast cancer cells. *Oncogene, 27*, 1019.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.

Wong, H. Y., Wang, G. M., Croessmann, S., Zabransky, D. J., Chu, D., Garay, J. P., et al. (2015). TMSB4Y is a candidate tumor suppressor on the Y chromosome and is deleted in male breast cancer. *Oncotarget, 6*(42), 44927.

Wu, C.-J., and Kasif, S. (2005). GEMS: a web server for biclustering analysis of expression data. *Nucleic Acids Research, 33*(suppl_2), W596-W599.

Wu, K., Jiao, X., Li, Z., Katiyar, S., Casimiro, M. C., Yang, W., et al. (2011). Cell fate determination factor Dachshund reprograms breast cancer stem cell function. *Journal of Biological Chemistry, 286*(3), 2132-2142.

Xia, S.-L., Wu, M.-L., Li, H., Wang, J.-H., Chen, N.-N., Chen, X.-Y., et al. (2015). CRABP-II- and FABP5-independent responsiveness of human glioblastoma cells to all-trans retinoic acid. *Oncotarget, 6*(8), 5889-5902.

Xiang, Y., Zhang, C.-Q., and Huang, K. (2012). *Predicting glioblastoma prognosis networks using weighted gene co-expression network analysis on TCGA data.* Paper presented at the BMC bioinformatics, S12.

Xu, J., Acharya, S., Sahin, O., Zhang, Q., Saito, Y., Yao, J., et al. (2015). 14-3-3ζ Turns TGF-β's Function from Tumor Suppressor to Metastasis Promoter in Breast Cancer by Contextual Changes of Smad Partners from p53 to Gli2. *Cancer Cell, 27*(2), 177-192.

Xu, J., Zhang, Q., and Shih, C.-K. (2006). V-cluster algorithm: A new algorithm for clustering molecules based upon numeric data. *Molecular diversity, 10*(3), 463-478.

Xu, M., and Setiono, R. (2015). Gene selection for cancer classification using a hybrid of univariate and multivariate feature selection methods. *arXiv preprint arXiv:1506.02085*.

Xu, T., Le, T. D., Liu, L., Wang, R., Sun, B., and Li, J. (2016). Identifying Cancer Subtypes from miRNA-TF-mRNA Regulatory Networks and Expression Data. *PLOS ONE, 11*(4), e0152792.

Xu, X. (2013). Enhancing gene expression clustering analysis using tangent transformation. *International Journal of Machine Learning and Cybernetics, 4*(1), 31-40.

Xu, Y., Gu, S., Bi, Y., Qi, X., Yan, Y., and Lou, M. (2018). Transcription factor PU. 1 is involved in the progression of glioma. *Oncology letters, 15*(3), 3753-3759.

Yachie-Kinoshita, A., and Kaizu, K. (2019). Cell Modeling and Simulation. In S. Ranganathan, M. Gribskov, K. Nakai and C. Schönbach (Eds.), *Encyclopedia of Bioinformatics and Computational Biology* (pp. 864-873). Oxford: Academic Press.

Yan, J., Risacher, S. L., Shen, L., and Saykin, A. J. (2017). Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Briefings in Bioinformatics, 19*(6), 1370-1381.

Yang, S., and Naiman, D. Q. (2014). Multiclass cancer classification based on gene expression comparison. *Statistical applications in genetics and molecular biology, 13*(4), 477-496.

Yang, W.-H., Dai, D.-Q., and Yan, H. (2011a). Finding correlated biclusters from gene expression data. *IEEE Transactions on Knowledge and Data Engineering, 23*(4), 568-584.

Yang, X. R., Chang-Claude, J., Goode, E. L., Couch, F. J., Nevanlinna, H., Milne, R. L., et al. (2011b). Associations of Breast Cancer Risk Factors With Tumor Subtypes: A Pooled Analysis From the Breast Cancer Association Consortium Studies. *JNCI: Journal of the National Cancer Institute, 103*(3), 250-263.

Yang, X. R., Sherman, M. E., Rimm, D. L., Lissowska, J., Brinton, L. A., Peplonska, B., et al. (2007). Differences in Risk Factors for Breast Cancer Molecular Subtypes in a Population-Based Study. *Cancer Epidemiology Biomarkers & Prevention, 16*(3), 439-443.

Yang, Y., Xing, Y., Liang, C., Hu, L., Xu, F., and Chen, Y. (2015). Crucial microRNAs and genes of human primary breast cancer explored by microRNA-mRNA integrated analysis. *Tumor Biology, 36*(7), 5571-5579.

Yao, Y., Liu, Y., Lv, X., Dong, B., Wang, F., Li, J., et al. (2016). Down-regulation of ribosomal protein S15A inhibits proliferation of human glioblastoma cells in vivo and in vitro via AKT pathway. *Tumor Biology, 37*(4), 4979-4990.

Yee, N. S., Ignatenko, N., Finnberg, N., Lee, N., and Stairs, D. (2015). ANIMAL MODELS OF CANCER BIOLOGY. *Cancer Growth and Metastasis, 8*(Suppl 1), 115-118.

Yersal, O., and Barutca, S. (2014). Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World journal of clinical oncology, 5*(3), 412-424.

Yeung, K. Y., Bumgarner, R. E., and Raftery, A. E. (2005). Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics, 21*(10), 2394-2402.

Yin, D., Ogawa, S., Kawamata, N., Tunici, P., Finocchiaro, G., Eoli, M., et al. (2009). High-Resolution Genomic Copy Number Profiling of Glioblastoma Multiforme by Single Nucleotide Polymorphism DNA Microarray. *Molecular Cancer Research*, 1541-7786.MCR-1508-0270.

Yin, L., and Liu, Y. (2018). Ensemble biclustering gene expression data based on the spectral clustering. *Neural Computing and Applications, 30*(8), 2403-2416.

Yip, K. Y., Cheung, D. W., and Ng, M. K. (2004). Harp: A practical projected clustering algorithm. *IEEE Transactions on knowledge and data engineering, 16*(11), 1387-1397.

Yousef, M., Showe, L., and Showe, M. (2009). A study of microRNAs in silico and in vivo: bioinformatics approaches to microRNA discovery and target identification. *The FEBS journal, 276*(8), 2150-2156.

Yu, F., and Fu, W. M. (2015). Identification of differential splicing genes in gliomas using exon expression profiling. *Molecular medicine reports, 11*(2), 843-850.

Yu, G., Yu, X., and Wang, J. (2017a). Network-aided Bi-Clustering for discovering cancer subtypes. *Scientific Reports, 7*(1), 1046.

Yu, L., and Liu, H. (2003). *Feature selection for high-dimensional data: A fast correlation-based filter solution.* Paper presented at the Proceedings of the 20th international conference on machine learning (ICML-03), 856-863.

Yu, S., Li, A., Liu, Q., Li, T., Yuan, X., Han, X., et al. (2017b). Chimeric antigen receptor T cells: a novel therapy for solid tumors. *Journal of Hematology & Oncology, 10*(1), 78.

Zang, H., Li, N., Pan, Y., and Hao, J. (2017). Identification of upstream transcription factors (TFs) for expression signature genes in breast cancer. *Gynecological Endocrinology, 33*(3), 193-198.

Zhai, Y., Bai, J., Wang, S., Li, M., Wang, F., Li, C., et al. (2018). Aberrant Expression of Extracellular Signal-Regulated Kinase and 15-Hydroxyprostaglandin Dehydrogenase Indicates Radiation Resistance and Poor Prognosis for Patients with Clival Chordomas. *World Neurosurgery, 115*, e146-e151.

Zhang, B., Pan, X., Wang, Q., Cobb, G. P., and Anderson, T. A. (2006). Computational identification of microRNAs and their targets. *Computational Biology and Chemistry, 30*(6), 395-407.

Zhang, C., Fu, J., Xue, F., Ryu, B., Zhang, T., Zhang, S., et al. (2016). Knockdown of ribosomal protein S15A induces human glioblastoma cell apoptosis. *World Journal of Surgical Oncology, 14*(1), 129.

Zhang, C., Liu, J., Wu, R., Liang, Y., Lin, M., Liu, J., et al. (2014). Tumor suppressor p53 negatively regulates glycolysis stimulated by hypoxia through its target RRAD. *Oncotarget, 5*(14), 5535-5546.

Zhang, J., Chen, L., Han, L., Shi, Z., Zhang, J., Pu, P., et al. (2015). EZH2 is a negative prognostic factor and exhibits pro-oncogenic activity in glioblastoma. *Cancer Letters, 356*(2, Part B), 929-936.

Zhang, Y., Xu, Y., Ma, J., Pang, X., and Dong, M. (2017). Adrenomedullin promotes angiogenesis in epithelial ovarian cancer through upregulating hypoxia-

inducible factor-1α and vascular endothelial growth factor. *Scientific Reports, 7*, 40524.

Zhang, Z., Tang, H., Wang, Z., Zhang, B., Liu, W., Lu, H., et al. (2011). MiR-185 targets the DNA methyltransferases 1 and regulates global DNA methylation in human glioma. *Molecular cancer, 10*(1), 124.

Zhao, L., Zhao, H., and Yan, H. (2018). Gene expression profiling of 1200 pancreatic ductal adenocarcinoma reveals novel subtypes. *BMC Cancer, 18*(1), 603.

Zhao, M., and Zhao, Z. (2016). Concordance of copy number loss and down-regulation of tumor suppressor genes: a pan-cancer study. *BMC Genomics, 17*(7), 532.

Zheng, B., Yoon, S. W., and Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications, 41*(4), 1476-1482.

Zhong, S., Wu, B., Han, Y., Cao, Y., Yang, L., Luo, S. X., et al. (2017). Identification of driver genes and key pathways of pediatric brain tumors and comparison of molecular pathogenesis based on pathologic types. *World neurosurgery, 107*, 990-1000.

Zhou, L., Tang, H., Wang, F., Chen, L., Ou, S., Wu, T., et al. (2018). Bioinformatics analyses of significant genes, related pathways and candidate prognostic biomarkers in glioblastoma. *Molecular medicine reports, 18*(5), 4185-4196.

Zhou, Q., Hong, W., Luo, L., and Yang, F. (2010). Gene selection using random forest and proximity differences criterion on DNA microarray data. *Journal of Convergence Information Technology, 5*(6), 161-170.

Zhou, X., and Tuck, D. P. (2007). MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data. *Bioinformatics, 23*(9), 1106-1114.

Zhu, C.-H., and Domann, F. E. (2002). Dominant negative interference of transcription factor AP-2 causes inhibition of ErbB-3 expression and suppresses malignant cell growth. *Breast cancer research and treatment, 71*(1), 47-57.

Zhu, D., Hunter, S. B., Vertino, P. M., and Van Meir, E. G. (2011). Overexpression of MBD2 in Glioblastoma Maintains Epigenetic Silencing and Inhibits the Anti-Angiogenic Function of the Tumor Suppressor Gene BAI1. *Cancer research*, canres. 1157.2011.