Chapman University

Chapman University Digital Commons

Pharmaceutical Sciences (MS) Theses

Dissertations and Theses

Summer 8-2021

Characterization of The Growth Factor Receptor Network Oncogenes in Lung Cancer

Ashley Duche
Chapman University, duche@chapman.edu

Recommended Citation

Duche, A. *Characterization of The Growth Factor Receptor Network Oncogenes in Lung Cancer*. [master's thesis]. Irvine, CA: Chapman University; 2021. https://doi.org/10.36837/chapman.000306

CHARACTERIZATION OF THE GROWTH FACTOR RECEPTOR NETWORK ONCOGENES IN LUNG CANCER

A Thesis by

Ashley H. Duche

Chapman University

Irvine, CA

School of Pharmacy

Submitted in partial fulfillment of the requirements for the degree of

Master of Science in Pharmaceuticals Sciences

August 2021

Committee in charge

Dr. Moom R. Roosan, PharmD, Ph.D

Dr. Rennolds Ostrom, Ph.D

Dr. Ajay Sharma, Ph.D

The thesis of Ashley H. Duche is approved.

Moom Roosan

Moom R. Roosan, PharmD, Ph.D.

Rennolds Ostrom, Ph.D.

Ajay Sharma

Ajay Sharma, Ph.D.

April 2021

CHARACTERIZATION OF THE GROWTH FACTOR RECEPTOR NETWORK

ONCOGENES IN LUNG CANCER

by Ashley H. Duche

ACKNOWLEDGEMENTS

To start, I would first like to thank my mentor Dr. Roosan R. Moom for her continuous support throughout my time here at Chapman University in the School of Pharmacy. I am beyond grateful to have had a mentor willing to teach me the very basics of R programming language and guide me throughout my progression to provide me with the bioinformatics and pharmacogenomics skills I have today. Throughout this experience she has given me various opportunities to apply my knowledge and allowed me the freedom to explore this process to be able to develop my own style of research. I cannot stress enough how much I admire her drive and knowledge within this field and would not be where I am today without her. Next, I would like to thank my lab member Yue Lie, for taking the time my first few weeks of research to teach me the very basics of how to read and dissect a research paper, onto performing advance bioinformatics analysis. I am privileged to have had the chance to learn and work with someone as intelligent and driven as her and strive to progress to her level of skill. I would also like to thank my lab member, Michael Sayer, for helping me throughout my research and providing me with the knowledge of how bioinformatics research can be applied in the clinical aspect. This has allowed me to progress my thinking beyond computational analysis and how research is applied beyond analysis. In addition, I cannot thank my committee members Dr. Rennolds Ostrom and Dr. Ajay Sharma, enough for their guidance and support. Throughout my research they have caused me to think critically and further develop my understanding of how research across various areas of focus can be interlinked and used to advance our comprehensive knowledge. Lastly, I would like to thank Chapman University School of Pharmacy for allowing me to perform my research and giving me the opportunity to further advance my knowledge.

ABSTRACT

CHARACTERIZATION OF THE GROWTH FACTOR RECEPTOR NETWORK ONCOGENES IN LUNG CANCER

by Ashley H. Duche

Lung cancer remains the leading cause of cancer related deaths worldwide, reportedly contributing to 1.8 million of the 10.0 million mortalities documented in the year 2020. Although advancements have been made in therapeutics and diagnostic methods, formulation of effective treatments and development of drug resistance continues to be a challenge. These challenges arise from our lack of understanding of intricate signaling pathways, such as the Growth Factor Receptor Network (GFRN), which contributes to complex lung tumor heterogeneity allowing for drug resistance development. In this study, gene expression signatures of six GFRN oncogenes overexpressed in human mammary epithelial cells (HMECs) were generated to interrogate this pathway's downstream crosstalk, beyond initial mutation status. Utilization of this method may reveal novel phenotypic patterns that could be used to improve targeted therapies for lung cancer. Thus, using computational analysis tools, gene expression signatures were generated of B*AD (BAD), HER2 (ERBB2), IGF1R (IGF1R), RAF (RAF1),* and *KRAS (G12V),* using the Bioconductor package, *Adaptive Signature Selection and InteGratioN (ASSIGN)*. Gene lists of various lengths were generated ranging from 5 to 500 genes produced in 25 gene increments. Pathway activation estimates were predicted in 541 lung adenocarcinoma (LUAD) tumors acquired from The Cancer Genome Atlas (TCGA). Each gene signature underwent validation using proteomics data from The Cancer Proteome Atlas (TCPA) and gene expression. Following thorough analysis, optimal gene signatures were determined for the genes *BAD, HER2, IGF1R, RAF* and *KRAS.* In all, the

optimized GFRN pathway-specific gene signatures were able to distinguish upregulated pathway activity within TCGA patient tumor samples. With the use of drug response data, novel phenotypic patterns may be revealed identifying drug targets to improve individualized drug targeted therapy for lung cancer.

Dedicated to my parents Tina Duche, and Dave Duche
As well as my grandparents.

“The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge.”
Stephen Hawking

TABLE OF CONTENTS

CHAPTERS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Lung cancer remains the leading cause of cancer related deaths despite progressive advancements in therapeutic and diagnostic methods worldwide. According to the American Cancer Society (ACS), it is estimated that of the 608,570 cancer related mortalities projected to occur in the United States in 2021, 131,880 cases will be due to lung cancer [1]. Similar to other cancers, lung tumors develop due to epigenetic factors causing genetic alterations, such as somatic mutations, gene amplifications and chromosomal rearrangements/translocation, affecting a cell's regulatory mechanisms and normal functions[1, 2]. With traditional methods, lung tumors can be classified into two major types, including small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC)[3]. Within NSCLC there are three main subtypes - squamous cell carcinoma, adenocarcinoma, and large cell carcinoma. Although through the advancements of diagnostic methods with the incorporation of molecular profiling, further tumor heterogeneity has emerged revealing diversification of lung tumors within the same histological subtype [2]. Such molecular profiling methods include immunohistochemistry (IHC), chromogenic/fluorescence in situ hybridization (CISH/FISH), next-generation sequencing, sanger and pyrosequencing, as well as quantitative polymerase chain reaction (qPCR) and fragment analysis (FA/Frag.Analysis) [4]. These methods allow for specific genetic alterations, referred to as biomarkers, to be identified within a tumor and used to make improved diagnosis, prognosis and therapeutic treatments. Although, a challenge continually faced is targeted mutations do not always respond to oncological treatments and consequently form mechanisms that allow resistance to therapeutic treatments [5]. This can result from unknown downstream signaling that remains uncharacterized in complex oncogenic networks such as the Growth Factor Receptor Network (GFRN) [7].

1.1 Overview

The GFRN is a known driving oncogenic network in lung cancer consisting of parallel signaling pathways responsible for regulating developmental and growth processes within the cell (Figure 1.1) [6]. Two stimulated growth factor pathways comprising of this network include the phosphatidylinositol-3-kinase (PI3K)/protein kinase B (AKT)/ mechanistic target of rapamycin kinase (mTOR) as well as the RAS/serine-threonine protein kinase (RAF)/ mitogen-activated protein kinase (MAPK) pathway [7, 8]. The PI3K/AKT/mTOR pathway is commonly associated with NSCLC responsible for controlling cell survival, metabolism and proliferation [7]. Within this pathway, upstream activation of receptor tyrosine kinases (RTKs) such as EGFR, HER2, and insulin-like growth factor receptor (IGF1R), initiates a complex signaling cascade leading to the activation of PI3K lipid kinases [7]. A signal is then relayed resulting in the activation of AKT, in turn activating serine/threonine (Ser/Thr) kinase mTOR [9]. Many negative feedback regulators are associated with this pathway such as the inactivation of AKT through phosphatase and tensin homolog (PTEN) tumor suppressor, as well as the inhibition of IGF1R signaling by downstream products of mTOR [9]. To bypass these negative feedback mechanisms, the PI3K/AKT/mTOR pathway interacts with the neighboring pathway RAS/RAF/MAPK [9, 10]. The RAS/RAF/MAPK pathway is also associated with tumorigenesis initiated through the phosphorylation of RTKs, such as EGFR [9]. Following receptor mediated activation, a signaling cascade is initiated activating the GTPase protein KRAS, transmitting a signal activating the Ser/Thr-protein kinase RAF1, also known as c-RAF [10]. Subsequent activation leads to phosphorylation of MEK1/2 resulting in activation of Ser/Thr kinases, ERK1/ERK[8, 10]. What ultimately makes this pathway difficult for formulation of effective drug targeted treatments is the alternate pathway activation that can occur between these parallel signaling pathways. For instance, alternate pathway activation of PI3K can be transduced through RAS signaling, mTOR can be activated through ERK, and AKT can inhibit activation of RAF as well as BAD (BCL2 Associated Agonist of Cell Death)[11, 12]. Therefore, simultaneous characterization of the GFRN is warranted for applying targeted therapies in lung cancer.

To begin to characterize the network of complex signaling pathways within lung cancer, gene expression signatures can be utilized to interrogate GFRN activity within lung tumors. A gene expression signature is a gene, or a combined group of genes expressing aberrant or normal pathway activity associated with causing a disease or biological process [13, 14]. Signatures consist of selected genes quantitatively expressing varying levels of gene expression in respect to the biological state of the pathway being explored [13, 14]. They can be used to represent a single pathway or be leveraged in conjunction to explore multiple activated pathways simultaneously [5]. This allows for a comprehensive profile of interconnecting signaling networks to be explored which can potentially be used to make improved prognostic, diagnostic, and therapeutic treatment decisions [6].

In summary, the utilization of generated gene expression signatures can be leveraged to explore complex signaling pathways using selected genes of possible significance to reveal underlying molecular mechanisms of a disease. Applying this concept, the objective of my research was to generate GFRN pathway-specific gene expression signatures of the pathways *BAD (BAD), HER2 (ERBB2), IGF1R (IGF1R), RAF (RAF1),* and *KRAS (KRAS, G12V* mutation*)*. It was hypothesized that if pathway-specific gene expression signatures of GFRN activity can be generated, representing the oncogenic state of that pathway, GFRN activity can be characterized within lung tumors to reveal novel phenotypic patterns to make drug response predictions.

1.2 Relevance of Exploration for Selected GFRN Oncogenes

Proven by previous studies, the GFRN has played a critical role in driving oncogenic processes leading to lung tumor formation. As referenced in Figure 1, the pro-apoptotic protein BAD, is one of the many signaling pathways comprising this network. BAD plays an important role in promoting apoptotic cell death, which has made it a predictive biomarker within lung cancer[11, 12]. Low levels of BAD expression have been associated with tumorigenesis across many other cancers as well, indicating its importance in anti-cancer cellular functions [11]. Inhibition of this pathway, as previously mentioned, stems from the activation of PI3K signaling activating AKT, which in turn inhibits the pro-apoptotic protein [6, 7, 12]. Having the knowledge of BAD's anticancer characteristics, and its role in tumor progression, studies have suggested that overexpression of this protein can also allow BAD to act as a tumor suppressor [11, 12]. This makes BAD a promising target for future use of formulating effective therapeutic treatments.

Another associated GFRN pathway is the protein tyrosine kinase HER2. HER2 is a cell surface receptor associated with PI3K pathway activation initiating tumorigenesis [15]. In recent studies, the presence of HER2 mutations within NSCLC patients may be correlated with lower survival rates [15, 16]. Additionally, utilization of molecular profiling methods may have revealed further intrinsic subtypes, showing a correlation with HER2 mutations with the presence of EGFR mutations, and ALK translocations[9, 15, 16]. Although there has been conflicting evidence of HER2's involvement in lung cancer making further exploration of this pathway essential.

In addition to BAD and HER2, another GFRN pathway associated with lung tumor development is IGFR. This RTK has shown correlations of overexpression linked to increased cell survival and proliferation of malignant cells [17]. Acquired resistance to therapies such as gefitinib and erlotinib have been observed with possible intrinsic subtypes such as the presence of EGFR mutations as well as ALK arrangements, similar to HER2 [17]. Additionally, IGF1R intrinsic subtypes may have also been correlated with the development of resistance to EGFR targeted treatments[17]. Benefits of further exploration of this pathway may lead to the development of effective therapeutic treatments against EGFR drug resistance mechanisms using molecular profiling to reveal cancer promoting cellular mechanisms.

As revealed in prior studies, the proto-oncogene RAF, has shown associations with the RAS signaling pathway within the GFRN [10]. Also known as RAF1 or c-RAF, the full characterization of this pathway's activation remains unclear, as well as its role in lung tumor development [10]. Although, studies have supported c-RAF activation is required for the initiation of tumorigenesis through KRAS transduction [10]. Within lung cancer, the development of KRAS drug resistance has continually been a challenge due to the ineffectiveness of current therapeutic treatments, as well as efficacy issues with targeted treatments of ERK/MAPK inhibition[10]. Possible leverage of targeting the c-RAF pathway, as well has further exploration revealing its molecular mechanisms, mays be used to develop novel effective treatments targeting KRAS with reduced drug resistance development.

Previously mentioned, a common mutation associated with lung cancer development is the RAS Family, proto-oncogene KRAS. Various variants of KRAS mutations have been identified including G12C, G12B, and G12V, classified based upon their amino acid substitution. The significant prevalence of this mutation within lung cancer presses the need for effective therapeutic treatments. Although, due to the complex signaling and alternate pathway activations, formulation of effective therapeutic treatments continues to be a challenge (Figure 1.1)[10, 18]. In an attempt to formulate targeting treatments for KRAS combating drug resistance development, exploration of coinciding mutations has been performed in previous studies[18]. Possible associations between the presence of KRAS coinciding with EGFR was revealed but little significance was observed pertaining to prognosis [18]. Although, additional studies have showed promise applying this method leading to further subtyping of KRAS using co-existing mutations revealing novel drug susceptible targets.

In all, our lack of understanding of underlying GFRN molecular mechanisms and intricate signaling pathways, stems our need for enhanced characterization methods such as gene expression signature exploration. Through the utilization of this method, a comprehensive profile of the GFRN, beyond initial mutation status, can begin to be developed and utilized to improve current therapeutic treatments to fight the development of drug resistance observed in lung cancer.

CHAPTER 2

METHODS

2.1 Generation of GFRN-Specific Gene Expression Data

To begin GFRN pathway analysis, previously processed RNA sequencing gene expression data generated from a published study was acquired [6]. Briefly, the cells used to produce the biological replicates were human mammary epithelial cells (HMECs) acquired from non-cancerous breast tissue. HMECs were transfected using recombinant adenovirus of GFRN-specific oncogenes *BAD (BAD), HER2 (ERBB2), IGF1R (IGF1R), RAF (RAF1*), and *KRAS (G12V)* to capture a transcriptional profile of aberrant pathway activity. Cells used to produce the biological replicates were produced using 0.25% serum-free mammary epithelial basal medium (MEBM) in conjunction with a Lonza "bullet kit" as referenced in the protocol [7]. HMECs expressing GFRN oncogenes *BAD (BAD), HER2 (ERBB2), IGF1R (IGF1R), RAF (RAF1)* or *GFP (control)* were incubated for 18 hours to capture the initial transcriptional profile. HMECs transfected with *KRAS (G12V)* along with its *GFP* respective controls were treated for 36 hours. Western blot analysis was then performed using corresponding protein antibodies to each GFRN oncogene to ensure successful overexpression of GFRN oncogenes within HMECs. Following validation mRNA was extracted from cells to generate 6 biological replicates for *BAD (BAD), IGF1R (IGF1R),* and RAF *(RAF1),* with 5 produced for *HER2 (ERBB2)*. For the separately treated HMECs expressing *KRAS (G12V),* 9 biological replicates were produced along with 9 GFP respective controls. The generated biological replicates of the overexpressed GFRN oncogenes from HMECs were then sequenced and aligned computationally using *Rsubread* R package (Version 1.14.2) to produce the gene expression RNA-Seq datasets.

2.2 Obtained RNA Sequencing Datasets

To begin gene signature generation and analysis, various databases were used to acquire the publicly available RNA-sequencing data (Table 2.1). From the National Center for Biotechnology Information (NCBI), Gene Expression Omnibus (GEO), the previously mentioned gene expression data was collected containing the 6 overexpressed GFRN oncogenes and their respective controls from 2 separate datasets [7]. The first dataset included the genes *BAD (BAD), HER2 (ERBB2), IGF1R (IGF1R), RAF (RAF1)*, with the *GFP* samples (the control) treated for 18 hours (GSE83083). The second dataset included the gene KRAS (G12V) with *GFP* samples (the control) treated for 30 hours (GSE83083). From TCGA, 541 LUAD patient tumor samples were collected along with a separate dataset used to classify and specify the cancer type (GSM1536837, GSE62944). Lastly, to perform validation, proteomics data was collected from TCPA.

2.3 Data Refinement

Utilizing the prcomp function from the *stats* R package, the collected gene expression data along with the TCGA patient tumor samples, were visualized using Principal Component Analysis (PCA) within Rstudio (Version 1.2.5019) (Figure 2.1 a-d). PCA is a statistical procedure used to produce principal components representative of the greatest variation occurring in the multidimensional data [12]. The first principal component produced represents the greatest variation, while the second represents the second greatest variation in the multidimensional data and so on (Figure 2.1 a and c) [12]. Due to the datasets being separately processed, significant batch effects and confounding variables were observed (Figure 2.1 a-b). This could be due to many external factors, such as tissue mishandling when producing the samples, varying lab protocols and conditions, as well as human error. Such variability can negativity affect the generation of our signatures and its ability to predict pathway activity within the tumor samples. To begin to reduce variations, the datasets underwent refinement to remove technical artifacts from the gene expression datasets. This included filtering of rows containing a certain percentage of zero values to capture genes with most variance in the dataset. PCA was then utilized throughout the study to ensure optimization of the data and signature generation.

2.4 Batch Adjustment

Following refinement of the RNA seq. data, the significant variances and confounding batch effects were adjusted for using the ComBat function from *sva* R package (Version 3.34.0) and visualized using PCA (Figure 2.1 c-d). This included specifying the gene expression data and patient tumor samples into 3 separate batches and performing a two-step batch adjustment. First, the appropriate training model was specified which included the 6 biological replicates for each oncogene including *BAD (BAD), IGF1R (IGF1R), RAF (RAF1)* and 5 for *HER2 (ERBB2);* along with its 12 *GFP* controls treated for 18 hours (control). The second batch, also specified as the training data, included the 9 biological replicates for *KRAS (G12V)* with its respective 9 *GFP* replicates, pre-treated for 36 hours (control). The first batch adjustment was then performed only including the training data, with the first batch specified as the reference used to compare and optimize data similarity. Following the first adjustment, the third batch was then specified as the 541 LUAD patient tumor samples from TCGA, classified as the test data. The second combat adjustment was then performed using the combat adjusted gene expression data (training data) combined with the TCGA patient tumor sample (test data) with the first batch selected as the reference batch. A PCA was then performed to confirm variances and confounding batch effects were removed to improve data similarity (Figure 2.1 c-d).

2.5 Gene Expression Signature Generation

With the adjusted data, gene expression signatures were generated representing pathway specific GFRN activity. This was performed using the "All-in-one" assign.wrapper function from the "semi-supervised pathway profiling toolkit", *Adaptive Signature and InteGratioN* (*ASSIGN*; Version 1.9.1). Within each pathway-specific gene expression signature, genes quantitatively expressing varying levels of expression were selected by *ASSIGN* to define a phenotypic pattern representative of aberrant GFRN-specific pathway activity. This included creating two distinctive patterns of expression within the signature to represent pathway activity turned on versus pathway activity turned off. For each GFRN specific pathway, this was produced internally by comparing the GFP gene expression data (control) to the specified overexpressed oncogene expression data.

2.6 ASSIGN Gene Expression Signature Output

Various gene lists of specified lengths were then generated ranging from lengths of 5 to 500 genes produced in 5 or 25 gene increments using the assign.wrapper function; utilizing a single pathway setting. The Bayesian variable selection approach was used to select genes expressing the greatest fold-change of differential expression from normal pathway activity to generate the signature. These genes selected displayed the highest signal strength and signal weights representing their possible contribution to the overall development of the disease. Additionally, an anchor gene was selected for the genes as follows *BAD (BAD)*, *HER2 (ERBB2)*, *IGF1R (IGF1R) RAF(RAF1),* and K*RAS (KRAS).* This ensures the overexpressed oncogene specific to the pathway being investigated is included in each gene signature output. Additional *ASSIGN* criteria were also specified including adaptive signature background parameters. This included the adaptive_B = TRUE, default parameter, which allows *ASSIGN* to adjust the test data baseline measures. Next, adaptive_S = FALSE was specified, preventing the adaptability of the gene signatures to adhere to the test data. Additional default parameters were also included specifying probability measures such as p_beta = 0.01, theta0=0.05, theta1=0.9. Next, the iteration was increased from the default parameter of iter = 2,000 to iter = 100,000 to increase the number of Markov Chain Monte Carlo (MCMC) simulations. Lastly, the number of burn-in iterations was increased from the default of burn_in = 1,000 to burn_in = 50,000 to optimize gene signature output. From the produced output, those that passed the internal leave-one-out cross validation (LOOCV) then underwent external validation using proteomics and gene expression data.

2.7 External Validation

Using the cor.test function from the *stats* package (Version 4.0.3) correlations were performed to validate the generated pathway activation estimates from ASSIGN. First, using proteomics data, Pearson pairwise correlations were calculated between Reverse Phase Protein Array (RPPA) data from The Cancer Proteome Atlas (TCPA) with the generated pathway activation estimates. This was performed using the cor.test function from the R *stats* package (Version 4.0.3), using the Pearson method. Pathway activation estimates were considered to be validated if the "Pearson's product moment", calculated using a 95% confidence interval, had a p-adjusted value of $\leq$ 0.002. The p-adjusted value was calculated due to the high quantity of TCGA patient tumor samples. The same parameters and cor.test function were used to validate the pathway activation predictions correlated to the TCGA patient tumor sample gene expression data. Lastly, using the function boxplot2 from the package *gplots* (Version 3.1.1), boxplots were produced expressing predicted pathway activity levels within the TCGA patient tumor samples. The data was first scaled to optimize boxplot generation along with specification of pathway activity levels by low, intermediate, and high percentiles. Samples with expression in the 10th percentile or below were classified as “low” expressing. Samples with expression in the 90th percentile or above were classified as “high” expressing. Samples with the expression above the 10th percentile and below the 90th percentile were classified as “intermediate” expressing samples. Pathway-specific boxplots were considered to be validated if higher predicted pathway activity could be seen within the patient tumor samples categorized in the “high” expressing percentile in comparison to the “intermediate” and “low” expressing percentiles.

CHAPTER 3

RESULTS

3.1 Pathway-Specific Gene Expression Signature Generation

With the use of RNA sequencing data of HMECs overexpressing GFRN oncogenes, gene expression signatures of varying gene list lengths were generated using Rstudio (Version 1.2.5019) (Table 3.1-3.5). Pathway activation estimates were also produced by projecting the signatures onto the 541 LUAD patient tumor samples to predict levels of pathway activity. These signatures were produced by comparing the overexpressing HMECs to its respective GFP (control) HMEC samples. To ensure the signatures’ ability to capture the levels of pathway activity are expressed within the HMEC samples, pathway-specific cross-validation scatterplots of the training data was assessed. Produced scatterplots of each GFRN pathway-specific oncogene that accurately displayed low levels or no level of pathway activity for GFP (control) versus high levels of activity for the overexpressed GFRN HMECs were considered to be internally validated. This included the gene lists lengths with the corresponding GFRN pathway being investigated as follows *BAD (BAD)*, 475; *HER2 (ERBB2)*, 5; *IGF1R (IGF1R)*, 25; *RAF (RAF1)*, 275; and *KRAS (KRAS, G12V)*, 500 (Table 3.1-3.5). External validation was then performed using proteomics and gene expression data to determine if the generated gene expression signatures accurately predicted levels of pathway activity within the LUAD patient tumor samples from TCGA.

3.2 Proteomics Validation

First, using proteomics data from TCPA pathway activation estimates were validated through statistical analysis. This included performing Pearson pairwise correlations between the produced pathway-specific gene expression signatures and their predicted pathway activity to RPPA protein expression data from TCPA (Table 3.6). For the signature validation of BAD, the TCPA protein expression of PDK1_pS241 phosphoprotein was correlated to the predicted levels of pathway activation for BAD. Due to the upstream signaling of PDK1 leading to the activation AKT which in turn inhibits BAD, negative correlations were observed as anticipated. Strongest negative correlations for BAD were most optimally seen using the 475-gene signature list (cor = -0.247206, p-value = 1.63E-06, optimal gene list = 475). For the signature validation of HER2, the phosphoprotein HER2_pY1248 showed a strong positive correlation to the predicted pathway activity using the 5-gene signature list (cor = 0.3180165, p-value =4.54E-10, optimal gene list = 5). Next, for RAF the phosphoprotein of CRAF_pS338 showed a significant positive correlation using the 275-signature gene list (cor = 0.3176497, p-value = 4.77E-10, optimal gene list = 275). Lastly, for the signature validation of KRAS the phospho-protein MEK1_pS217S221 was utilized due to downstream activation of MEK1 as a consequence of KRAS upstream activation. The highest positive correlation was observed using the KRAS 500-gene signature list (cor =

0.1643924, p-value = 0.001577, optimal gene list = 500). All gene expression signatures were able to be validated using protein expression levels, except for IGF1R, as referenced in Table 3.6.

3.3 Gene expression Validation

Next, Pearson pairwise correlations were performed between the signature predicted pathway activity of the respective GFRN pathway to the expression levels of the gene of interest within the LUAD patient tumor samples from TCGA (Table 3.7). For the validation of BAD, the estimated pathway levels predicted by the BAD 475- gene signature showed a positive correlation to the patient samples expressing higher levels of bad activity indicating accurate signature predictability (cor = 0.1127843, p-value = 0.008649, optimal gene list = 475). Next, for HER2 validation, the 5-gene signature showed a strong positive correlation to HER2 mutated levels of activity within the patient tumor samples (cor = 0.4114047, p-value = < 2.2e-16, optimal gene list = 5). Lastly, IGF1R was validated using the IGF1R oncogene test gene expression with the strongest positive correlation being seen using the 25-gene signature list (cor = 0.178464, p-value = 2.98E-05, optimal gene list = 25). Overall, with the corresponding oncogene expression from the patient tumor samples, the pathways BAD, HER2, and IGF1R were validated with the exceptions of RAF and KRAS, summarized in Table 3.7.

3.4 Gene Expression Boxplot Validation

Additionally, gene expression box plots were generated to distinguish levels of pathway activity within patient tumor samples using the predicted pathway activity levels from the gene expression signatures (Figure 3.1). As mentioned, prior, patient tumor samples were classified into “low”, “intermediate”, and “high” percentiles based upon their levels of expression. As summarized in Table 3.8 and Figure 3.1, this method was able to validate the GFRN pathways BAD, HER2, IGF1R with the exception of RAF and KRAS.

3.5 Optimal Gene Signature Selection

In all, optimal gene list lengths were determined through statistical analysis by cross referencing proteomics and gene expression correlations (Table 3.8). For the GFRN pathway BAD, proteomics, gene expression, and gene expression box plots validated the 475-signature gene list (Table 3.1). For HER2, all three methods were also used to validate the HER2's 5-signature gene list (Table 3.2). Next, for IGF1R, only the gene expression and generated gene expression boxplot was used for validation of the 25-signature gene list (Table 3.3). For the GFRN RAF, only protein expression was used for the validation of its 275- signature gene list (Table 3.4). Lastly, for KRAS, only protein expression was used for the validation of the 500-signature gene list (Table 3.5).

CHAPTER 4

DISCUSSION

4.1 Significance of Findings and Future Implications

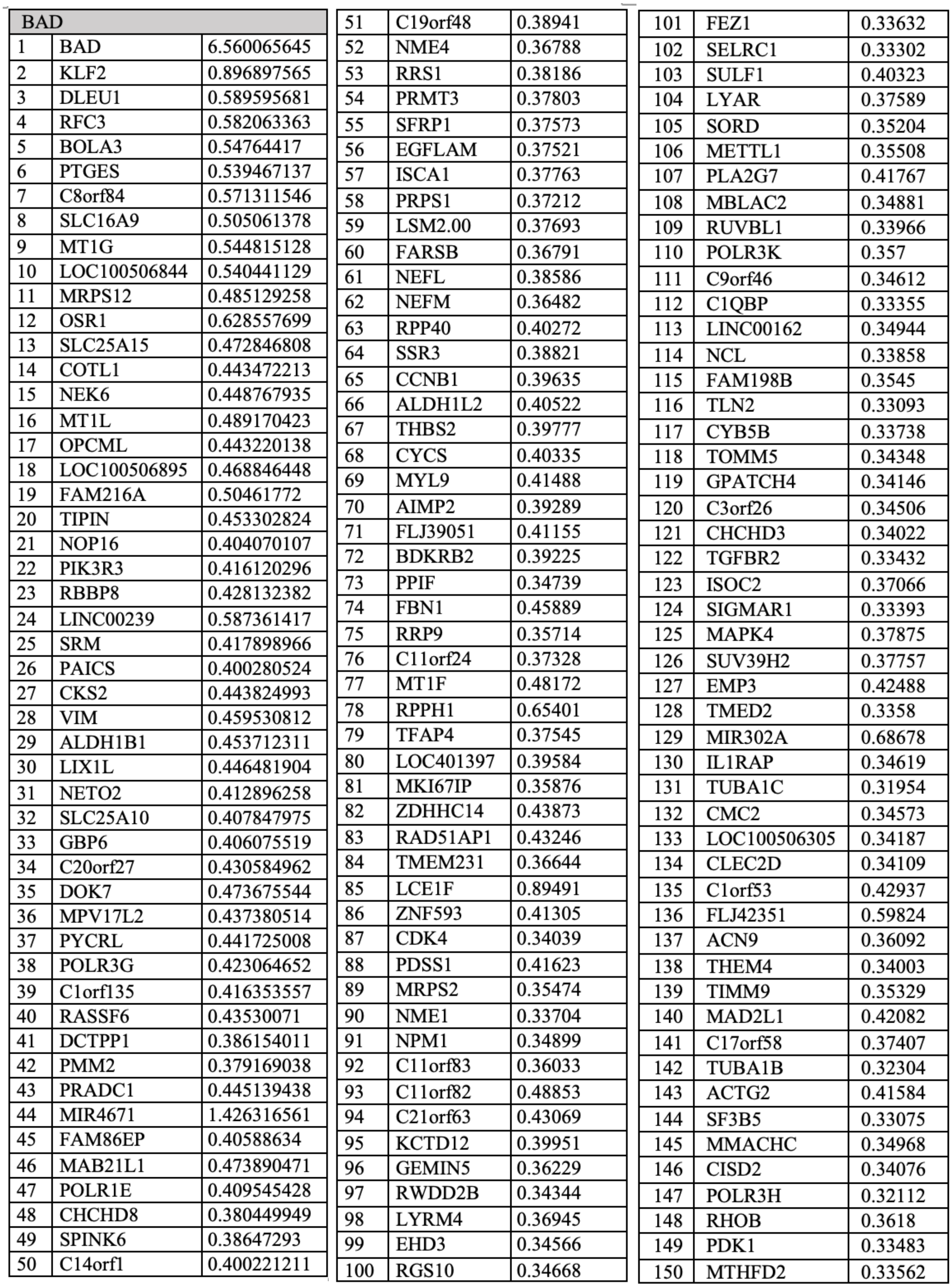
In this study, GFRN-specific gene expression signatures, represented of aberrant pathway activity, were generated to interrogate GFRN pathway activity within lung tumors. Optimal gene expression signatures were then determined for the GFRN pathways *BAD (BAD), HER2 (ERBB2), IGF1R (IGF1R), RAF (RAF1),* and *KRAS (G12V)* using proteomics and gene expression data (Figure 4.1). For the signatures *HER2(ERBB2), IGF1R(IGF1R), RAF(RAF1)* and *KRAS(G12V),* predicted pathway activity showed a positive correlation with downstream protein expression levels, indicating downstream pathway activation of the investigated pathways. For the signature BAD, protein expression representing downstream activation of the AKT pathway, activated upstream by PDK1, showed corresponding negative correlations indicating inhibition of the BAD pathway activity, as anticipated. Next, corresponding higher levels of gene expression were observed in HER2 and IGF1R when correlated with mutated levels of gene expression supporting aberrant pathway activation of the two pathways. In addition, upregulated levels of AKT pathway activity were used to validate BAD's signature representing abnormal pathway activity, in which negative correlations were seen, accurately depicting the inhibition of BAD by AKT activation. In addition, boxplots were used to validate signature generation for the pathways BAD, HER2, and IGF1R. A percentage of the tumor samples were distinguished to have higher levels of pathway activity signifying the gene expression signatures ability to characterize mutated levels of pathway activity. In all, it was concluded that the generated GFRN-pathway specific gene expression signatures, representative of aberrant GFRN activity, accurately distinguished higher levels of pathway activity within LUAD patient tumor samples.

In future studies, a multiple pathway analysis will be performed using the generated gene expression signatures to begin to comprehend underlying molecular mechanisms of the GFRN. Through the projection of these signatures, simultaneously onto lung cancer cell lines, hierarchical clustering can be utilized to reveal patterns of gene expression. These gene expression patterns, or phenotypic patterns, can be characterized to reveal drug sensitive or resistant phenotypes by performing drug response predictions. Potential intrinsic subtypes could also be revealed exposing sensitivity patterns within this complex network. Overall, with the use of multiple-pathway analysis with the GFRN pathway-specific gene expression signatures, a potential comprehensive profile of the GFRN can be built to reveal novel phenotypic patterns and identify drug sensitivities. This in turn, can be used to enhance prognostic, diagnostic, and therapeutic treatment decisions against lung cancer, overall enhancing precision medicine approaches to combat drug resistance development.

Table 2.1 Publicly available datasets acquired for gene signature generation and analysis consisting of gene expression signature data along with LUAD patient tumor samples and proteomics validation dataset.

Dataset	**Source**	**Content**
Accession GSE83083	NCBI GEO	Gene expression data of overexpressed HMECs • GFP18: 6 controls IGF1R: 6 samples • BAD (BAD): 6 samples RAF (RAF1): 6 samples • HER2(ERBB2): 5 samples
Accession GSE83083	NCBI GEO	Gene expression data of overexpressed HMECs • GFP30: 9 controls • KRAS_GV (G12V): 9 samples
Accession GSE59765	NCBI GEO	Gene expression data of overexpressed HMECs: • Control: 6 EGFR controls • EGFR (EGR1): 6 samples
Accession GSM1536837	NCBI GEO	TCGA Patient Tumor Samples gene expression: • LUAD: 541 samples
Accession GSE62944	NCBI GEO	TCGA Cancer Type Samples TCGA tumor sample barcode with corresponding sample classification.
______	TCPA	Proteomics expression levels of corresponding GFRN downstream pathway activations.

Table 3.1 Optimal signature gene list generated for BAD pathway listing all 475 genes and their associated weight in the signature in predicting BAD pathway activity.



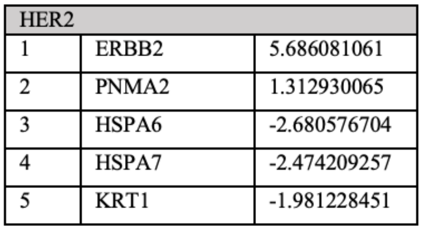
BAD		
1	BAD	6.560065645
2	KLF2	0.896897565
3	DLEU1	0.589595681
4	RFC3	0.582063363
5	BOLA3	0.54764417
6	PTGES	0.539467137
7	C8orf84	0.571311546
8	SLC16A9	0.505061378
9	MT1G	0.544815128
10	LOC100506844	0.540441129
11	MRPS12	0.485129258
12	OSR1	0.628557699
13	SLC25A15	0.472846808
14	COTL1	0.443472213
15	NEK6	0.448767935
16	MT1L	0.489170423
17	OPCML	0.443220138
18	LOC100506895	0.468846448
19	FAM216A	0.50461772
20	TIPIN	0.453302824
21	NOP16	0.404070107
22	PIK3R3	0.416120296
23	RBBP8	0.428132382
24	LINC00239	0.587361417
25	SRM	0.417898966
26	PAICS	0.400280524
27	CKS2	0.443824993
28	VIM	0.459530812
29	ALDH1B1	0.453712311
30	LIX1L	0.446481904
31	NETO2	0.412896258
32	SLC25A10	0.407847975
33	GBP6	0.406075519
34	C20orf27	0.430584962
35	DOK7	0.473675544
36	MPV17L2	0.437380514
37	PYCRL	0.441725008
38	POLR3G	0.423064652
39	C1orf135	0.416353557
40	RASSF6	0.43530071
41	DCTPP1	0.386154011
42	PMM2	0.379169038
43	PRADC1	0.445139438
44	MIR4671	1.426316561
45	FAM86EP	0.40588634
46	MAB21L1	0.473890471
47	POLR1E	0.409545428
48	CHCHD8	0.380449949
49	SPINK6	0.38647293
50	C14orf1	0.400221211
51	C19orf48	0.38941
52	NME4	0.36788
53	RRS1	0.38186
54	PRMT3	0.37803
55	SFRP1	0.37573
56	EGFLAM	0.37521
57	ISCA1	0.37763
58	PRPS1	0.37212
59	LSM2.00	0.37693
60	FARSB	0.36791
61	NEFL	0.38586
62	NEFM	0.36482
63	RPP40	0.40272
64	SSR3	0.38821
65	CCNB1	0.39635
66	ALDH1L2	0.40522
67	THBS2	0.39777
68	CYCS	0.40335
69	MYL9	0.41488
70	AIMP2	0.39289
71	FLJ39051	0.41155
72	BDKRB2	0.39225
73	PPIF	0.34739
74	FBN1	0.45889
75	RRP9	0.35714
76	C11orf24	0.37328
77	MT1F	0.48172
78	RPPH1	0.65401
79	TFAP4	0.37545
80	LOC401397	0.39584
81	MKI67IP	0.35876
82	ZDHHC14	0.43873
83	RAD51AP1	0.43246
84	TMEM231	0.36644
85	LCE1F	0.89491
86	ZNF593	0.41305
87	CDK4	0.34039
88	PDSS1	0.41623
89	MRPS2	0.35474
90	NME1	0.33704
91	NPM1	0.34899
92	C11orf83	0.36033
93	C11orf82	0.48853
94	C21orf63	0.43069
95	KCTD12	0.39951
96	GEMIN5	0.36229
97	RWDD2B	0.34344
98	LYRM4	0.36945
99	EHD3	0.34566
100	RGS10	0.34668
101	FEZ1	0.33632
102	SELRC1	0.33302
103	SULF1	0.40323
104	LYAR	0.37589
105	SORD	0.35204
106	METTL1	0.35508
107	PLA2G7	0.41767
108	MBLAC2	0.34881
109	RUVBL1	0.33966
110	POLR3K	0.357
111	C9orf46	0.34612
112	C1QBP	0.33355
113	LINC00162	0.34944
114	NCL	0.33858
115	FAM198B	0.3545
116	TLN2	0.33093
117	CYB5B	0.33738
118	TOMM5	0.34348
119	GPATCH4	0.34146
120	C3orf26	0.34506
121	CHCHD3	0.34022
122	TGFBR2	0.33432
123	ISOC2	0.37066
124	SIGMAR1	0.33393
125	MAPK4	0.37875
126	SUV39H2	0.37757
127	EMP3	0.42488
128	TMED2	0.3358
129	MIR302A	0.68678
130	IL1RAP	0.34619
131	TUBA1C	0.31954
132	CMC2	0.34573
133	LOC100506305	0.34187
134	CLEC2D	0.34109
135	C1orf53	0.42937
136	FLJ42351	0.59824
137	ACN9	0.36092
138	THEM4	0.34003
139	TIMM9	0.35329
140	MAD2L1	0.42082
141	C17orf58	0.37407
142	TUBA1B	0.32304
143	ACTG2	0.41584
144	SF3B5	0.33075
145	MMACHC	0.34968
146	CISD2	0.34076
147	POLR3H	0.32112
148	RHOB	0.3618
149	PDK1	0.33483
150	MTHFD2	0.33562

BAD		
151	FKSG29	0.33031
152	GAPDH	0.30838
153	CDC20	0.41351
154	LHFP	0.34696
155	POP7	0.30358
156	COQ2	0.33965
157	CDT1	0.36604
158	ORC6	0.40206
159	MRPL17	0.30942
160	CT62	0.42331
161	RWDD1	0.34693
162	RHOJ	0.34622
163	PPP1R14A	0.41562
164	RPSAP52	0.44641
165	MYL7	0.39286
166	PPAT	0.34922
167	MRPL3	0.31069
168	TMEM241	0.3439
169	NDUFAF2	0.33534
170	TMEM5	0.33865
171	ERVMER34-1	0.35836
172	RANBP1	0.30885
173	PRKCDBP	0.46157
174	PDXP	0.34874
175	NAT14	0.32723
176	SUMO3	0.30041
177	NME2	0.37872
178	EIF4EBP1	0.33727
179	LRRTM2	0.36414
180	DCBLD2	0.32848
181	ENC1	0.30618
182	CLN6	0.30249
183	NOB1	0.3145
184	TRIB3	0.30554
185	NEGR1	0.36164
186	TIMM17A	0.32168
187	GJA5	0.34432
188	PFDN2	0.30095
189	SLC25A38	0.30684
190	ZNF689	0.30237
191	RABEPK	0.30724
192	C1orf51	0.34836
193	ACAT2	0.31639
194	LSM4.00	0.30677
195	UBIAD1	0.30868
196	MRPL30	0.30089
197	UBE2N	0.31953
198	PADI3	0.37054
199	IMP4	0.29509
200	MEST	0.31219
201	RRP15	0.31651
202	KDELR2	0.29509
203	PNO1	0.33598
204	METTL5	0.29712
205	LTV1	0.31921
206	MRPL12	0.35476
207	SNRPF	0.33101
208	APRT	0.28466
209	LPAR1	0.33898
210	ATP5E	0.4379
211	IGFBP5	0.3668
212	FAM58A	0.30059
213	GTF3A	0.2931
214	ARHGAP18	0.3147
215	CBY1	0.3006
216	GEMIN6	0.33669
217	NHP2L1	0.28701
218	NOL10	0.27936
219	PPIL1	0.30875
220	TUBB	0.28837
221	STAMBPL1	0.3485
222	CKS1B	0.38696
223	RAB36	0.31147
224	PRR11	0.34222
225	ZNF556	0.42574
226	TIMM13	0.29779
227	PUS7	0.29613
228	B7H6	0.30915
229	CHAC2	0.32236
230	RRM2	0.3558
231	PEMT	0.31289
232	THAP4	0.30744
233	MRRF	0.29402
234	ECE2	0.30208
235	PEX3	0.29952
236	PINX1	0.33205
237	TSPAN1	0.28017
238	HSPA7	-2.8076
239	IL8	-2.3871
240	HSPA1A	-2.4556
241	DNAJA4	-2.1219
242	HSPA1B	-2.0748
243	KRT1	-1.6717
244	CCL20	-1.5741
245	FOXQ1	-1.5098
246	GDF15	-1.488
247	CXCL5	-1.4311
248	KRTDAP	-1.4195
249	CRYAB	-1.4108
250	SBSN	-1.424
251	KRT10	-1.3727
252	IL17C	-1.3045
253	KRT23	-1.227
254	DSG1	-1.2722
255	CFB	-1.2321
256	TNFAIP2	-1.2318
257	EGR1	-1.1879
258	DUSP2	-1.1708
259	FOS	-1.1559
260	CXCL2	-1.1492
261	SAA2	-1.1292
262	NPR3.00	-1.1545
263	BNIPL	-1.0906
264	GRHL1	-1.0445
265	S100A7	-1.0601
266	ATF3	-1.0605
267	DLC1	-0.9997
268	NFKBIZ	-0.9868
269	FAM25A	-1.0538
270	CCL2	-0.9873
271	SGPP2	-0.9625
272	MYO5C	-0.9297
273	CXCL3	-0.9168
274	PRSS22	-0.9152
275	BMF	-0.8954
276	LCE3D	-1.012
277	C10orf99	-0.9154
278	ERRFI1	-0.8882
279	MMP3	-0.886
280	LMO1	-0.8753
281	LIF	-0.8663
282	ATHL1	-0.8646
283	IGFBP3	-0.8796
284	FOSB	-0.8346
285	TMEM45B	-0.8517
286	GABRE	-0.8566
287	CDRT1	-0.8513
288	RRAD	-0.8617
289	STON1	-0.8495
290	AKAP12	-0.8401
291	EGR3	-0.8248
292	TMPRSS13	-0.8337
293	CXCL6	-0.8438
294	LYPD3	-0.8503
295	INHBA	-0.8284
296	DUSP1	-0.821
297	GSDMC	-0.8151
298	IFNK	-0.8217
299	IL20	-0.8786
300	EPHB6	-0.8072

BAD		
301	DSC1	-0.8057
302	PDZK1IP1	-0.8027
303	HSP90AA1	-0.8045
304	CXCL1	-0.804
305	ZFAND2A	-0.7836
306	MMP7	-0.788
307	PLA2G4F	-0.8051
308	GRB7	-0.7733
309	HMOX1	-0.7868
310	SELENBP1	-0.7571
311	GSDMB	-0.763
312	BIRC3	-0.7478
313	OVOL1	-0.7565
314	PIM1	-0.7501
315	SLC34A2	-0.7465
316	GAB2	-0.732
317	PPP2R2C	-0.7468
318	NPNT	-0.7415
319	LTF	-0.728
320	HSPH1	-0.7268
321	HSP90AA4P	-0.7413
322	FERMT3	-0.7261
323	LCN2	-0.7174
324	AQP3	-0.7242
325	KLHL24	-0.7114
326	GLCCI1	-0.7006
327	BAG3	-0.7075
328	DEDD2	-0.6949
329	DAPK1	-0.6973
330	HSPB8	-0.7052
331	KRT80	-0.7012
332	TNFRSF11B	-0.6894
333	DNAJB4	-0.6871
334	NRARP	-0.6816
335	DUSP6	-0.6732
336	MUM1L1	-0.6726
337	TIAM2	-0.6649
338	CDKN1A	-0.664
339	OLFM4	-0.6646
340	EEF1A2	-0.6776
341	ID4	-0.6527
342	BCORL1	-0.6469
343	PLA2G4C	-0.6518
344	TLR2	-0.6475
345	C1orf63	-0.6463
346	SLC28A3	-0.6888
347	PRDM1	-0.6399
348	LOC100288077	-0.6733
349	ETS1	-0.637
350	OXTR	-0.6451
351	DFNB31	-0.6395
352	OLFML2A	-0.6429
353	IFRD1	-0.6401
354	CAPNS2	-0.6442
355	FBXW10	-0.6439
356	PVRL4	-0.6327
357	STARD13	-0.6356
358	GGT6	-0.6407
359	SLCO4A1	-0.6252
360	TGM1	-0.6278
361	MMP13	-0.7688
362	LOC146880	-0.6184
363	C17orf103	-0.6124
364	NFKBID	-0.6401
365	IER5	-0.6137
366	SLC5A1	-0.6063
367	C3	-0.598
368	PNLDC1	-0.6361
369	IER3	-0.5992
370	BIK	-0.6147
371	DUSP5	-0.5985
372	GDF6	-0.5888
373	ERBB3	-0.6068
374	FAM43A	-0.7264
375	FNIP2	-0.5858
376	SAA1	-0.5939
377	EDN2	-0.8265
378	ALDH2	-0.6367
379	DNER	-0.5793
380	ZC3H12A	-0.5738
381	OTUD1	-0.6074
382	TNFSF14	-0.5767
383	GPRC5A	-0.5718
384	NYNRIN	-0.5679
385	ENGASE	-0.5653
386	PDZD2	-0.5557
387	PPP1R15A	-0.5572
388	NCF2	-0.6592
389	MIR614	-0.7163
390	PARM1	-0.588
391	SATB1	-0.5729
392	C5orf41	-0.5719
393	NLRP10	-0.6026
394	MIR5047	-0.6988
395	LY6D	-0.654
396	PTGS2	-0.6109
397	TRERF1	-0.5597
398	GM2A	-0.5555
399	HERC2P2	-0.5816
400	MXD1	-0.5437
401	CITED2	-0.546512043
402	DNAJC6	-0.54131458
403	KLF6	-0.543920213
404	ANG	-0.572609921
405	TMEM2	-0.547173357
406	ABCA1	-0.590029574
407	PLEKHM1P	-0.541015423
408	IL7R	-0.560042503
409	RB1CC1	-0.5416114
410	LIMCH1	-0.53602323
411	JHDM1D	-0.54938032
412	LOC283174	-0.538117539
413	SOCS3	-0.53187086
414	MAFF	-0.558654828
415	PLEKHA6	-0.523475627
416	C6orf141	-0.529266397
417	GCNT2	-0.539075823
418	SEMA6C	-0.606539153
419	CLDN4	-0.546292088
420	FBXL20	-0.531235542
421	DNAJB1	-0.521818797
422	TSLP	-0.589758944
423	MB21D1	-0.666220105
424	PRODH	-0.634864813
425	TBC1D8	-0.527704318
426	FAM59A	-0.558627458
427	PCDH7	-0.534398737
428	SERPINB4	-0.670101339
429	AGAP11	-0.51546475
430	PVRIG	-0.63926593
431	SEC31B	-0.578292306
432	SLC6A14	-0.565985459
433	TTC9	-0.530293846
434	CACHD1	-0.501482911
435	BTN2A3P	-0.52425662
436	CA8	-0.63590865
437	MGAT4A	-0.54686226
438	HBEGF	-0.539871734
439	INSR	-0.513491805
440	BMP6	-0.738044179
441	SLC24A6	-0.496981111
442	LPL	-0.501864709
443	MYH14	-0.562429327
444	FILIP1L	-0.52264068
445	CYP24A1	-0.511903219
446	ULK1	-0.633205267
447	PSAPL1	-0.53763688
448	EFS	-0.57825175
449	PROC	-0.563859271
450	ZNF488	-0.541178786

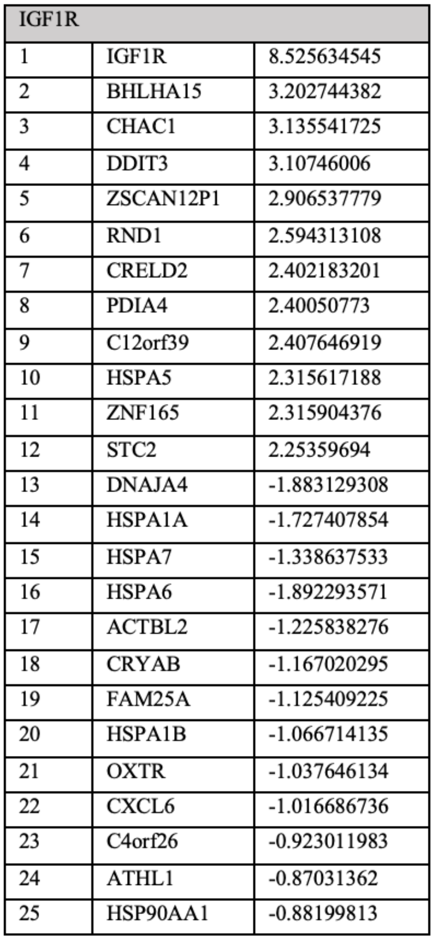
BAD		
451	VAV3	-0.492
452	JMY	-0.5054
453	CNNM3	-0.4932
454	HRH1	-0.5087
455	SLC38A2	-0.5098
456	CEBPA	-0.7024
457	DDIT3	-0.4963
458	ABTB2	-0.5134
459	ARID5B	-0.4783
460	IRAK2	-0.5479
461	BRD3	-0.4808
462	SOD2	-0.485
463	LRG1	-0.5531
464	FGF2	-0.5402
465	DNASE1L2	-0.5879
466	ARHGEF10L	-0.5666
467	ZNF217	-0.476
468	LOC100292680	-0.4751
469	EPHA4	-0.4689
470	IL17RB	-0.5387
471	C7orf53	-0.5596
472	ARHGAP19	-0.5219
473	ZSWIM4	-0.4857
474	YPEL3	-0.559
475	RAD21	-0.4625

Table 3.2 Optimal signature gene list generated for HER2 pathway listing all 5 genes and their associated weight in the signature in predicting HER2 pathway activity.



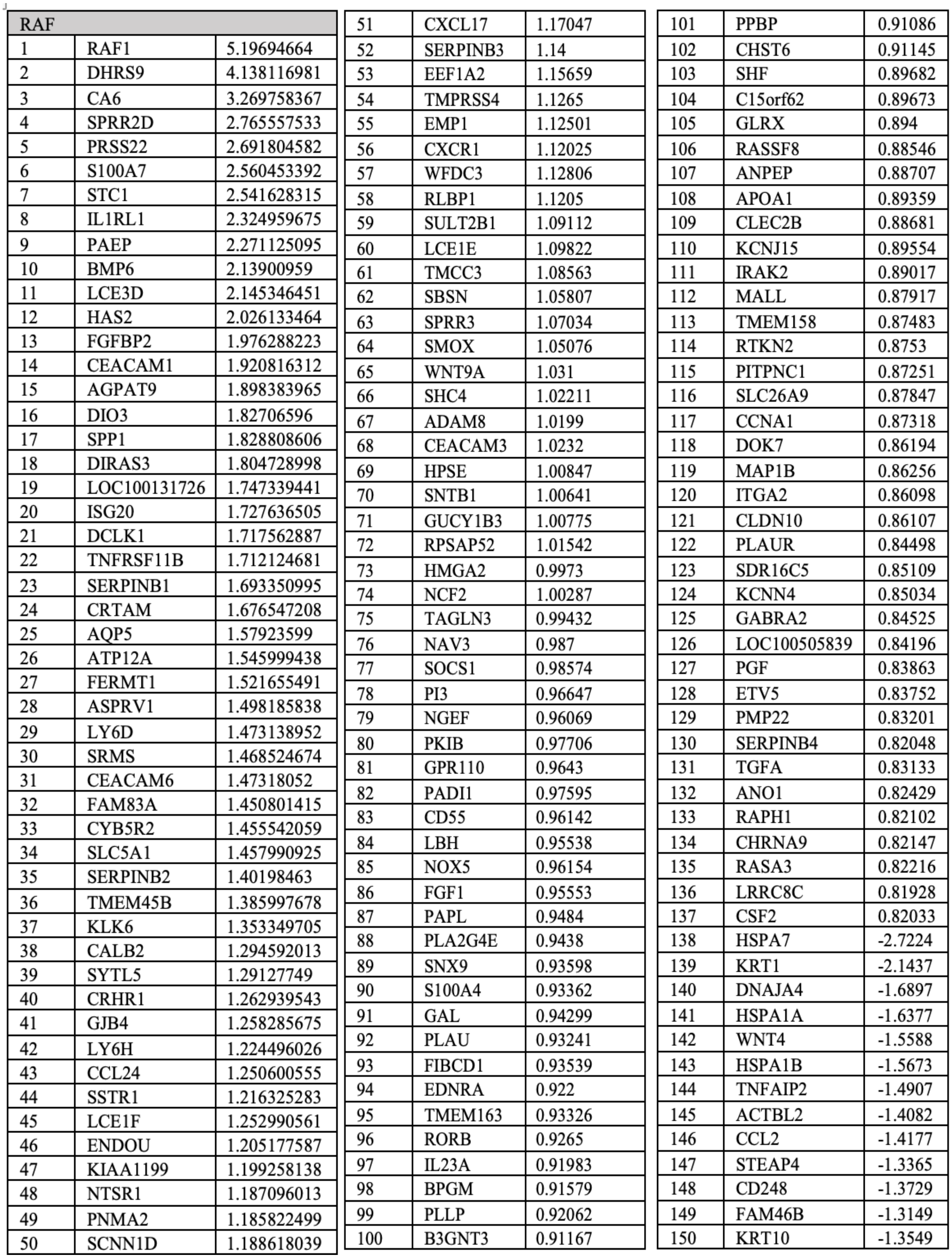
HER2		
1	ERBB2	5.686081061
2	PNMA2	1.312930065
3	HSPA6	-2.680576704
4	HSPA7	-2.474209257
5	KRT1	-1.981228451

Table 3.3 Optimal signature gene list generated for IGF1R pathway listing all 25 genes and their associated weight in the signature in predicting IGF1R pathway activity.



IGF1R		
1	IGF1R	8.525634545
2	BHLHA15	3.202744382
3	CHAC1	3.135541725
4	DDIT3	3.10746006
5	ZSCAN12P1	2.906537779
6	RND1	2.594313108
7	CRELD2	2.402183201
8	PDIA4	2.40050773
9	C12orf39	2.407646919
10	HSPA5	2.315617188
11	ZNF165	2.315904376
12	STC2	2.25359694
13	DNAJA4	-1.883129308
14	HSPA1A	-1.727407854
15	HSPA7	-1.338637533
16	HSPA6	-1.892293571
17	ACTBL2	-1.225838276
18	CRYAB	-1.167020295
19	FAM25A	-1.125409225
20	HSPA1B	-1.066714135
21	OXTR	-1.037646134
22	CXCL6	-1.016686736
23	C4orf26	-0.923011983
24	ATHL1	-0.87031362
25	HSP90AA1	-0.88199813

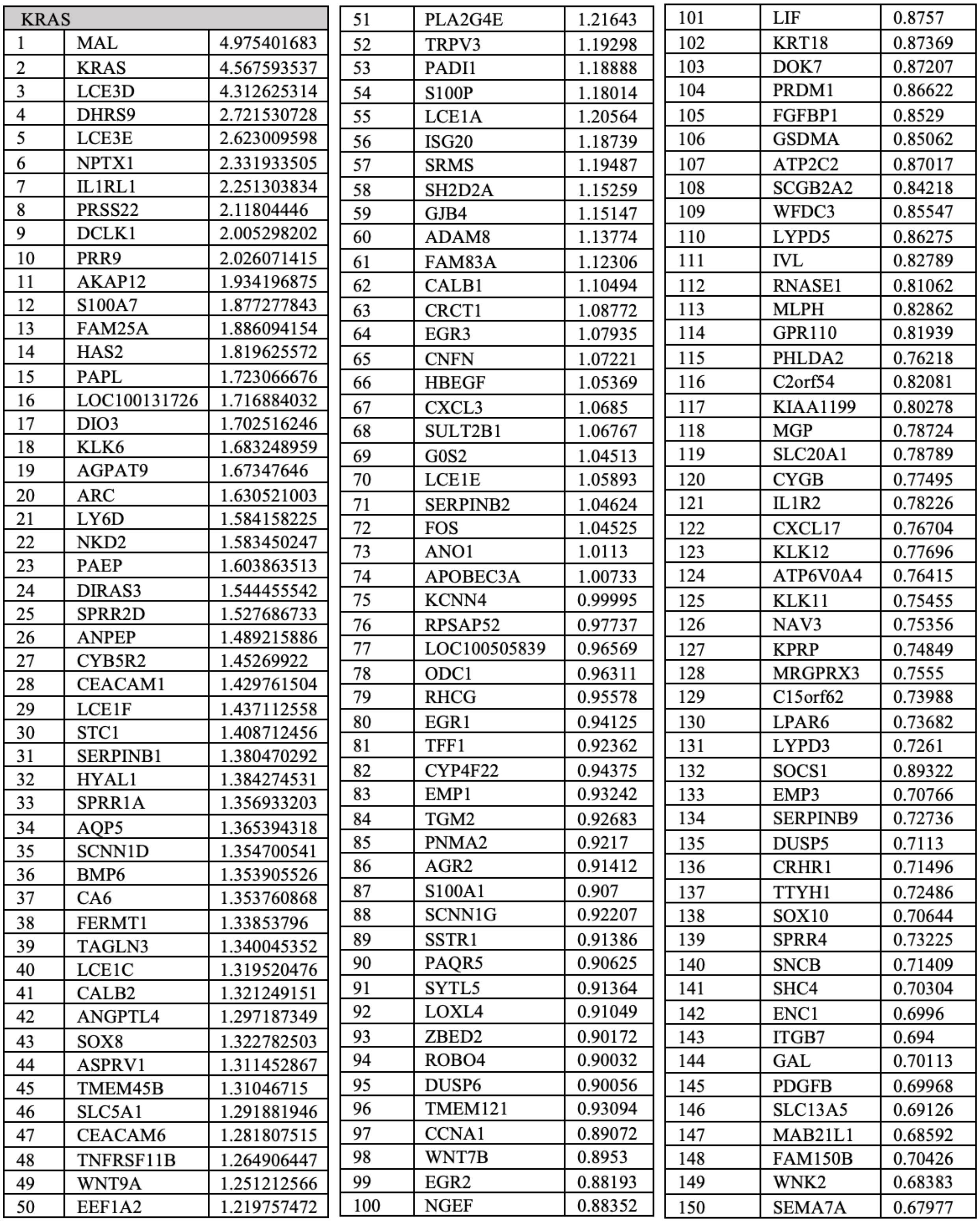
Table 3.4 Optimal signature gene list generated for RAF pathway listing all 275 genes and their associated weight in the signature in predicting RAF pathway activity.



RAF		
1	RAF1	5.19694664
2	DHRS9	4.138116981
3	CA6	3.269758367
4	SPRR2D	2.765557533
5	PRSS22	2.691804582
6	S100A7	2.560453392
7	STC1	2.541628315
8	IL1RL1	2.324959675
9	PAEP	2.271125095
10	BMP6	2.13900959
11	LCE3D	2.145346451
12	HAS2	2.026133464
13	FGFBP2	1.976288223
14	CEACAM1	1.920816312
15	AGPAT9	1.898383965
16	DIO3	1.82706596
17	SPP1	1.828808606
18	DIRAS3	1.804728998
19	LOC100131726	1.747339441
20	ISG20	1.727636505
21	DCLK1	1.717562887
22	TNFRSF11B	1.712124681
23	SERPINB1	1.693350995
24	CRTAM	1.676547208
25	AQP5	1.57923599
26	ATP12A	1.545999438
27	FERMT1	1.521655491
28	ASPRV1	1.498185838
29	LY6D	1.473138952
30	SRMS	1.468524674
31	CEACAM6	1.47318052
32	FAM83A	1.450801415
33	CYB5R2	1.455542059
34	SLC5A1	1.457990925
35	SERPINB2	1.40198463
36	TMEM45B	1.385997678
37	KLK6	1.353349705
38	CALB2	1.294592013
39	SYTL5	1.29127749
40	CRHR1	1.262939543
41	GJB4	1.258285675
42	LY6H	1.224496026
43	CCL24	1.250600555
44	SSTR1	1.216325283
45	LCE1F	1.252990561
46	ENDOU	1.205177587
47	KIAA1199	1.199258138
48	NTSR1	1.187096013
49	PNMA2	1.185822499
50	SCNN1D	1.188618039
51	CXCL17	1.17047
52	SERPINB3	1.14
53	EEF1A2	1.15659
54	TMPRSS4	1.1265
55	EMP1	1.12501
56	CXCR1	1.12025
57	WFDC3	1.12806
58	RLBP1	1.1205
59	SULT2B1	1.09112
60	LCE1E	1.09822
61	TMCC3	1.08563
62	SBSN	1.05807
63	SPRR3	1.07034
64	SMOX	1.05076
65	WNT9A	1.031
66	SHC4	1.02211
67	ADAM8	1.0199
68	CEACAM3	1.0232
69	HPSE	1.00847
70	SNTB1	1.00641
71	GUCY1B3	1.00775
72	RPSAP52	1.01542
73	HMGA2	0.9973
74	NCF2	1.00287
75	TAGLN3	0.99432
76	NAV3	0.987
77	SOCS1	0.98574
78	PI3	0.96647
79	NGEF	0.96069
80	PKIB	0.97706
81	GPR110	0.9643
82	PADI1	0.97595
83	CD55	0.96142
84	LBH	0.95538
85	NOX5	0.96154
86	FGF1	0.95553
87	PAPL	0.9484
88	PLA2G4E	0.9438
89	SNX9	0.93598
90	S100A4	0.93362
91	GAL	0.94299
92	PLAU	0.93241
93	FIBCD1	0.93539
94	EDNRA	0.922
95	TMEM163	0.93326
96	RORB	0.9265
97	IL23A	0.91983
98	BPGM	0.91579
99	PLLP	0.92062
100	B3GNT3	0.91167
101	PPBP	0.91086
102	CHST6	0.91145
103	SHF	0.89682
104	C15orf62	0.89673
105	GLRX	0.894
106	RASSF8	0.88546
107	ANPEP	0.88707
108	APOA1	0.89359
109	CLEC2B	0.88681
110	KCNJ15	0.89554
111	IRAK2	0.89017
112	MALL	0.87917
113	TMEM158	0.87483
114	RTKN2	0.8753
115	PITPNC1	0.87251
116	SLC26A9	0.87847
117	CCNA1	0.87318
118	DOK7	0.86194
119	MAP1B	0.86256
120	ITGA2	0.86098
121	CLDN10	0.86107
122	PLAUR	0.84498
123	SDR16C5	0.85109
124	KCNN4	0.85034
125	GABRA2	0.84525
126	LOC100505839	0.84196
127	PGF	0.83863
128	ETV5	0.83752
129	PMP22	0.83201
130	SERPINB4	0.82048
131	TGFA	0.83133
132	ANO1	0.82429
133	RAPH1	0.82102
134	CHRNA9	0.82147
135	RASA3	0.82216
136	LRRC8C	0.81928
137	CSF2	0.82033
138	HSPA7	-2.7224
139	KRT1	-2.1437
140	DNAJA4	-1.6897
141	HSPA1A	-1.6377
142	WNT4	-1.5588
143	HSPA1B	-1.5673
144	TNFAIP2	-1.4907
145	ACTBL2	-1.4082
146	CCL2	-1.4177
147	STEAP4	-1.3365
148	CD248	-1.3729
149	FAM46B	-1.3149
150	KRT10	-1.3549

RAF		
151	MGC16121	-1.3131
152	ATF3	-1.2822
153	PIK3C2B	-1.2554
154	RASD2	-1.2697
155	CRYAB	-1.2914
156	IFIT1	-1.4294
157	POU3F1	-1.2468
158	EDN2	-1.2541
159	EPGN	-1.2129
160	FILIP1L	-1.2057
161	EPHA4	-1.205
162	ELF3	-1.2139
163	SLC34A2	-1.2206
164	BBOX1	-1.1913
165	CCL28	-1.1796
166	USP2	-1.1623
167	HSPB8	-1.1746
168	SLC47A2	-1.1735
169	ETV7	-1.1448
170	CXCR7	-1.1514
171	HS3ST6	-1.1452
172	CFB	-1.1101
173	C10orf81	-1.0963
174	ANGPTL7	-1.094
175	EVPLL	-1.0891
176	IFI44	-1.1376
177	IGFBP5	-1.0965
178	LOC285629	-1.0669
179	GPR1	-1.0508
180	CA2	-1.0388
181	SAA2	-1.0501
182	EPSTI1	-1.0642
183	EDN1	-1.032
184	USH1G	-1.0443
185	LIMCH1	-1.0004
186	KLHDC7B	-1.0207
187	EPHA3	-1.0033
188	CXCL12	-0.9996
189	SERPINB13	-0.9904
190	RARRES3	-1.027
191	GRAMD2	-0.9765
192	OTUD1	-0.9896
193	ADAP2	-1.0208
194	CYP1B1	-0.9691
195	PAQR7	-0.9605
196	RARB	-0.953
197	ATHL1	-0.9529
198	APCDD1	-0.9478
199	GABRE	-0.9547
200	DAPK1	-0.937
201	CDRT1	-0.9213
202	SLC27A2	-0.9218
203	LMO1	-0.9236
204	NPR3.00	-0.9329
205	PDZK1IP1	-0.9137
206	RASD1	-0.9046
207	KIT	-0.9074
208	CXCL2	-0.9116
209	MYO18B	-0.9189
210	IFI44L	-0.9916
211	OXTR	-0.8897
212	NFE2	-0.9028
213	ZDHHC8P1	-0.8793
214	EGR1	-0.8874
215	KANK4	-0.8568
216	KMO	-0.852
217	DSC1	-0.8745
218	NEFM	-0.8483
219	AMOT	-0.8381
220	IL6	-0.8617
221	KCNJ5	-0.8354
222	FERMT3	-0.825
223	PPP1R3C	-0.8402
224	TNNI2	-0.8364
225	PRR15L	-0.8205
226	TRIM22	-0.8087
227	C10orf67	-0.7944
228	FBXO32	-0.79
229	SYNM	-0.7877
230	RECK	-0.7943
231	SPINK1	-0.7852
232	ADM	-0.7824
233	NOTCH1	-0.798
234	PROM1	-0.7671
235	CD180	-0.7806
236	MX1	-1.1257
237	DNAJC6	-0.7725
238	NKX1-2	-0.7663
239	SLC30A10	-0.7534
240	SEMA5B	-0.7764
241	MAF	-0.7579
242	TMCC2	-0.7483
243	DNAJB4	-0.7584
244	MTUS1	-0.7496
245	PLD6	-0.7497
246	ST6GALNAC5	-0.7573
247	VAV3	-0.7467
248	SYBU	-0.7441
249	GBP6	-0.7421
250	BST2	-0.7377
251	MIR17HG	-0.7415
252	TLR1	-0.7351
253	PCDH19	-0.7218
254	FBXW10	-0.7239
255	TRIM6	-0.725
256	EFNA5	-0.7317
257	PARP9	-0.73
258	DUSP2	-0.7177
259	SYTL2	-0.7181
260	ADAMTS1	-0.7336
261	FOSL2	-0.7178
262	ENGASE	-0.7113
263	NPNT	-0.7146
264	ZNF488	-0.7149
265	MTSS1L	-0.708
266	TNS3	-0.7078
267	VGLL3	-0.7182
268	EGFL6	-0.7187
269	SOSTDC1	-0.7119
270	LRRN1	-0.7005
271	CORO6	-0.7068
272	FSTL4	-0.692
273	ANKRD2	-0.7042
274	ASAP3	-0.6875
275	FABP5	-0.6994

Table 3.5 Optimal signature gene list generated for KRAS pathway listing all 500 genes and their associated weight in the signature in predicting KRAS pathway activity.



KRAS		
1	MAL	4.975401683
2	KRAS	4.567593537
3	LCE3D	4.312625314
4	DHRS9	2.721530728
5	LCE3E	2.623009598
6	NPTX1	2.331933505
7	IL1RL1	2.251303834
8	PRSS22	2.11804446
9	DCLK1	2.005298202
10	PRR9	2.026071415
11	AKAP12	1.934196875
12	S100A7	1.877277843
13	FAM25A	1.886094154
14	HAS2	1.819625572
15	PAPL	1.723066676
16	LOC100131726	1.716884032
17	DIO3	1.702516246
18	KLK6	1.683248959
19	AGPAT9	1.67347646
20	ARC	1.630521003
21	LY6D	1.584158225
22	NKD2	1.583450247
23	PAEP	1.603863513
24	DIRAS3	1.544455542
25	SPRR2D	1.527686733
26	ANPEP	1.489215886
27	CYB5R2	1.45269922
28	CEACAM1	1.429761504
29	LCE1F	1.437112558
30	STC1	1.408712456
31	SERPINB1	1.380470292
32	HYAL1	1.384274531
33	SPRR1A	1.356933203
34	AQP5	1.365394318
35	SCNN1D	1.354700541
36	BMP6	1.353905526
37	CA6	1.353760868
38	FERMT1	1.33853796
39	TAGLN3	1.340045352
40	LCE1C	1.319520476
41	CALB2	1.321249151
42	ANGPTL4	1.297187349
43	SOX8	1.322782503
44	ASPRV1	1.311452867
45	TMEM45B	1.31046715
46	SLC5A1	1.291881946
47	CEACAM6	1.281807515
48	TNFRSF11B	1.264906447
49	WNT9A	1.251212566
50	EEF1A2	1.219757472
51	PLA2G4E	1.21643
52	TRPV3	1.19298
53	PADI1	1.18888
54	S100P	1.18014
55	LCE1A	1.20564
56	ISG20	1.18739
57	SRMS	1.19487
58	SH2D2A	1.15259
59	GJB4	1.15147
60	ADAM8	1.13774
61	FAM83A	1.12306
62	CALB1	1.10494
63	CRCT1	1.08772
64	EGR3	1.07935
65	CNFN	1.07221
66	HBEGF	1.05369
67	CXCL3	1.0685
68	SULT2B1	1.06767
69	G0S2	1.04513
70	LCE1E	1.05893
71	SERPINB2	1.04624
72	FOS	1.04525
73	ANO1	1.0113
74	APOBEC3A	1.00733
75	KCNN4	0.99995
76	RPSAP52	0.97737
77	LOC100505839	0.96569
78	ODC1	0.96311
79	RHCG	0.95578
80	EGR1	0.94125
81	TFF1	0.92362
82	CYP4F22	0.94375
83	EMP1	0.93242
84	TGM2	0.92683
85	PNMA2	0.9217
86	AGR2	0.91412
87	S100A1	0.907
88	SCNN1G	0.92207
89	SSTR1	0.91386
90	PAQR5	0.90625
91	SYTL5	0.91364
92	LOXL4	0.91049
93	ZBED2	0.90172
94	ROBO4	0.90032
95	DUSP6	0.90056
96	TMEM121	0.93094
97	CCNA1	0.89072
98	WNT7B	0.8953
99	EGR2	0.88193
100	NGEF	0.88352
101	LIF	0.8757
102	KRT18	0.87369
103	DOK7	0.87207
104	PRDM1	0.86622
105	FGFBP1	0.8529
106	GSDMA	0.85062
107	ATP2C2	0.87017
108	SCGB2A2	0.84218
109	WFDC3	0.85547
110	LYPD5	0.86275
111	IVL	0.82789
112	RNASE1	0.81062
113	MLPH	0.82862
114	GPR110	0.81939
115	PHLDA2	0.76218
116	C2orf54	0.82081
117	KIAA1199	0.80278
118	MGP	0.78724
119	SLC20A1	0.78789
120	CYGB	0.77495
121	IL1R2	0.78226
122	CXCL17	0.76704
123	KLK12	0.77696
124	ATP6V0A4	0.76415
125	KLK11	0.75455
126	NAV3	0.75356
127	KPRP	0.74849
128	MRGPRX3	0.7555
129	C15orf62	0.73988
130	LPAR6	0.73682
131	LYPD3	0.7261
132	SOCS1	0.89322
133	EMP3	0.70766
134	SERPINB9	0.72736
135	DUSP5	0.7113
136	CRHR1	0.71496
137	TTYH1	0.72486
138	SOX10	0.70644
139	SPRR4	0.73225
140	SNCB	0.71409
141	SHC4	0.70304
142	ENC1	0.6996
143	ITGB7	0.694
144	GAL	0.70113
145	PDGFB	0.69968
146	SLC13A5	0.69126
147	MAB21L1	0.68592
148	FAM150B	0.70426
149	WNK2	0.68383
150	SEMA7A	0.67977

KRAS		
151	TMPRSS4	0.6788
152	CDC20	0.67643
153	PTPN22	0.6847
154	NR4A1	0.67316
155	OSR1	0.65648
156	KIAA0754	0.6587
157	CD55	0.64813
158	GDPD3	0.70112
159	NT5E	0.65586
160	PPP1R1B	0.64674
161	FZD8	0.63076
162	S100A4	0.64993
163	COL13A1	0.66295
164	TMEM163	0.65854
165	COL6A2	0.61233
166	PTK6	0.6349
167	EDNRB	0.63641
168	MALL	0.65604
169	SPRY4	0.63402
170	LOC646329	0.63555
171	PLEK2	0.62729
172	SMOX	0.63367
173	OLAH	0.63771
174	IGFN1	0.62946
175	GABRA2	0.63677
176	EREG	0.63136
177	SERPINB3	0.62045
178	RASA3	0.62401
179	TMCC3	0.62502
180	GJB3	0.61735
181	SH3TC1	0.60283
182	SHF	0.61369
183	RUNDC3B	0.60537
184	TMEM169	0.6107
185	TMC7	0.60315
186	FOSL1	0.59723
187	CYP27B1	0.59918
188	HMGA2	0.59463
189	PLAUR	0.59469
190	FAM19A5	0.59652
191	FIBCD1	0.5978
192	C6orf15	0.59348
193	ZNF114	0.60719
194	PPAPDC1A	0.57577
195	THBS1	0.57095
196	PMP22	0.57659
197	SLCO4A1	0.60859
198	PYGB	0.57893
199	KRT19	0.58324
200	TGFA	0.57524
201	ATG16L1	0.5798
202	PTHLH	0.56531
203	C16orf74	0.56848
204	FGF18	0.60103
205	C17orf28	0.57244
206	OSBP2	0.5679
207	EMR2	0.55752
208	ATP12A	0.57511
209	CDC42EP2	0.56594
210	PLCXD1	0.55885
211	PLAT	0.54353
212	DUSP4	0.54897
213	GPRC5A	0.54712
214	TRPV4	0.54794
215	FGFBP2	0.59485
216	SPON1	0.55303
217	KRT8	0.54167
218	TMPRSS11E	0.5529
219	PI3	0.52609
220	BMP2	0.54315
221	SLC9A3R2	0.53703
222	SERPINF2	0.56519
223	CLDN7	0.53619
224	MFI2	0.53145
225	SLC10A6	0.65379
226	RAPH1	0.51314
227	CXCL1	0.5187
228	SPTBN5	0.52699
229	MMP1	0.56046
230	ALDH1A3	0.53567
231	TCN1	0.51371
232	UCA1	0.55714
233	S100A6	0.5188
234	CXCR1	0.54057
235	PTPRE	0.50595
236	PLAU	0.49263
237	GLRX	0.52822
238	RAET1L	0.5207
239	BAIAP2L2	0.547
240	SLC16A3	0.51242
241	C8orf84	0.51158
242	ENDOU	0.52965
243	PDE1C	0.50194
244	SLC1A1	0.48819
245	C12orf35	0.46702
246	IL1A	0.49519
247	KIF1A	0.52765
248	RFX8	0.53032
249	LANCL3	0.51748
250	RASAL1	0.52654
251	HSPA1A	-3.8795
252	HSPA1B	-3.5699
253	HSPA7	-3.1361
254	DNAJA4	-2.6933
255	CCL26	-2.6199
256	CRYAB	-2.2305
257	BAG3	-1.6267
258	HSPB8	-1.608
259	HSP90AA1	-1.5994
260	HSP90AA4P	-1.5663
261	DNAJB1	-1.3923
262	ATF3	-1.376
263	OXTR	-1.344
264	SH3BGR	-1.2492
265	DNAJB4	-1.2326
266	CCL2	-1.2195
267	HSP90AA6P	-1.2173
268	ACTBL2	-1.1013
269	HMOX1	-1.0554
270	IFI6	-1.1112
271	CHAC1	-1.009
272	ZFAND2A	-0.9981
273	IL7R	-0.9941
274	ULBP1	-0.976
275	UBB	-0.9525
276	DNAJA1	-0.9606
277	GLYATL2	-0.9777
278	CDRT1	-0.9394
279	UBC	-0.9349
280	EPSTI1	-0.9392
281	FAM49A	-0.9062
282	BST2	-0.9168
283	HSPA8	-0.8668
284	HSPD1	-0.8857
285	LOC100130238	-0.8746
286	ID4	-0.8648
287	TNFAIP2	-0.8629
288	LOC645638	-0.8519
289	MGC16121	-0.8717
290	MB21D1	-0.8361
291	DUSP8	-0.838
292	DLC1	-0.8165
293	FILIP1L	-0.8024
294	SESN2	-0.802
295	CHORDC1	-0.8019
296	LOC727896	-0.7814
297	LAMP3	-0.8295
298	HSPE1	-0.7823
299	KRT10	-0.7951
300	LOC285629	-0.7953

KRAS		
301	BEX1	-0.7987
302	IFI44L	-0.899
303	ZNF323	-0.7778
304	CACYBP	-0.75
305	GBP1	-0.7587
306	BBOX1	-0.7477
307	METTL7A	-0.7421
308	FERMT3	-0.7515
309	C21orf7	-0.7405
310	TMEM27	-0.7571
311	IFRD1	-0.74
312	ABHD3	-0.7351
313	IFI44	-0.7679
314	MORC4	-0.7196
315	GREM1	-0.7107
316	LIMCH1	-0.6971
317	CFB	-0.6969
318	ENGASE	-0.6905
319	C4orf49	-0.6949
320	CCDC117	-0.6959
321	ANGPTL7	-0.701
322	LGR5	-0.6956
323	DFNB31	-0.6909
324	DSC1	-0.6824
325	LCN10	-0.662
326	SLC34A2	-0.6744
327	HERC5	-0.6515
328	CLU	-0.6615
329	DIO2	-0.6581
330	SLC16A14	-0.6468
331	ALOXE3	-0.6422
332	CYFIP2	-0.6379
333	MMP13	-0.6513
334	ASAP3	-0.623
335	OLFM4	-0.6314
336	COL1A2	-0.6319
337	ARHGAP24	-0.6169
338	SLC40A1	-0.6076
339	ATHL1	-0.6182
340	SECTM1	-0.6161
341	MARVELD3	-0.6019
342	NPNT	-0.6104
343	CYP1B1	-0.6078
344	CSRP2	-0.6076
345	PSG6	-0.6086
346	MICB	-0.59
347	FAM46B	-0.6171
348	LHFPL2	-0.5937
349	ZNF761	-0.5881
350	FAM26E	-0.5893
351	COL1A1	-0.6445
352	SOD2	-0.5675
353	FKBP4	-0.5686
354	ABCB1	-0.5975
355	MITF	-0.5592
356	SLC2A12	-0.5643
357	NECAB2	-0.6291
358	SERPINH1	-0.5687
359	C8orf47	-0.6309
360	DSEL	-0.5611
361	JDP2	-0.561
362	EFNA5	-0.5601
363	BPIFC	-0.5819
364	CAP2	-0.5584
365	GDF5	-0.6444
366	STIP1	-0.5483
367	TRIM22	-0.5568
368	SERPINB13	-0.5588
369	NKIRAS2	-0.5452
370	SOWAHB	-0.573
371	GBP6	-0.546
372	ADM2	-0.6457
373	GSR	-0.5492
374	HERC6	-0.5556
375	ISG15	-0.7862
376	EML5	-0.5436
377	ABHD4	-0.5544
378	TSPYL2	-0.5362
379	CPT1C	-0.5357
380	MEX3B	-0.5493
381	IFI27	-0.7756
382	FAM83D	-0.5375
383	IER5	-0.5233
384	TOX	-0.5587
385	PLAC2	-0.5151
386	HSPH1	-1.4223
387	DNHD1	-0.5029
388	DDIT3	-0.5528
389	RND3	-0.5134
390	DEDD2	-0.5079
391	FAM46A	-0.5158
392	MX1	-1.0652
393	GABRE	-0.6457
394	CDKN2B	-0.4942
395	OSGIN1	-0.5185
396	CAPN6	-0.5395
397	SYNM	-0.5351
398	GAL3ST4	-0.5265
399	EN1	-0.5318
400	HSPA1L	-0.5122
401	ZSCAN16	-0.5478
402	AMOT	-0.4965
403	GM2A	-0.4842
404	NTN4	-0.4917
405	PDGFD	-0.5244
406	PARM1	-0.5043
407	GKAP1	-0.5756
408	AP3B2	-0.4847
409	EPGN	-0.5116
410	BBC3	-0.565
411	FTL	-0.4825
412	GPR75	-0.506
413	MRPL18	-0.4768
414	OLFML2A	-0.5122
415	FNIP2	-0.4888
416	FABP5	-0.4984
417	LAYN	-0.5021
418	NID1	-0.4798
419	TRIM16	-0.4543
420	MAP7	-0.4519
421	JUN	-0.5599
422	MME	-0.5016
423	DAPK1	-0.4714
424	ASNS	-0.5247
425	GAB2	-0.459
426	RGS2	-0.5889
427	CXCL12	-0.489
428	CCRN4L	-0.457
429	ETV7	-0.4869
430	IL32	-0.5757
431	MDK	-0.5372
432	HIST1H4H	-0.5219
433	GPNMB	-0.4927
434	NCALD	-0.5094
435	KRT80	-0.4766
436	SYNPO2	-0.448
437	IFIH1	-0.5497
438	TRIM16L	-0.4592
439	GRHL3	-0.5
440	LOC100507495	-0.5215
441	MALAT1	-0.5
442	OAS3	-0.5378
443	MAP2	-0.4459
444	PTGES3	-0.4531
445	TRIM17	-0.5172
446	ZNF711	-0.5192
447	PLXNA2	-0.4481
448	BHLHB9	-0.4385
449	CALCA	-0.5853
450	BLNK	-0.6235

KRAS		
451	XAF1	-0.5916
452	NFIL3	-0.4577
453	B3GNT1	-0.4826
454	CREG2	-0.5158
455	DNAJC6	-0.4191
456	GRAMD2	-0.4335
457	HIST1H3D	-0.5556
458	PLD1	-0.4275
459	STOX2	-0.4571
460	SIAH1	-0.4665
461	HSP90AB1	-0.4022
462	ANK3	-0.4245
463	FAM129A	-0.4753
464	GAMT	-0.4972
465	MBNL2	-0.4265
466	VAV3	-0.4183
467	BRD3	-0.4179
468	TAF15	-0.4281
469	SFMBT2	-0.4139
470	KLHL25	-0.4533
471	ADRBK2	-0.4163
472	INPP5D	-0.4079
473	ELOVL5	-0.4129
474	EDN1	-0.4641
475	IL6	-0.5401
476	C1R	-0.4441
477	CCDC84	-0.526
478	NME5	-0.4401
479	MB21D2	-0.434
480	DOCK10	-0.4252
481	LOC653513	-0.5228
482	MAPK4	-0.4509
483	NBPF1	-0.448
484	NGF	-0.6122
485	DDX60	-0.6192
486	STC2	-0.4182
487	ZNF117	-0.4531
488	GPR1	-0.4633
489	RBM24	-0.4383
490	CYP39A1	-0.4766
491	PIK3C2B	-0.4126
492	FBXW10	-0.4436
493	HMGN3	-0.4005
494	SAMHD1	-0.4831
495	BTN2A2	-0.4324
496	ST13P4	-0.428
497	PPP1R15A	-0.3806
498	HSP90AB3P	-0.4065
499	LOC284837	-0.4228
500	PTN	-0.4857

Table 3.6 Optimal gene list selection using proteomics validation calculated with Pearson pairwise correlations between predicted pathway activations and TCPA protein expression levels.

Pathway	List length	Antibody	cor	p-value
BAD	475	PDK1_pS241	-0.247206	1.63E-06
HER2	5	HER2_pY1248	0.3180165	4.54E-10
IGF1R	25	IGF1R_pY1135Y1136	x	x
RAF	275	CRAF_pS338	0.3176497	4.77E-10
KRAS	500	MEK1_pS217S221	0.1643924	0.001577

Table 3.7 Optimal gene list selection using gene expression validation calculated with Pearson pairwise correlations between predicted pathway activations and TCGA patient tumor expression levels.

Pathway	List length	Validation Gene	cor	p-value
BAD	475	BAD	x	x
HER2	5	ERBB2	0.4114047	< 2.2e-16
IGF1R	25	IGF1R	0.178464	2.98E-05
RAF	275	RAF1	x	x
KRAS	500	KRAS	x	x

Table 3.8 Summary table of gene signature selection and methods used for validation.

Pathway	Oncogene	List length	Proteomics	Gene	Box plot
BAD	BAD	475	PDK1_pS241	BAD	✔
HER2	ERBB2	5	HER2_pY1248	ERBB2	✔
IGF1R	IGF1R	25	x	IGF1R	✔
RAF	RAF1	275	CRAF_pS338	x	x
KRAS	G12V	500	MEK1_pS217S221	x	x

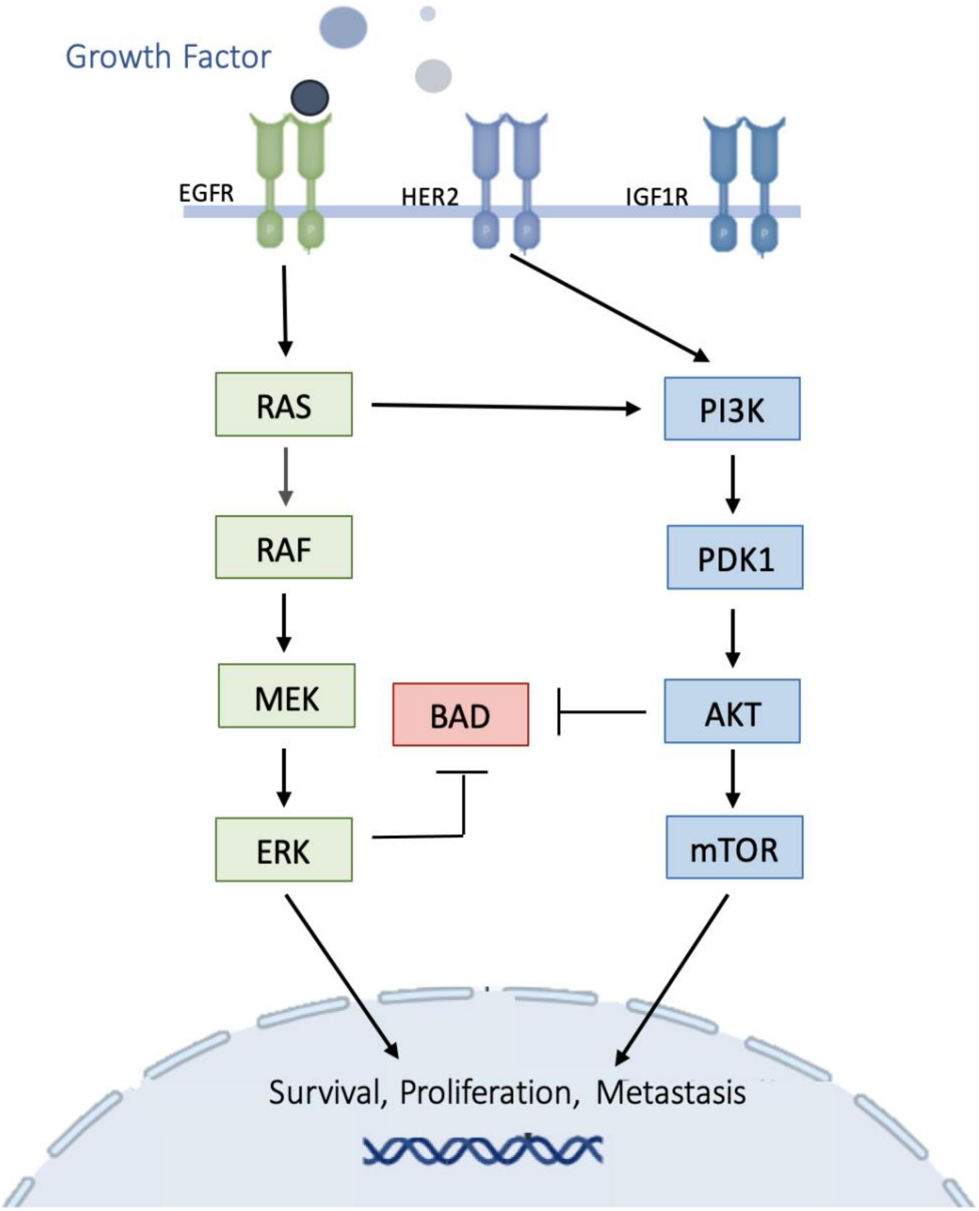


Figure 1.1 Schematic overview of the driving oncogenic Growth Factor Receptor Network (GFRN) responsible for cell survival, growth, and metastasis. Consist of two intercommunicating parallel signaling pathways including RAS/RAF/MAPK pathway, shown in green, and the PI3K/AKT/mTOR, shown in blue. RAS pathway activation can be initiated by EGFR receptor mediated signaling leading to activation of RAF, in turn initiating MEK activation, as a result initiating tumorigenesis through ERK activation. Its neighboring pathway PI3K can be initiated through HER2(ERBB2) receptor mediated signaling as well as RAS activation. This then results in the inactivation of PDK1 activating AKT signaling which can inhibit the BAD pathway and/or lead to activation of mTOR resulting in tumorigenesis. Additional, signaling pathways can be initiated such as the inhibition of ERK leading to inhibition of RAF through mTOR activation. Although various alternate pathways of activation leading to drug resistance remain uncharacterized.

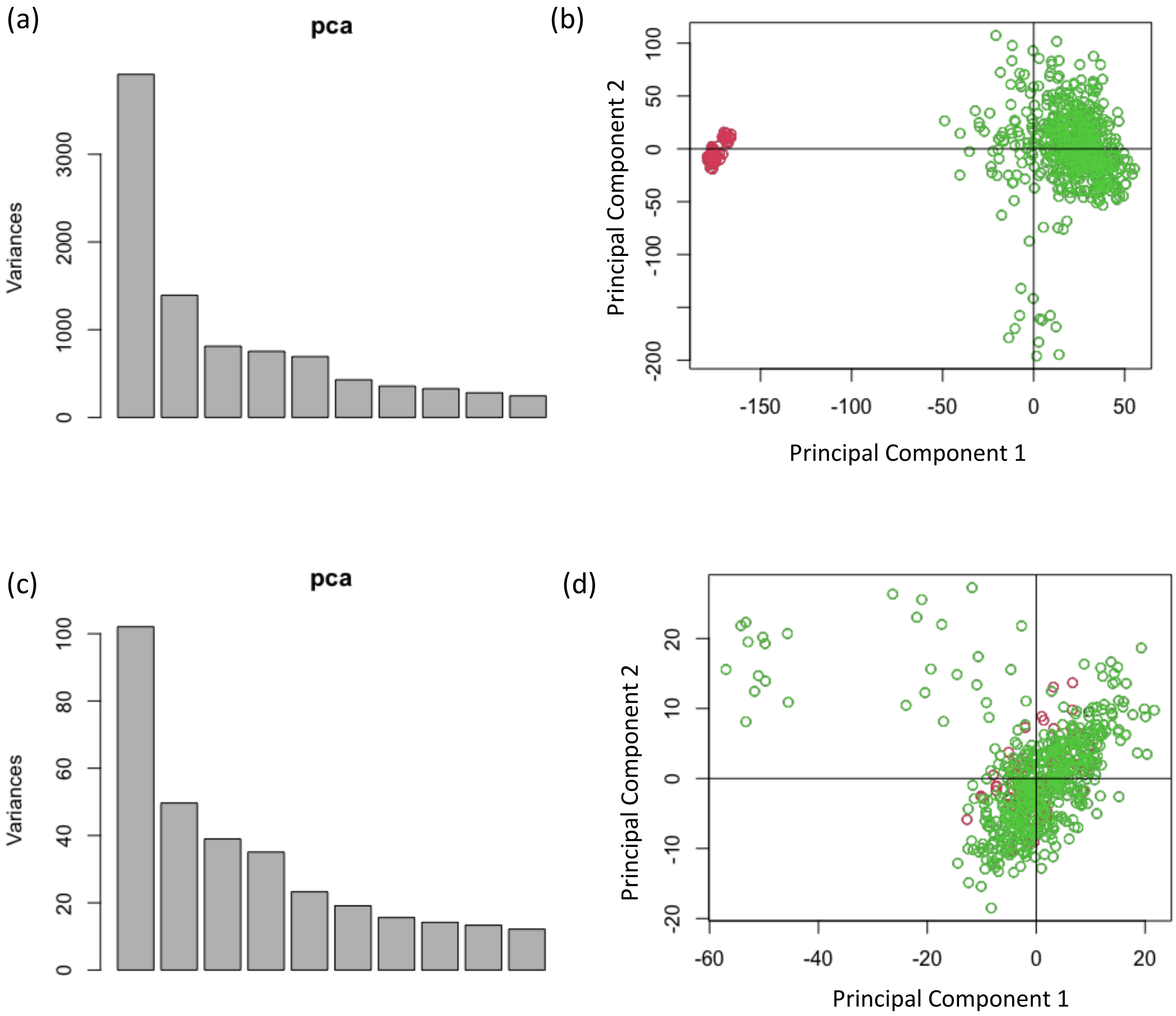


Figure 2.1 (a)Principal component Analysis (PCA) expressing the first two PCAs representing the greatest variations between the gene expression data and LUAD patient tumor samples from TCGA. Due to external factors significant variances and confounding batch effects are observed. (b) PCA scatter plot displaying the first two PCAs representing the greatest variations between the datasets. This included the gene expression data, shown in green, and LUAD patient tumor samples from TCGA, shown in red, in which significant confounding variables and variances were observed. (c) The PCA following adjustment and refinement of gene expression data and patient tumor samples using the *ComBat* function resulting in increased data similarity. (d) PCA scatter plot displaying the gene expression data, shown in red and LUAD patient tumor samples, shown in green, following *ComBat* adjustment displaying significant improvement in data similarity and reduction of variances and confounding batch effects.

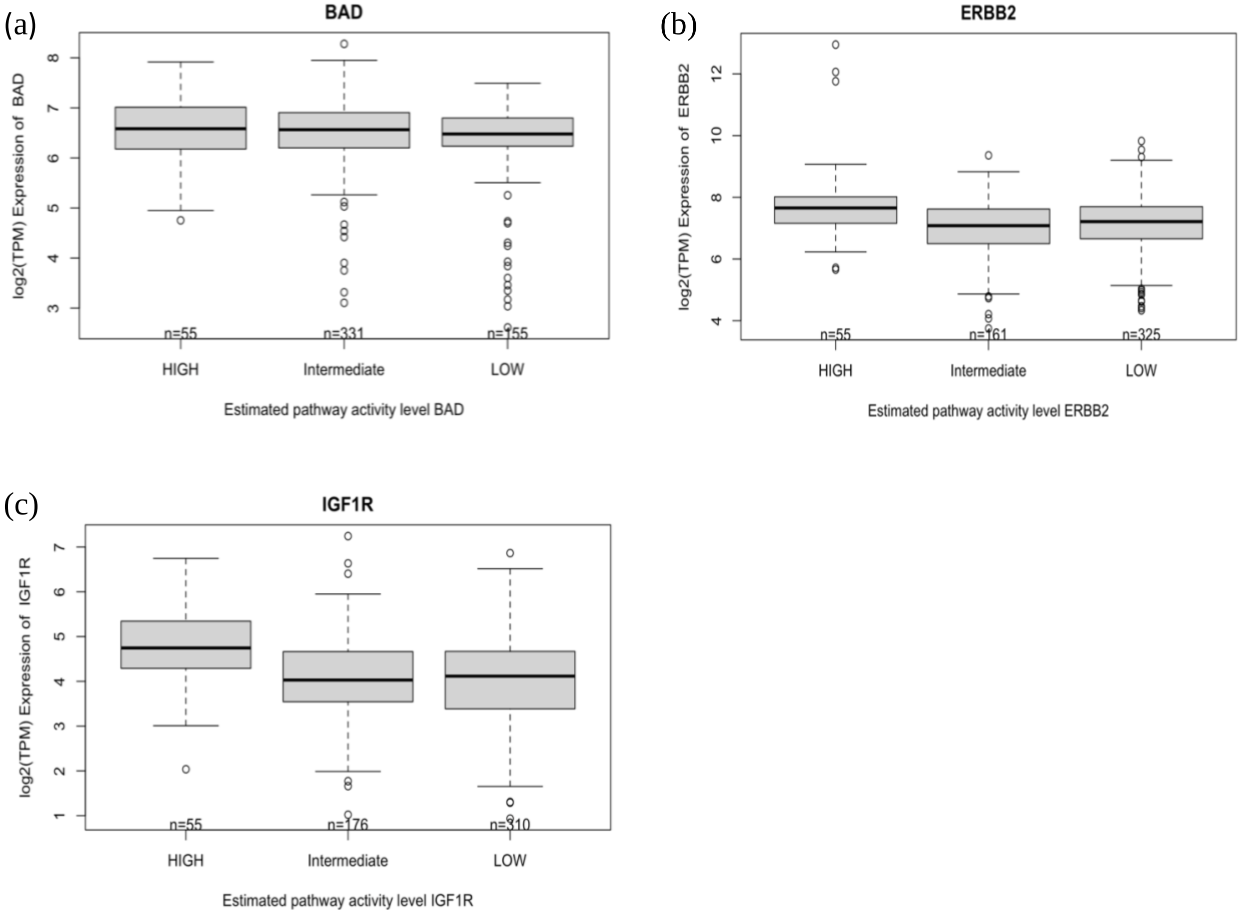


Figure 3.1 Gene expression box plots used for gene expression signature validation. (a) Generated box plot used for validation of BAD displaying the signature's ability, shown on the x-axis, to distinguish levels of pathway activity within LUAD patient tumor samples shown on the y-axis. As a result, higher levels of pathway activity were predicted in 55 samples classified as "HIGH" expressing, while 331 showed "intermediate" pathway activity, and 155 showed low levels of BAD pathway activation classified as "LOW" expressing samples. Concluding the signature's ability to distinguished levels of aberrant activity with TCGA samples. (b) In this figure, the generated gene expression signature of HER2 predicted higher levels of pathway activity in 55 patient tumor samples classified as high expressing, 161 intermediate expressing samples, and 125 low expressing samples distinguishing levels of pathway activity further validating the signature. (c) Lastly, the gene expression signature of IGF1R was able to distinguish levels of increased pathway activity within 55 patient tumor samples, classified as "HIGH" expressing, 176 were identified as "intermediate" expressing, and 310 were characterized as low expressing. In all, validating the signature's ability to distinguish accurate levels of pathway activity.

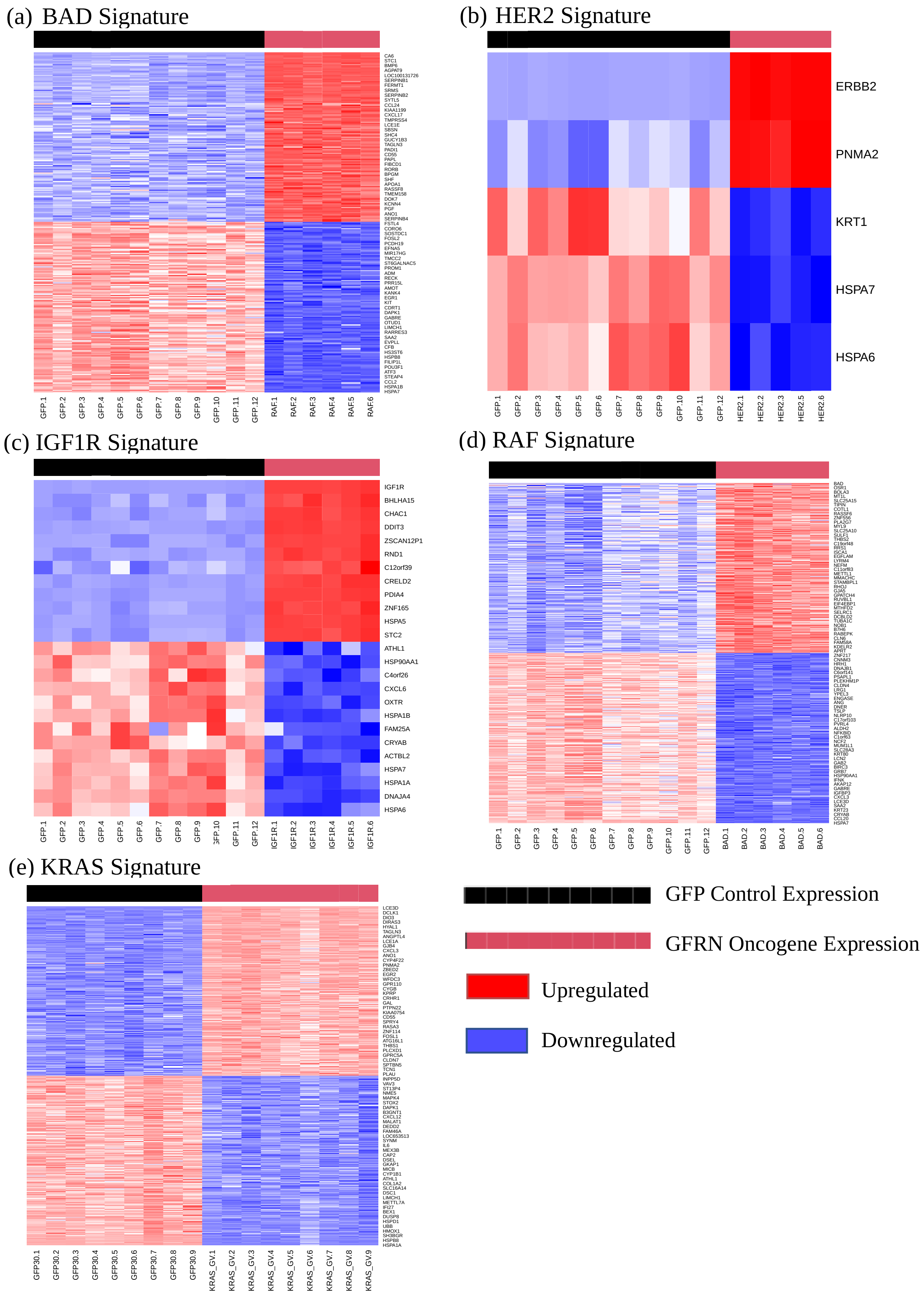


Figure 4.1 Complex heatmaps generated of optimized gene expression signatures representative of aberrant pathway activity for the GFRN pathways (a) BAD, 475-gene signature (b) HER2, 5- gene signature

(c)IGF1R, 25-gene signature (d) RAF, 275-gene signature, and (e) KRAS, 500 gene-signature. The black bar indicates normal pathway activity or the respective GFRN pathway turned off, expressed using the HMECs overexpressing GFP (control). The red bar is then used to represent aberrant pathway activity or pathway activity turned on, generated by the HMECS overexpressing the GFRN-pathways respective oncogene. Relative to the pathway's state of activation, genes comprising the signature are shown on the right expressing varying levels of activity, indicated in red or blue. Genes expressing upregulated levels of expression are represented in red, and the brighter the red, the higher levels of activity while blue indicates downregulated levels of activity and the darker the blue, the lower the level of activity.

REFERENCES

1. Siegel, R.L., et al., *Cancer Statistics, 2021.* CA Cancer J Clin, 2021. **71**(1): p. 7-33.
2. Zito Marino, F., et al., *Molecular heterogeneity in lung cancer: from mechanisms of origin to clinical implications.* International journal of medical sciences, 2019. **16**(7): p. 981-989.
3. Golub, T.R., et al., *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.* Science, 1999. **286**(5439): p. 531-7.
4. Shim, H.S., et al., *Molecular Testing of Lung Cancers.* Journal of Pathology and Translational Medicine, 2017. **51**(3): p. 242-254.
5. De Marco, C., et al., *Specific gene expression signatures induced by the multiple oncogenic alterations that occur within the PTEN/PI3K/AKT pathway in lung cancer.* PloS one, 2017. **12**(6): p. e0178865-e0178865.
6. Rahman, M., et al., *Activity of distinct growth factor receptor network components in breast tumors uncovers two biologically relevant subtypes.* Genome Medicine, 2017. **9**(1): p. 40.
7. Yang, J., et al., *Targeting PI3K in cancer: mechanisms and advances in clinical trials.* Molecular Cancer, 2019. **18**(1): p. 26.
8. Pradhan, R., et al., *MAPK pathway: a potential target for the treatment of non-small-cell lung carcinoma.* Future Medicinal Chemistry, 2019. **11**(8): p. 793-795.
9. Cairns, J., et al., *Differential roles of ERRFI1 in EGFR and AKT pathway regulation affect cancer proliferation.* EMBO reports, 2018. **19**(3): p. e44767.
10. Karreth, F., et al., *C-Raf Is Required for the Initiation of Lung Cancer by K-Ras(G12D).* Cancer discovery, 2011. **1**: p. 128-36.

11. Jiang, L., et al., *BAD overexpression inhibits cell growth and induces apoptosis via mitochondrial-dependent pathway in non-small cell lung cancer*. Cancer Cell International, 2013. **13**(1): p. 53.

12. Jin, X., et al., *Identification of key pathways and genes in lung carcinogenesis.* Oncol Lett, 2018. **16**(4): p. 4185-4192.

13. Chibon, F., *Cancer gene expression signatures - the rise and fall?* Eur J Cancer, 2013. **49**(8): p. 2000-9.

14. Tavassoly, I., et al., *Genomic signatures defining responsiveness to allopurinol and combination therapy for lung cancer identified by systems therapeutics analyses.* Molecular oncology, 2019. **13**(8): p. 1725-1743.

15. Singh, V., et al., *Characterization of ERBB2 alterations in non-small cell lung cancer*. Journal of Clinical Oncology, 2020. **38**(15_suppl): p. e21553-e21553.

16. Zhao, J. and Y. Xia, *Targeting HER2 Alterations in Non–Small-Cell Lung Cancer: A Comprehensive Review*. JCO Precision Oncology, 2020(4): p. 411-425.

17. Wang, R., et al., *Transient IGF-1R inhibition combined with osimertinib eradicates AXL-low expressing EGFR mutated lung cancer*. Nature Communications, 2020. **11**(1): p. 4607.

18. Fois, S.S., et al., *Molecular Epidemiology of the Main Druggable Genetic Alterations in Non-Small Cell Lung Cancer*. International Journal of Molecular Sciences, 2021. **22**(2): p. 612.