WEIGHTED GENE CO-EXPRESSION NETWORK ANALYSIS OF COLORECTAL PATIENTS TO IDENTIFY RIGHT DRUG-RIGHT TARGET FOR POTENT EFFICACY OF TARGETED THERAPY

Anamika Tripathi

Submitted to the faculty of the School of Informatics
in partial fulfillment of the requirements
for the degree of
Master of Science in Bioinformatics,
Indiana University
12/2017

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Anamika Tripathi

WEIGHTED GENE CO-EXPRESSION NETWORK ANALYSIS OF COLORECTAL PATIENTS TO IDENTIFY RIGHT DRUG-RIGHT TARGET FOR POTENT EFFICACY OF TARGETED THERAPY

Colon rectal cancer (CRC) is one of the most common cancers worldwide. It is characterized by the successive accumulation of mutations in genes controlling epithelial cell growth and differentiation leading to genomic instability. This results in the activation of proto-oncogene(K-ras), loss of tumor suppressor gene activity and abnormality in DNA repair genes. Targeted therapy is a new generation of cancer treatment in which drugs attack targets which are specific for the cancer cell and are critical for its survival or for its malignant behavior. Survival of metastatic CRC patients has approximately doubled due to the development of new combinations of standard chemotherapy, and the innovative targeted therapies, such as monoclonal antibodies against epidermal growth factor receptor (EGFR) or monoclonal antibodies against vascular endothelial growth factor (VEGFR).The study is to exhibit the need for right drug-right target and provides a proof of principle for potent efficacy of molecular targeted therapy for CRC. We have performed the weighted gene co-expression network analysis for three different patient cohort treated with different targeted therapy drugs. The results demonstrates the variation across different treatment regime in context of transcription factor networks. New significant transcription factors have been identified as potential biomarker for CRC cancer including EP300, STAT6, ATF3, ELK1, HNF4A, JUN, TAF1, IRF1, TP53, ELF1 and YY1. The results provides guidance for future omic study on CRC and additional validation work for potent biomarker for CRC.

CHAPTER ONE: INTRODUCTION & BACKGROUND

Colorectal Cancer (CRC) and Targeted Therapy

Colon rectal cancer (CRC) is the cancer affecting caecum, colon and rectum. Colon Adenocarcinoma (COAD) is the third leading cause of cancer death among men and women in the United States [1, 2]. In 2016, it was estimated that 70,820 new cases of colon and rectum cancer in men and 63,670 cases in women in the US. CRC related death in 2016 across USA were estimated to be 26,020 in men and 23,170 in women [3]. The life time risk of developing CRC is about: 1 in 21 (4.6%) for men and 1 in 24 (4.2%)for women.

Risk factors for CRC

The risk factors of CRC are age, smoking, poor diet, and family history. Carcinogenesis is characterized by the successive accumulation of mutations in genes controlling epithelial cell growth and differentiation leading to genomic instability whereby widespread loss of DNA integrity is perpetuated [4] , such as activation of proto-oncogene (K-ras), loss of tumor suppressor gene activity (APC,DCC) and abnormality in DNA repair genes (hMSH2, hMLH1) specially HNPCC syndrome.

Targeted Therapy

Targeted therapy is a new generation of cancer treatment which uses drugs or other substances to more precisely identify and attack cancer cells. As the name suggests, targeted therapies interfere with specific proteins involved in tumorigenesis. Targeted therapy drugs attack targets which are specific for the cancer cell and are critical for its survival or for its malignant behavior. In other words, targeted drugs target certain parts of cancer cells that make them different from other cells and they also target other cells that help cancer cells grow. Additionally, The drugs target certain parts of the cell and the signals that are needed for a cancer to develop and keep growing. Therefore, targeted therapy is a growing part of the treatment for many types of cancer.

CHAPTER TWO: LITERATURE REVIEW

Survival of metastatic CRC patients has approximately doubled [5]. This significant improvement is mainly due to the development of new combinations of standard chemotherapy, and with the introduction of new targeted therapies, such as monoclonal antibodies against epidermal growth factor receptor (EGFR) or monoclonal antibodies against vascular endothelial growth factor (VEGFR). The chimeric IgG1monoclonal antibody cetuximab has been proven efficient in irinotecan-resistant metastatic CRC with response rates ranging between 8.8% when used in monotherapy and 22.9% when combined with irinotecan [6-8]. EGFR has been validated as a therapeutic target in several human tumors, including CRC [9-11]. Ligand occupancy of the EGFR activates the RAS/ RAF/MAPK, STAT, and PI3K/AKT signaling pathways, which together modulate cellular proliferation, adhesion, angiogenesis, migration, and survival [12, 13]. The anti-EGFR targeted antibodies cetuximab and panitumumab administered as monotherapy in CRC have shown response and disease stabilization rates [9, 11].

KRAS, the human homolog of the Kirsten rat sarcoma-2 virus oncogene, encodes a small GTP binding protein that acts as a self-inactivating signal transducer by cycling from GDP- to GTP-bound states in response to stimulation of a cell surface receptor, including EGFR [14, 15]. *KRAS* can harbor oncogenic mutations that yield a constitutively active protein. Such mutations are found in approximately 30% to 50% of CRC tumors and are common in other tumor types [16-22]. Several studies have indicated that the presence of mutant *KRAS* in lung and CRC tumors is associated with lack of response to EGFR inhibitors [23, 24]. Vascular endothelial growth factor (VEGF) mediates numerous changes within the tumor vasculature, including endothelial cell proliferation, migration, invasion, survival, chemotaxis of bone marrow-derived progenitor cells, vascular permeability and vasodilatation. VEGF has several important functions that are independent of vascular processes, including autocrine effects on tumor cell function(survival, migration, invasion), immune suppression, and homing of bone marrow progenitors to ‘prepare’ an organ for subsequent metastasis [25].The anti- VEGF monoclonal antibody bevacizumab (Avastin, Genentech) and other VEGF-targeted therapies, showing

clinical benefit in patients with metastatic colorectal cancer (CRC),either as single agents or when combined with chemotherapy [26-28].

One of the outstanding issues in treating cancer is the vexing heterogeneity in patient course and response to therapy, even in apparently similar tumors as defined by conventional criteria. Due to this heterogeneity, an agent targeting one particular pathway is unlikely to be effective in all patients. Individualized therapies that are tailored to a patient's genetic composition and tests that can predict which therapy he/she will respond to will be of tremendous value for CRC. Molecularly targeted therapies are transforming the treatment of cancer [29]. Small molecule inhibitors that target key enzymes on which cancer cells depend, raise the possibility of rational approaches to cancer therapy. This study, demonstrates the need for right drug-right target and provides a proof of principle for potent efficacy of molecular targeted therapy for CRC. This study clearly demonstrates the variation across different treatment regime in context of transcription factor networks. Identifying new drug targets is a critical step. However, gene expression data can provide a key first step toward constructing a systems level view of the perturbed networks in cancer cells, which can potentially help to identify key genes, networks, or pathways that can be therapeutically targeted [30]. However, the identification of key molecular targets still remains a challenge. Recent work highlights the potential for uncovering oncogenic pathways and molecular targets, when genomic data are analyzed at the level of gene co-expression modules or meta-genes or when aggregated gene sets are used to assess modules enriched for key biological processes [29, 31, 32]. Integrating this type of data with studies in model systems in which modules can be studied in response to relevant molecular perturbations (e.g., oncogene over-expression or pharmacological inhibition) may further facilitate the identification and validation of novel molecular targets [33-35].

CHAPTER THREE: METHODOLOGY

Methodology Flow diagram

Figure 1 gives the overall methodology of this study.

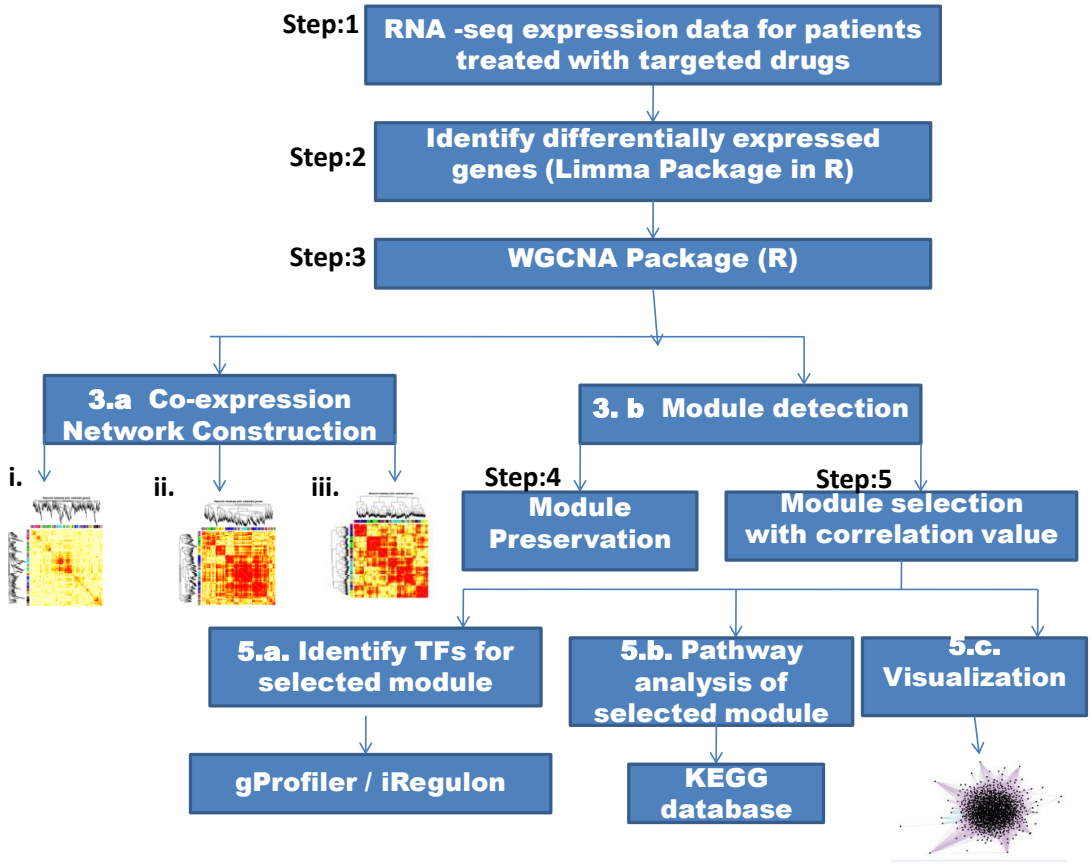


Figure 1: Flow Diagram of Methodology

Step 1: Gene expression data of CRC patients, treated with targeted therapy were obtained from Genomic data commons data portal website and these include (i) Bevacizumab- Stage III (5 samples); (ii) Bevacizumab – Stage IV (9 samples) and; (iii)Cetuximab (5 samples) and 10 normal samples (Table 1). These three cohorts are analyzed in seven steps as shown in flow diagram of methodology (Figure 1).

Table 1:Sample information treated with targeted therapy

Patient sample	Number of patients	Tumor stage	Normal sample
Bevacizumab treated stage-III patients	5	Stage-III	5
Bevacizumab treated stage-IV patients	9	Stage-IV	9
Cetuximab treated patients Stage-IV	5	Stage-IV	5

Step 2: Identify differentially expressed (DE) genes using Limma package in R - The GDC gene expression data were normalized, pre-processed and data matrix was designed. This data matrix includes a coefficient for the normal vs diseased type difference for two drugs cohorts. The core component of the limma package is the ability to fit gene-wise linear models to gene expression data in order to assess DE [36]. Each analysis begins with a matrix of expression levels, with probes/genes/exons in the rows and different samples (biological/technical replicates) in the columns. The linear modeling is performed in a row-wise fashion, with regression coefficients and standard errors either directly estimating the comparisons of interest or via contrasts [36]. Test-statistics are obtained for gene ranking that can be further summarized at the gene set level to perform gene signature/pathway-level ranking. All the linear model were fitted by using lmFit function. ebays function was used to squeeze gene wise variance towards the common or trended variance, which reduces the number of false positives for genes with very small variances and improves power to detect DE for genes with larger variances. Top scored

differentially expressed genes were obtained on the basis of log fold change at p-value 0.001(supplementary table a).

Step 3: Weighted gene co-expression network analysis - Co-expression networks have been found useful for describing the pair-wise relationships among gene transcripts. The co-expression networks, refer to nodes as 'genes' and to the node significance measure GS_i as the gene significance measure. WGCNA R-package was used to generate weighted gene co-expression networks [37]. The WGCNA package contains a comprehensive set of functions for performing a correlation network analysis of large, high-dimensional data sets. Co-expression network analysis was performed for three patients' cohorts:

I. bevacizumab treated stage III patients
II. bevacizumab treated stageIV patients
III. cetuximab treated patients

3.a: Co-expression network construction - A network is fully specified by its adjacency matrix a_{ij}, asymmetric $n \times n$ matrix with entries in [0, 1] whose component a_{ij} encodes the network connection strength between nodes i and j. To calculate the adjacency matrix, an intermediate quantity called the co-expression similarity s_{ij} is first defined. The default method defines the co-expression similarity s_{ij} as the absolute value of the correlation coefficient between the profiles of nodes i and j:

$$s_{ij} = |cor(x_i, x_j)| \qquad \textit{Equation 1}$$

Using a thresholding procedure, the co-expression similarity is transformed into the adjacency. An unweighted network adjacency a_{ij} between gene expression profiles x_i and x_j can be defined by hard thresholding the co-expression similarity s_{ij} as:

$$a_{ij} = \begin{cases} 1 & if\, s_{ij} \geq \tau; \\ 0 & otherwise, \end{cases} \qquad \textit{Equation 2}$$

where Ƭ is the "hard" threshold parameter. Thus, two genes are linked (a_{ij} = 1) if the absolute correlation between their expression profiles exceeds the (hard) threshold Ƭ . The hard-thresholding procedure is implemented in the function signumAdjacencyFunction. Weighted networks allow the adjacency to take on continuous values between 0 and 1. A weighed network adjacency can be defined by raising the co-expression similarity to a power [5,10].

$$\boldsymbol{a_{ij} = s_{ij}^{\beta},} \qquad \textit{Equation 3}$$

with β≥ 1. The function adjacency calculates the adjacency matrix from expression data. The weighted adjacency a_{ij} between two genes is proportional to their similarity on a logarithmic scale, $log(a_{ij}) = \beta \times log(s_{ij})$. Adjacency functions for both weighted and unweighted networks require the user to choose threshold parameters. In this study, undirected weighted gene co-expression networks were used. The nodes of such a network correspond to gene expressions, and edges between genes are determined by the pair wise Pearson correlations between expressions of genes. The soft threshold β = 6 was used for scale free topology criterion. The package provides functions pickSoftThreshold, pickHardThreshold that assist in choosing the parameters, as well as the function scaleFreePlot for evaluating whether the network exhibits a scale free topology. Appendix Figure (i) shows a plot identifying scale free topology in gene expression data of patients treated with cetuximab.

3.b: Module Detection: The WGCNA package [37] provides a robust set of R functions for constructing weighted co-expression networks. Once the networks are constructed, modules were detected for each cohorts. Modules are defined as clusters of densely interconnected genes. WGCNA identifies gene modules using unsupervised clustering, i.e. without the use of a priori defined gene sets. Standard R function hclust [38] was used for hierarchical clustering. The branches of the hierarchical clustering dendrogram correspond to modules and can be identified using any branch cutting method. In this study, dynamic tree cut method was applied. The height and shape parameters of the dynamic tree cut method provide improved flexibility for branch cutting and module detection. For module detection, a relatively large minimum module size of 30, and a medium sensitivity (deepSplit=2) to cluster splitting were selected. Unlike unweighted networks that use a hard threshold to dichotomize the correlation matrix, the soft thresholding of

weighted gene co-expression networks preserves the continuous nature of the gene co-expression information, leading to highly robust results. To organize genes (transcripts) into modules, the topological overlap measure as a robust measure of interconnectedness in a hierarchical cluster analysis were used [37]. The function blockwiseModules is designed to handle network construction and module detection in large data sets. The function first pre-clusters nodes into large clusters, referred to as blocks, using a variant of k-means clustering (function projectiveKMeans). Hierarchical clustering is applied to each block and modules are defined as branches of the resulting dendrogram. To synthesize the module detection results across blocks, an automatic module merging step (function mergeCloseModules) is performed that merges modules whose eigengenes are highly correlated.

Each modules were correlated with external traits and look for the most significant associations. Each association was color coded by the correlation value which gives a suitable graphical representation that helps in reading the table and see the trait association of modules. Associations of individual genes with trait of interest were defined by Gene Significance GS. For each module, a quantitative measure of module membership MM as the correlation of the module eigengene and the gene expression profile were defined. This allows to quantify the similarity of all genes on the array to every module. Using the GS and MM measures, genes that have a high significance as well as high module membership in each module were identified. A scatter plot of Gene Significance vs. Module Membership was plotted (Figure 5).

Step 4: Module preservation: Module preservation statistical tests [39] were used to assess how well network properties of a module in one reference data set were preserved in a test data set (modulePreservation function in WGCNA). Preservation statistics are influenced by a number of variables (module size, network size, etc). A composite preservation Z-score (Zsummary) was used to define preservation relative to a module of randomly assigned genes where values 5 > Z < 10 represent moderate preservation, while Z > 10 indicated high preservation. The composite statistic summarized density-based and connectivity-based preservation statistics (Eq. 4):

$$Z_{Summary} = \frac{Z_{density} + Z_{connectivity}}{2} \qquad \textit{Equation 4}$$

Density-based measures assessed whether module nodes remained densely connected in a test network; connectivity-based measures defined whether intra node connectivity patterns in the reference network were similar to those in the test network. A separate summary *p value* for module preservation, given as the median of the *log-p values* for the associated permutation *Z statistics*, was calculated. Permutation tests, where the module labels of the test network were randomly permuted, were employed to determine the significance of the observed preservations test statistics. A module of randomly assigned genes, "gold" (R21) module, was prepared as a sham module to evaluate bias in the module preservation across species.

Step 5: Module selection - Module selection were performed based on correlation value. For the present work modules were selected for all the three cohorts as they have high correlation values.

5.a: Identification of TFs - gProfiler [40] was used to identify TFs. g:Profiler studies flat and ranked gene lists and finds statistically significant Gene Ontology terms, pathways and other gene function related terms. TFs related genes were identified by using uniprot database [41] for top three modules of the three cohort. iRegulon, a cyoscape plug-in was used to identify top significant TFs on the basis of enrichment score.

5.b: Pathway analysis - KEGG [42] database was used for pathway analysis. Pathway analysis were performed for the TFs of selected modules for all three patient cohort.

5.c: Network visualization- Network visualization were done by cytoscape. Networks were exported in cytoscape [43] using function exportNetworkToCytoscape.

CHAPTER FOUR: RESULTS

Step 1: Data collection - Table 1 shows the profile of the three dataset used in the study.

Step 2: Differential gene analysis - Limma package in R was used to identify the differentially expressed genes at *p- value* 0.001(supplementary table a). Figure 2 shows the profile for the differentially expressed genes identified for three cohorts: 2,724 (bevacizumab treated stage- III), 10,817(bevacizumab treated stage- IV) and 14,590 (for cetuximab treated patients).

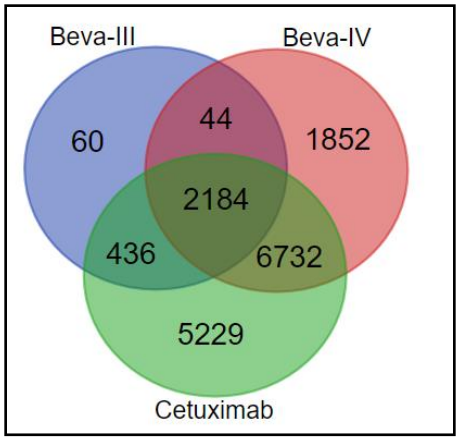


Figure 2: Venn diagram for differentially expressed gene for bevacizumab treated stage-III, Bevacizumab treated stage-IV and cetuximab treated patients at *p-value 0.001*

Step 3. Weighted gene co-expression network analysis - Each differentially expressed gene cohort was further analyzed for its coexpression network using WGCNA [37].

3.a: Co-expression network - In this study, WGCNA was applied to 2,724, 10,817 and 14,590 differentially expressed genes for bevacizumab treated stage-III patients', bevacizumab treated stage-III patients', and cetuximab treated patients' respectively. The co-expression networks of the selected genes were generated using an unsupervised co-expression relationship. This was initially built on the basis of the adjacency matrix of connection strengths by using Pearson's correlation coefficients for gene pairs . Figure 3 shows networks heatmap for three patient cohorts.

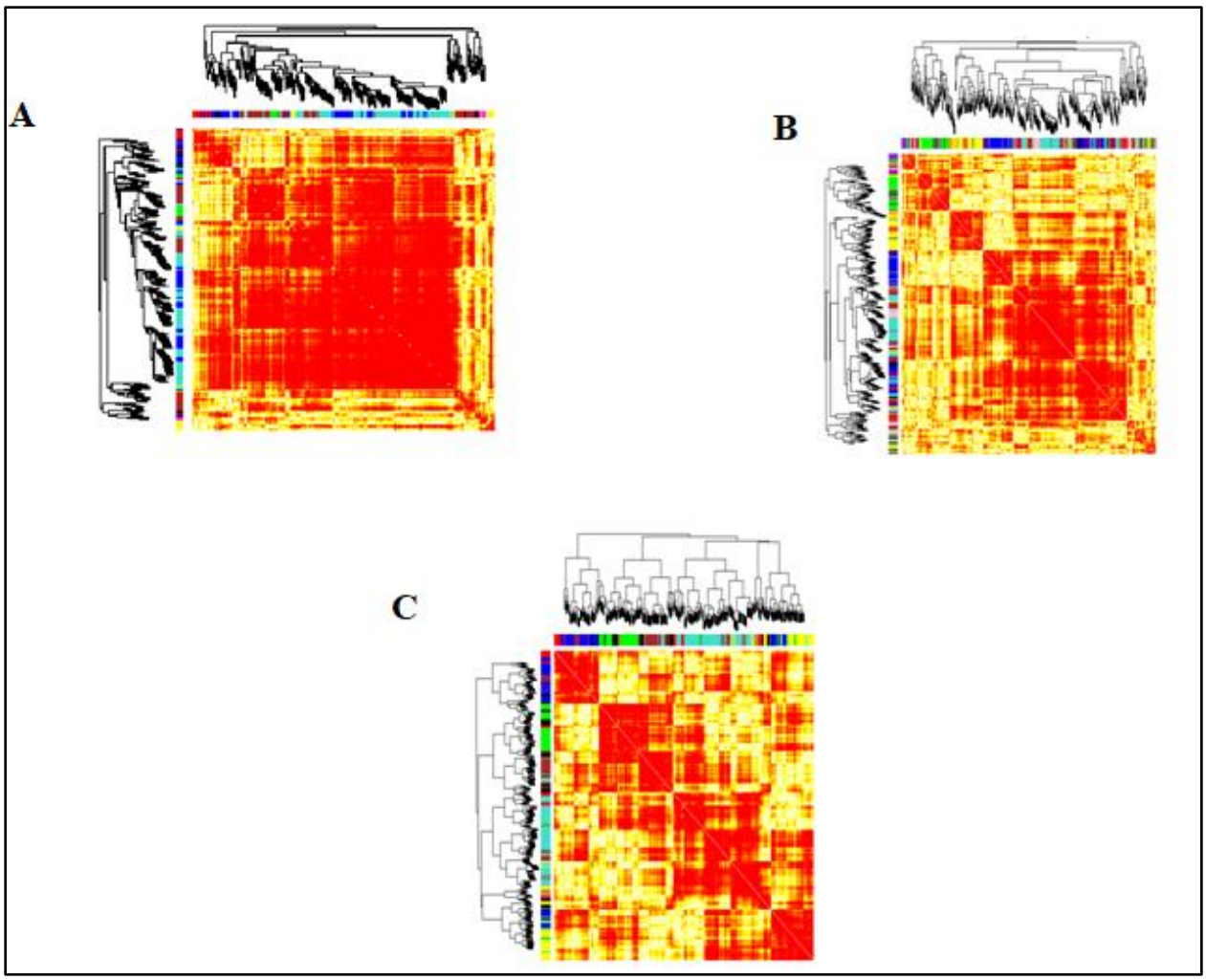


Figure 3: Network heatmap plots (A) bevacizumab treated stage-III patients, (B) bevacizumab treated stage-III patients, (C) Cetuximab treated patients

3.b: Detection of co-expression modules - Topological overlap measure was used for module detection. WGCNA identifies gene modules using unsupervised clustering. A co-expression module may reflect a true biological signal (e.g. a pathway) or it may reflect noise (e.g. a technical artifacts, tissue contamination, or a false positive). Table **2** shows the number of modules and their size range for all the three group of patients and Figure 4 shows the gene co-expression modules with different color for all the three patients cohorts. More information about the module color and each modules size are mentioned in appendix table i, ii, and iii for bevacizumab treated stage-III, bevacizumab stage-IV and cetuximab treated patients cohort respectively.

Table 2: Shows number of modules and modules size range for each cohort

Patients Cohort	***p-value***	**Number of modules**	**Modules size range**
Bevacizumab treated stage-III CRC Patients	0.05	9	46- 734
Bevacizumab treated stage-IV CRC Patients	0.05	53	30-968
Cetuximab treated CRC patients	0.05	8	1101 - 3860

Highly correlated module genes identified by WGCNA are represented and summarized by their first principal component. These are referred as the module eigengene [36]. The module eigengenes are used to define measures of module membership which quantify how close a gene is to a given module. Module membership measures allow to annotate all genes on the array and to screen for disease related intramodular hub genes[38].Gene significance measure as a function GS that assigns a non-negative number to each gene; the higher GS*i* the more biologically significant is gene *i* (Supplementary table b). Genes with high module membership in modules related to traits are natural candidates for further validation. Figure 5 shows scatter plot of Gene significance(GS) vs Module membership(MM) for mediumpurple module genes of bevacizumab treated stage-IV patients.

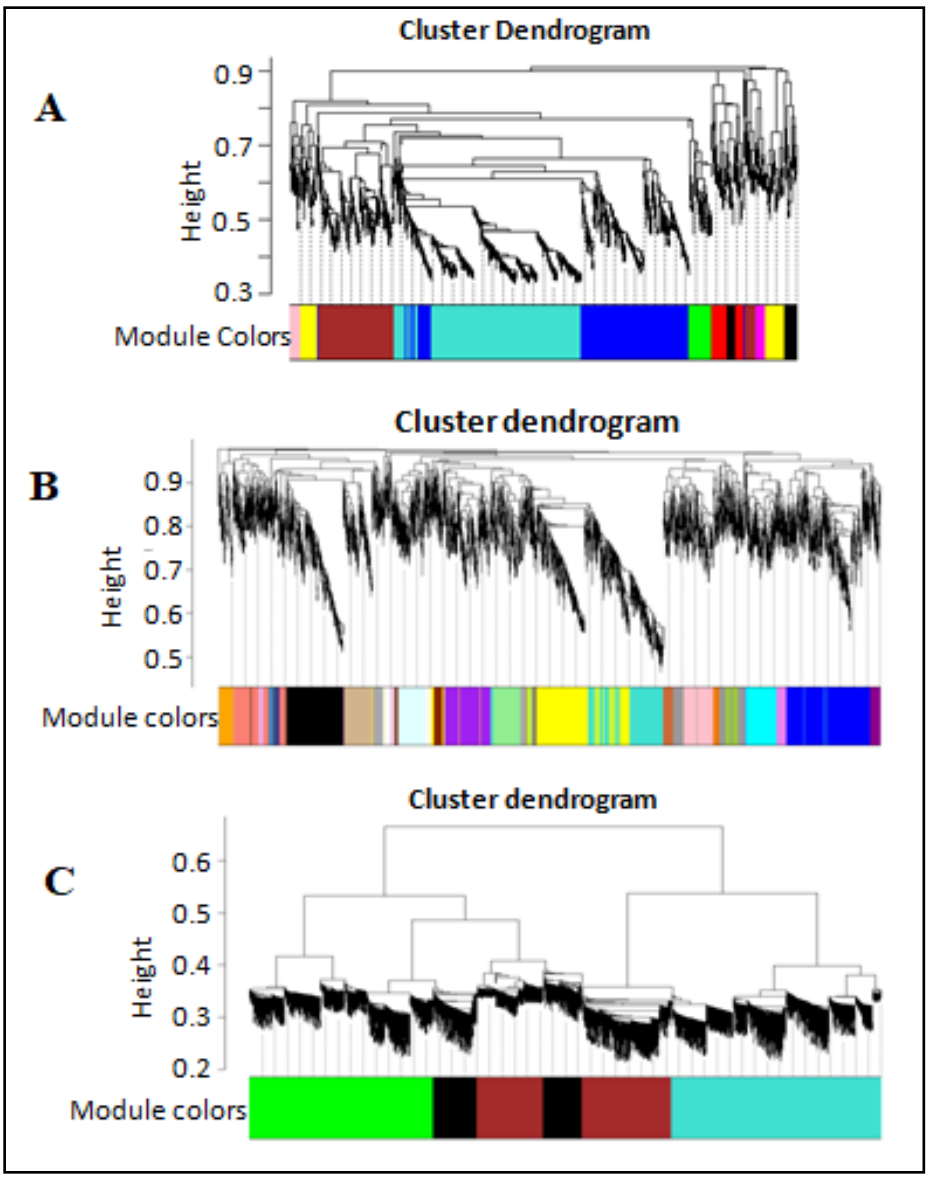


Figure 4: Gene co-expression modules color and dendrogram (A) For bevacizumab treated stage-III patients, (B) For bevacizumab treated stage-III patients, (C) For Cetuximab treated patients

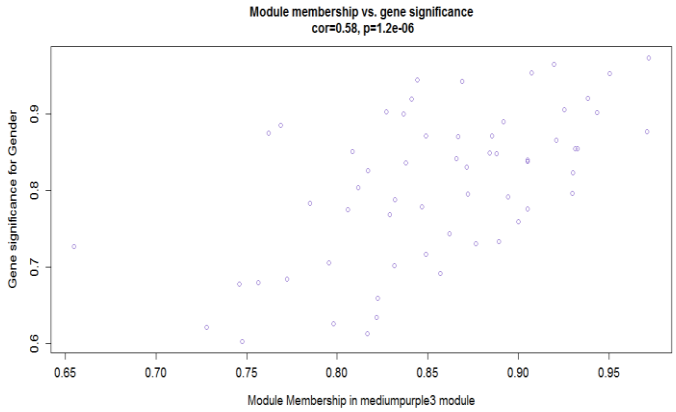


Figure 5: A scatterplot of Gene Significance (GS) for gender vs. Module Membership (MM) in the mediumpurple module of bevacizumab tretaed stage-IV patient. There is a highly significant correlation between GS and MM in this module

Step 4: Module preservation - To establish how well the modules defined in the larger reference set (each cohort) were preserved and reproducible in the test network(remaining two cohort), module preservation statistics were calculated for each reference-test module pair using a series of permutation tests for measures of module density and connectivity. Module preservation analysis were performed for all the three cohorts with respect to other cohorts. Module preservation statistics were calculated for each reference-test module pair using a series of permutation tests for measures of module density and connectivity.

Bevacacizumab treated stage -III cohort: Module preservation analysis for this cohort (9 modules) was performed with respect to bevacizumab treated stage-IV(52 modules) and cetuximab treated patients (7 modules). Only three modules of bevacizumab treated stage - III patients were preserved (summary Z-score>5)(appendix table iv). It was also observed that these three modules are strongly preserved within the cohorts (summary Z-score >10). Table 3 and Figure 6 show module preservation profile for bevacizumab treated stage-III patient cohort (test cohort) with respect to bevacizumab treated stage-IV and cetuximab treated patient cohort(reference cohorts).

Table 3: Module preservation of bevacizumab treated stage-III patients with respect to bevacizumab treated stage-IV and cetuximab cohort

Module color	Module preservation with	Module preservation with

	Bevacizumab treated stage-IV	Cetuximab treated patients
Black	7	7
Blue	2	2
Brown	3	3
Green	5	5
Magenta	9	9
Pink	8	8
Red	6	6
Turquoise	1	1
Yellow	4	4

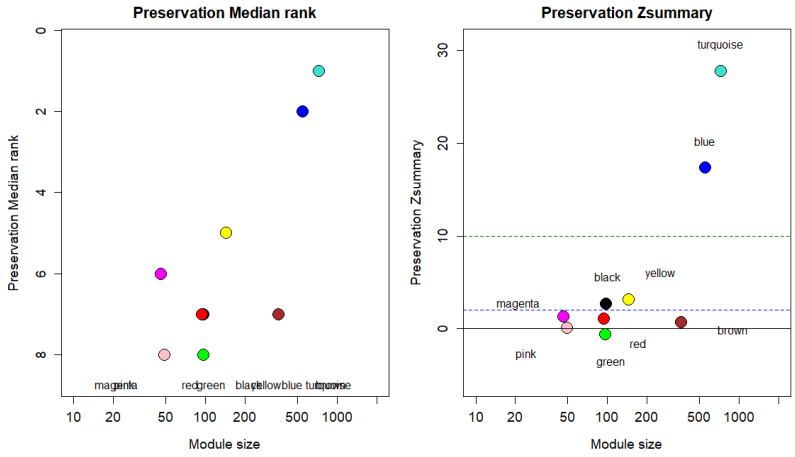


Figure 6: Preservation z-summary and preservation median rank for the modules of bevacizumab treated stage-III CRC patients

Bevacacizumab treated stage -IV cohort : Similarly, for this cohort, it was found that 47 out of 53 modules were preserved in test cohorts (summary Z-score>5), while some modules were appeared specific (black, blue, darkslateblue, grey, ivory and orange) to the this cohort network (Z-score <5) .(appendix table v). It was also observed that only 43 out of 53 modules were strongly preserved between the cohorts (summary Z-score > 10). Table 4 and Figure 7shows the module preservation profile for bevacizumab treated stage-IV.

Table 4: Module preservation of bevacizumab treated stage-IV patients with respect to bevacizumab treated stage-III and cetuximab cohort

Module color	Module preservation with Bevacizumab treated stage-III	Module preservation with Cetuximab treated patients
Bisque4	47	47
Black	7	7
Blue	2	2
Brown	3	3
Brown4	46	46
Cyan	14	14
Darkgreen	22	22
Darkgrey	24	24
Darkmagenta	34	34
Darkolivegreen	33	33
Darkorange	26	26
Darkorange2	45	45
Darkred	21	21
Darkslateblue	48	48
Darkturquoise	23	23
Floralwhite	44	44
Green	5	5
Greenyellow	11	11
Grey60	17	17
Ivory	43	43
Lightcyan	16	16
Lightcyan1	42	42
Lightgreen	18	18
Lightsteelblue1	41	41
Lightyellow	19	19
Magenta	9	9
mediumpurple3	40	40
Midnightblue	15	15
Orange	25	25

Orangered4	39	39
Paleturquoise	31	31
Pink	8	8
Plum1	38	38
Plum2	49	49
Purple	10	10
Red	6	6
Royalblue	20	20
Saddlebrown	29	29
Salmon	13	13
Salmon4	NA	NA
Sienna3	35	35
Skyblue	28	28
Skyblue3	37	37
Steelblue	30	30
Tan	12	12
Thistle1	NA	NA
Thistle2	50	50
Turquoise	1	1
Violet	32	32
White	27	27
Yellow	4	4
Yellowgreen	36	36

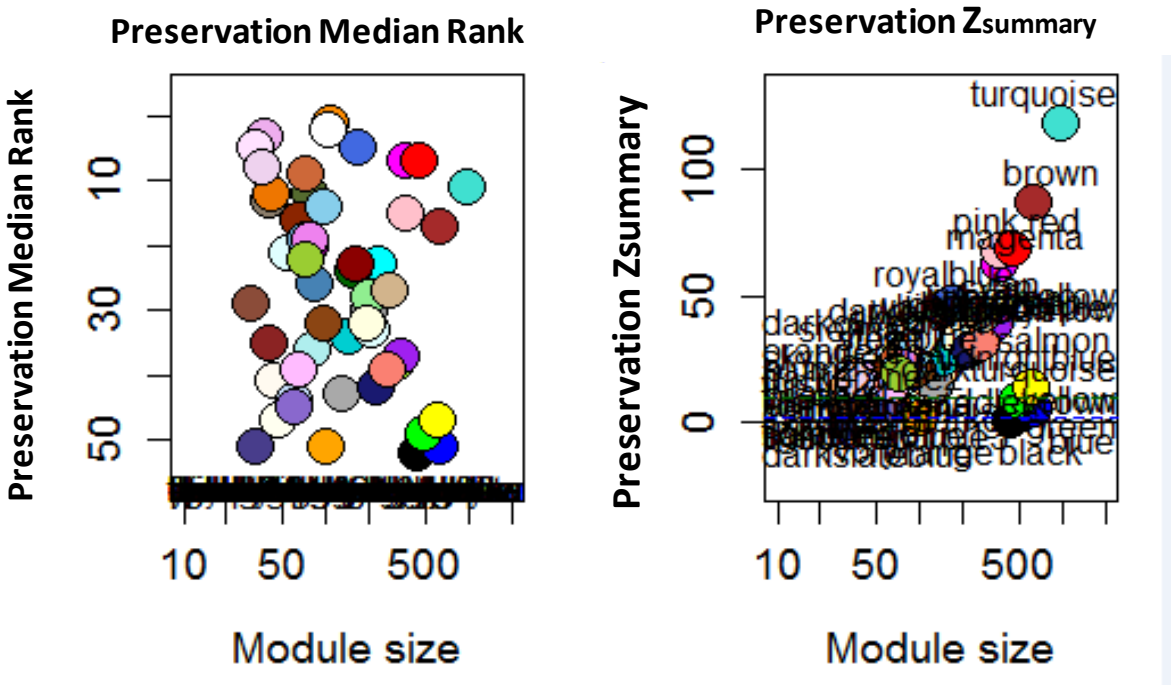


Figure 7: Preservation z-summary and preservation median rank for the modules of bevacizumab treated stage-IV CRC patients

Cetuximab treated CRC patients: For this cohort of patients, it was found that all the modules were shown to have well defined counterparts with bevacizumab treated stage-III patients' cohorts(summary Z-score>5), while only six modules out of eight were showing preservation with bevacizumab treated patients cohort (Z-score >5) (appendix table vi). Table 5 and Figure 8 shows the module preservation profile for cetuximab treated patients.

Table 5: Module preservation of Cetuximab treated patients with respect to bevacizumab treated stage-III and Bevacizumab treated stage-IV cohort

Module color	Module preservation with Bevacizumab treated stage-III	Module preservation with Bevacizumab treated stage-IV
Black	7	7
Blue	2	2
Brown	3	3
Green	5	5
Red	6	6
Turquoise	1	1
Yellow	4	4

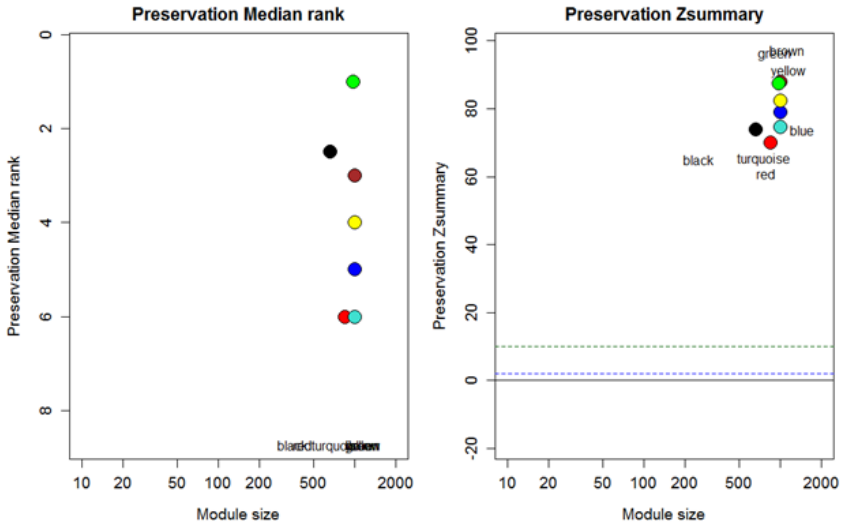


Figure 8: Preservation z-summary and preservation median rank for the modules of cetuximab treated CRC patients

Step 5: Module selection - To reduce the complexity only top three modules with highest correlation value were selected for further study. For each cohort selected modules , their size range and correlation values are summarized in Table 6.

Table 6: Shows selected modules, modules size and their correlation values

Patients cohort	Module color	Module size	Correlation value
Bevacizumab treated stage-III	Pink	43	0.88
	Yellow	135	0.53
	Brown	339	0.48
Bevacizumab treated stage-IV	Mediumpurple3	57	0.94
	DarkRed	144	0.63
	Yellow	530	0.62
Cetuximab treated patients	Blue	1822	0.99
	Brown	1860	0.67
	Red	1218	0.66

Step 5.a: Identification of TFs - gProfiler was applied to identify TFs of selected module of the three cohorts and TF related genes were identified by using Uniprot database. Figure 9 and Figure 10 shows the overlap between TFs across modules of each cohort and TF overlapping across all three patients' cohort .

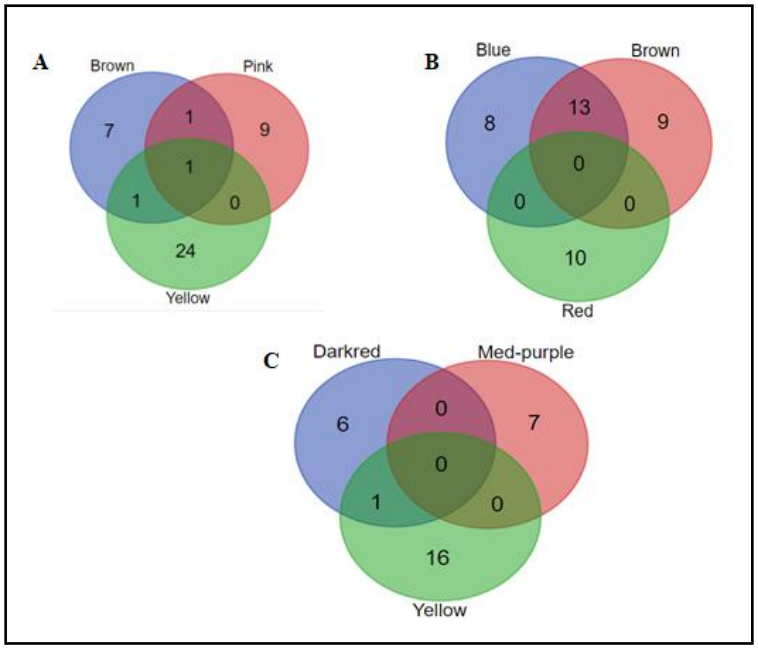


Figure 9: Common TFs among selected modules (A) Venn diagram of common TFs of bevacizumab treated stage-III cohort (B) Venn diagram of common TFs of cetuximab treated patients, (C) Venn diagram of common TFs of Cetuximab treated patients

It was observed that each modules consists of large number of TFs. iRegulon a cytoscape plug in was used to indentify most significant TFs for each modules. iRegulon ranked the significant TFs on the basis of enrichment score.

Table 7, Table 8 and Table 9 shows the significant TFs of selected modules of each patient cohort. Top three TFs in each modules were further analyzed for their significance in CRC.

Table 7: Significant TFs for selected modules of bevacizumab treated stage-III patients

Modules	TF	Enrichment score
Brown	**ATF3**	4.53
	SP2	4.50

	ELK1	3.98
	MEF2A	3.88
	ELF3	3.86
	YOD1	3.6
	STAT5A	3.39
	IRF5	3.35
	GABPA	3.31
Pink	**EP300**	6.7
	STAT6	5.94
	ZNF652	5.58
	IRF1	4.29
	PRDM1	4.13
	MEF2C	4.11
	EBF1	4.10
	MEF2A	3.7
	YY1	3.66
	USF1	3.45
Yellow	**HNF4A**	4.85
	BRF1	4.27
	GRHL1	4.12
	STAT5A	3.80
	GMEB1	3.73
	POLR3A	3.71
	ELF1	3.70
	GATA1	3.44
	MEF2A	3.40

Table 8: Significant TFs for selected modules of bevacizumab treated stage-IV patients

Modules	TF	Enrichment score
DarkRed	**RFX5**	5.31

	USF2	3.58
	JUN	3.43
	ELF1	3.38
	STAT1	3.36
	NKX3-2	3.35
	PITX3	3.34
	POU5F1	3.29
	EGR1	3.21
	IRF1	3.19
Mediumpurple	**CEBPB**	4.44
	TAF1	4.36
	YY1	3.98
	CCNT2	3.50
	TGIF2	3.49
	JUND	3.28
	NRF1	3.24
	IRF5	3.20
	CYB5R1	3.10
	SREBF1	3.04
Yellow	**IRF1**	3.72
	ELF1	3.10
	STAT1	3.08
	TEAD1	3.05
	PAX3	3.04
	ETV6	3.01
	CRX	3.00
	SOX14	2.99
	ASAP3	2.98

Table 9: Significant TFs for selected modules of Cetuximab treated patients

Modules	TF	Enrichment score
Blue	**GABPA**	3.78
	TP53	3.06
	YY1	3.03
	GATA1	3.02
	ESRRB	3.01
	FOXA3	3.00
Brown	**NFIC**	4.51
	SPI1	4.40
	CEBPB	3.25
	ZNF354	3.24
	RHOXF1	3.23
Red	**E2F4**	5.90
	BRCA1	4.96
	CEBPB	3.73
	UBTF	3.49
	KAT2A	3.25
	MYC	3.24
	POLR2A	3.07
	NR3C1	3.07
	USF1	3.04

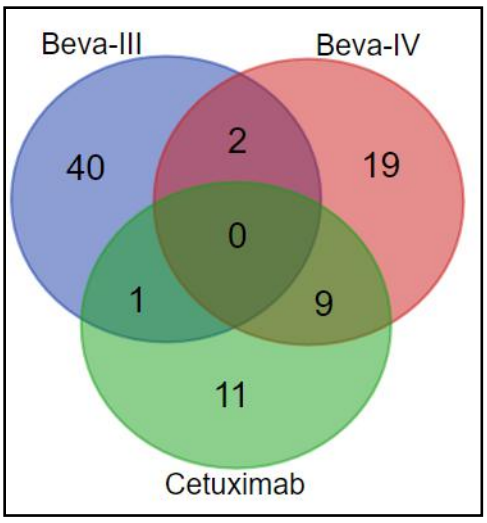


Figure 10: Venn diagram for common TFs among all three patients cohorts

On TFs analysis it was observed that only two TFs were common among bavacizumab treated stage-III patients and bevacizumab treated stage- IV patients, one TF was common between bevacizumab treated stage-III and cetuximab treated patients and nine TFs was common among bevacizumab treated stage-IV and cetuximab treated patients. .

Step 5.b Pathway analysis of TFs - Pathway analysis was performed for all TFs of selected modules in each of the three cohorts, It was observed that three most significant pathways across all cohort are as : (i) transcriptional mis-regulation, (ii) pathways in cancer and (iii) MAPK signaling pathway.

Table 10 shows the pathway analysis result for each of the three cohort.

Table 10: Pathway analysis result for TFs for all the three cohort

Cohorts	Pathway	Number of TFs	TFs
Bevacizumab stage-III	Transcriptional mis-regulation in cancer	5	*CEBPA, RARA, RXRA, SP1, TFE3*
	Pathways in cancer	2	*EP300, STAT5A*
	MAPK signaling pathway	2	*ELK1, MEF2C*
Bevacizumab stage-IV	Transcriptional mis-regulation in cancer	4	*CNT2, CEBPB, ETV6, PAX3*
	Pathways in cancer	2	*JUN, STAT1*
	MAPK signaling pathway	2	*JUN, JUND*
	Colorectal cancer genes	1	*JUN*
Cetuximab Stage-IV	Transcriptional mis-regulation in cancer	4	*CEBPB, MYC, SPI1, TP53*
	Pathways in cancer	3	*MYC, SPI1, TP53*
	MAPK signaling pathway	2	*MYC, TP53*
	Colorectal cancer genes	2	*MYC, TP53*

5.c : Visualizing the gene network - Further, genes of selected modules for each of the cohort were exported in cytoscape and networks were generated for all the three cohorts. Figure 11 shows the network of brown, yellow and pink module of bevacizumab treated stage-III module.

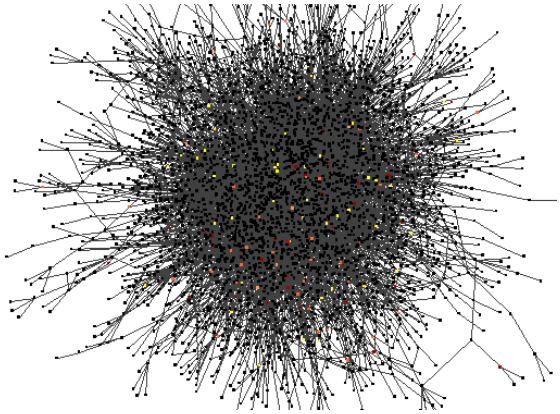


Figure 11: Network of Brown, yellow and pink module of bevacizumab treated stage-III CRC patients

CHAPTER FIVE: DISCUSSION AND CONCLUSION

Discussion

A total of 30 gene expression samples belonging to three different treatment regime and control were downloaded from the GDC website. A large number of DE genes (*p-value <0.001*) were identified for both bevacizumab-IV and cetuximab treated patients cohorts (10,817 and 14,590 respectively). Whereas only 2,724 DE genes were identified for bevacizumab treated stage-III cohort. A total of 6,732 DE genes overlapped between bevacizumab-IV and cetuximab treated patients cohorts. A total of 2,184 DE genes were common across bevacizumab treated stage-III, bevacizumab treated stage-IV and cetuximab treated patients' cohort. Additionally, 44 genes were common in between bevacizumab treated stage-III and bevacizumab treated stage-IV cohort and 436 genes were common to bevacizumab treated stage -III and cetuximab treated patients' cohort. Each cohort has unique DE genes 60, 1852 and 5229 for bevacizumab stage-III, bevacizumab stage-IV and Cetuximab treated cohort respectively. This shows that the DE genes profiles are common as well as unique to each treatment regime.

.Co-expression module analysis was performed for three different cohort and it was hypothesized that tightly co-expressed gene modules, enriched in shared functional annotation, would provide the most fruitful predictions of candidate gene. Co-expression networks heatmap were generated using WGCNA function. In the heatmap, high co-expression interconnectedness is indicated by progressively more saturated yellow and red colors (Figure 3). Number of modules were identified for bevacizumab treated stage-III, bevacizumab treated stage-IV and cetuximab treated patient cohorts was nine, 53 and eight with their size ranges from 46-734, 30-968 and 1101-3860 respectively. Some modules have large number of genes , some have small number of genes indicating modules size varies across and between the cohorts. On evaluating the modules preservation across all the cohorts, it was observed that for bevacizumab treated stage-III CRC patients, all the modules are preserved across other two cohort (Z-summary>5), while in the case of bevacizumab treated stage-IV patients(52 modules), only 47 modules are preserved(Z-summary>5). module preservation was performed for each cohort with respect to other two remaining cohort. This indicates that bevacizumab treated stage-IV patients have high number of DE genes and as a result the modules with this late DE genes are not identified in early stage.

For cetuximab treated CRC patients all the seven modules were preserved within and between the cohort. The modules identified were analyzed for their correlation with CRC. Each modules was annotated with its correlation score computed by WGCNA. The top three high correlation value modules were used for further analysis, i.e., to understand their significance in CRC. The modules were annotated with colors and scores.

From the literature, it is known that TFs play an important role in CRC. On analyzing top three modules of each cohort, it was identified that each of these modules consist of large number of genes. As for each of the cohort, module consists of large number of genes , to reduce the complexity the TFs in these cohort were further analyzed for their significance in CRC. It was observed that there was very less overlap between TFs of top three modules in each cohorts depicting that these modules were unique in content of TFs and consecutively showing the different treatment regime. iRegulon was used to rank the TFs in context of their enrichment score and targets. Using literature top three TFs from each modules were further analyzed for their correlation in CRC.

Table **11** shows the top three TFs and their association with CRC for each cohort.

The TFs were also analyzed for their pathway enrichment. It was observed that TFs in all these modules enriched in well known CRC pathways as well as some unique pathways. The common CRC pathways enriched were; transcriptional mis-regulation in cancer, pathways in cancer and MAPK signaling pathway. Top three pathways and related TFs are reported in the Table 10 . When TFs were analyzed in respect of colorectal cancer, it was found that only one TF *(JUN*) is associated with CRC pathway for bevacizumab treated stage- IV patients. In addition to that *JUN* was also associated with other pathways such as: pathways in cancer and MAPK signaling pathway for bevacizumab treated stage-IV patients. *JUN* is a proto-oncogene which produce *c-jun* protein. Over-expressed c-jun act as an oncogene [44]. It increase the angiogenic activity of tumor cells, by increasing expression of a positive regulator of angiogenesis, such as VEGF, by decreasing expression of an angiogenesis inhibitor, such as thrombospondin-1, or by both mechanisms [45]. Therefore, targeting c-Jun could be candidates for novel angiogenesis inhibitors for use in cancer or in other angiogenesis-dependent diseases, such as macular degeneration or psoriasis. Further, for cetuximab tretaed stage-IV patients, only two TFs*(MYC and TP53)* are associated with CRC pathway and these two TFs are also associated with other pathways such as: transcriptional mis-regulation, pathways in cancer and MAPK signaling

pathway. Myc is a proto oncogene , contributes to the genesis of many human cancer. It activates the transription of growth related genes. Its expression and function have led to new cancer therapeutic opportunities. Myc's activation could be inhibited by drug-like molecules, resulting in tumor inhibition in vivo [46]. Therefore, it could be a potent new biomarker for CRC patients. On contrary, no CRC associated TFs were identified for bevacizumab treated stage-III patients. Most probably it is due to lesser number of mutated genes in stage-III patients in comparison to stage-IV patients. On pathways analysis of TFs, it was observed that CEBPs were common among all the patient cohort. C/EBPs are DNA binding proteins. The CEBP encoded proteins are essential for terminal differentiation and functional maturation of granulocyte progenitor cells. The main function of CEBPB are the regulation of expression of genes involves in immune and inflammatory response[47]. So, CEBP could also be a new biomarker for CRC.

Table **11** shows that each treatment regime identified TFs known to be involved in CRC and in other cancers. In addition, each treatment regime also identified new TFs that can be further validated in labs(wet-lab) such as: (i) for bevacizumab stage-III patients, top known TFs in CRC are [EP300, STAT6,ATF3,SP2 and HNFA]. The new TFs that are not validated as biomarker or reported in CRC are (ELK1 and GRHL1). (ii) for bevacizumab stage-IV patients, the known TFs in CRC are RFX5,IRF1, STAT1 and ELF1. The new TFs which could be potent biomarker are CEBPs, YY1 and JUN.(iii) for cetuximab treated patient cohort, known TFs are TP53 and E2F4. The new TFs are GABPA, YY1, NFIC, BRCA1 and CEBPE.

Table 11: Significant TFs function and their association with cancer

Bevacizumab treated stage-III patients				
Module color	**TFs**	**Function**	**Cancer**	**Biomarker**
Pink	EP300	Regulates transcription via chromatin remodeling and important in the processes of cell proliferation and differentiation. It is a co-activator of HIF1A (hypoxia-inducible factor 1 alpha), and thus plays a role in the stimulation of hypoxia-induced genes such as VEGF [48]. Diseases associated with EP300 include Rubinstein-Taybi syndrome2 and Colorectal cancer [49].	CRC Stomach Breast Pancreas Prostrate	Yes
	STAT6	Promote proliferation and inhibit apoptosis. STAT6 is over-expressed and active in numerous malignancies including	CRC, Breast Glioblastoma Urothelial	Yes

		prostate and colon cancer [50].	Lymphoma Prostrate	
	ZNF652	A transcriptional repressor protein. It directly repressed key drivers of invasion and metastasis, such as TGFB1, TGFB2, TGFBR2, EGFR, SMAD2 and VIM. Loss of ZNF652 expression in primary breast tumors was significantly correlated with increased local invasion and defined a population of breast cancer patients with metastatic tumors [51] .	Breast Melanoma Endometrial Thyroid	New?
Brown	ATF3	ATF3 up regulates the expression of several genes in the tumor necrosis factor pathway in cancer development [52]. Up regulation of ATF3 in lung cancer promotes cell proliferation, migration, and invasion, and may represent a novel therapeutic target for lung cancer [53]. ATF3 may play a dichotomous role in apoptosis and metastasis in human colorectal cancer cells [54] .	Colon Breast Lung Pancreas	Yes
	SP2	Over expression of Sp2 Inhibits Epidermal Differentiation and Increases Susceptibility to Wound- and Carcinogen-Induced Tumor genesis [55]. Sp family proteins are required for endogenous expression of VEGF in pancreatic cancer cells. COX-2 inhibitors and other nonsteroidal anti-inflammatory drugs (NSAIDs) inhibit development and growth of colon cancer by down regulating sp-proteins. COX-2 inhibitors decreased VEGF protein, mRNA, and reporter gene expression, and this was accompanied by down-regulation of both Sp1 and Sp4 but not Sp3 or Sp2 proteins [56].	Colon Carcinoid Ovarian Stomach	Yes
	ELK1	It forms a ternary complex by binding to the serum response factor and the serum response element in the promoter of the c-fos proto-oncogene. The protein encoded by this gene is a nuclear target for the ras-raf-MAPK signaling cascade.	Colon Breast Prostrate Gastric	New?
Yellow	HNF4A	It is expressed in the liver, pancreas, kidney, stomach, small intestine and colon, where it regulates many important aspects of epithelial cell morphogenesis and function. HNF4α plays a protective role against inflammatory bowel disease, an important risk factor for colorectal cancer [57].	CRC Liver Stomach	Yes

	BRF1	Brf1 (TFIIB-related factor 1) plays a crucial role in cell transformation and tumorigenesis. Brf1 is a new biomarker of heptocellular carcinoma(HCC) [58]. Germline mutations in BRF1 associated with predisposition to CRC [59].	Prostrate Heaptocellular carcinoma (HCC)	
	GRHL1	.GRHL1 is a tumor suppressor in the squamous cell carcinoma (SCC) of the skin. GRHL proteins play important role in cancer development [60].	Skin Cervical Urethral	New?
Bevacizumab tretaed stage-IV				
Module color	**TFs**	**Function**	**Cancer**	**Biomarker**
Darkred	RFX5	RFX5 mutations were related to lack of or faint HLA class II antigen expression IN CRC. Somatic mutations of the RFX5 gene represent a novel mechanism of loss of HLA class II antigen expression in tumor cells, potentially contributing to immune evasion in MSI-H CRCs [61].	CRC Endometrial Lymphoma Stomach Lungs Head and neck	Yes
	USF2	It regulates cellular growth and proliferation. USF2 involves in prostate tumorigenesis [77]. USF2 associated with reoccurrence of CRC [62].	Breast Lung ovarian	New?
	JUN	c-Jun potentially plays an important role in carcinogenesis and cancer progression. c-Jun activation is associated with proliferation and angiogenesis in invasive breast cancer [63]. Jun protein family levels has been reported higher in CRC patients [64].	CRC, Breast, Prostrate, Skin Lung, Ovarian Bladder	New?
Mediumpurple	CEBP	(C/EBPs) are a family of leucine-zipper TFs that regulate gene expression to control cellular proliferation, differentiation, inflammation and metabolism. Encoded by an intronless gene. C/EBP-β-activated miR-223 contributes to tumor growth by targeting RASA1 in CRC and miR-223-targeted inhibitors may have clinical promise for CRC treatment via suppression of miR-223[65].	CRC Liver Lung Testis	New?
	TAF1	It encodes a protein that functions as a transcriptional co-activator. The protein is also involved in the recognition of transcriptional promoters and the modification of general transcription factors (GTFs).	Breast CRC Renal Liver	New?
	YY1	Yin Yang 1 (YY1) is highly expressed in various types of cancers and regulates tumor genesis through multiple pathways.	Breast Prostrate CRC	New?

		YY1 is generally over expressed in breast cancer cells and tissues. Yin Yang 1 plays an essential role in breast cancer and negatively regulates p27 [66]. Role of YY1 in repression of dominant negative LEF-1(Lymphoid enhancer factor 1) expression in colon cancer [67].		
Yellow	IRF1	Some tumors escape the antitumor effects of IFN-gamma by cellular changes reflected in IRF-1 and IRF-2 expression. IRF-1 expression is consistent with its role as a tumor suppressor [68]. IRF1 involves in CRC tumorigenesis. A positive feedback loop between IRF1 and miR-29b may contribute to the sensitivity of CRC cells to IFN-γ [69].	Breast Stomach Cervical CRC	Yes
	STAT1	Transcription factors STAT3 and STAT1, both downstream effectors of interleukin (IL)-6 and its receptor, are involved in growth and developmental control of CRC cells [70].	CRC Melanoma Breast Skin	Yes
	ELF1	It involved in cellular growth and differentiation. Enhanced expression of Elf-1 has been reported in prostate cancer, breast cancer, and osteosarcoma [71]. Ets factors show altered expression in colon cancer, where they regulate pathways relevant to tumor progression [72].	CRC Breast Pancretic Skin	Yes
Cetuximab treated patients				
Module color	**TFs**	**Function**	**Cancer**	**Biomarker for CRC**
Blue	GABPA	It involved in activation of cytochrome oxidase expression and nuclear control of mitochondrial function. Abnormal expression of GABPA is associated with tumor development and progression[73]. GABP is essential for eIF6 promoter activity which indicates GABP is a global regulator of ribosome biosynthesis.	Carcinoid Heaptocellular Breast CRC	New?
	TP53	Somatic *TP53* mutations occur in almost every type of cancer at rates from 38%–50% in ovarian, esophageal, colorectal, head and neck, larynx, and lung cancers to about 5% in primary leukemia, sarcoma, testicular cancer, malignant melanoma, and cervical cancer[74].	Ovarian Esophagal CRC Breast Head and neck Melanoma cervical	Yes
	YY1	Yin Yang 1 (YY1) is highly expressed in various types of cancers and regulates tumor genesis through multiple pathways.	Breast Prostrate CRC	New?

		YY1 is generally over expressed in breast cancer cells and tissues. Yin Yang 1 plays an essential role in breast cancer and negatively regulates p27 [66].		
Brown	NFIC	Downstream factors of NFI-C, such as KLF4 and E-cadherin play roles in epithelial-to-mesenchymal transition (EMT). NFI-C is expressed in normal mammary gland or noninvasive breast cancer cells. NFI-C play important role in regulating KLF4 during tumorigenesis [75]. NFIC also involved in CRC [76].	Breast Stomach Endometrial CRC Ovarian	New?
	SPI1	Activates gene expression during myeloid and B-lymphoid cell development. The nuclear protein binds to a purine-rich sequence known as the PU-box, and regulates their expression in coordination with other transcription factors and cofactors.	Glioma Lung Stomach Lymphoma pancreatic	New?
	CEBPE	C/EBPs) are a family of leucine-zipper transcription factors that regulate gene expression to control cellular proliferation, differentiation, inflammation and metabolism. Encoded by an intronless gene. Its related pathways are transcriptional mis-regulation [77].	CRC Liver Lung Testis	Yes
Red	E2F4	E2F4 expression slowed down G1/S phase transition and the proliferation rate of normal human intestinal epithelial cells (HIEC) and of colon cancer cells[78].	Breast CRC Prostrate Lung	Yes
	BRCA1	BRCA1 predicted survival in all colorectal cancer patients. Low expression of BRCA1 was associated with loss of MLH1 or MSH2 expression[79]. BRCA1 is a TF involved in CRC[76] .	Breast Prostrate Ovarian CRC	New?
	CEBPE	C/EBPs) are a family of leucine-zipper transcription factors that regulate gene expression to control cellular proliferation, differentiation, inflammation and metabolism. Encoded by an intronless gene. Its related pathways are transcriptional mis-regulation [80].	CRC Liver Lung Testis	Yes

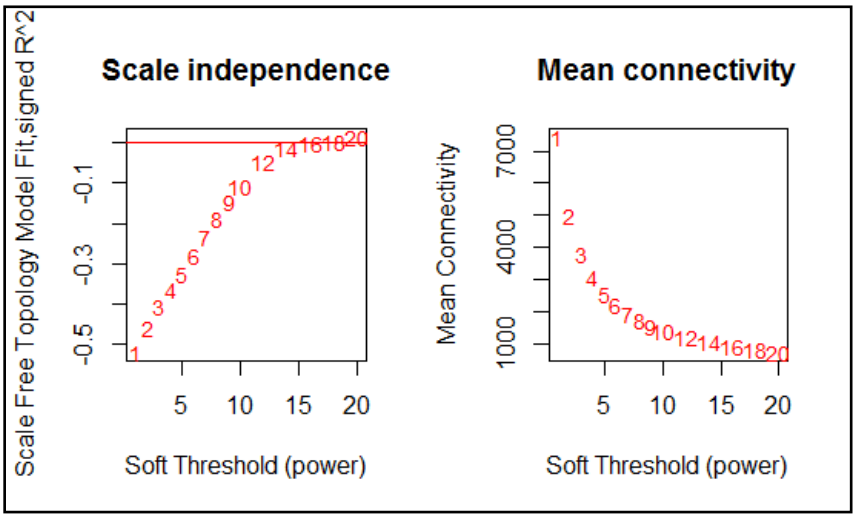
Conclusion

This brief analysis shows that different treatment regime do target some common and known CRC related genes and TFs . But at the same time they are unique in context of their targets. It also underlies the knowledge that there are genotypical variation in cancer across the patient. It is essential to identify the patient profile initially and then go for right target for particular patient. The above analysis shows that though there is overlap between the DEs across the different regimes, still each of the regimes are unique in their context which is showing a clear necessity for targeted treatment for right patient.

Supplementary Files

a) Differentially expressed transcripts for each of the cohorts i.e bevacizumab treated patients (stage-III and Stage-IV) and Cetuximab treated CRC patients.
b) Gene significance and Gene module membership for all three cohorts.
c) GO Enrichment Analysis result of all three cohorts.

Appendix



Appendix figure i : Scale free topology for co-expression network of patients treated with cetuximab

Table i : Modules color and size of 14 identified modules for bevacizumab treated stage-III CRC patients

Module color	Module size
Pink	43
Magenta	44
Red	88
Black	91
Green	95
Yellow	135
Brown	339
Blue	510
turquoise	682

Table ii : Module color and size of 53 identified modules of bevacizumab trated stage-IV CRC patients

Module color	Module size
Grey	1
salmon4	23
thistle1	24
darkslateblue	28
plum2	33
thistle2	33
bisque4	36
darkorange2	36
floralwhite	36
brown4	38
Ivory	39
lightcyan1	48
plum1	49
orangered4	53
lightsteelblue1	54
mediumpurple3	57
skyblue3	58
yellowgreen	60
darkmagenta	64
darkolivegreen	66
steelblue	67
Violet	68
sienna3	69
paleturquoise	76
skyblue	79
orange	82
saddlebrown	82

White	89
darkorange	91
darkgrey	114
darkturquoise	127
darkgreen	135
darkred	144
royalblue	150
grey60	174
lightgreen	179
lightyellow	179
lightcyan	185
midnightblue	204
Cyan	217
Tan	223
salmon	236
greenyellow	278
Purple	294
Pink	307
magenta	310
Black	378
Red	385
Green	429
Brown	526
Yellow	530
Blue	566
turquoise	843

Table iii : Seven identified modules of cetuximab trated CRC patients

Module color	Module size
Black	916
Red	1218
Green	1392
Yellow	1723
Blue	1822
Brown	1860
turquoise	3301

Table iv: Bevacizumab treated stage-III patients module preservation Z-summary

Bevacizumab stage-III and Cetuximab:

	medianRank.pres	medianRank.qual	Zsummary.pres	Zsummary.qual
black	7	7	2.70	5.90
blue	2	5	17.00	19.00
brown	7	9	0.70	11.00
gold	3	10	23.00	0.62
green	8	1	-0.57	9.70
magenta	6	2	1.30	7.30
pink	8	4	0.10	6.10
red	7	6	1.10	6.20
turquoise	1	3	28.00	33.00
yellow	5	8	3.20	7.00

Bevacizumab stage-III to Bevacizumab stage-IV:

	medianRank.pres	medianRank.qual	Zsummary.pres	Zsummary.qual
black	7	7	2.70	5.90
blue	2	5	17.00	19.00
brown	7	9	0.70	11.00
gold	3	10	23.00	0.62
green	8	1	-0.57	9.70
magenta	6	2	1.30	7.30
pink	8	4	0.10	6.10
red	7	6	1.10	6.20
turquoise	1	3	28.00	33.00
yellow	5	8	3.20	7.00

Table v : Bevacizumab treated stage-IV module preservation Z-summary

Bevacizumab stage-IV and Cetuximab

	medianRank.pres	medianRank.qual	Zsummary.pres	Zsummary.qual
bisque4	12	1.0	17.0	27
black	52	22.0	1.6	72
blue	51	49.5	4.9	73
brown	16	50.5	78.0	75
brown4	36	10.0	10.0	22
cyan	24	21.0	42.0	64
darkgreen	23	29.5	32.0	45
darkgrey	43	40.0	15.0	43
darkmagenta	19	15.0	21.0	34
darkolivegreen	12	45.0	28.0	28
darkorange	1	13.5	33.0	48
darkorange2	11	4.0	16.0	30
darkred	21	42.0	32.0	38
darkslateblue	49	13.5	1.9	21
darkturquoise	35	23.5	26.0	48
floralwhite	41	32.0	9.1	21
gold	48	53.0	18.0	-1
green	48	46.0	7.4	73
greenyellow	38	43.0	40.0	58
grey	NA	NA	NaN	NA
grey60	31	49.5	32.0	37
ivory	47	41.0	2.3	20
lightcyan	34	37.5	31.0	46
lightcyan1	21	16.0	16.0	32
lightgreen	28	33.0	34.0	53
lightsteelblue1	45	31.0	7.1	24
lightyellow	31	48.0	30.0	39
magenta	6	25.0	60.0	74
mediumpurple3	45	11.0	8.1	33
midnightblue	42	44.0	26.0	48
orange	50	28.5	2.1	33
orangered4	17	20.5	20.0	26
paleturquoise	33	47.0	18.0	26
pink	15	27.5	61.0	76
plum1	39	39.0	14.0	26
plum2	4	3.0	15.0	22
purple	37	37.0	38.0	61
red	7	36.5	63.0	69
royalblue	5	19.5	46.0	47
saddlebrown	31	17.0	19.0	38
salmon	39	52.0	28.0	50
salmon4	28	2.0	8.9	25
sienna3	9	5.0	24.0	35
skyblue	15	34.5	24.0	30
skyblue3	18	12.0	20.0	33
steelblue	24	23.5	20.0	32
tan	29	37.0	37.0	56
thistle1	3	6.5	15.0	23
thistle2	8	8.5	16.0	22
turquoise	13	28.5	110.0	110

violet	19	8.5	21.0	37
white	2	6.5	29.0	41
yellow	47	26.5	14.0	85
yellowgreen	23	18.0	19.0	32

Bevacizumab stage-IV to Bevacizumab stage-III

	medianRank.pres	medianRank.qual	Zsummary.pres	Zsummary.qual
bisque4	13	1.0	16.00	30.0
black	52	20.0	1.40	99.0
blue	51	51.0	5.00	76.0
brown	17	47.0	87.00	83.0
brown4	35	10.0	12.00	25.0
cyan	23	22.5	42.00	63.0
darkgreen	24	32.5	33.00	45.0
darkgrey	43	40.5	15.00	36.0
darkmagenta	20	16.0	21.00	37.0
darkolivegreen	12	44.0	28.00	28.0
darkorange	1	14.0	33.00	46.0
darkorange2	12	4.0	17.00	31.0
darkred	23	40.5	35.00	44.0
darkslateblue	51	11.0	0.45	21.0
darkturquoise	34	24.0	25.00	47.0
floralwhite	41	32.0	9.50	21.0
gold	46	53.0	25.00	-1.3
green	49	46.0	8.70	78.0
greenyellow	38	43.0	39.00	60.0
grey	NA	NA	NaN	NA
grey60	30	50.0	36.00	40.0
ivory	47	41.5	2.50	22.0
lightcyan	33	38.0	38.00	53.0
lightcyan1	21	17.5	17.00	27.0
lightgreen	28	32.0	36.00	55.0
lightsteelblue1	44	28.5	8.90	28.0
lightyellow	32	49.0	34.00	42.0
magenta	7	25.0	63.00	79.0
mediumpurple3	45	12.0	7.40	36.0
midnightblue	42	45.0	29.00	51.0
orange	51	27.0	1.80	40.0
orangered4	16	21.0	22.00	30.0
paleturquoise	36	48.0	18.00	27.0
pink	15	30.5	68.00	81.0
plum1	39	39.0	15.00	31.0
plum2	3	3.0	18.00	28.0
purple	37	37.0	39.00	69.0
red	7	34.5	69.00	81.0
royalblue	5	17.0	48.00	56.0
saddlebrown	32	15.0	21.00	39.0
salmon	39	52.0	32.00	46.0

salmon4	29	2.0	8.60	24.0
sienna3	9	5.5	25.00	38.0
skyblue	14	35.5	30.00	33.0
skyblue3	19	13.0	21.00	35.0
steelblue	26	23.5	22.00	39.0
tan	27	34.5	43.00	68.0
thistle1	5	7.0	15.00	27.0
thistle2	8	8.0	17.00	28.0
turquoise	11	28.5	120.00	130.0
violet	19	9.0	22.00	42.0
white	2	5.5	34.00	52.0
yellow	47	26.0	14.00	100.0

Table vi: Cetuximab treated CRC patients module preservation Z-summary

Cetuximab to Bevacizumab stage-III

	medianRank.pres	medianRank.qual	Zsummary.pres	Zsummary.qual
black	2.5	3	74	120.00
blue	5.0	5	79	150.00
brown	3.0	2	88	150.00
gold	8.0	8	34	0.41
green	1.0	1	87	160.00
red	6.0	6	70	120.00
turquoise	6.0	7	75	130.00
yellow	4.0	4	82	140.00

Cetuximab to Bevacizumab stage-IV

	medianRank.pres	medianRank.qual	Zsummary.pres	Zsummary.qual
black	7	3	1.7	120.00
blue	1	5	23.0	150.00
brown	2	2	13.0	150.00
gold	5	8	2.1	0.41
green	5	1	11.0	160.00
red	4	6	7.3	120.00
turquoise	3	7	7.0	130.00
yellow	6	4	7.3	140.00

Table vii : Modules and their correlation values for bevacizumab treated stage-III patients

Module color	Correlation value
Pink	0.88
Magenta	0.19
Red	-0.48
Black	0.05
Green	0.15
Yellow	0.53
Brown	0.48
Blue	0.07
turquoise	-0.39

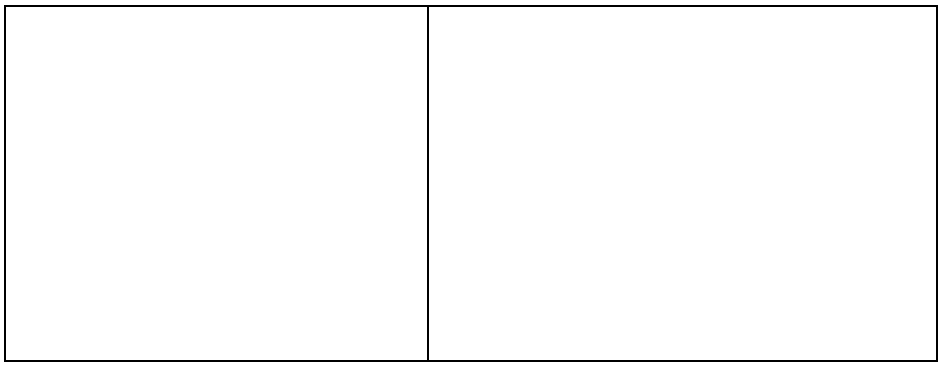


Table viii:Modules and their correlation values for bevacizumab treated stage-IV patients

Module color	Correlation value	Module color	Correlation value
Brown4	-0.754	Greenyellow	-0.631
Plum1	-0.576	Pink	-0.49
Darkgrey	-0.46	Saddlebrown	-0.441
Lightsteelblue1	-0.44	Skyblue3	-0.425
Tan	-0.41	Lightyellow	-0.389
Skyblue	-0.359	Grey60	-0.334
Orangered4	-0.331	Brown	-0.277
Lightcyan1	-0.256	Magenta	-0.249
Orange	-0.248	Red	-0.183
White	-0.162	Green	-0.153
Black	-0.101	Salmon	-0.094
Midnightblue	-0.069	Darkturquoise	-0.063
Darkorange	-0.057	Darkgreen	-0.052
Salmon4	-0.045	Royalblue	-0.021
Violet	-0.001	Ivory	0.007
Plum2	0.014	Cyan	0.019
Darkorange2	0.024	Grey	0.036
Thistle1	0.061	Bisque4	0.099
Steelblue	0.124	Turquoise	0.177
Darkolivegreen	0.184	Floralwhite	0.262
Yellowgreen	0.301	Darkslateblue	0.304

Thistle2	0.318	Lightcyan	0.339
Purple	0.35	Blue	0.365
Darkmagenta	0.386	Lightgreen	0.407
Sienna3	0.439	Darkred	0.584
Paleturquoise	0.63	Yellow	0.62
Mediumpurple3	0.943		

Table ix :Modules and their correlation values for cetuximab treated patients

Module color	Correlation value
Green	-0.31
Blue	0.99
Brown	0.66
Black	-0.42
Yellow	0.16
Red	0.67
Turquoise	-0.28

References

1. *Cancer Facts and Statistics.* American Cancer Society, Atlanta, Georgia, 1947. **1**.
2. *Colorectal Cancer Facts & Figures.* American Cancer Society, Atlanta, Georgia, 2014-2016.
3. Miller, K.D., et al., *Cancer treatment and survivorship statistics, 2016.* CA Cancer J Clin, 2016. **66**(4): p. 271-89.
4. Goel, A., et al., *Characterization of Sporadic Colon Cancer by Patterns of Genomic Instability.* Cancer Research, 2003. **63**(7): p. 1608-1614.
5. Lièvre, A., et al., *KRAS Mutation Status Is Predictive of Response to Cetuximab Therapy in Colorectal Cancer.* Cancer Research, 2006. **66**(8): p. 3992-3995.
6. Shepherd, F.A., et al., *Erlotinib in previously treated non-small-cell lung cancer.* N Engl J Med, 2005. **353**(2): p. 123-32.
7. Hurwitz, H., et al., *Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer.* N Engl J Med, 2004. **350**(23): p. 2335-42.
8. Saltz, L.B., et al., *Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor.* J Clin Oncol, 2004. **22**(7): p. 1201-8.
9. Van Cutsem, E., et al., *Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer.* J Clin Oncol, 2007. **25**(13): p. 1658-64.
10. Moore, M.J., et al., *Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group.* J Clin Oncol, 2007. **25**(15): p. 1960-6.
11. Cunningham, D., et al., *Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer.* N Engl J Med, 2004. **351**(4): p. 337-45.
12. Mendelsohn, J. and J. Baselga, *Epidermal growth factor receptor targeting in cancer.* Semin Oncol, 2006. **33**(4): p. 369-85.

13. Hynes, N.E. and H.A. Lane, *ERBB receptors and cancer: the complexity of targeted inhibitors.* Nat Rev Cancer, 2005. **5**(5): p. 341-54.
14. Downward, J., *Targeting RAS signalling pathways in cancer therapy.* Nat Rev Cancer, 2003. **3**(1): p. 11-22.
15. Schubbert, S., K. Shannon, and G. Bollag, *Hyperactive Ras in developmental disorders and cancer.* Nat Rev Cancer, 2007. **7**(4): p. 295-308.
16. Bos, J.L., *ras oncogenes in human cancer: a review.* Cancer Res, 1989. **49**(17): p. 4682-9.
17. Lievre, A., et al., *KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer.* Cancer Res, 2006. **66**(8): p. 3992-5.
18. Ince, W.L., et al., *Association of k-ras, b-raf, and p53 status with the treatment effect of bevacizumab.* J Natl Cancer Inst, 2005. **97**(13): p. 981-9.
19. Esteller, M., et al., *K-ras and p16 aberrations confer poor prognosis in human colorectal cancer.* J Clin Oncol, 2001. **19**(2): p. 299-304.
20. Di Fiore, F., et al., *Clinical relevance of KRAS mutation detection in metastatic colorectal cancer treated by Cetuximab plus chemotherapy.* Br J Cancer, 2007. **96**(8): p. 1166-9.
21. Benvenuti, S., et al., *Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies.* Cancer Res, 2007. **67**(6): p. 2643-8.
22. Andreyev, H.J., et al., *Kirsten ras mutations in patients with colorectal cancer: the 'RASCAL II' study.* Br J Cancer, 2001. **85**(5): p. 692-6.
23. Pao, W., et al., *KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib.* PLoS Med, 2005. **2**(1): p. e17.
24. De Roock, W., et al., *KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab.* Ann Oncol, 2008. **19**(3): p. 508-15.
25. Kaplan, R.N., et al., *VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche.* Nature, 2005. **438**(7069): p. 820-7.
26. Sandler, A., et al., *Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer.* N Engl J Med, 2006. **355**(24): p. 2542-50.

27. Motzer, R.J., et al., *Sunitinib versus interferon alfa in metastatic renal-cell carcinoma.* N Engl J Med, 2007. **356**(2): p. 115-24.
28. Miller, K., et al., *Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer.* N Engl J Med, 2007. **357**(26): p. 2666-76.
29. Segal, E., et al., *A module map showing conditional activity of expression modules in cancer.* Nat Genet, 2004. **36**(10): p. 1090-8.
30. Hood, L., et al., *Systems biology and new technologies enable predictive and preventative medicine.* Science, 2004. **306**(5696): p. 640-3.
31. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
32. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-73.
33. Lamb, J., et al., *A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer.* Cell, 2003. **114**(3): p. 323-34.
34. Kang, Y., *Functional genomic analysis of cancer metastasis: biologic insights and clinical implications.* Expert Review of Molecular Diagnostics, 2005. **5**(3): p. 385-395.
35. Bild, A.H., et al., *Oncogenic pathway signatures in human cancers as a guide to targeted therapies.* Nature, 2006. **439**(7074): p. 353-7.
36. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.
37. Zhang, B. and S. Horvath, *A general framework for weighted gene co-expression network analysis.* Stat Appl Genet Mol Biol, 2005. **4**: p. Article17.
38. Leonard Kaufman, P.J.R., *Finding Groups in Data: An Introduction to Cluster Analysis*. 27 MAY 2008.
39. Mueller, A.J., et al., *Cross-species gene modules emerge from a systems biology approach to osteoarthritis.* npj Systems Biology and Applications, 2017. **3**(1): p. 13.
40. Reimand, J., et al., *g:Profiler-a web server for functional interpretation of gene lists (2016 update).* Nucleic Acids Res, 2016. **44**(W1): p. W83-9.

41. Apweiler, R., et al., *UniProt: the Universal Protein knowledgebase.* Nucleic Acids Research, 2004. **32**(Database issue): p. D115-D119.
42. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.
43. Shannon, P., et al., *Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.* Genome Research, 2003. **13**(11): p. 2498-2504.
44. Folkman, J., *Angiogenesis and c-Jun.* JNCI: Journal of the National Cancer Institute, 2004. **96**(9): p. 644-644.
45. Rak, J., et al., *Oncogenes and angiogenesis: signaling three-dimensional tumor growth.* J Investig Dermatol Symp Proc, 2000. **5**(1): p. 24-33.
46. Dang, C.V., *MYC on the Path to Cancer.* Cell, 2012. **149**(1): p. 22-35.
47. *Uniprot/CEBPB.*
48. cancerindex/EP3000.
49. *Genecards/EP3000.*
50. Li, B.H., et al., *IL-4/Stat6 activities correlate with apoptosis and metastasis in colon cancer cells.* Biochem Biophys Res Commun, 2008. **369**(2): p. 554-60.
51. Neilsen, P.M., et al., *Mutant p53 drives invasion in breast tumors through up-regulation of miR-155.* Oncogene, 2013. **32**(24): p. 2992-3000.
52. Yin, X., J.W. Dewille, and T. Hai, *A potential dichotomous role of ATF3, an adaptive-response gene, in cancer development.* Oncogene, 2008. **27**(15): p. 2118-27.
53. Li, X., et al., *Activating transcription factor 3 promotes malignance of lung cancer cells in vitro.* Thoracic Cancer, 2017. **8**(3): p. 181-191.
54. Jiang, X., et al., *Potential Dual Role of Activating Transcription Factor 3 in Colorectal Cancer.* Anticancer Res, 2016. **36**(2): p. 509-16.
55. Kim, T.-H., et al., *Overexpression of Transcription Factor Sp2 Inhibits Epidermal Differentiation and Increases Susceptibility to Wound- and Carcinogen-Induced Tumorigenesis.* Cancer Research, 2010.

56. Abdelrahim, M. and S. Safe, *Cyclooxygenase-2 Inhibitors Decrease Vascular Endothelial Growth Factor Expression in Colon Cancer Cells by Enhanced Degradation of Sp1 and Sp4 Proteins.* Molecular Pharmacology, 2005. **68**(2): p. 317-329.

57. Chellappa, K., G.R. Robertson, and F.M. Sladek, *HNF4α: a new biomarker in colon cancer?* Biomarkers in medicine, 2012. **6**(3): p. 297-300.

58. Zhong, Q., et al., *The significance of Brf1 overexpression in human hepatocellular carcinoma.* Oncotarget, 2016. **7**(5): p. 6243-6254.

59. Bellido, F., et al., *Association Between Germline Mutations in BRF1, a subunit of the RNA Polymerase III Transcription Complex, and Hereditary Colorectal Cancer.* Gastroenterology, 2017.

60. Mlacki, M., et al., *Recent discoveries concerning the involvement of transcription factors from the Grainyhead-like family in cancer.* Experimental Biology and Medicine, 2015. **240**(11): p. 1396-1401.

61. Michel, S., et al., *Lack of HLA class II antigen expression in microsatellite unstable colorectal carcinomas is caused by mutations in HLA class II regulatory genes.* International journal of cancer. Journal international du cancer, 2010. **127**(4): p. 889-898.

62. Christensen, L.L., et al., *MiRNA-362-3p induces cell cycle arrest through targeting of E2F1, USF2 and PTPN1 and is associated with recurrence of colorectal cancer.* Int J Cancer, 2013. **133**(1): p. 67-78.

63. Vleugel, M.M., et al., *c-Jun activation is associated with proliferation and angiogenesis in invasive breast cancer.* Hum Pathol, 2006. **37**(6): p. 668-74.

64. Wang, H., M. Birkenbach, and J. Hart, *Expression of Jun family members in human colorectal adenocarcinoma.* Carcinogenesis, 2000. **21**(7): p. 1313-1317.

65. Sun, D., et al., *C/EBP-beta-activated microRNA-223 promotes tumour growth through targeting RASA1 in human colorectal cancer.* Br J Cancer, 2015. **112**(9): p. 1491-500.

66. Wan, M., et al., *Yin Yang 1 plays an essential role in breast cancer and negatively regulates p27.* Am J Pathol, 2012. **180**(5): p. 2120-33.

67. Yokoyama, N.N., et al., *A role for YY1 in repression of dominant negative LEF-1 expression in colon cancer.* Nucleic Acids Res, 2010. **38**(19): p. 6375-88.

68. Connett, J.M., et al., *Interferon regulatory factor 1 (IRF-1) and IRF-2 expression in breast cancer tissue microarrays.* J Interferon Cytokine Res, 2005. **25**(10): p. 587-94.

69. Yuan, L., et al., *IFN-gamma-mediated IRF1/miR-29b feedback loop suppresses colorectal cancer cell growth and metastasis by repressing IGF1.* Cancer Lett, 2015. **359**(1): p. 136-47.

70. Gordziel, C., et al., *Both STAT1 and STAT3 are favourable prognostic determinants in colorectal carcinoma.* British Journal of Cancer, 2013. **109**(1): p. 138-146.

71. Takai, N., et al., *The significance of Elf-1 expression in epithelial ovarian carcinoma.* Int J Mol Med, 2003. **12**(3): p. 349-54.

72. Jedlicka, P. and A. Gutierrez-Hartmann, *Ets transcription factors in intestinal morphogenesis, homeostasis and disease.* Histology and histopathology, 2008. **23**(11): p. 1417-1424.

73. Zhang, S., et al., *GABPA predicts prognosis and inhibits metastasis of hepatocellular carcinoma.* BMC Cancer, 2017. **17**: p. 380.

74. Petitjean, A., et al., *Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database.* Hum Mutat, 2007. **28**(6): p. 622-9.

75. Lee, H.-K., D.-S. Lee, and J.-C. Park, *Nuclear factor I-C regulates E-cadherin via control of KLF4 in breast cancer.* BMC Cancer, 2015. **15**: p. 113.

76. Liu, H.Y. and C.J. Zhang, *Identification of differentially expressed genes and their upstream regulators in colorectal cancer.* Cancer Gene Ther, 2017. **24**(6): p. 244-250.

77. *Genecard/CEBPE.*

78. Garneau, H., et al., *E2F4 expression is required for cell cycle progression of normal intestinal crypt cells and colorectal cancer cells.* J Cell Physiol, 2009. **221**(2): p. 350-8.

79. Grabsch, H., et al., *Expression of DNA Double-Strand Break Repair Proteins ATM and BRCA1 Predicts Survival in Colorectal Cancer.* Clinical Cancer Research, 2006. **12**(5): p. 1494-1500.

80. *Genecards/CEBPE.*