The identification of highly upregulated genes in claudin-low breast cancer through an integrative bioinformatics approach

Hatem Zayed

PII: S0010-4825(20)30174-8

DOI: https://doi.org/10.1016/j.compbiomed.2020.103806

Reference: CBM 103806

To appear in: *Computers in Biology and Medicine*

Received Date: 19 February 2020

Revised Date: 2 May 2020

Accepted Date: 2 May 2020

Please cite this article as: H. Zayed, The identification of highly upregulated genes in claudin-low breast cancer through an integrative bioinformatics approach, *Computers in Biology and Medicine* (2020), doi: https://doi.org/10.1016/j.compbiomed.2020.103806.

THE IDENTIFICATION OF HIGHLY UPREGULATED GENES IN CLAUDIN-LOW BREAST CANCER THROUGH AN INTEGRATIVE BIOINFORMATICS APPROACH

[1*]Hatem Zayed

[1]Department of Biomedical Sciences, College of Health Sciences, QU Health, Qatar University, Doha, Qatar

[*]**Correspondence to** Hatem Zayed, Department of Biomedical Sciences, College of Health and Sciences, QU Health, Qatar University, Doha, Qatar, hatem.zayed@qu.edu.qa

Abstract

Breast cancer (BC) is one of the leading causes of cancer-related death among women worldwide, and claudin-low breast cancer (CLBC) is a subtype of BC that remains poorly described. This study aimed to identify upregulated genes and significant pathways involved in CLBC. The SUM159 cell line is derived from human CLBC tissue; the GSE50697 dataset contains three replicates of SUM159 cells treated with pBabe puro miR-203 and three replicates of control SUM159 cells (pBabe puro). The data were normalized and upregulated, and downregulated genes were identified based on the logFC values. Gene Ontology (GO) and pathway analysis identified the most significant pathways and genes involved in CLBC pathogenesis. A total of 156 significant genes were identified (69 upregulated genes and 64 downregulated genes). From the pathway analysis, the senescence-associated secretory phenotype, which involves the *CXCL8*, *IL1A,* and *IL6* genes, was found to be mapped through more than one pathway (WikiPathways and Reactome). From the refined GO analysis, the IL-13 signaling pathway was identified; this pathway includes the *IL6, CXCL8, VEGF-C, NRG1,* and *EREG* genes, which were mapped as hub genes in several pathogenesis pathways. From the survival analysis, high levels of *IL6, CXCL8,* and *EREG* were related to high survival rates, and low levels of *VEGFC* and *NRG1* were related to high survival rates. The *IL6* and *CXCL8* genes were the most significant and the most highly represented in the GO and refined GO analyses. This study might provide a potential biomarker for the treatment of CLBC.

Keywords: Breast cancer, Microarray, Claudin-low breast cancer, Biomarkers, Functional enrichment analysis, Differential gene expression

Introduction

Cancer is a broad set of diseases wherein cells divide, grow, and invade other parts of the body abnormally and without control. Classification of cancer depends primarily on the cell type and origin of the tumor; cancer can be classified as carcinoma, sarcoma, or lymphoma based on whether the tumor is derived from epithelial cells, connective tissue, or lymph nodes, respectively. Cancers with the highest incidence are lung, breast, stomach, prostate, colorectal, and uterine cancer (Rahman and Zayed 2018; Sidenna et al. 2019; Siegel, Miller, and Jemal 2019; Thirumal Kumar et al. 2019; Younes and Zayed 2018). Among all the different types of cancers identified to date, breast cancer (BC) accounts for 14% of cancers in women and is the second most commonly occurring cancer worldwide (Ferlay et al. 2013; Thirumal Kumar and George Priya Doss 2016b, 2017). BC is thought to be a genetic disorder that is caused by mutations in different genes that control metabolic pathways and the cell cycle (De and Kuppusamy 2019; Sudhakar et al. 2016). Although the majority of patients present with symptoms of BC, which includes a lump in the breast, a distortion in breast shape, dimpling of the skin of the breast, fluid impending from the nipple, a newly inverted nipple or a red or scaly patch of skin of the breast, approximately 40% cases are diagnosed by the NHS breast screening program when they are asymptomatic (Sibbering and Courtney 2016). This UK-based screening method, according to the evidence provided by Threlfall et al reduces BC mortality (Threlfall, Collins, and Woodman 2003).

Family history and genetics are significant risk factors for the development of BC (Gazalla Ayub et al. 2014; Neamatzadeh, Shiryazdi, and Kalantar 2015). Approximately 3% to 10% of BCs and approximately 30% of all early-onset BCs are caused by hereditary factors (Calderón-Garcidueñas et al. n.d.). Germline mutations in BC genes (BRCA1 and BRCA2) are considered

to be the primary gene changes associated with breast and ovarian cancers (Kwong et al. 2011; Mehrgou and Akouchekian 2016). BRCA1 is positioned at chromosome 17q, and BRCA2 is located at chromosome 13q. These two genes are tumor suppressor genes, and mutations in these genes lead to approximately 30% of all breast and ovarian cancers (Martínez-Ferrandis et al. 2003; Pohlreich et al. 2005). A precise human gene locus comprising 15 kallikrein genes on chromosome 19q13.4 is recognized as being the leading continuous gene group of serine proteases in the human genome. Kallikreins are seen in epithelial and endocrine tissues and, therefore, are probable serum biomarkers in ovarian, breast, and prostate cancers (Coticchia et al. 2009).

In addition to these common cancer types, there is a cancer type known as claudin-low breast cancer (CLBC), which is a molecular subtype of BC that is associated with poor prognosis and has no specific treatment so far. Thus, treatments and diagnostic biomarkers for this type of BC are needed. Microarray analysis is a recent technique that is widely used in the analysis of patient samples and the identification of disease biomarkers. This technique is most commonly used in the treatment of cancer and bacterial infections (Kaliyappan et al. 2012). A recent study by Xie et al. (2019) evaluated three different BC cell lines from different GEO datasets to study the functions of the intersectin-1 (*ITSN1*) gene in BC. The dataset included the GEO dataset with ID GSE50697 (Xie et al. 2019). This dataset contains samples from CLBC tissue treated with pBabe puro miR-203. The microarray data of the CLBC cell line expressing microRNA-203 were used for this study (Taube et al. 2013). Epithelial-Mesenchymal Transition (EMT) facilitates the migration and invasion of cancer cells and increases the ability of cancer cells to grow in a secondary site by promoting their survival in the blood circulation (Kalluri and Weinberg 2009; Mani et al. 2008; O'Regan et al. 2017; Polyak and Weinberg 2009; Thiery 2002). MiR-200

targets the transcription factors that induce EMT: Zeb1 and Zeb2 (Bracken et al. 2008; Gregory et al. 2008; Park et al. 2008). The activity of Zeb1 and Zeb2 is suppressed by histone modification and DNA methylation, which then promotes EMT at the initial stage of carcinogenesis (Neves et al. 2010; Tellez et al. 2011; Vrba et al. 2010). The expression of miR-203 in mesenchymal cells reduces the migratory and invasive capacities of cells *in vitro* and results in tumor initiation and metastasis *in vivo* (Taube et al. 2013). Expression of miR-203 reduces B-catenin levels by enhancing the expression of DKK1, which plays a significant role as an inhibitor of Wnt signaling. It affects the stemness of adjacent cells (Li et al. 2010). Increased levels of miR-203 expression may result in the inhibition of metastasis. This study aimed to use a computational analysis pipeline to identify the significant biological pathways and genes involved in CLBC, thereby identifying biomarkers for the treatment of CLBC.

Materials and Methods

Dataset

The microarray dataset was retrieved from the GEO database with GEO accession number GSE50697 (Clough and Barrett 2016; Taube et al. 2013). This dataset contains six samples in CEL format with accession numbers GSM1226581, GSM1226582, GSM1226583, GSM1226584, GSM1226585, and GSM1226586 corresponding to SUM159 control reps 1, 2, 3 and SUM159 miR-203 reps 1, 2, 3, respectively.

Data normalization and quality control

Chipster, a user-friendly software used for analyzing high-throughput data such as NGS and microarrays, was used in our study to analyze the CLBC dataset (Kallio et al. 2011). The normalization of the CEL files estimates the expression and call values for the genes. The Robust Multichip Averaging (RMA) normalization method with original Affymetrix annotations was used for normalization. The Quality Control (QC) stat, RNA degradation, and spike-in performance plots were obtained as the output. Further, quality control was assessed using the Affymetrix primary method.

Preprocessing of the normalized data

The normalized files were classified into two groups using the Phenodata editor of the Chipster package (https://chipster.csc.fi/). The control samples were grouped under number 1, the samples treated with pBabe puro miR-203 were grouped under number 2, and a standard deviation value of 3 SDs (99.7%) was set as the base value to filter the significant genes.

Statistical analysis and annotations

Since the retrieved sample dataset contains two groups, a two-group statistical test was performed with the default empirical Bayes test and p-value adjusted with the Benjamini-Hochberg correction (BH) method (Benjamini and Hochberg 1995). The p-value threshold for significance was set to 0.05. Further, the annotation was performed using the Affymetrix gene list.

Clustering and pathway enrichment analysis of DEGs

Gene clustering was performed using the Pearson distance with the average tree method (VII. Note on regression and inheritance in the case of two parents 1895). A total of 1000 iterations

were generated to identify the most appropriate clusters. The pathways were identified using the gene set test against the KEGG pathway with a minimum pathway size of 5 and a p-value of 0.05 (Kanehisa 2000). The multiple testing correlation was analyzed using the BH method. The hypergeometric test for GO was performed to classify genes based on various ontologies, such as biological process, molecular function, and cellular component, with a minimum of five populations and overrepresentation against the AmiGO 2 database. Finally, the hypergeometric test for ConsensusPathDB (http://cpdb.molgen.mpg.de/) was analyzed with a p-value threshold of 0.05 with a gene symbol as an identifier (Kamburov et al. 2011).

Fold change calculation & PPI network construction

The fold change between the two groups was calculated as the geometric mean with a scale of log2. The Cytoscape standalone package was used to build the interactions between the identified significant genes (Shannon et al. 2003). The 'stringApp' plugin of Cytoscape was used to retrieve the interacting genes with the identified significant genes with a confidence cutoff of 0.40. This plugin extracts the pool of interacting genes based on the data from the online STRING database.

Refined GeneGo Analysis

The significant DEGs were further examined in MetaCore, Cortellis solution software. GeneGo empowers the quick and easy analysis of protein networks, metabolic pathways, and maps for the list of genes/proteins (MetaCore Login|Clarivate Analytics). The pathway maps tool was used to identify the enriched pathways involving DEGs in terms of the hypergeometric distribution, and the p-values were analyzed by using the default database. The graphical depictions of the interactions were generated based on a significant p-value < 0.05.

Survival and Expression DIY analysis

The survival and expression DIY analysis were performed using the online GEPIA2 server (http://gepia2.cancer-pku.cn/) with the BC dataset selected for the analysis. For survival analysis, the samples were divided into high and low expression groups according to the 50% cutoff value. The confidence interval was maintained at 95%. For the box plot analysis, the |Log2FC| cutoff was set to 1, and the p-value cutoff was set to 0.01.

Results

Data Normalization and Quality Control

Data normalization and quality control were performed using the robust multichip averaging (RMA) method to obtain the QC stat, RNA degradation, and spike-in performance plots. The QC data aids in the understanding of the number of probesets in the present flag and backgrounds in the chip. The percent in probesets in the microarray001.cel, microarray002.cel, microarray003.cel, microarray004.cel, microarray005.cel, and microarray006.cel chips were 41.54%, 42.11%, 40.82, 39.4%, 39.57%, and 41.36%, respectively. Additionally, the average background on the chip was found to be 46.99, 49.29, 49.97, 50.88, 53.95, and 53.67, respectively. The GAPDH3/GAPDH5 scaling factor/ratios were found to be within 1.25-fold, as observed by the blue color (Supplementary Figure 1). Further, RNA degradation plots and spike-in performance plots were generated and showed that the slopes and profiles were similar and stable across the plot and provided confidence that the samples were suitable for further analysis (Supplementary Figure 2A-B).

Preprocessing the normalized data and fold change (FC) calculation

The samples were mapped into control and disease using the phenodata editor package, and preprocessing was initiated with a standard deviation of 0.997, which is 99.7% significance or 3 SDs. The filtration process demonstrated 165 genes that satisfied this condition. The symbols of 17 genes were not identified during the process; hence, they are noted as 'NA' in the description (Supplementary Table 1). As there were two groups (control and miR-203), two groups' tests were performed with the BH, p-value adjustment method and p-value threshold of 0.05 with normalized gene data and phenodata as input. From the statistical analysis, 156 genes were found to be significant, out of which 17 lacked gene annotation (Supplementary Table 2). The gene expression heat map is shown in Supplementary Figure 3.

Further, hierarchical clustering was analyzed with these genes with the Pearson distance average tree method with 1000 replicates. The obtained hierarchical clustering is shown in Supplementary Figure 4. The fold change calculation was calculated between the two groups with the geometric mean and scale of log2 for the 156 DEGs. The genes with FC values above 1 were identified to be upregulated, and the genes with FC values below -1 were identified to be downregulated. From the FC values, 69 genes were found to be upregulated, 64 genes were found to be downregulated, and 17 unidentified genes were excluded from the study (Supplementary Table 3). The volcano plot was generated with the corresponding data: the upregulated genes are mapped in red, and the downregulated genes are mapped in green. In contrast, the genes that did not change are mapped in black (Supplementary Figure 5).

PPI network construction and pathway analysis

The interacting network of upregulated and downregulated genes is shown using Cytoscape software, and the data retrieved from the STRING database are provided in Supplementary Table

4; 69 nodes and 126 edges were obtained as the result of the interaction (Figure 1). The network of interacting upregulated genes is shown in Figure 2. Further, the pathways and gene list were analyzed among the groups with a p-value threshold of 0.05 and BH multiple testing correction methods. The analysis of upregulated genes revealed involvement in 211 pathways, of which toxoplasmosis, lysine degradation, glycerolipid metabolism, pathways in cancer, and mTOR signaling ranked highest, with networks of 5145, 310, 561, 5200, and 4150 genes, respectively (Supplementary Table 5). A positive and negative correlation between the dysregulated genes and the top 5 pathways are shown in Supplementary Figure 6A-4E. Further, the gene annotation was performed using the Affymetrix gene list parameter, including Probe, Symbol, Description, Chromosome, Chromosome, Location, GenBank, Gene, Cytoband, UniGene, PubMed, Gene Ontology, and Pathway for the 156 significant genes. The detailed data with cross-references against NCBI and KEGG database hyperlinks are provided in Supplementary File 1.

Pathway enrichment analysis of DEGs

Hypergeometric Gene Ontology (GO) was performed to identify the various processes, such as biological process, molecular function, and cellular component. They yielded 165, 6, and 7 GO terms for biological process, molecular function, and cellular component, respectively. (Supplementary Table 6). The hypergeometric test for ConsensusPathDB was performed against ConsensusPathDB (http://cpdb.molgen.mpg.de/) with humans as a reference to map the genes to the respective pathways. This analysis identified 193 different pathways, and the respective genes were mapped to the identified pathways. Senescence-associated secretory phenotype (SASP) was associated with *CXCL8*, *IL1A,* and *IL6* genes and was mapped through WikiPathways as well as through Reactome predictions (Supplementary Table 7).

Refined GeneGo Analysis

Refined GeneGo analysis was performed using MetaCore software to identify the top 10 pathway maps, GO processes, process networks, and diseases (by biomarkers). From the pathway analysis, immune response IL-13 signaling via JAK-STAT, G protein-coupled receptor signaling in lung cancer, and cell adhesion ECM remodeling were found to be ranked in the top 3 in the analysis. Second-messenger-mediated signaling, anatomical structure development, and multicellular development were found to be the top 3 ranked GO processes. Inflammation-related IL-13 signaling, cell adhesion cell-matrix interaction, and inflammation histamine interactions were found to be the top 3 in terms of process network. Based on disease biomarkers, carcinoma, adenocarcinoma, and colonic diseases were ranked in the top 3. The top 10 list of each analysis is provided in Figure 3A-D. The detailed list of genes involved in pathway maps, GO processes, process networks, and diseases (by biomarkers) are given in Tables 1-3 and Supplementary Table 8. The top 3 pathways (immune response IL-13 signaling via JAK-STAT, G protein-coupled receptor signaling in lung cancer, and cell adhesion ECM remodeling) with a top-scored map (map with the lowest p-value) based on the enrichment distribution sorted by 'statistically significant maps' are shown in Figure 4A-C. Finally, the analyze networks algorithm was employed with the default settings to prioritize the networks based on the number of fragments of canonical pathways in the network. From the analysis, three significant networks and their respective processes were identified. The major network included *IL8*, which is in the regulation of cell proliferation (90.0%); *IL6*, which is involved in the positive regulation of intracellular signal transduction (72.0%); *VEGFC*, which is involved in the positive regulation of protein metabolic processes (80.0%); *NRG1* (neuregulin 1), which is

involved in the positive regulation of multicellular organismal processes (82.0%); and *EREG* (epiregulin), which is involved in response to hormone (74.0%) functions (Table 4).

Survival and Expression DIY analysis

From the overall survival analysis, it was found that the hazard ratios of the identified significant genes *IL6, CXCL8, VEGF-C, NRG1,* and *EREG* were 0.93, 0.92, 1.1, 0.88, and 0.75, respectively. Further, the box plot showed considerable changes in gene expression. The *IL6* and *NRG1* genes showed a higher significance than the other three genes (Figure 5A-J). A separate analysis was performed on the four hub genes. The interrelations between the hub genes identified from the pathways are shown in Figure 6.

Discussion

This work investigated the gene expression of the SUM159 CLBC cell line expressing microRNA-203. The study contained six samples: three control samples and three samples treated with miR-203. The samples were normalized using RMA, and the quality of the samples was analyzed with the Affymetrix basics tool embedded within Chipster. QC analysis plays a crucial role in any scientific work that generates huge data. This analysis aids in the understanding of the quality of a microarray experiment and, in particular, helps to identify outlier samples, thus revealing highly sensitive data for analysis (Freue et al. 2007; Kauffmann and Huber 2010; Shieh and Hung 2009). Our QC analysis revealed that the samples were highly sensitive and appropriate for further analysis (Supplementary Figure 2A and 2B).

The genes were filtered using the criteria of 99.7% significance, and 165 genes were found to satisfy the significance criteria. The symbols of 17 genes were not identified during the process (Supplementary Table 2). From the analysis of genes with a logFC value of 2, 69 genes and 64 genes were found to be upregulated and downregulated, respectively (Supplementary Table 3). The entire list of 165 genes was subjected to interaction analysis using the STRING database, and the interactions were visualized using Cytoscape. In total, 69 nodes and 126 edges were identified to be involved in the interactions (Supplementary Table 4 & Figure 1). To identify the interactions of genes that are involved in the upregulation, a separate plot was generated (Figure 2). These upregulated genes were found to play a significant role in 211 different pathways. Of those, toxoplasmosis, lysine degradation, glycerolipid metabolism, pathways in cancer, and mTOR signaling were found to rank as the most significant (Supplementary Table 5 and Supplementary Figure 6A-E). Interestingly, all these pathways were found to play a significant role in BC in previous literature (Chen et al. 2018; Hynes and Boulay 2006; Kalantari et al. 2017; Luo et al. 2017). Further gene annotation was processed using the Affymetrix gene list parameters. The entire annotated list with hyperlinks is shown in Supplementary File 1. The hypergeometric test for ConsensusPathDB was performed against the ConsensusPathDB database with humans as a reference to map the genes to respective pathways. From the analysis, 193 different pathways and the corresponding genes were mapped accordingly. Biocarta, EHMN, INOH, KEGG, PharmGKM, PID, Reactome, Signalink, SMPDB, and WikiPathways were employed to identify the pathways to correlate the genes. The senescence-associated secretory phenotype (SASP) pathway, which involves the *CXCL8*, *IL1A*, and *IL6* genes, was mapped through WikiPathways as well as through Reactome (Supplementary Table 7). To increase the confidence of the above findings, a refined GeneGo analysis was performed to

identify the top pathway maps, GO processes, process networks, and diseases (by biomarkers). Immune response IL-13 signaling via JAK-STAT, G protein-coupled receptor signaling in lung cancer, and cell adhesion ECM remodeling were found to be the top pathways, and the interactions of these pathways are shown in Figures 3 and 4. Similar to the pathway maps, the IL-13 signaling pathway was also found to be top-ranked in the process networks, signifying that the pathway IL-13 signaling pathway could be the more significant pathway in the disease (Figure 3A-D). This association of interleukin with BC was reported in several earlier research works, which supported our findings (Cao et al. 2016; DeNardo et al. 2009; Fasoulakis et al. 2018).

The analyze networks algorithm was employed to identify the major networks and genes involved in the pathway. From the analysis, three major networks and their respective processes were identified. This result was again found to agree with our earlier findings, where the major genes among the main networks were found to be *IL-8* (*CXCL8*), which is involved in the regulation of cell proliferation (90.0%); *IL-6*, which is involved in the positive regulation of intracellular signal transduction (72.0%); *VEGFC*, which is involved in the positive regulation of protein metabolic processes (80.0%); neuregulin 1, which involved in the positive regulation of multicellular organismal processes (82.0%); and epiregulin, which is involved in response to hormone (74.0%) functions (Table 4). *IL-8* and *IL-6* were previously identified by two different pathway identifiers (WikiPathways and Reactome) to be involved in the senescence-associated secretory phenotype (SASP). SASP biomarkers are well-studied tumor suppressors in cancers, including BC (Campisi 2001; Coppé et al. 2008; Perrott et al. 2017). In particular, Coppé et al., 2008 stressed the involvement of IL-6 and IL-8 in the involvement of tumor suppressor actions, which supports our findings. There are also suitable previous studies that support the role of our

identified genes (*VEGF-C*, *NRG1*, and *EREG*) in the involvement of BC (Farooqui et al. 2015; Jeong et al. 2014; Skobe et al. 2001). The survival analysis revealed that high levels of *IL6*, *CXCL8*, and *EREG* were associated with high survival rates; in contrast, low levels of *VEGFC* and *NRG1* were associated with high survival rates. The *IL6* and *NRG1* genes were expressed at significantly higher levels than the other three genes (Figure 5). Finally, correlation analysis between the identified hub genes (*IL6*, *CXCL8*, *VEGF-C*, *NRG1*, and *EREG*) was performed using the STRING bioinformatics tool. The association between these genes was mainly derived from text mining and coexpression analysis involving the *EREG*, *CXCL8*, *IL6*, and *VEGF-C* genes. The identified hub genes were shown to be associated with various signaling pathways by interacting with each other (Figure 6). The results depict that *CXCL8* was found to be involved in the regulation of signaling receptor activity biological process (GO), while *VEGF-C* was involved in the regulation of signaling receptor activity and positive regulation of peptidyl-tyrosine phosphorylation biological processes (GO). In addition, from the UniProt Keywords search, *VEGF-C* was also found to be involved in mitogen and angiogenesis. The gene *IL-6* was shown to be involved in the positive regulation of peptidyl-tyrosine phosphorylation and the regulation of signaling receptor activity biological processes (GO). In addition, it was found to have an action in MAPK1/MAPK3 signaling by the Reactome pathway analysis. The *NRG1* gene was found to play a significant role in the downregulation of ERBB2 signaling, MAPK1/MAPK3 signaling, and in PI3K events in ERBB2 signaling in the Reactome pathway analysis. In terms of the biological process (GO), they were found to be involved in the positive regulation of peptidyl-tyrosine phosphorylation and the regulation of signaling receptor activity. Finally, from the search against SMART protein domains, they were also found to play a role in the epidermal growth factor-like domain. *EREG* was also found to possess all the characteristics

of the *NRG1* gene, and in addition, they were also found to be involved in mitogen and angiogenesis. These findings were well supported in several previous studies on breast cancer-causing genes and pathways (Arora et al. 2016; Coticchia et al. 2009; Sudhakar et al. 2016; Thirumal Kumar and George Priya Doss 2016a). Thus, this study suggests that the *IL6, CXCL8, VEGF-C, NRG1*, and *EREG* genes might be suitable biomarkers in the treatment of CLBC.

Conclusion

A comprehensive bioinformatics approach was performed to identify the pathways and genes that were significantly enriched between the SUM159 CLBC cell line expressing microRNA-203 and control cells. The microarray data were obtained from the GEO database with ID GSE50697. A total of 165 genes were found to be differentially expressed. Based on the logFC values, 69 genes were found to be upregulated, and 64 genes were classified as downregulated. The upregulated genes were prioritized for GO and refined GO analyses using the built-in packages of Chipster and MetaCore, respectively. Pathway analysis identified 193 pathways, of which the inflammatory IL-13 signaling pathway was found to be the most significantly enriched. Five upregulated genes (*IL6, CXCL8, VEGF-C, NRG1*, and *EREG*) were mapped as hubs, indicating that they might play crucial roles in CLBC. High levels of *IL6, CXCL8,* and *EREG* and low levels of *VEGFC* and *NRG1* were found to be related to high survival rates through survival analysis. Finally, through box plot analysis, the expression levels of *IL6* and *NRG1* were found to be significantly higher than those of the other genes. This study suggests that the five genes *IL6, CXCL8, VEGF-C, NRG1*, and *EREG* might be potential biomarkers for CLBC.

Conflict of Interest

None

Figure Legends

Figure 1. Gene interaction network of significant genes obtained via the STRING database.

Figure 2. Gene interaction network of only the significantly upregulated genes obtained via the STRING database.

Figure 3A-D. (A) Top 10 pathway profiles; (B) top 10 GO processes; (C) top 10 process networks; (D) top 10 diseases according to biomarkers.

Figure 4A. Pathway of the immune response-related IL-13 signaling via JAK-STAT.

Figure 4B. Pathway of G protein-coupled receptor signaling in lung cancer.

Figure 4C. Pathway of cell adhesion ECM remodeling with a top-scored map.

Figure 5. Kaplan–Meier overall survival and box plot analysis of the hub genes expressed in the SUM159 breast cancer cell line: (A and B) IL6 gene, (C and D) CXCL8 gene, (E and F) VEGFC gene, (G and H) NRG1 gene, and (I and J) EREG gene.

Figure 6. Network visualization showing the correlation between the identified hub genes (IL6, CXCL8, VEGF-C, NRG1, and EREG). The network was visualized using the online STRING server. Color codes: cyan – positive

regulation of peptidyl-tyrosine phosphorylation, brown – regulation of signaling receptor activity, dark green – downregulation of ERBB2 signaling, yellow – MAPK1/MAPK3 signaling, red – mitogen, violet – angiogenesis, light green – epidermal growth factor-like domain.

Table Legends

Table 1. Top 10 maps identified from the refined GO study using MetaCore and the list of genes mapped from the network objects from active data.

Table 2. Top 10 processes identified from the refined GO using MetaCore study and the list of genes mapped from the network objects from active data.

Table 3. Top 10 networks identified from the refined GO using MetaCore study and the list of genes mapped from the network objects from active data.

Table 4. List of top 3 networks and their processes identified using the refined GO process.

References

Arora, Deepika et al. 2016. "Evaluation and Physiological Correlation of Plasma Proteomic Fingerprints for Deltamethrin-Induced Hepatotoxicity in Wistar Rats." *Life Sciences* 160: 72–83.

Benjamini, Y, and Y Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society B* 57(1): 289–300. http://www.stat.purdue.edu/~doerge/BIOINFORM.D/FALL06/Benjamini and Y FDR.pdf (December 23, 2019).

Bracken, Cameron P. et al. 2008. "A Double-Negative Feedback Loop between ZEB1-SIP1 and the MicroRNA-200 Family Regulates Epithelial-Mesenchymal Transition." *Cancer Research* 68(19): 7846–54.

Calderón-Garcidueñas, Ana Laura, Pablo Ruiz-Flores, Ricardo M Cerda-Flores, and Hugo A Barrera-Saldaña. "Clinical Follow up of Mexican Women with Early Onset of Breast Cancer and Mutations in the BRCA1 and BRCA2 Genes." *Salud publica de Mexico* 47(2): 110–15. http://www.ncbi.nlm.nih.gov/pubmed/15889636 (December 23, 2019).

Campisi, Judith. 2001. "Cellular Senescence as a Tumor-Suppressor Mechanism." *Trends in Cell Biology* 11(11).

Cao, Hui et al. 2016. "IL-13/STAT6 Signaling Plays a Critical Role in the Epithelialmesenchymal Transition of Colorectal Cancer Cells." *Oncotarget* 7(38): 61183–98.

Chen, Li et al. 2018. "Pan-Cancer Analysis Reveals the Functional Importance of Protein Lysine Modification in Cancer Development." *Frontiers in Genetics* 9(JUL).

Clough, Emily, and Tanya Barrett. 2016. "The Gene Expression Omnibus Database." In *Methods in Molecular Biology*, Humana Press Inc., 93–110.

Coppé, Jean Philippe et al. 2008. "Senescence-Associated Secretory Phenotypes Reveal Cell-

Nonautonomous Functions of Oncogenic RAS and the P53 Tumor Suppressor." *PLoS biology* 6(12).

Coticchia, Christine M. et al. 2009. "Calmodulin Modulates Akt Activity in Human Breast Cancer Cell Lines." *Breast Cancer Research and Treatment* 115(3): 545–60.

De, Anindita, and Gowthamarajan Kuppusamy. 2019. "Metformin in Breast Cancer: Preclinical and Clinical Evidence." *Current Problems in Cancer*.

DeNardo, David G. et al. 2009. "CD4+ T Cells Regulate Pulmonary Metastasis of Mammary Carcinomas by Enhancing Protumor Properties of Macrophages." *Cancer Cell* 16(2): 91–102.

Farooqui, Mariya et al. 2015. "Epiregulin Contributes to Breast Tumorigenesis through Regulating Matrix Metalloproteinase 1 and Promoting Cell Survival." *Molecular Cancer* 14(1).

Fasoulakis, Zacharias, George Kolios, Valentinos Papamanolis, and Emmanuel N Kontomanolis. 2018. "Interleukins Associated with Breast Cancer." *Cureus*.

Ferlay, J. et al. 2013. "Cancer Incidence and Mortality Patterns in Europe: Estimates for 40 Countries in 2012." *European Journal of Cancer* 49(6): 1374–1403. http://dx.doi.org/10.1016/j.ejca.2012.12.027.

Freue, Gabriela V.Cohen et al. 2007. "MDQC: A New Quality Assessment Method for Microarrays Based on Quality Control Reports." *Bioinformatics* 23(23): 3162–69.

Gazalla Ayub, Shiekh et al. 2014. "Mutational Analysis of the BRCA2 Gene in Breast Carcinoma Patients of Kashmiri Descent." *Molecular Medicine Reports* 9(2): 749–53.

Gregory, Philip A. et al. 2008. "The MiR-200 Family and MiR-205 Regulate Epithelial to Mesenchymal Transition by Targeting ZEB1 and SIP1." *Nature Cell Biology* 10(5): 593–601.

Hynes, Nancy E., and Anne Boulay. 2006. "The MTOR Pathway in Breast Cancer." *Journal of Mammary Gland Biology and Neoplasia* 11(1): 53–61.

Jeong, Hoiseon et al. 2014. "Neuregulin-1 Induces Cancer Stem Cell Characteristics in Breast Cancer Cell Lines." *Oncology Reports* 32(3): 1218–24.

Kalantari, Narges et al. 2017. "Detection of Toxoplasma Gondii DNA in Malignant Breast Tissues in Breast Cancer Patients." *International Journal of Molecular and Cellular Medicine* 6(3): 190–96.

Kaliyappan, Karunakaran, Murugesan Palanisamy, Rajeshwar Govindarajan, and Jeyapradha Duraiyan. 2012. "Microarray and Its Applications." *Journal of Pharmacy and Bioallied Sciences* 4(6): 310.

Kallio, M Aleksi et al. 2011. "Chipster: User-Friendly Analysis Software for Microarray and Other High-Throughput Data." *BMC Genomics* 12(1): 507. http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-12-507 (December 23, 2019).

Kalluri, Raghu, and Robert A. Weinberg. 2009. "The Basics of Epithelial-Mesenchymal Transition." *Journal of Clinical Investigation* 119(6): 1420–28.

Kamburov, Atanas et al. 2011. "ConsensusPathDB: Toward a More Complete Picture of Cell Biology." *Nucleic Acids Research* 39(SUPPL. 1).

Kanehisa, M. 2000. "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 28(1): 27–30.

Kauffmann, Audrey, and Wolfgang Huber. 2010. "Microarray Data Quality Control Improves the Detection of Differentially Expressed Genes." *Genomics* 95(3): 138–42.

Kwong, Ava et al. 2011. "A Novel de Novo BRCA1 Mutation in a Chinese Woman with Early Onset Breast Cancer." *Familial Cancer* 10(2): 233–37.

Li, Yonghe et al. 2010. "Dkk1 Stabilizes Wnt Co-Receptor LRP6: Implication for Wnt Ligand-Induced LRP6 Down-Regulation" ed. Cheng-Xin Gong. *PLoS ONE* 5(6): e11014. https://dx.plos.org/10.1371/journal.pone.0011014 (December 23, 2019).

Luo, Xiangjian et al. 2017. "Emerging Roles of Lipid Metabolism in Cancer Metastasis." *Molecular Cancer* 16(1): 76. http://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-017-0646-3 (December 19, 2019).

Mani, Sendurai A. et al. 2008. "The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells." *Cell* 133(4): 704–15.

Martínez-Ferrandis, J.I. et al. 2003. "Mutational Analysis of BRCA1 and BRCA2 in Mediterranean Spanish Women with Early-Onset Breast Cancer: Identification of Three Novel Pathogenic Mutations ." *Human Mutation* 22(5): 417–18.

Mehrgou, Amir, and Mansoureh Akouchekian. 2016. "The Importance of BRCA1 and BRCA2 Genes Mutations in Breast Cancer Development." *Medical Journal of the Islamic Republic of Iran* 30(1).

Neamatzadeh, Hossein, Seyed Mostafa Shiryazdi, and Seyed Mahdi Kalantar. 2015. "BRCA1

and BRCA2 Mutations in Iranian Breast Cancer Patients: A Systematic Review." *Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences* 20(3): 284–93. http://www.ncbi.nlm.nih.gov/pubmed/26109977 (December 23, 2019).

Neves, Rui et al. 2010. "Role of DNA Methylation in MiR-200c/141 Cluster Silencing in Invasive Breast Cancer Cells." *BMC Research Notes* 3(1): 219. https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-3-219 (December 23, 2019).

O'Regan, Grace, Ruth-Mary deSouza, Roberta Balestrino, and Anthony H. Schapira. 2017. "Glucocerebrosidase Mutations in Parkinson Disease." *Journal of Parkinson's Disease* 7(3): 411–22. http://www.ncbi.nlm.nih.gov/pubmed/28598856 (July 3, 2018).

Park, Sun Mi, Arti B. Gaur, Ernst Lengyel, and Marcus E. Peter. 2008. "The MiR-200 Family Determines the Epithelial Phenotype of Cancer Cells by Targeting the E-Cadherin Repressors ZEB1 and ZEB2." *Genes and Development* 22(7): 894–907.

Perrott, Kevin M., Christopher D. Wiley, Pierre-Yves Desprez, and Judith Campisi. 2017. "Apigenin Suppresses the Senescence-Associated Secretory Phenotype and Paracrine Effects on Breast Cancer Cells." *GeroScience* 39(2): 161–73. http://link.springer.com/10.1007/s11357-017-9970-1 (December 19, 2019).

Pohlreich, Petr et al. 2005. "High Proportion of Recurrent Germline Mutations in the BRCA1 Gene in Breast and Ovarian Cancer Patients from the Prague Area." *Breast Cancer Research* 7(5).

Polyak, Kornelia, and Robert A. Weinberg. 2009. "Transitions between Epithelial and

Mesenchymal States: Acquisition of Malignant and Stem Cell Traits." *Nature Reviews Cancer* 9(4): 265–73.

Rahman, Sumaya, and Hatem Zayed. 2018. "Breast Cancer in the GCC Countries: A Focus on BRCA1/2 and Non-BRCA1/2 Genes." *Gene* 668: 73–76.

Shannon, Paul et al. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." *Genome Research* 13(11): 2498–2504.

Shieh, Albert D., and Yeung Sam Hung. 2009. "Detecting Outlier Samples in Microarray Data." *Statistical Applications in Genetics and Molecular Biology* 8(1).

Sibbering, Mark, and Carol Ann Courtney. 2016. "Management of Breast Cancer: Basic Principles." *Surgery (United Kingdom)* 34(1): 25–31.

Sidenna, Mariem et al. 2019. "Association of Genetic Variants with Colorectal Cancer in the Extended MENA Region: A Systematic Review." *Current Molecular Medicine* 20(4): 286–98.

Siegel, Rebecca L, Kimberly D Miller, and Ahmedin Jemal. 2019. “Cancer Statistics, 2019: {Cancer} {Statistics}, 2019.” *CA: A Cancer Journal for Clinicians*.

Skobe, Mihaela et al. 2001. "Induction of Tumor Lymphangiogenesis by VEGF-C Promotes Breast Cancer Metastasis." *Nature Medicine* 7(2): 192–98.

Sudhakar, N et al. 2016. "Deciphering the Impact of Somatic Mutations in Exon 20 and Exon 9 of PIK3CA Gene in Breast Tumors among Indian Women through Molecular Dynamics Approach." *Journal of biomolecular structure & dynamics* 34(1): 29–41. http://www.ncbi.nlm.nih.gov/pubmed/25679319 (September 12, 2016).

Taube, Joseph H. et al. 2013. "Epigenetic Silencing of MicroRNA-203 Is Required for EMT and Cancer Stem Cell Properties." *Scientific Reports* 3.

Tellez, Carmen S. et al. 2011. "EMT and Stem Cell-like Properties Associated with MiR-205 and MiR-200 Epigenetic Silencing Are Early Manifestations during Carcinogen-Induced Transformation of Human Lung Epithelial Cells." *Cancer Research* 71(8): 3087–97.

Thiery, Jean Paul. 2002. "Epithelial–Mesenchymal Transitions in Tumour Progression." *Nature Reviews Cancer* 2(6): 442–54.

Thirumal Kumar, D. et al. 2019. "A Computational Approach for Investigating the Mutational Landscape of RAC-Alpha Serine/Threonine-Protein Kinase (AKT1) and Screening Inhibitors against the Oncogenic E17K Mutation Causing Breast Cancer." *Computers in Biology and Medicine* 115.

Thirumal Kumar, D., and C. George Priya Doss. 2016a. "Investigating the Inhibitory Effect of Wortmannin in the Hotspot Mutation at Codon 1047 of PIK3CA Kinase Domain: A Molecular Docking and Molecular Dynamics Approach." *Advances in Protein Chemistry and Structural Biology* 102: 267–97.

———. 2016b. "Investigating the Inhibitory Effect of Wortmannin in the Hotspot Mutation at Codon 1047 of PIK3CA Kinase Domain." In *Advances in Protein Chemistry and Structural Biology*, , 267–97. http://www.ncbi.nlm.nih.gov/pubmed/26827608 (February 19, 2018).

Thirumal Kumar, D, and C George Priya Doss. 2017. "Role of E542 and E545 Missense Mutations of PIK3CA in Breast Cancer: A Comparative Computational Approach." *Journal of Biomolecular Structure and Dynamics* 35(12): 2745–57.

http://www.ncbi.nlm.nih.gov/pubmed/27581627 (October 26, 2016).

Threlfall, A. G., S. Collins, and C. B.J. Woodman. 2003. "Impact of NHS Breast Screening on Advanced Disease and Mortality from Breast Cancer in the North West of England." *British Journal of Cancer* 89(1): 77–80.

"VII. Note on Regression and Inheritance in the Case of Two Parents." 1895. *Proceedings of the Royal Society of London* 58(347–352): 240–42.

Vrba, Lukas et al. 2010. "Role for DNA Methylation in the Regulation of MiR-200c and MiR-141 Expression in Normal and Cancer Cells" ed. Catherine M. Suter. *PLoS ONE* 5(1): e8697. http://dx.plos.org/10.1371/journal.pone.0008697 (December 23, 2019).

Xie, Chen et al. 2019. "Intersectin 1 (ITSN1) Identified by Comprehensive Bioinformatic Analysis and Experimental Validation as a Key Candidate Biological Target in Breast Cancer." *OncoTargets and Therapy* 12: 7079–93.

Younes, Nadin, and Hatem Zayed. 2018. "Genetic Epidemiology of Ovarian Cancer in the 22 Arab Countries: A Systematic Review." *Gene*. https://linkinghub.elsevier.com/retrieve/pii/S0378111918310837 (October 27, 2018).

1

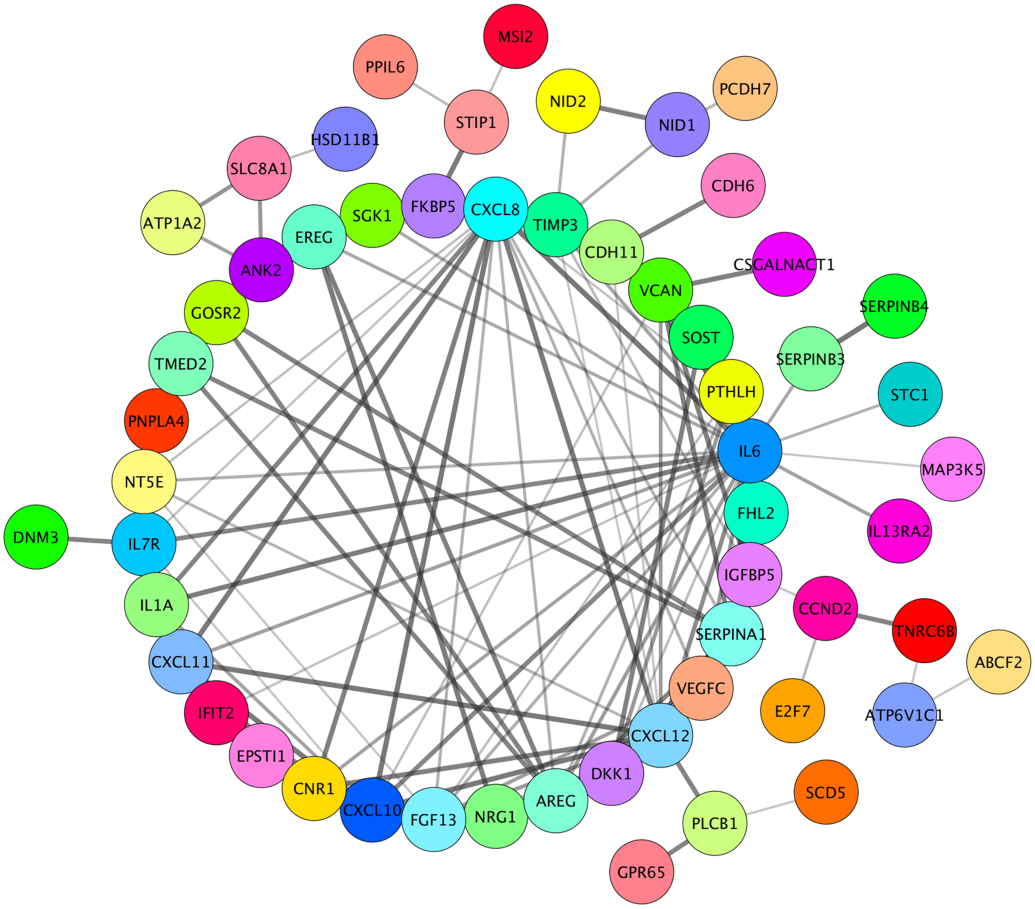
#	Maps	Network Objects fromActive Data
1	Immune response_IL-13 signaling via JAK-STAT	IL13RA2, FOXJ1, SCCA-2, iNOS, SCCA-1
2	G protein-coupled receptors signaling in lung cancer	CNR1, Galpha(i)-specific cannabis GPSRs, VIP receptor 1, IL-8, HB-EGF(mature), HB-EGF
3	Cell adhesion_ECM remodeling	Collagen III, IL-8, HB-EGF, Versican, TIMP3
4	Maturation and migration of dendritic cells in skin sensitization	IL-6, IL-8, ASK1 (MAP3K5), MHC class II alpha chain
5	Multiple myeloma (general schema)	IL-6, WHSC1, DKK1
6	Neuroendocrine transdifferentiation in Prostate Cancer	IL-6, IL-8, HB-EGF, PTHrP
7	PDE4 regulation of cyto/chemokine expression in inflammatory skin diseases	IL-6, IL-8, Adenylate cyclase, iNOS
8	Role of fibroblasts in the sensitization phase of allergic contact dermatitis	IL-6, Collagen III, IL-8
9	Immune response_MIF-mediated glucocorticoid regulation	IL-6, IL-8, iNOS
10	ERBB family and HGF signaling in gastric cancer	Neuregulin 1, IL-8, HB-EGF, Epiregulin

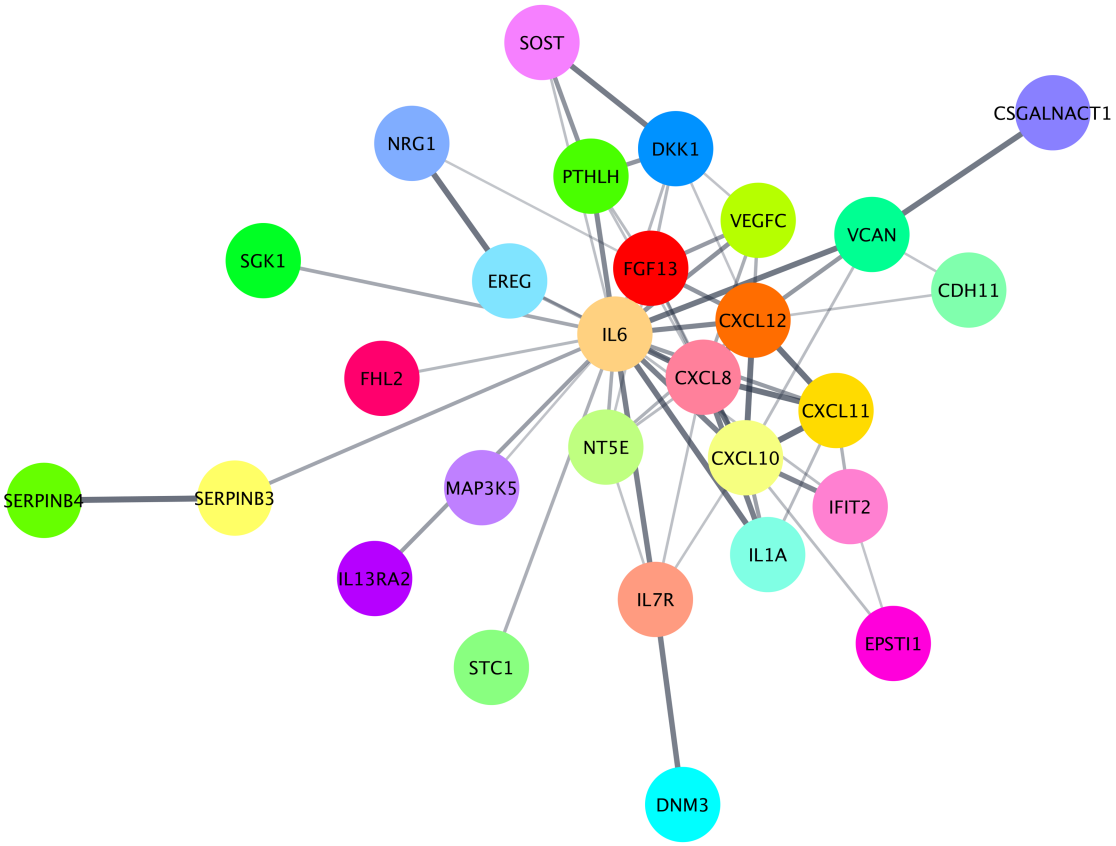
#	Processes	Network Objects fromActive Data
1	second-messenger-mediated signaling	CCL20, Olfactory receptor, GPR110, Galpha(s)-specific class A orphan/other GPCRs, GPR64, IL-8, AMPK alpha 1 subunit, AMPK alpha subunit, GPR65, NF-AT3(NFATC4), NF-AT, ATP1A2, ATP1alpha subunit, OA1, Adenylate cyclase type II, Adenylate cyclase, G-protein gamma, iNOS, Ankyrin-B, PTHrP
2	anatomical structure development	FHL2, LUZP1, GPR18, ACTL8, CNR1, Galpha(i)-specific cannabis GPSRs, RABGAP1L, RBG10, Keratin HB6, Gamma crystallin C, Keratin 17, LPP3, PPAP2, ECM2/SC1, IL-6, Olfactory receptor, GPR110, Neuregulin 1, Dynein, axonemal, heavy chains, Galpha(s)-specific class A orphan/other GPCRs, HMGA2, TRIM15, G3ST1, Collagen III, NPAS2, WHSC1, DLG5(P-dlg), MAZR, MPV17L, Podoplanin, COX VIIb-1, COX VIIb, PCDH17, MSI2, LAMB3, ITGB8, MST4, TNNT1, Troponin T, skeletal, Cdc42 subfamily, Rho GTPase, IL-8, BACE2, DKK1, HB-EGF, COL9A3, Dynamin-3, Dynamin, GPR65, MAP7(EMAP115), 5'-NTD, FGF13, FOXJ1, PHLDA1, KCRU, SGK1, Myomesin 2, Alpha 1-antitrypsin, ZNF420, PTPR-sigma, HSD11B1, RGS2, Formin, NF-AT3(NFATC4), NF-AT, INTU, FGF1, TACC2, ATP1alpha subunit, MKP-3, ANGPTL6, Aggrecanase-2, Adenylate cyclase, SNF2L1, K(+) channel, subfamily J, Kir1.1, E2F7, Epiregulin, MHC class II alpha chain, TMEFF2, G-protein gamma, KRT81, HIC1, HIC1/2, Versican, Versican proteoglycan, C4orf34, iNOS, VEGF-C, TIMP3, SIX6, CHST2, Carbohydrate sulfotransferases, NEBL, UMODL1, OTX2, Ankyrin-B, IBP5, IBP, PTHrP, CUTL2
3	multicellular organism development	FHL2, LUZP1, GPR18, CNR1, Galpha(i)-specific cannabis GPSRs, RABGAP1L, RBG10, Keratin HB6, Gamma crystallin C, Keratin 17, LPP3, PPAP2, ECM2/SC1, IL-6, Olfactory receptor, GPR110, Neuregulin 1, Dynein, axonemal, heavy chains, Galpha(s)-specific class A orphan/other GPCRs, HMGA2, TRIM15, G3ST1, Collagen III, NPAS2, WHSC1, DLG5(P-dlg), MAZR, MPV17L, Podoplanin, COX VIIb-1, COX VIIb, PCDH17, LAMB3, ITGB8, MST4, Cdc42 subfamily, Rho GTPase, IL-8, BACE2, DKK1, HB-EGF, COL9A3, Dynamin-3, Dynamin, GPR65, MAP7(EMAP115), 5'-NTD, FGF13, FOXJ1, PHLDA1, KCRU, SGK1, Myomesin 2, Alpha 1-antitrypsin, ZNF420, PTPR-sigma, HSD11B1, RGS2, Formin, NF-AT3(NFATC4), NF-AT, INTU, FGF1, TACC2, ATP1alpha subunit, MKP-3, ANGPTL6, Aggrecanase-2, Adenylate cyclase, SNF2L1, K(+) channel, subfamily J, Kir1.1, E2F7, Epiregulin, MHC class II alpha chain, G-protein gamma, KRT81, HIC1, HIC1/2, Versican, Versican proteoglycan, C4orf34, iNOS, VEGF-C, TIMP3, SIX6, CHST2, Carbohydrate sulfotransferases, NEBL, UMODL1, OTX2, Ankyrin-B, IBP5, IBP, PTHrP, CUTL2
4	regulation of multicellular organismal process	CCL20, GPR18, CNR1, Galpha(i)-specific cannabis GPSRs, Keratin 17, LPP3, PPAP2, IL-6, Olfactory receptor, Neuregulin 1, Galpha(s)-specific class A orphan/other GPCRs, HMGA2, TRIM15, Collagen III, WHSC1, DLG5(P-dlg), MAZR, Podoplanin, AMIGO2, ITGB8, KCNMB4, LCAT, TNNT1, Troponin T, skeletal, Cdc42 subfamily, Rho GTPase, IL-8, DKK1, AMPK alpha 1 subunit, AMPK alpha subunit, HB-EGF, Dynamin-3, Dynamin, FGF13, FOXJ1, LLIR, SGK1, PTPR-sigma, RGS2, NF-AT3(NFATC4), NF-AT, INTU, DOCK4, FGF1, ATP1A2, ATP1alpha subunit, RNF125, Sclerostin, MKP-3, Aggrecanase-2, Adenylate cyclase, K(+) channel, subfamily J, Epiregulin, HLA-DQA1, MHC class II alpha chain, HLA-DQA, G-protein gamma, KIAA0748, iNOS, VEGF-C, SCCA-1, UMODL1, OTX2, Ankyrin-B, IBP5, IBP, PTHrP, CUTL2
5	cellular response to hormone stimulus	FHL2, PPAP2, IL-6, Olfactory receptor, Galpha(s)-specific class A orphan/other GPCRs, ATP6V0A, ATP6V0A2, Cdc42 subfamily, Rho GTPase, AMPK alpha 1 subunit, AMPK alpha subunit, FOXJ1, SGK1, HSD11B1, p67-phox, ATP1A2, ATP1alpha subunit, Sclerostin, Adenylate cyclase type II, Adenylate cyclase, HLA-DQA1, MHC class II alpha chain, HLA-DQA, G-protein gamma 11, G-protein gamma, ATP6V1C, ATP6V1C1, UMODL1, IBP
6	developmental process	FHL2, LUZP1, GPR18, ACTL8, CNR1, Galpha(i)-specific cannabis GPSRs, STYK1, RABGAP1L, RBG10, Keratin HB6, Gamma crystallin C, Keratin 17, LPP3, PPAP2, ECM2/SC1, IL-6, Olfactory receptor, GPR110, Neuregulin 1, Dynein, axonemal, heavy chains, Galpha(s)-specific class A orphan/other GPCRs, HMGA2, TRIM15, G3ST1, Collagen III, NPAS2, WHSC1, DLG5(P-dlg), MAZR, MPV17L, Podoplanin, COX VIIb-1, COX VIIb, PCDH17, MSI2, LAMB3, ITGB8, MST4, TNNT1, Troponin T, skeletal, Cdc42 subfamily, Rho GTPase, IL-8, BACE2, DKK1, HB-EGF, COL9A3, Dynamin-3, Dynamin, GPR65, MAP7(EMAP115), 5'-NTD, FGF13, FOXJ1, PHLDA1, KCRU, SGK1, Myomesin 2, Alpha 1-antitrypsin, ZNF420, PTPR-sigma, HSD11B1, RGS2, Formin, NF-AT3(NFATC4), NF-AT, INTU, FGF1, p67-phox, TACC2, ATP1alpha subunit, MKP-3, ANGPTL6, Aggrecanase-2, Adenylate cyclase, SNF2L1, K(+) channel, subfamily J, Kir1.1, E2F7, Epiregulin, MHC class II alpha chain, TMEFF2, G-protein gamma, KRT81, HIC1, HIC1/2, Versican, Versican proteoglycan, C4orf34, iNOS, VEGF-C, TIMP3, SIX6, KYNU, CHST2, Carbohydrate sulfotransferases, NEBL, UMODL1, OTX2, Ankyrin-B, IBP5, IBP, PTHrP, CUTL2
7	response to hormone	FHL2, PPAP2, IL-6, Olfactory receptor, Neuregulin 1, Galpha(s)-specific class A orphan/other GPCRs, ATP6V0A, ATP6V0A2, LCAT, Cdc42 subfamily, Rho GTPase, AMPK alpha 1 subunit, AMPK alpha subunit, FOXJ1, SGK1, Alpha 1-antitrypsin, MTAP, HSD11B1, p67-phox, ATP1A2, ATP1alpha subunit, Sclerostin, Adenylate cyclase type II, Adenylate cyclase, K(+) channel, subfamily J, Epiregulin, HLA-DQA1, MHC class II alpha chain, HLA-DQA, G-protein gamma 11, G-protein gamma, ATP6V1C, ATP6V1C1, iNOS, TIMP3, UMODL1, IBP5, IBP
8	regulation of cellular component movement	CCL20, GPR18, LPP3, PPAP2, IL-6, Neuregulin 1, Dynein, axonemal, heavy chains, Galpha(s)-specific class A orphan/other GPCRs, Collagen III, DLG5(P-dlg), MAZR, Podoplanin, MST4, Cdc42 subfamily, Rho GTPase, IL-8, HB-EGF, FGF13, SGK1, DOCK4, FGF1, ATP1A2, ATP1alpha subunit, K(+) channel, subfamily J, Epiregulin, TMEFF2, VEGF-C, Carbohydrate sulfotransferases, SCCA-1, Ankyrin-B, IBP5, IBP

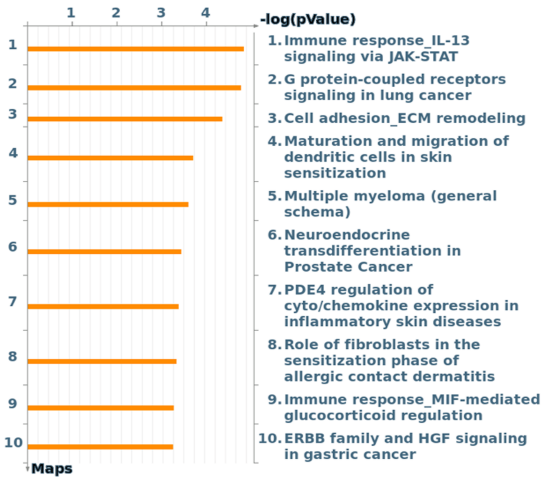
9	regulation of response to stimulus	CCL20, FHL2, GPR18, CNR1, Galpha(i)-specific cannabis GPSRs, STYK1, NUP210, SLC39A10, LPP3, PPAP2, IL-6, Olfactory receptor, Neuregulin 1, Galpha(s)-specific class A orphan/other GPCRs, HMGA2, PARG1, TRIM15, Collagen III, NPAS2, WHSC1, DLG5(P-dlg), MPV17L, BCAP, SNX25, Podoplanin, IL13RA2, MST4, TNNT1, Troponin T, skeletal, WRCH-1, Cdc42 subfamily, Rho GTPase, IL-8, Klhl15, DKK1, AMPK alpha 1 subunit, AMPK alpha subunit, HB-EGF, Dynamin, GPR65, 5'-NTD, FGF13, FOXJ1, LLIR, SCCA-2, PTPR-sigma, RGS2, NF-AT3(NFATC4), NF-AT, INTU, FGF1, C1r, ATP1A2, ATP1alpha subunit, RNF125, Sclerostin, MKP-3, LASP1, OA1, ASK1 (MAP3K5), Adenylate cyclase, MUCL1, Epiregulin, HLA-DQA1, MHC class II alpha chain, HLA-DQA, G-protein gamma, ATP6V1C, HIC1, HIC1/2, KIAA0748, RSG5, iNOS, VEGF-C, TIMP3, SCCA-1, OTX2, Mucin 12, IBP5, IBP, PTHrP, CUTL2
10	anatomical structure morphogenesis	FHL2, LUZP1, Keratin 17, LPP3, PPAP2, ECM2/SC1, IL-6, Olfactory receptor, Neuregulin 1, Dynein, axonemal, heavy chains, Galpha(s)-specific class A orphan/other GPCRs, HMGA2, TRIM15, Collagen III, WHSC1, DLG5(P-dlg), Podoplanin, LAMB3, ITGB8, MST4, TNNT1, Troponin T, skeletal, Cdc42 subfamily, Rho GTPase, IL-8, DKK1, HB-EGF, Dynamin, MAP7(EMAP115), FOXJ1, SGK1, Myomesin 2, Formin, NF-AT3(NFATC4), NF-AT, INTU, FGF1, ANGPTL6, Aggrecanase-2, Adenylate cyclase, E2F7, Epiregulin, TMEFF2, G-protein gamma, VEGF-C, SIX6, NEBL, OTX2, Ankyrin-B, IBP5, IBP, PTHrP

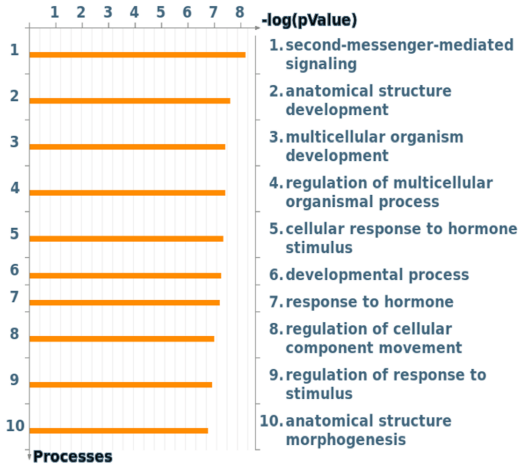
#	Networks	Network Objects fromActive Data
1	Inflammation_IL-13 signaling pathway	IL13RA2, FOXJ1, SCCA-2, Adenylate cyclase type II, iNOS, SCCA-1
2	Cell adhesion_Cell-matrix interactions	ECM2/SC1, Collagen III, LAMB3, COL9A3, ITGA9, Aggrecanase-2, Versican, TIMP3
3	Inflammation_Histamine signaling	CCL20, IL-6, IL-8, p67-phox, Adenylate cyclase type II, Adenylate cyclase, iNOS
4	Immune response_Innate immune response to RNA viral infection	CCL20, IL-6, IL-8, iNOS
5	Signal transduction_ESR1-membrane pathway	Neuregulin 1, HB-EGF, Adenylate cyclase type II, Adenylate cyclase
6	Inflammation_MIF signaling	IL-6, IL-8, Adenylate cyclase type II, Adenylate cyclase, iNOS
7	Immune response_Th17-derived cytokines	CCL20, IL-6, IL-8, iNOS
8	Inflammation_Neutrophil activation	IL-6, IL-8, p67-phox, Adenylate cyclase type II, Adenylate cyclase, iNOS
9	Inflammation_IL-4 signaling	IL-6, IL13RA2, IL-8, HLA-DQA1
10	Proteolysis_Connective tissue degradation	Collagen III, Alpha 1-antitrypsin, Aggrecanase-2, TIMP3

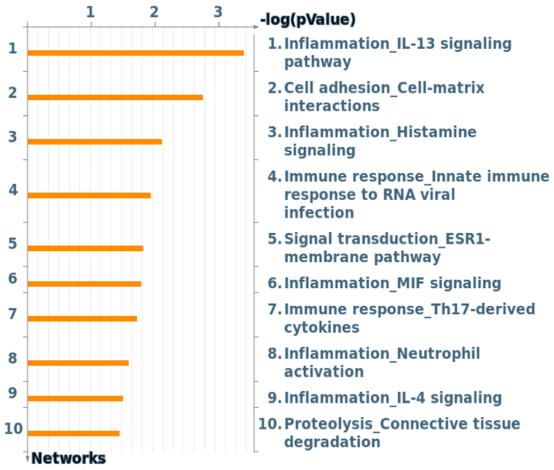
No	Network name	Processes
1	IL-8, IL-6, VEGF-C, Neuregulin 1, Epiregulin	regulation of cell proliferation (90.0%), positive regulation of intracellular signal transduction (72.0%), positive regulation of protein metabolic process (80.0%), positive regulation of multicellular organismal process (82.0%), response to hormone (74.0%)
2	WHSC1, Alpha 1-antitrypsin, WRCH-1, Tetraspanin-7, GPR65	response to peptide (40.0%), response to organic cyclic compound (48.9%), response to peptide hormone (35.6%), response to organonitrogen compound (46.7%), intracellular signal transduction (51.1%)
3	BCAP, RNF125, ANKRD18B, SCCA-1, LAMB3	positive regulation of CD8-positive, alpha-beta T cell proliferation (52.1%), regulation of CD8-positive, alpha-beta T cell proliferation (52.1%), antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-independent (50.0%), positive regulation of tolerance induction to nonself antigen (47.9%), regulation of tolerance induction to nonself antigen (47.9%)

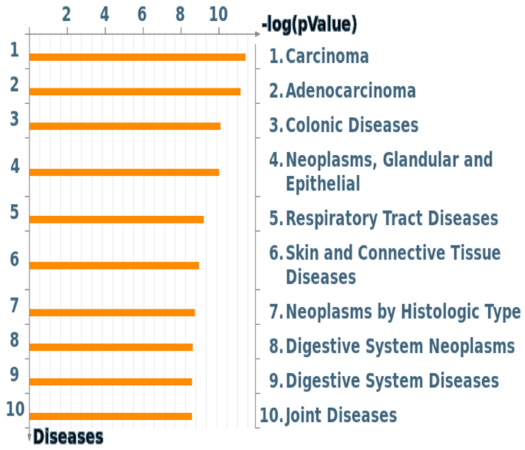

MSI2
PPIL6
STIP1
NID2
NID1
PCDH7
HSD11B1
SLC8A1
ATP1A2
SGK1
FKBP5
CXCL8
TIMP3
CDH6
EREG
CDH11
ANK2
CSGALNACT1
VCAN
GOSR2
SERPINB4
SOST
TMED2
SERPINB3
PTHLH
STC1
PNPLA4
IL6
MAP3K5
NT5E
FHL2
DNM3
IL7R
IL13RA2
IGFBP5
IL1A
CCND2
TNRC6B
SERPINA1
CXCL11
ABCF2
VEGFC
IFIT2
E2F7
ATP6V1C1
CXCL12
EPSTI1
DKK1
CNR1
SCD5
CXCL10
FGF13
NRG1
AREG
PLCB1
GPR65

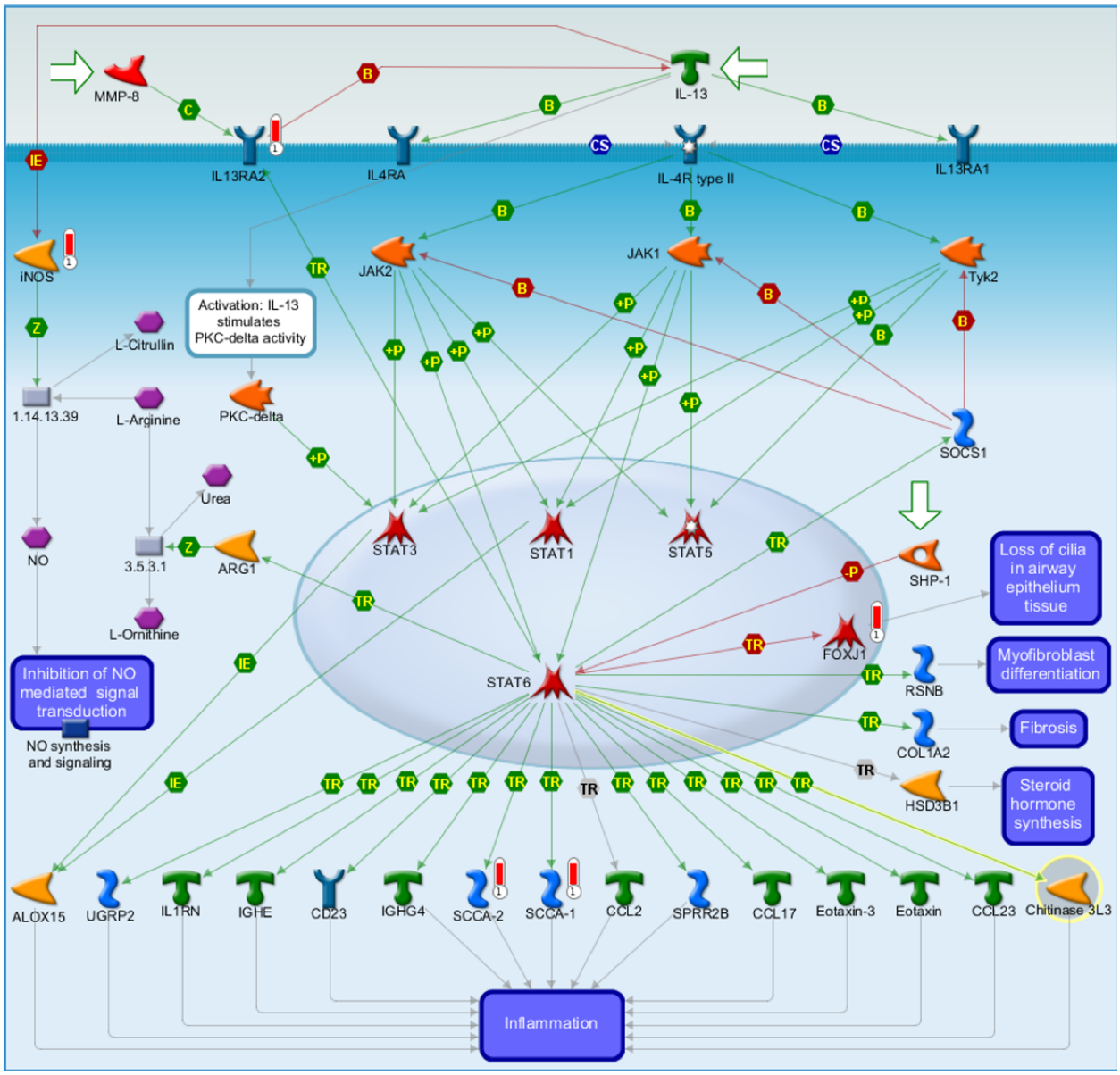

SOST
CSGALNACT1
NRG1
DKK1
PTHLH
VEGFC
VCAN
SGK1
EREG
FGF13
CDH11
CXCL12
IL6
FHL2
CXCL8
CXCL11
NT5E
CXCL10
SERPINB4
SERPINB3
MAP3K5
IFIT2
IL1A
IL13RA2
IL7R
EPSTI1
STC1
DNM3

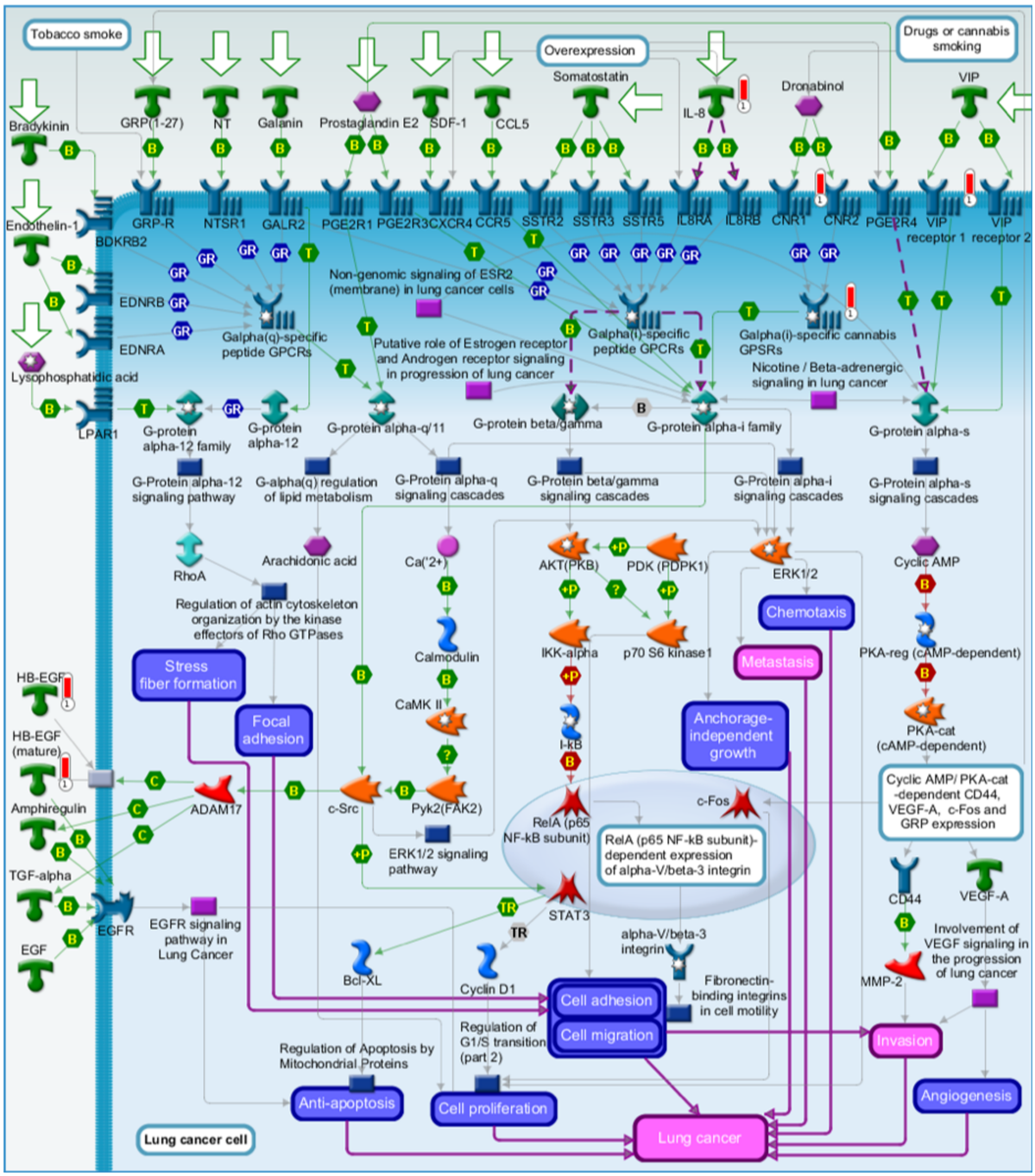

1 2 3 4
-log(pValue)
1. Immune response_IL-13 signaling via JAK-STAT
2. G protein-coupled receptors signaling in lung cancer
3. Cell adhesion_ECM remodeling
4. Maturation and migration of dendritic cells in skin sensitization
5. Multiple myeloma (general schema)
6. Neuroendocrine transdifferentiation in Prostate Cancer
7. PDE4 regulation of cyto/chemokine expression in inflammatory skin diseases
8. Role of fibroblasts in the sensitization phase of allergic contact dermatitis
9. Immune response_MIF-mediated glucocorticoid regulation
10. ERBB family and HGF signaling in gastric cancer
Maps

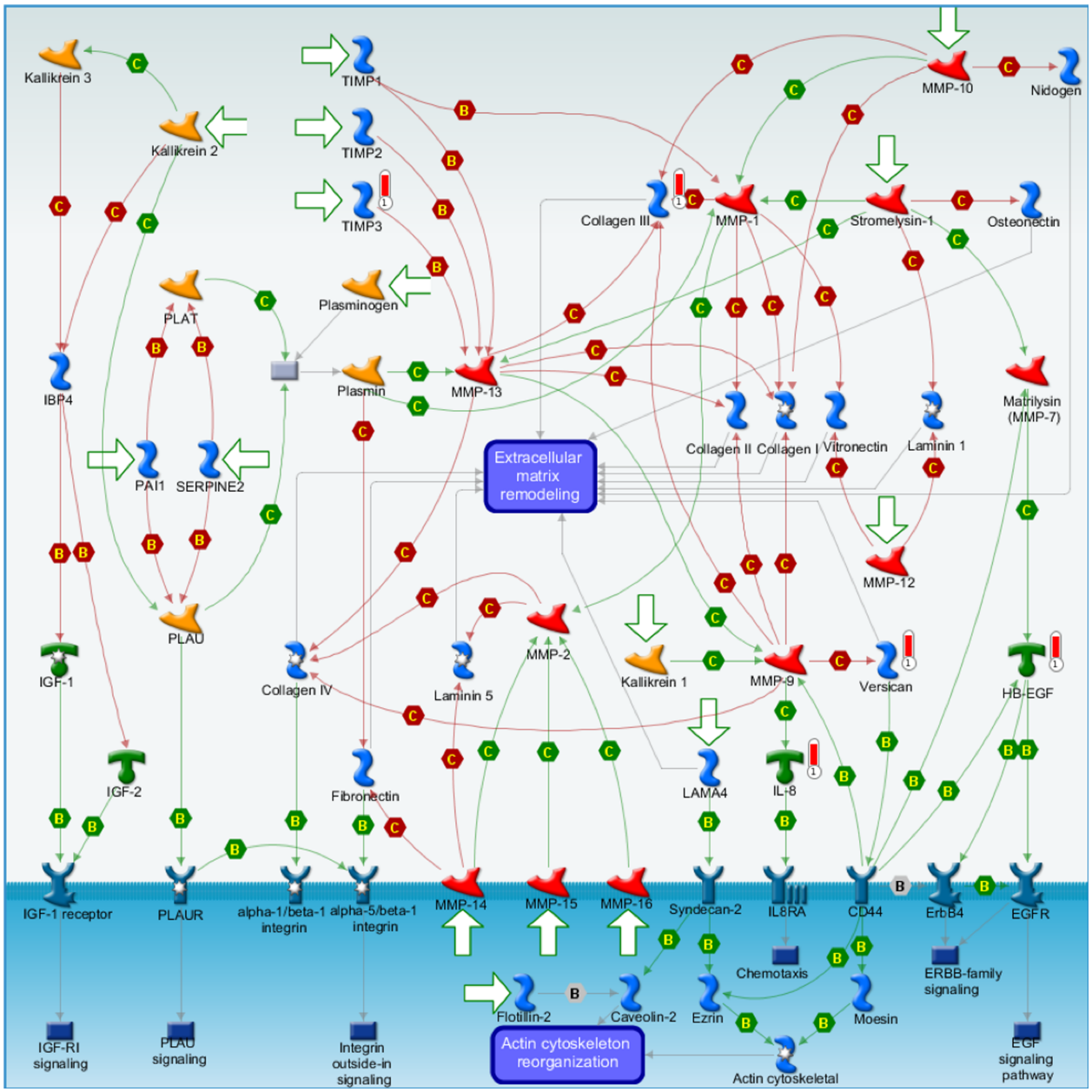

1 2 3 4 5 6 7 8
-log(pValue)
1. second-messenger-mediated signaling
2. anatomical structure development
3. multicellular organism development
4. regulation of multicellular organismal process
5. cellular response to hormone stimulus
6. developmental process
7. response to hormone
8. regulation of cellular component movement
9. regulation of response to stimulus
10. anatomical structure morphogenesis
Processes

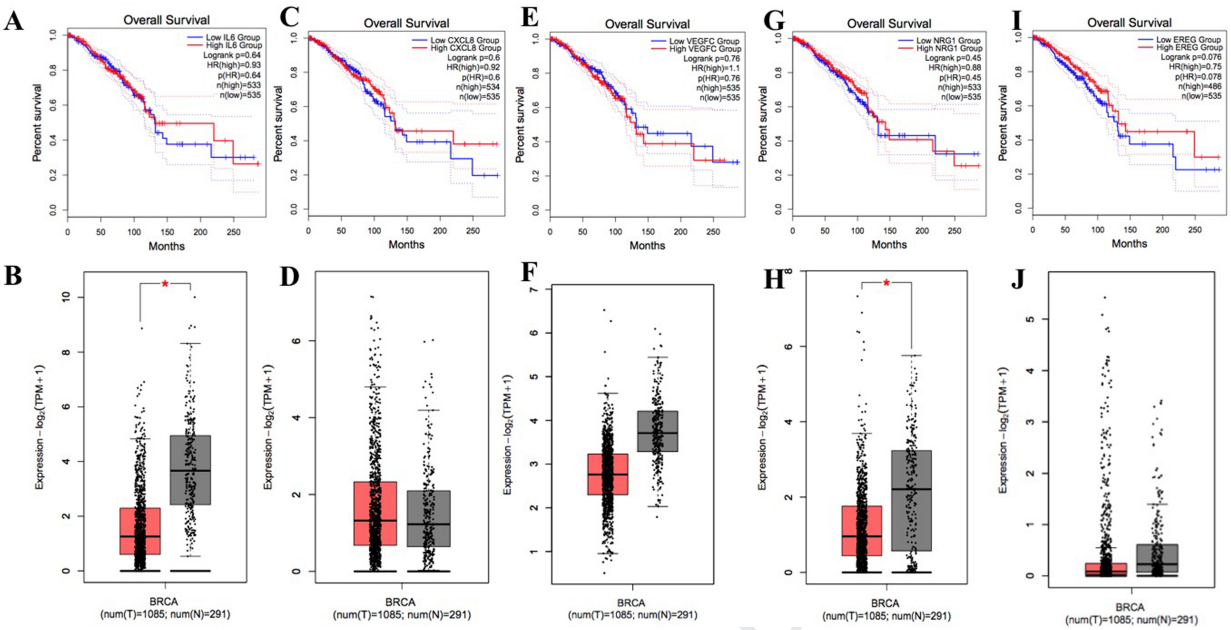

1 2 3
-log(pValue)
1. Inflammation_IL-13 signaling pathway
2. Cell adhesion_Cell-matrix interactions
3. Inflammation_Histamine signaling
4. Immune response_Innate immune response to RNA viral infection
5. Signal transduction_ESR1-membrane pathway
6. Inflammation_MIF signaling
7. Immune response_Th17-derived cytokines
8. Inflammation_Neutrophil activation
9. Inflammation_IL-4 signaling
10. Proteolysis_Connective tissue degradation
Networks

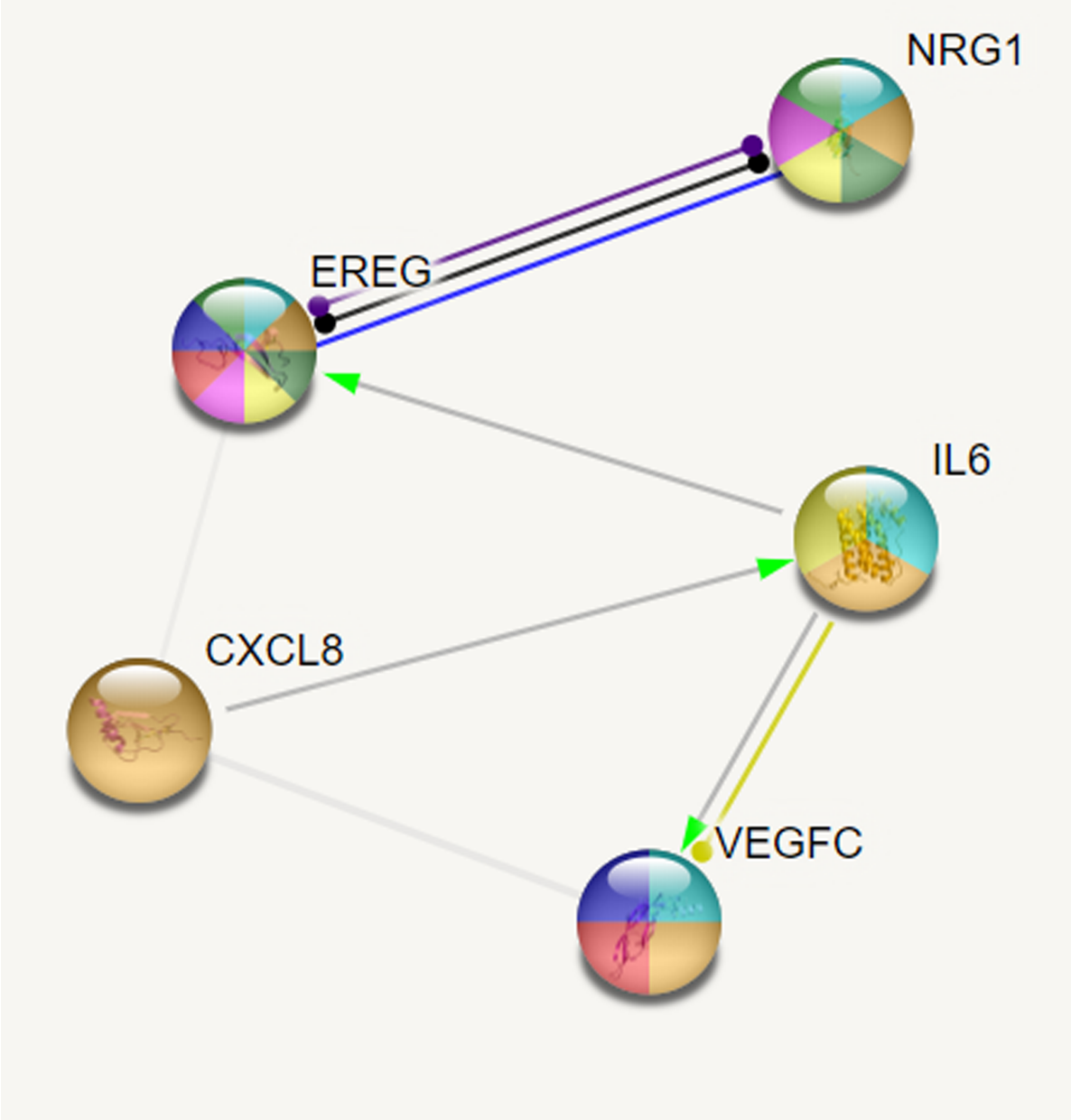

2 4 6 8 10
-log(pValue)
1. Carcinoma
2. Adenocarcinoma
3. Colonic Diseases
4. Neoplasms, Glandular and Epithelial
5. Respiratory Tract Diseases
6. Skin and Connective Tissue Diseases
7. Neoplasms by Histologic Type
8. Digestive System Neoplasms
9. Digestive System Diseases
10. Joint Diseases
Diseases


MMP-8
IL-13
IL13RA2
IL4RA
IL-4R type II
IL13RA1
iNOS
JAK2
JAK1
Tyk2
Activation: IL-13 stimulates PKC-delta activity
L-Citrullin
1.14.13.39
L-Arginine
PKC-delta
SOCS1
Urea
NO
3.5.3.1
ARG1
L-Ornithine
STAT3
STAT1
STAT5
SHP-1
Loss of cilia in airway epithelium tissue
FOXJ1
Inhibition of NO mediated signal transduction
NO synthesis and signaling
STAT6
RSNB
Myofibroblast differentiation
COL1A2
Fibrosis
HSD3B1
Steroid hormone synthesis
ALOX15
UGRP2
IL1RN
IGHE
CD23
IGHG4
SCCA-2
SCCA-1
CCL2
SPRR2B
CCL17
Eotaxin-3
Eotaxin
CCL23
Chitinase 3L3
Inflammation


Tobacco smoke
Drugs or cannabis smoking
Overexpression
Somatostatin
Dronabinol
VIP
Bradykinin
GRP(1-27)
NT
Galanin
Prostaglandin E2
SDF-1
CCL5
IL-8
Endothelin-1
GRP-R
NTSR1
GALR2
PGE2R1
PGE2R3
CXCR4
CCR5
SSTR2
SSTR3
SSTR5
IL8RA
IL8RB
CNR1
CNR2
PGE2R4
VIP receptor 1
VIP receptor 2
BDKRB2
EDNRB
EDNRA
Lysophosphatidic acid
LPAR1
Galpha(q)-specific peptide GPCRs
Non-genomic signaling of ESR2 (membrane) in lung cancer cells
Putative role of Estrogen receptor and Androgen receptor signaling in progression of lung cancer
Galpha(i)-specific peptide GPCRs
Galpha(i)-specific cannabis GPSRs
Nicotine / Beta-adrenergic signaling in lung cancer
G-protein alpha-12 family
G-protein alpha-12
G-protein alpha-q/11
G-protein beta/gamma
G-protein alpha-i family
G-protein alpha-s
G-Protein alpha-12 signaling pathway
G-alpha(q) regulation of lipid metabolism
G-Protein alpha-q signaling cascades
G-Protein beta/gamma signaling cascades
G-Protein alpha-i signaling cascades
G-Protein alpha-s signaling cascades
RhoA
Arachidonic acid
Ca(2+)
AKT(PKB)
PDK (PDPK1)
ERK1/2
Cyclic AMP
Regulation of actin cytoskeleton organization by the kinase effectors of Rho GTPases
Calmodulin
IKK-alpha
p70 S6 kinase1
Chemotaxis
Metastasis
PKA-reg (cAMP-dependent)
Stress fiber formation
Focal adhesion
CaMK II
I-kB
Anchorage-independent growth
PKA-cat (cAMP-dependent)
HB-EGF
HB-EGF (mature)
Amphiregulin
TGF-alpha
EGF
EGFR
ADAM17
c-Src
Pyk2(FAK2)
ERK1/2 signaling pathway
RelA (p65 NF-kB subunit)
c-Fos
RelA (p65 NF-kB subunit)-dependent expression of alpha-V/beta-3 integrin
Cyclic AMP/ PKA-cat -dependent CD44, VEGF-A, c-Fos and GRP expression
CD44
VEGF-A
EGFR signaling pathway in Lung Cancer
STAT3
alpha-V/beta-3 integrin
Involvement of VEGF signaling in the progression of lung cancer
Bcl-XL
Cyclin D1
Fibronectin-binding integrins in cell motility
MMP-2
Cell adhesion
Cell migration
Invasion
Regulation of Apoptosis by Mitochondrial Proteins
Regulation of G1/S transition (part 2)
Anti-apoptosis
Cell proliferation
Angiogenesis
Lung cancer cell
Lung cancer


Kallikrein 3
Kallikrein 2
TIMP1
TIMP2
TIMP3
MMP-10
Nidogen
Collagen III
MMP-1
Stromelysin-1
Osteonectin
PLAT
Plasminogen
IBP4
Plasmin
MMP-13
Matrilysin (MMP-7)
Collagen II
Collagen I
Vitronectin
Laminin 1
PAI1
SERPINE2
Extracellular matrix remodeling
MMP-12
PLAU
IGF-1
Collagen IV
Laminin 5
MMP-2
Kallikrein 1
MMP-9
Versican
HB-EGF
IGF-2
Fibronectin
LAMA4
IL-8
IGF-1 receptor
PLAUR
alpha-1/beta-1 integrin
alpha-5/beta-1 integrin
MMP-14
MMP-15
MMP-16
Syndecan-2
IL8RA
CD44
ErbB4
EGFR
Chemotaxis
ERBB-family signaling
Flotillin-2
Caveolin-2
Ezrin
Moesin
IGF-RI signaling
PLAU signaling
Integrin outside-in signaling
Actin cytoskeleton reorganization
Actin cytoskeletal
EGF signaling pathway




NRG1
EREG
IL6
CXCL8
VEGFC

THE IDENTIFICATION OF HIGHLY UPREGULATED GENES IN CLAUDIN-LOW BREAST CANCER THROUGH AN INTEGRATIVE BIOINFORMATICS APPROACH

CONFLICTS OF INTEREST

NONE DECLARED