Functional analysis of cancer gene subtype from co-clustering and classification

Logenthiran Machap[1], Afnizanfaizal Abdullah[2], Zuraini Ali Shah[3]
[1,2]Synthetic Biology Research Group, Universiti Teknologi Malaysia, Malaysia
[3]Artificial Intelligence and Bioinformatics Group, School of Computing, Faculty of Engineering,
Universiti Teknologi Malaysia, Malaysia

Article Info	ABSTRACT
	Cancer is a heterogeneity genetic disease with huge phenotypic alterations among dissimilar cancers types or even between same cancer types. Recent expansions of genome-wide profiling technologies offer a chance to explore molecular changes variations throughout advancement of cancer. Therefore, various statistical and machine learning algorithms have been designed and developed for the handling and interpretation of high-throughput microarray molecular data. Discovery of molecular subtypes studies have permitted the cancer to be allocated into similar groups that are deliberated to port similar molecular and clinical characteristics. Thus, the main objective of this research is to discover cancer gene subtypes and classify genes to obtain higher accuracy. In particular improved co-clustering algorithm used to discover cancer subtypes. And then supervised infinite feature selection gene selection method was combined with multi class SVM for classification of selected genes and further biological analysis. The analysis on breast cancer and glioblastoma multiforme evidences that top genes involved in cancer and the pathways present in both cancer top genes. The functional analysis is useful in medical and pharmaceutical field for cancer diagnosis and prognosis.

Corresponding Author:

Logenthiran Machap,
School of Computing, Faculty of Engineering,
Universiti Teknologi Malaysia, 81310 Skudai Johor Malaysia.
Email: logmac_87@yahoo.com

1. INTRODUCTION

Abnormalities of cancer genome can be observed through basic researches which have been used to categorize patients with respect to enhance their clinical decision making and implement more efficient treatments. Even though this types of categorization have enhanced the efficiency of treatment of various cancers, but the heterogeneity among the populations still remains as a main challenge. The advancement of DNA microarray technology has permitted an extensive understanding of genes especially in oncology field for start, diagnosis and prognosis of cancers. These various diagnostics are useful for different types of cancer, which lead to individual treatment plans and accurate clinical outcomes estimation [1, 2].

As the initial stage in organizing and investigating high-throughput gene expression datasets is through artificial intelligence in deep machine learning approach by grouping them together (cluster) according to similar biological features (gene) or conditions (samples) conferred on some similarity measures [3-5]. Meanwhile for both features and conditions are typically inadequate with prior knowledge, the clustering process is conducted as an unsupervised process via grouping features and conditions [6]. The conventional clustering is not said to be an ideal method for complicated and heterogeneous cancers. This is because, there are only certain genes in a subset of samples are expressed as a cancer genes in cellular processes among the similar clinical types of cancer in a specific tissue. Hence, it has been found a limitation that a single gene might play role in regulating and participating in numerous clusters and pathways of different conditions [7].

Therefore, Cheng and Church [8] has been the pioneer in implementation bi-clustering approach on gene expression datasets. Fundamentally, co-clustering [or bi-clustering] simultaneously clusters genes and samples to discover subtypes or subgroups of genes which displays similar patterns within certain conditions subset of experiments [9-11]. On top of that, this application has provided unique opportunities and challenges to classify tumors and discover more tumor subtype mechanisms. In addition to this, the researchers are also trying to identify multiple bi-clusters at a time. For example, statistical method [12], information theory [13], matrix factorization [14], and graph based bi-clustering [15]. Many co-clustering methods were developed by researchers. Cho and Dhillon [16] proposed a minimum sum-squared residue co-clustering (MSSRCC) to identify coherent bi-clusters. Modular Singular Value Decomposition (Mod-SVD) was proposed by Aradhya, Masulli [17] to discover bi-clusters from SVD computation. Huang, Sun [18] developed a modified fuzzy co-clustering (MFCC) while Hussain and Ramazan [19] proposed a method based on co-similarity measure between genes (and conditions).

Therefore, in this research we proposed to improve network assisted co-clustering for the identification of cancer subtypes (iNCIS). Generally this method utilizes gene network prior knowledge to be integrated with gene expression data to obtain bi-clusters.

2.   RESEARCH METHOD
2.1.   Dataset and Tools

Two cancer microarray datasets were used in this research. They are Breast Cancer (BRCA) and Glioblastoma Multiforme (GBM). In addition, both of these datasets are publicly available at The Cancer Genome Atlas (TCGA) where BRCA obtained from the Network [20] meanwhile GBM is from Verhaak, Hoadley [21]. The cancer datasets were in text file format and had been pre-formatted to be used as an input for the software. These datasets mainly comprise of numerical values; rows representing the genes and columns representing the samples/patients but the classes are unknown. BRCA contains 17814 genes and 547 samples while GBM contains 11861 genes and 202 samples. Co-clustering and classification are performed using MATLAB 2014a, while Feature selection library (FSLib) [22] is used to combine for classification.

2.2.   Co-clustering and Validation

In this stage, the first step is to assign weights to each genes using modified PageRank algorithm. Then the beginning of co-clustering (iNCIS) algorithm where the objectives function is improved to minimize the sum-squared residues and optimizes matrix X. The selections of parameters are based on cophenetic correlation coefficient and some of them are default [23, 24]. For the validation, silhouette analysis was done [25]. The larger the silhouette value is better the clustering. Beside this, subnetworks are obtained for a particular gene for both cancer and validated.

2.3.   Classification and Validation

The results of number of classes for both datasets were used further for feature selection and classification. Supervised infinite feature selection (SinFS) technique is combining to multiclass support vector machine (mSVM). Hence it is required the number of classes for each genes belongs to which we obtained from the co-clustering method. The genes with highest ranking and lowest threshold p-value are selected for classification further. And then, the selected top ten genes are analyze for functional analysis where the gene is belong to a specific cancer subtypes, biomarkers, oncogenes, transcription factors, tumor suppressor or antigen in addition their pathways involved.

3.   RESULTS AND ANALYSIS

This section deeply discussing about the overall results obtained from implementation of co-clustering and classification of two cancer gene expression datasets.

3.1.   Cancer Subtypes and Subnetworks of BRCA

From the implementation of BRCA dataset to iNCIS, there are five different cancer subtypes has been identified which are tabulated in Table 1 and visualize in Figure 1 (a) to (e). Gene ABCC8's subnetwork was produced for these five subtypes that obtained from iNCIS. Subnetwork visualization which shows the expression level and the weight of the gene will be helpful in order to demonstrate the difference of the five subtypes. Primarily, this subnetwork was chosen because it has connected with small number of genes and is easy to present visibly.

From BRCA data set, a small subnetwork (Figure 1) ABCC8 taken as an example to demonstrate the differences of five subtypes. There is total 16 genes presence in this subnetwork. Gene ABCC8 has been highly

expressed in Luminal A and B subtypes, moderately expressed in normal-like subtype and triple-negative and HER2 subtype shows low expression. Besides this, KLK11, KLK13, HDAC5 and RRAD genes are expressed moderate to high level in all subtypes.

Table 1. Number of samples for BRCA subtypes

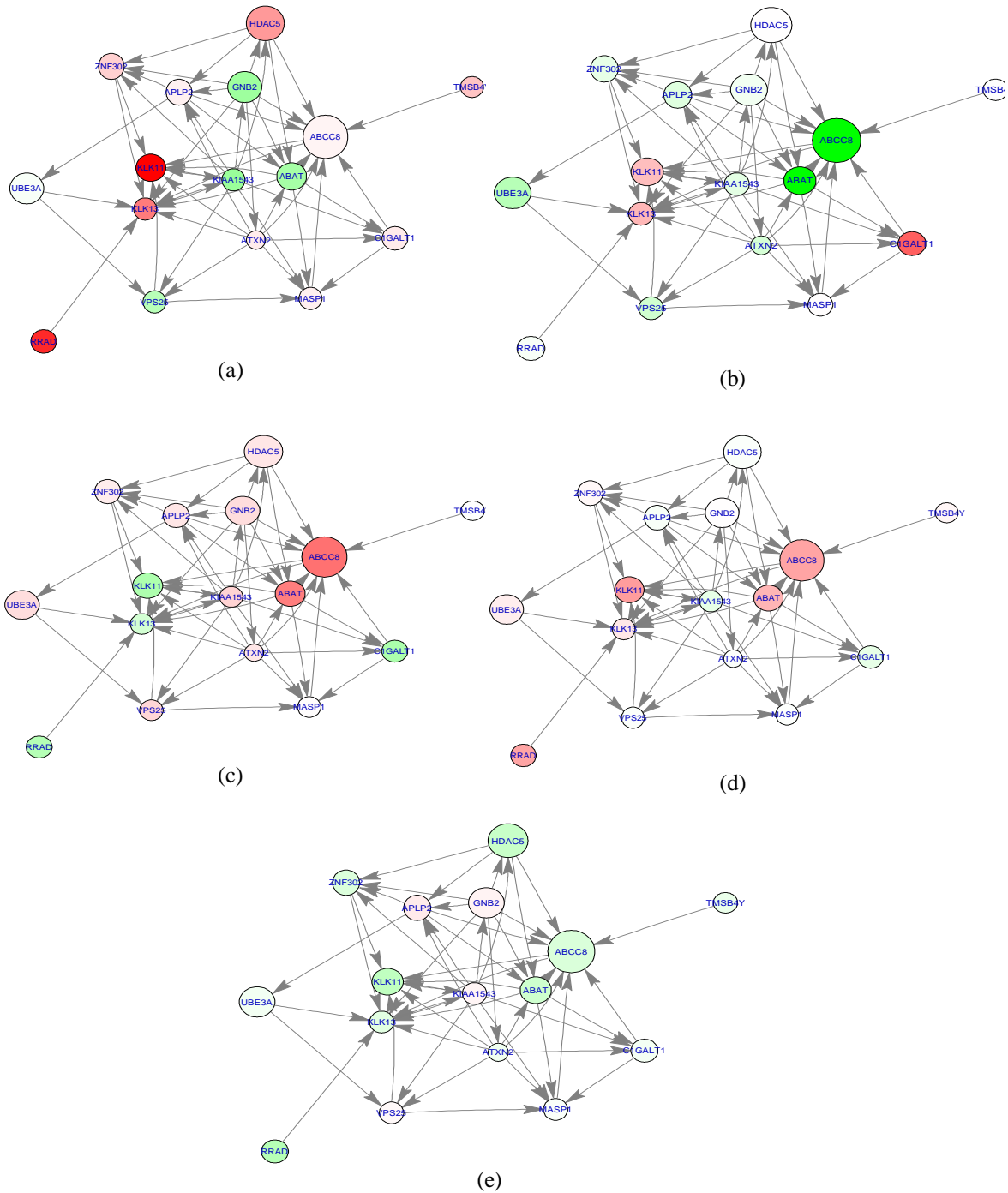| Normal-like	Triple negative/ basal-like	Luminal A	Luminal B	HER2-enriched	TOTAL
24	99	160	146	118	547

Figure 1. ABCC8 subnetwork in BRCA subtypes expression patterns. Direct connected genes to ABCC8 and genes targeting ABCC8's downstream are shown. Circle colour shows the gene expression level; circle size is based on gene weight. (a) Normal-like; (b) Basal-like; (c) Luminal A; (d) Luminal B; (e) HER2-enriched

3.2. Cancer Subtypes and Subnetworks of GBM

In this study, four subtypes have been identified through implantation of GBM in the co-clustering algorithm. Table 2 shows the obtained subtypes and its number of samples. Scientists from TCGA have published the finding of four distinct subtypes of GBM. They are Proneural, Neural, Mesenchymal and Classical subtypes [21, 26].

Table 2. Number of Samples for GBM Subtypes

Proneural	Neural	Mesenchymal	Classical	TOTAL
51	64	43	44	202

According to this implementation, gene NPTX1 act as the target gene, and the subnetwork was produced to interpret the relationship among genes. Figure 2 (a) to (d) shows all the four subtypes generated. Figure 2 (a) concluded as Proneural subtypes. This subtype is very common among the young adults which are normally characterised by IDH/TP53 positivity [27, 28]. From iNCIS, 51 samples of Proneural subtype were obtained. In addition, proneural subtype mostly deriving from low-grade gliomas which are associated with better prognosis [27].
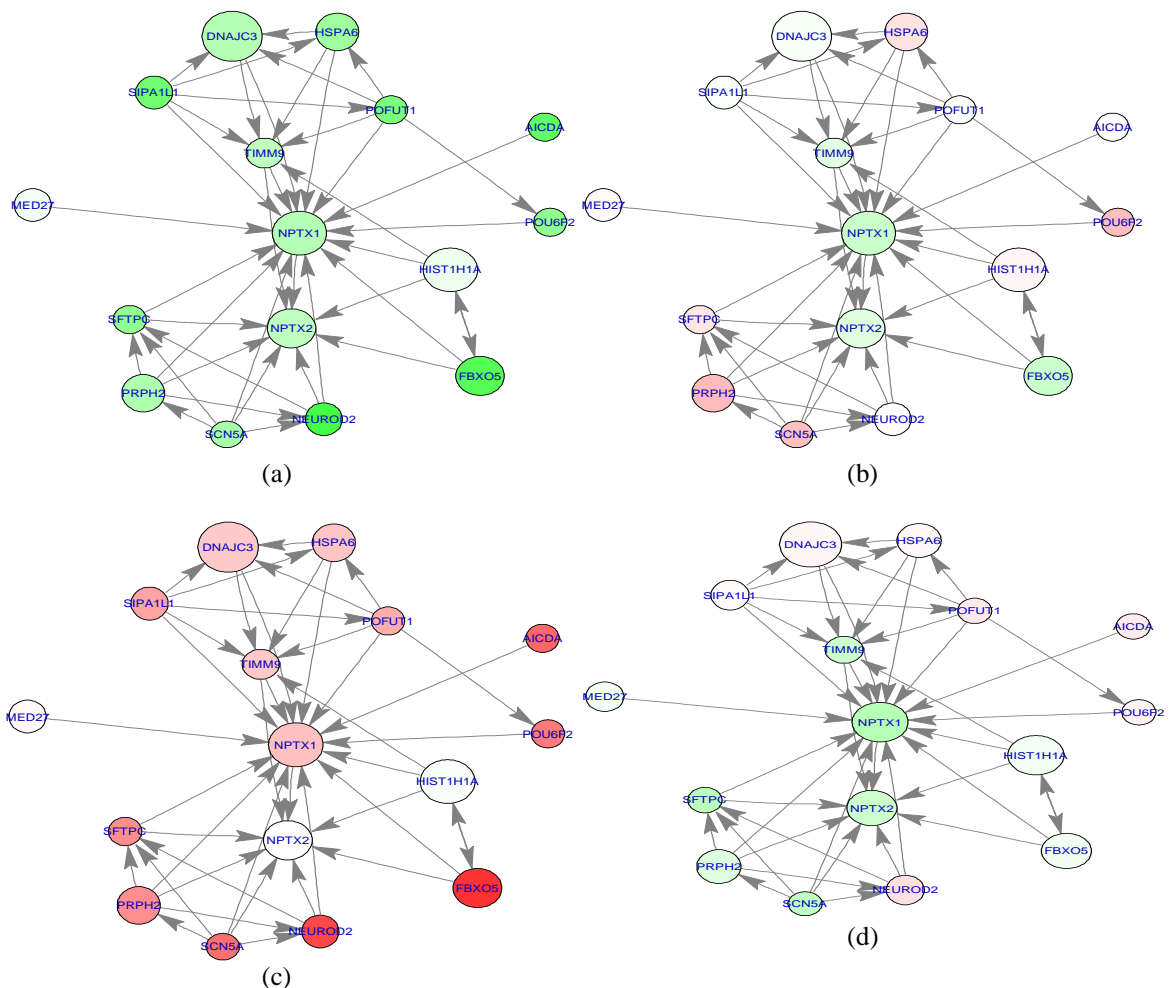

Figure 2. NPTX1 subnetwork in GBM subtypes expression patterns. Direct connected genes to NPTX1 and genes targeting NPTX1's downstream are shown. Circle colour shows the gene expression level; circle size is based on gene weight. (a) Proneural; (b) Neural; (c) Mesenchymal; (d) Classical

On average, Figure 3 shows optimal clusters. Hence the iodentification of five cancer subtypes of BRCA are concluded to be true. On average, Figure 4 shows optimal clusters. Hence the iodentification of four cancer subtypes of GBM are determined to be true.
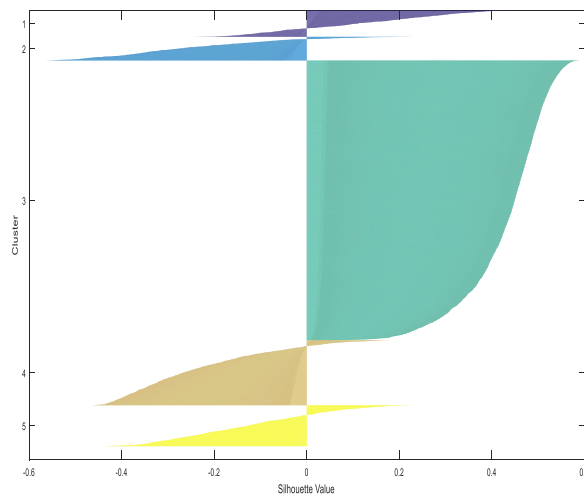

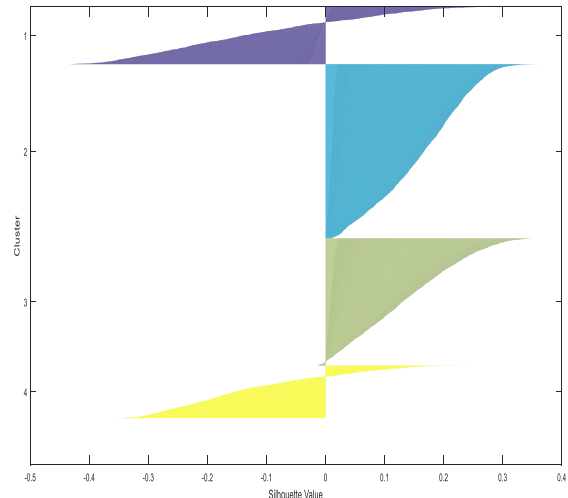Figure 3. Silhoutte plot for BRCA 5 subtype. Each subtypes shows different values

Figure 4. Silhoutte plot for GBM 4 subtype. Each subtypes shows different

3.3. Gene Analysis for BRCA and GBM

Apart from that, for the gene selection process, gene ranking plays an important role. The ranking is given to these genes based on proposed SinfFS-mSVM method. Although a produced gene has obtained highest rank among the gene list, it does not promise that this gene will be selected as one of the best gene.

The top ten (10) ranked genes listed in Table 3 and Table 4 are produced from gene ranking calculation as reported in proposed SinfFS+mSVM method. In addition, genes ranked from this list compared to genes' *p*-value which is generated from t-test and ANOVA. The *p*-value from ANOVA is calculated for the whole data set of BRCA and GBM. Hence the obtained genes' *p*-value is compared and selects the lowest *p*-value gene as the threshold value is 0.05 with the highest gene weight score.

Table 3. Top 10 Genes for BRCA subtypes

Overall	Normal-like	Triple negative	Luminal A	Luminal B	HER2
NPY1R	COL17A1	BCL11A	CCL13	IGF1*	HOXB13*
CEACAM6	KRT5	ABCC8*	EGFR	JAM2	MYBL2
TFAP2B	SFRP1	NAT1	ESR1*	LAMA2	PKMYT1
UGT2B11	ID4	GRPR	IL18RAP	NDN	CDKN3
SCGB2A2	NFIB	MLPH*	LCK	SLIT2*	E2F1
CBLN2	KRT17	CA12	TBX21*	RUNX1T1	PLK1*
ROPN1B	OSR1	SCUBE2	PLA2G4A	JAM3	AURKA*
AREG	TRIM29	GFRA1*	SLAMF1	BMX	KIF4B*
ROPN1	EGFR	ESR1	PTX3	CXCL12	GTSE1
PDZK1	BIRC5	ERBB4	BCL11A	COL14A1	CX3CR1

*Red: significant genes; * found particularly in the subtypes*

Table 4. Top 10 genes for GBM subtypes

Overall	Proneural	Neural	Classical	Mesenchymal
RPS4Y1	CHD3	PTPRJ	ACTA2	POLB
LTF	PGM1	ACY1*	LHX1	CEP76
IL8	ADARB1	CEP27	ASGR1*	NFATC4
DKK1	DMD	CPS1	AMFR*	MAP3K14*
EGFR	OXT	PF4	RPS15A	ACTA2*
PTX3	ANGPTL4	CNTNAP2	CYBA	DNAJB5
IL13RA2	CYP7A1	DNAJB5*	CCT6A	PPARD
FABP5	EIF2B5	THAP11	INHA*	RAB17
CHI3L1	BCL2L10	PHKB	RNGTT	MYH2
MOXD1	PKLR	RPL19	RGS14*	TCERG1

*Red: significant genes; * found particularly in the subtypes*

Functional analysis of cancer gene subtype from co-clustering and classification (Logenthiran Machap)

The first column of both Table 3 and Table 4 which is named as overall set of gene is produced through comparison of lowest ANOVA *p*-value and highest weight score. Whereas, the gene list for each subtype is attained from comparison of lowest *p*-value from t-test and highest weight score.
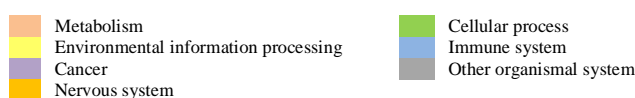
3.4. Pathway Analysis for BRCA and GBM

The pathways involved in each subtype of BRCA and GBM investigated. The pathway analysis conducted on gene selected for each subtype as shown Table 5 and 6 for BRCA and GBM respectively. Hence, GeneCards (https://www.genecards.org) and KEGG: Kyoto Encyclopaedia of Genes and Genomes (https://www.genome.jp/kegg/) used to identify the top significant pathways for each subtypes. The pathway analysis was conducted for the selected overall genes on both data sets.

Table 5 and 6, shows the top enriched pathways of the common genes and subtype-specific genes of each subtype from BRCA and GBM correspondingly. The pathways for both data sets have been classified into seven different types, which involve (i) cellular process, (2) metabolism, (3) environmental information processing, (4) nervous system, (5) cancers, (6) immune system, and (7) other organismal systems. The first column in both table shows the pathways obtained based on overall gene from BRCA and GBM data sets.

It can be noticed that the BRCA pathways are relatively different among different subtypes. The pathways for overall genes and samples (in the first column of the Table 5) are related to metabolism and cellular process. Pathways for triple-negative subtype are associated to cancers and metabolism which infers the growth of breast cell tumors. The pathways in subtype Luminal A related to cancers, metabolism and immune systems. Pathways in subtype Luminal B are highly related to cancers and metabolism type. While pathways from HER2 subtype are linked with cellular process and metabolism. Most of these pathways are having linked with breast cancer in breadth and depth.

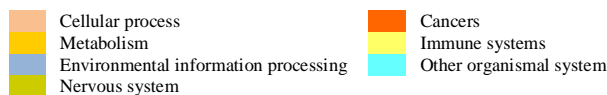Table 5. Top 10 Pathways for BRCA Subtypes

Overall	Triple-negative	Luminal A	Luminal B	HER2
Signaling by GPCR.	Development_Leptin signaling via PI3K-dependent pathway	Akt Signaling	ERK Signaling	Regulation of Androgen receptor activity
Hematopoietic Stem Cell Differentiation Pathways and Lineage-specific Markers	Integrated Breast Cancer Pathway	NF-kappa B signaling	Focal adhesion	Class I MHC mediated antigen processing and presentation
Apoptosis and Autophagy	Type II diabetes mellitus.	ERK Signaling	Blood-Brain Barrier and Immune Cell Transmigration: VCAM-1/CD106 Signaling	Gene Expression
Integrated Breast Cancer Pathway.	Metabolism	Hematopoietic Stem Cell Differentiation	Endometrial cancer	EGF/EGFR Signaling
Synaptogenesis induction.	Pharmacokinetics.	Integrated Breast Cancer Pathway.	Pathways in cancer	Cyclin A/B1 associated events during G2/M transition
PKA-dependent signaling processes	Signaling by GPCR	IL12 signaling mediated by STAT4	Adipogenesis and Cytokine Signaling in Immune system	Cell Cycle, Mitotic.
Signaling by Rho GTPases	Akt Signaling ERK signalling	Immune response IL-23 signaling	Developmental Biology	Integrated Breast Cancer Pathway.
ERK Signaling	Deregulation of Rab and Rab Effector Genes in Bladder Cancer	T cell receptor signaling	Signaling by Robo receptor.	Golgi-to-ER retrograde transport
Regulation of CFTR activity (norm and CF) and Uricosurics Pathway, Pharmacodynamics.	Translational Control	Glucocorticoid receptor regulatory network	G-Beta Gamma Signaling	Developmental Biology
Activation of cAMP-Dependent PKA.	CDK-mediated phosphorylation and removal of Cdc6	p38 MAPK Signaling	Apoptotic cleavage of cellular proteins	Regulation of beta-cell development

▉ Metabolism		▉ Cellular process	
▉ Environmental information processing		▉ Immune system	
▉ Cancer		▉ Other organismal system	
▉ Nervous system			

From Table 6, it can be concluded that pathways for overall top genes from GBM are linked with mixed types such as immune system, metabolism, nervous system and cellular process on cell growth. Proneural subtypes involved highly in cellular process, metabolism and some immune system types. Beside this, neural, classical and mesenchymal subtypes have greatly involved in cellular processing pathways and significant metabolism pathways. Moreover, mesenchymal subtype having many immune system pathways implicates. On top of that, most of these pathways are associated with glioblastoma multiforme cancer which is incorporate directly or indirectly.

Table 6. Top 10 Pathways for GBM Subtypes

Overall	Proneural	Neural	Classical	Mesenchymal
Defensins	Activation of the mRNA upon binding of the cap-binding complex and eIFs	Activated PKN1 stimulates transcription of AR (androgen receptor) regulated genes KLK2 and KLK3	ERK Signaling	Telomere C-strand (Lagging Strand) Synthesis
Innate Immune System	Defensins	Gene Expression	Apelin signaling pathway	Cell Cycle, Mitotic
Akt Signaling	Rheumatoid arthritis	Galactose metabolism	Ectoderm Differentiation	Organelle biogenesis and maintenance.
Reelin Pathway	Akt signalling	Pyrimidine metabolism (KEGG).	Neural Crest Differentiation	GPCR Pathway
Gene Expression	Reelin Pathway	C6 deamination of adenosine	Metabolism of proteins	T cell receptor signaling pathway
Lipoprotein metabolism Metabolism.	ERK Signaling	Formation of the Editosome	Aldosterone synthesis and secretion	Bacterial infections in CF airways
ERK Signaling	Gene expression	Dilated cardiomyopathy (DCM)	Calnexin/calreticulin cycle	ERK Signaling
Post NMDA receptor activation events	Innate Immune System	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	Protein processing in endoplasmic reticulum.	Apelin signaling pathway
Signaling by PTK6	Lung fibrosis	Peptide ligand-binding receptors	TNFR1 Pathway	Gene Expression
NAD metabolism	Lipoprotein metabolism	Lipoprotein metabolism	Class I MHC mediated antigen processing and presentation	Adipogenesis

▮	Cellular process	▮	Cancers
▮	Metabolism	▮	Immune systems
▮	Environmental information processing	▮	Other organismal system
▮	Nervous system		

4.　CONCLUSION

From the results, it can be concluded that, there are five subtypes from BRCA and four subtypes from GBM were successfully identified. The iNCIS algorithm is able to produce simple subnetwork to show gene expression in each subtype. By the feature selection and classification, it has been able to prioritize significant genes for each subtype of both data sets which are analyzed for disease prognosis and diagnosis.

ACKNOWLEDGEMENTS

REFERENCES

[1]. Barrier A, Boelle P-Y, Roser F, Gregg J, Tse C, Brault D, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. *Journal of Clinical Oncology*. 2006; 24(29):4685-91.
[2]. Zhang S. Computational Methods for Subtyping of Tumors and Their Applications for Deciphering Tumor Heterogeneity. *Cancer Bioinformatics: Springer*; 2019. p. 193-207.
[3]. Abdullah A, Deris S, Hashim SZM, Mohamad MS, Arjunan SNV, editors. *An improved local best searching in Particle Swarm Optimization using Differential Evolution*. 2011 11th International Conference on Hybrid Intelligent Systems (HIS); 2011 5-8 Dec. 2011.
[4]. Abdullah A, Deris S, Mohamad MS, Anwar S. An improved swarm optimization for parameter estimation and biological model selection. *PLoS One*. 2013; 8(4):e61258.
[5]. Algfoor ZA, Sunar MS, Abdullah A. A new weighted pathfinding algorithms to reduce the search time on grid maps. *Expert Systems with Applications*. 2017; 71:319-31.
[6]. Eisen MB, Spellman PT, Brown PO, Botstein D. *Cluster analysis and display of genome-wide expression patterns*. Proceedings of the National Academy of Sciences. 1998; 95(25):14863-8.
[7]. Singh A, Bhanot G, Khiabanian H. TuBA: Tunable Biclustering Algorithm Reveals Clinically Relevant Tumor Transcriptional Profiles in Breast Cancer. *BioRxiv*. 2018:245712.
[8]. Cheng Y, Church GM. *Biclustering of Expression Data*. Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology. 660833: AAAI Press; 2000. p. 93-103.
[9]. Liu Y, Gu Q, Hou J, Han J, Ma J. A network-assisted co-clustering algorithm to discover cancer subtypes based on gene expression. *BMC Bioinformatics*. 2014; 15(1):37.
[10]. Yu G, Yu X, Wang J. Network-aided Bi-Clustering for discovering cancer subtypes. *Scientific Reports*. 2017; 7(1):1046.

[11]. Xu T, Le TD, Liu L, Wang R, Sun B, Li J. Identifying Cancer Subtypes from miRNA-TF-mRNA Regulatory Networks and Expression Data. *PLOS ONE*. 2016; 11(4):e0152792.
[12]. Shan H, Banerjee A, editors. *Bayesian Co-clustering*. 2008 Eighth IEEE International Conference on Data Mining; 2008 15-19 Dec. 2008.
[13]. Dhillon IS, Mallela S, Modha DS. *Information-theoretic co-clustering*. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining; Washington, D.C. 956764: ACM; 2003. p. 89-98.
[14]. Carmona-Saez P, Pascual-Marqui RD, Tirado F, Carazo JM, Pascual-Montano A. Biclustering of gene expression data by non-smooth non-negative matrix factorization. *BMC Bioinformatics*. 2006; 7(1):78.
[15]. Denitto M, Farinelli A, Bicego M, editors. Biclustering Gene Expressions Using Factor Graphs and the Max-Sum Algorithm. *IJCAI*; 2015.
[16]. Cho H, Dhillon IS. Coclustering of Human Cancer Microarrays Using Minimum Sum-Squared Residue Coclustering. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2008; 5(3):385-400.
[17]. Aradhya VNM, Masulli F, Rovetta S, editors. A Novel Approach for Biclustering Gene Expression Data Using Modular Singular Value Decomposition2010; Berlin, *Heidelberg: Springer Berlin Heidelberg*.
[18]. Huang S-Y, Sun H-J, Huang C-D, Chung I-F, Su C-H, editors. *A modified fuzzy co-clustering (MFCC) approach for microarray data analysis. Fuzzy Systems (FUZZ-IEEE)*, 2014 IEEE International Conference on; 2014: IEEE.
[19]. Hussain SF, Ramazan M. Biclustering of human cancer microarray data using co-similarity based co-clustering. *Expert Systems with Applications*. 2016; 55(Supplement C):520-31.
[20]. Network CGA. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490(7418):61.
[21]. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010; 17(1):98-110.
[22]. Roffo G, Melzi S, Cristani M, editors. *Infinite feature selection*. Proceedings of the IEEE International Conference on Computer Vision; 2015.
[23]. Ismail MA, Deris S, Mohamad MS, Abdullah A. A Newton Cooperative Genetic Algorithm Method for In Silico Optimization of Metabolic Pathway Production. *PLOS ONE*. 2015; 10(5):e0126199.
[24]. Ismail MA, Deris S, Mohamad MS, Isa MA, Abdullah A, Remli MA, et al. A hybrid of optimization method for multiobjective constraint optimization of biochemical system production. *Journal of Theoretical & Applied Information Technology*. 2015; 81(3).
[25]. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987; 20:53-65.
[26]. Network CGAR. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *New England Journal of Medicine*. 2015; 372(26):2481-98.
[27]. Li M, Xiao A, Floyd D, Olmez I, Lee J, Godlewski J, et al. CDK4/6 inhibition is more active against the glioblastoma proneural subtype. *Oncotarget*. 2017; 8(33):55319.
[28]. Olar A, Aldape KD. Using the molecular classification of glioblastoma to inform personalized treatment. *The Journal of pathology*. 2014; 232(2):165-77.

BIOGRAPHIES OF AUTHORS

Logenthiran Machap finished B.Sc. Bioinformatics from National University of Malaysia (UKM), Master of Computer Science from University of Technical Malaysia Melaka (UTeM). He is a PhD candidate of computer science from University of Technology Malaysia (UTM). He is currently working on co-clustering algorithm and its application on cancer microarray gene expression data. His research interests include data mining, machine learning, artificial intelligence and bioinformatics.

Afnizanfaizal Abdullah holds B.Sc., M.Sc., and PhD in computer science from University of Technology Malaysia (UTM). He's a Senior Lecturer at School of Computing in Faculty of Engineering at UTM. Besides this, he is a Deputy Director of UTM Centre for Student Innovation & Technological Entrepreneurship and served as Chair for IEEE Young Professionals (Malaysia Section). With background of computing and computer science, his proficiency involved around Data Science comprise Machine learning approaches for vast data analysis; Computational Systems and Synthetic Biology in designing and improving biological data models.

Zuraini Ali Shah is a Senior Lecturer in Department of Software Engineering, School of Computing, Faculty of Engineering at UTM, where she has been since 2000. She received a B.Sc and M.Sc from UTM. She received her Ph.D. in Computer Science in 2012. Her research interests include Computational Intelligence in Pattern recognition, Machine learning in Bioinformatics and others.