

**GENOME WIDE INTEGRATION ANALYSIS - A NEW APPROACH TO PREDICT
SUSCEPTIBLE GENE(S) FOR LUNG CANCER.****Muthusamy Chinnasamy¹ and Thirunalasundari Thiyagarajan^{*2}**¹Research Scholar, Department of Industrial Biotechnology, Bharathidasan University, Tiruchirappalli,
Tamil Nadu – 620 024.²Professor & Head, Department of Industrial Biotechnology, Bharathidasan University, Tiruchirappalli,
Tamil Nadu – 620 024.***Correspondence for Author: Dr. Thirunalasundari Thiyagarajan**

Professor & Head, Department of Industrial Biotechnology, Bharathidasan University, Tiruchirappalli, Tamil Nadu – 620 024.

Article Received on 22/02/2016

Article Revised on 13/03/2016

Article Accepted on 03/04/2016

ABSTRACT

Genome wide integration study (GWIS) is the new branch of science that deals with complete genome of the cell type with a specific condition. It is a comprehensive approach to test the hypothesis that, multiple genes work together and account for disease. Diagnosis of the disease at early stage is essential to cure the cancer. Methods available for cancer diagnosis as of now are all lab based and they are time bound and economically ineffective. Hence, a comprehensive analytical method is an urgent need of the hour to diagnose cancer. This study was aimed to find out a new approach for the detection of susceptible gene(s) for cancer using tensor algorithm. Three datasets were downloaded from GEO database. They are i) GSE18454- normal lung cell (NLC), ii) GDS1204 - lung cancer cell (LCC) and iii) GDS1204 -lung cancer cell treated with Mgd (LCCTD). Pre-processing of the data leads to 7,000 valid genes that are similar in all these 3 datasets. Tensor algorithm was implemented to analyze the data. Results revealed that 93% of genes were classified in to different categories, and the rest 7% genes were insignificant. Of the 93% genes classifiable 56% of input genes are down regulated after treatment, 4% of genes are up regulated. Mgd treatment restores the genes expression successfully in 4%. 29% of genes were found to be further up-regulated after treatment. As an outcome a tool, will be provided to medical personale and using that they can analyse genomic data and predict lung cancer gene target(s).

KEYWORDS: Lung cancer, Tensor, Genome Wide Analysis (GWA), PHP, MySQL, Mgd.**INTRODUCTION**

Genome wide integration study (GWIS) is a new branch of science that deals with complete genome of the cell type with a specific condition. It is a comprehensive approach to test the hypothesis that, multiple genes work together and account for disease. By performing GWIS it is possible to address many important questions such as, genes that are down-regulated or up-regulated, impact of a gene(s) on other gene(s) expression and list of co-expression network profiles of different cell types. It helps in the development of personalized treatment strategies. In addition, it helps to discover the association of gen(s) in cancer.^[1] Further, it revolutionizes the method of genetic analysis and look for genes that influences complex diseases or genetic disorders like P⁵³, pRb, PTEN, BRCA1, APC, CD95, ST5 and YPEL3 etc.,^[2] Recent advances in high-throughput technologies produces massive amount of high resolution data, that enable us to monitor molecular signals on genomic scale, such as gene expression levels, protein – DNA binding and occupancy levels, that corresponds to multiple events of cellular systems i.e., DNA replication, transcription and translation.^[3,4] Integrative analysis of these global signals promises to give new insights into

cellular mechanisms of global cellular activities. However, analyzing these biological data requires efficient tool that can handle large amount of data sets in terms of a common frame. This may reduce the complexity of data and makes them comprehensible.

Making use of computational and mathematical sciences, WGIS has discovered hundreds of disease-associated networks in human and provided insights into the molecular architecture of complex diseases. Traditionally, several algorithms are used to implement WGIS. Some of them are Mutual information, Bayesian inference, Correlation, PCA, Neural Network, DFT and Dynamical System.^[5] Right now these methods are employed in studies of cancer, diabetics and heart disease.^[6] These algorithms facilitate the study of normal and disease or affected tissues.

Cancer is the state, which loses its normal control over cell growth and division. In normal conditions, a cell maintains a specific state by tightly controlling various molecules using a variety of regulatory mechanisms (mitosis and cell death). Certainly in critical conditions, a cell adjusts its regulatory mechanism(s) accordingly. The

damaged/alterd regulatory mechanism may cause the cells transition into another state that significantly differs from the normal. Cancer is of more than 100 different types based on the cell that is initially affected. Broadly cancer is divided into i) Carcinoma – cancers of epithelial cells that covers 90% of human cancers, ii) Sarcomas – tumors of connective tissue that forms rarely and iii) Leukemia and lymphoma – cancers of blood cells that comprise 8% of human cancers. Further classifications of cancers are based on the organ affected or area affected and that include lung cancer, breast carcinoma, squamous cell carcinoma, adenocarcinoma, retinoblastoma and neuroblastoma.

Lung cancer (cancer of the lung) is one of the most common cancers all over the world that caused 15,90,000 deaths in 2012 (19% of the total cancers).^[7] Lung cancers found in 8 out of 10 cases were in people over the age of 60, usually the smokers. In India, the number of new cases increased from around 65,000 to 90,000 between 2009 & 2013, registering 15-20% increase annually. It is possible to reduce only when it is diagnosed and treated early. As of now lung cancer is diagnosed with chest X-ray, which is a simple and quick test that shows changes like abnormal shadowing. However, this cannot confirm the cancer. Use of computerized tomography (CT) scan helps to confirm the existence of lung cancer. Early analysis of cancer is essential for treatment. Analysis of cancer genome has already led to new targets for cancer therapy. In addition genome wide analysis has given new insights into specific genetic mutations and clinical response, as well as new approaches useful for diagnosis and prognosis. Methods of cancer treatment include surgery, chemotherapy and radiotherapy. The treatment is based on the site of primary tumor, type of cancer, stage of the cancer and general health. About 80% of patients died due to poor diagnosis and inadequate treatment

strategies. Hence, there is an urgent need to develop diagnostic tools/methods and new medicines to treat cancer.

AIM

This study was aimed to find out a new approach for the detection of susceptible gene(s) for cancer using tensor algorithm on large-scale genome level expression of human lung cancer cell gene(s).

MATERIAL AND METHODS

(i) Collection of data

Development of innovative technologies leads to accumulation of huge amount of biological data (raw) and they are free to access worldwide through internet. Since it is a crude data, it is hard to understand the hidden facts. Hence, an attempt was made to annotate the gene expression data by making use of computational methods particularly 3D Tensor. To initiate the study microarray expression data was collected from Gene Expression Omnibus (GEO) database. Since this study is designed to work on human lung cancer, gene expression of human normal lung cell (NLC), human lung cancer cell (LCC), and lung cancer cell treated with Mgd (LCCTD) were collected. Normal lung cell acted as the control.

In this study two kinds of data were used, i) Standard data - defined as manually annotated data that are known/characterized genes that can causes cancer both in suppression or induction pathway. ii) Raw data are (Table.1) un-annotated data. Standard data contains a total of 820 genes (Supp. Table.1) that are collected manually from the literature and databases. Standard genes were used to check the algorithm designed. Raw data is crude and un-annotated data downloaded from GEO database.

Table: 1 Dataset used in the study.

Description	NLC	LCC	CTD
Dataset	GSE18454	GDS1204	GDS1204
Subset ID	GSM459746	GSM39816	GSM39813
	GSM459747	GSM39817	GSM39814
	GSM459748	GSM39818	GSM39815
Microarray Count	54613	22283	22283
Microarray Platform	GPL570	GPL96	GPL96

(ii) Pre-processing of data

SOFT files were downloaded from GEO database, which are delimited file and are hard to readout. In order to make it more understandable Structured Query Language tool (My SQL) was used to remove entries that are not common in NLC, LCC and LCCTD. In addition, the null values were also removed in NLC/LCC/LCCTD datasets. Since, the designed algorithm accepts input only in MS Excel format, all processed gene datasets were converted in to Excel (Supp. Table. 2).

(iii) Software design

Tensor is a highly complex algorithm/concept (Fig. 1). It is defined as matrices or arrays that represent a physical quantity mathematically. It is also referred as multi-dimensional array of numbers; also called an n-way or n-mode array. For example, $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ \mathcal{A} is a third-order tensor. To implement this multi dimensional concept Hypertext Preprocessor (PHP) language is used. PHP is a server-side scripting language designed for web development. It is also used as a general-purpose programming language. To implement 3D Tensor the algorithm is divided in to two portions i) Data Input (processing) & ii) Data Calculation.

PHP is a user-friendly computer language, and hence it is used as frontend. In order to store and retrieve the data My SQL is used as backend. My SQL can be administered with php My Admin (MAMP).

RESULTS

Three datasets were downloaded from GEO database. They are i) GSE18454 – normal lung cell (NLC), ii) GDS1204 – lung cancer cell (LCC), and iii) GDS1204 – lung cancer cell treated with Mgd (LCCTD) (Table. 1). Normal lung cell (NLC) gene expression profile consists of 54,613 spots; whereas as the gene expression of lung cancer cell (LCC) and lung cancer cell treated with Mgd (LCCTD) consist of approximately 22,000 spots.

All these data were preprocessed carefully. After processing, it was observed that only 6,462 genes are common and are valid out of 22,000 genes. Hence, these 6,462 genes were considered for further study and served as input data for algorithm. The algorithm was designed carefully so that it can handle huge volume of data in less period of time. The advantage of this algorithm is that it takes less than a minute to process 6,462 genes. The designed algorithm was checked with standard/training dataset. Fig.1 shows the method of tensor implementation, where Php My Admin is used to connect frontend (PHP) and backend (MySQL). The designed algorithm classifies the genes in different groups (Fig. 2).

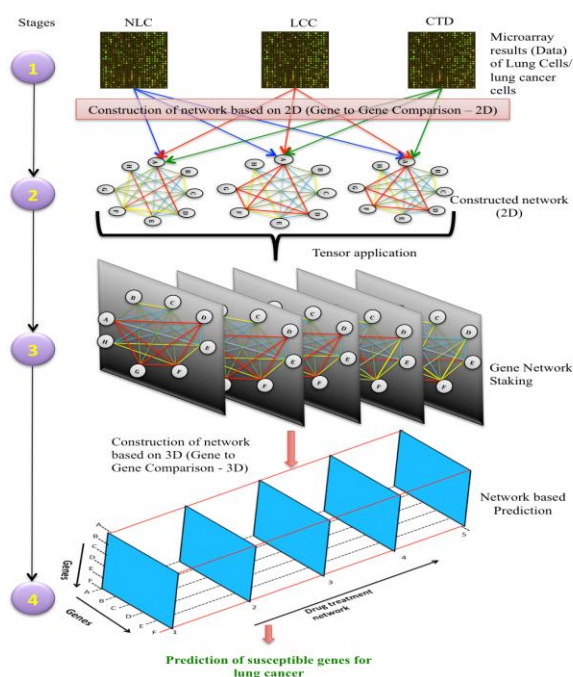
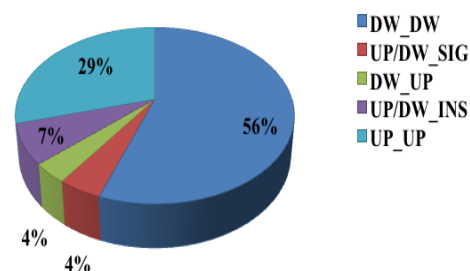


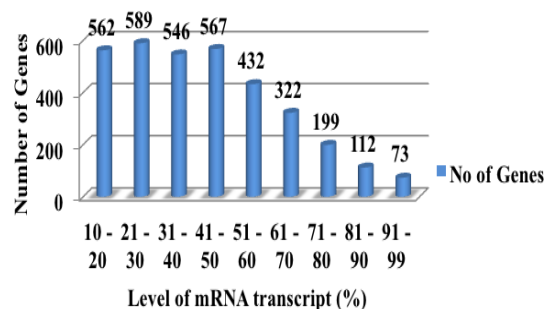
Fig. 1 Tensor & its role in prediction of susceptible genes for cancer.

Fig. 2 Differential expression of genes in lung cancer cell (before and after treatment)



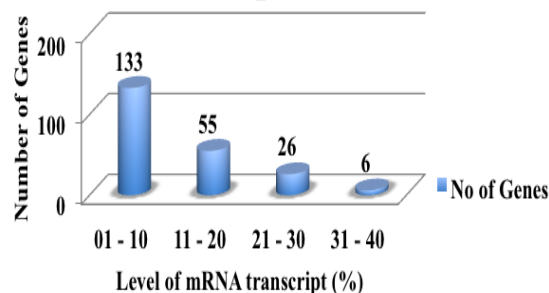
Results revealed that many genes are down-regulated in lung cancer. However, 56% (3402 genes out of 22,000) of genes are further down-regulated after treatment (Fig. 3 & Table. 2).

Fig. 3 Differential expression of Downregulated Genes (DW_DW)

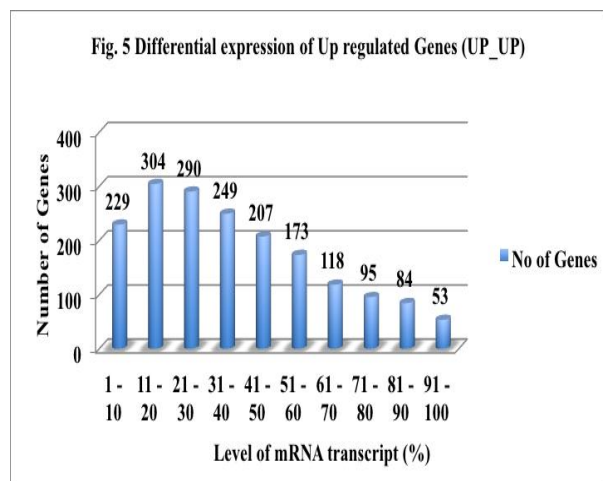


Differential expression of down-regulated genes ranges between 10 to 99%. Maximum number of genes found to be differentially expressed and ranges between 10 to 50%. Whereas only 73 genes are down-regulated in 90 to 100% range. In addition 4% of genes where up-regulated after treatment (Fig. 4), that covers 220 genes. Differential expressions were observed in 1 to 40% of genes, and the expression was insignificant. 10% of differential expressions were observed for maximum number of genes (133).

Fig. 4 Differential expression of up regulated Genes (DW_UP)



In general many genes are up-regulated in cancer. In this study it is found that 29% (1802 genes) are further up-regulated after Mgd treatment (Fig. 5 & Table. 3). Differential expression ranges between 1 to 100%. However, maximum numbers of genes are up-regulated in 1 to 50% range.



7% (448 out of 7,000 genes) of genes were found to be insignificant for Mgd treatment. Differential expressions of these genes were able to classify by comparing with normal lung cell, lung cancer cell and lung cell treated with Mgd.

It was found that only 4% (226 genes) of genes that are

positively responded to Mgd treatment (Fig. 6 & Table. 4). Interestingly, 90% to 100% of gene expression was recovered back. Of which 26 genes were completely restored.

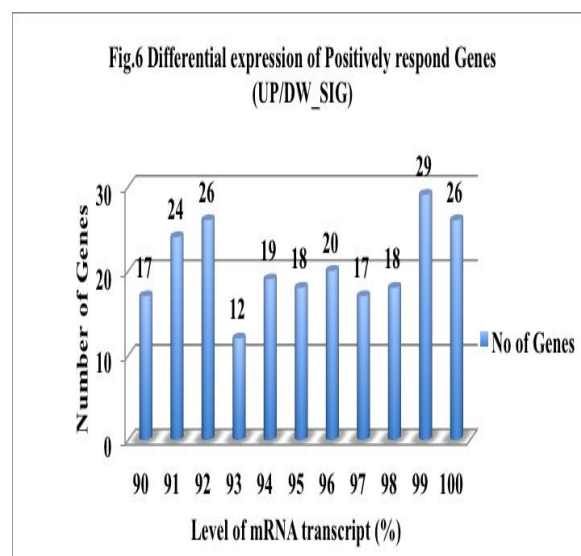


Table. 2 List of genes that are down regulated before and after drug treatment (DW_DW: n=3402).

S. No	Id_ref (Spot ID)	Identifier (Gene Symbol)	Role of genes	Differential expression (%)
1	202286_s_at	TACSTD2	Tumor-associated calcium signal transducer 2	99%
2	203256_at	CDH3	Cadherin 3, type 1, P-cadherin (placental)	98%
3	203571_s_at	ADIRF	Adipogenesis regulatory factor	97%
4	200953_s_at	CCND2	Cyclin D2	97%
5	201694_s_at	EGR1	Early growth response 1	95%
6	201288_at	ARHGDI1B	Rho GDP dissociation inhibitor (GDI) beta	95%
7	205240_at	GPSM2	G-protein signaling modulator 2	94%
8	204475_at	MMP1	Matrix metalloproteinase 1	94%
9	202769_at	CCNG2	Cyclin G2	93%
10	201842_s_at	EFEMP1	EGF containing fibulin-like extracellular matrix protein 1	92%
11	202870_s_at	CDC20	Cell division cycle 20	92%
12	203362_s_at	MAD2L1	MAD2 mitotic arrest deficient-like 1 (yeast)	91%
13	205239_at	AREG	Amphiregulin	91%
14	201983_s_at	EGFR	Epidermal growth factor receptor	90%
15	203132_at	RB1	Retinoblastoma 1	90%
16	200952_s_at	CCND2	Cyclin D2	89%
17	202770_s_at	CCNG2	Cyclin G2	88%
18	202823_at	TCEB1	Transcription elongation factor B (SIII), polypeptide 1 (15kDa, elongin C)	86%
19	202177_at	GAS6	Growth arrest-specific 6	83%
20	203418_at	CCNA2	cyclin A2	82%

Table. 3 Genes that has up regulated both in cancer and treated (UP_UP: n=1802).

S. No	Id_ref (Spot ID)	Identifier (Gene Symbol)	Role of genes	Differential expression (%)
1	204927_at	RASSF7	Ras association (RalGDS/AF-6) domain family (N-terminal) member 7	13%
2	205932_s_at	MSX1	Msh homeobox 1	19%
3	202586_at	POLR2L	Polymerase (RNA) II (DNA directed) polypeptide L, 7.6kDa	17%
4	205396_at	SMAD3	SMAD family member 3	16%
5	206706_at	NTF3	Neurotrophin 3	16%
6	203692_s_at	E2F3	E2F transcription factor 3	15%
7	203931_s_at	MRPL12	Mitochondrial ribosomal protein L12	11%
8	203171_s_at	RRP8	Ribosomal RNA processing 8, methyltransferase, homolog	9%
9	202979_s_at	CREBZF	CREB/ATF bZIP transcription factor	9%
10	202996_at	POLD4	Polymerase (DNA-directed), delta 4, accessory subunit	7%
11	205085_at	ORC1	Origin recognition complex, subunit 1	7%
12	201067_at	PSMC2	Proteasome (prosome, macropain) 26S subunit, ATPase, 2	5%

Table. 4 List of genes that are significant for drug treatment (UP/DW_SIG: n=265)

S. No	Id_ref (Spot ID)	Identifier (Gene Symbol)	Role of genes	Differential Expression (%)
1	200644_at	MARCKSL1	MARCKS-like 1	100%
2	200800_s_at	HSPA1B	Heat shock 70kDa protein 1B	100%
3	200814_at	PSME1	Proteasome (prosome, macropain) activator subunit 1	100%
4	201640_x_at	CLPTM1	Cleft lip and palate associated transmembrane protein 1	100%
5	203168_at	ATF6B	Activating transcription factor 6 beta	100%
6	204473_s_at	ZNF592	Zinc finger protein 592	100%
7	206813_at	CTF1	Cardiotrophin 1	100%
8	201159_s_at	NMT1	N-myristoyltransferase 1	99%
9	202889_x_at	MAP7	Microtubule-associated protein 7	99%
10	203589_s_at	TFDP2	Transcription factor Dp-2	99%
11	205296_at	RBL1	Retinoblastoma-like 1	99%
12	205585_at	ETV6	Ets variant 6	99%
13	202136_at	ZMYND11	Zinc finger, MYND-type containing 11	98%
14	203004_s_at	MEF2D	Myocyte enhancer factor 2D	98%
15	205249_at	EGR2	Early growth response 2	98%
16	205635_at	KALRN	Kalirin, RhoGEF kinase	98%
17	206928_at	ZNF124	Zinc finger protein 124	98%
18	203050_at	TP53BP1	Tumor protein p53 binding protein 1	95%
19	203737_s_at	PPRC1	Peroxisome proliferator-activated receptor gamma, coactivator-related 1	95%
20	204973_at	GJB1	Gap junction protein, beta 1, 32kDa	95%
21	205702_at	PHTF1	Putative homeodomain transcription factor 1	95%
22	204475_at	MMP1	Matrix metalloproteinase 1	94%
23	202227_s_at	BRD8	Bromodomain containing 8	93%
24	205250_s_at	CEP290	Centrosomal protein 290kDa	92%
25	200881_s_at	DNAJA1	DnaJ (Hsp40) homolog, subfamily A, member 1	91%
26	205861_at	SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	91%
27	206654_s_at	POLR3G	Polymerase (RNA) III (DNA directed) polypeptide G	91%
28	200719_at	SKP1	S-phase kinase-associated protein 1	90%
29	206238_s_at	YAF2	YY1 associated factor 2	90%

DISCUSSION

Analysis of genomic data being a taunting process, annotation is much needed to understand hidden facts.^[8]

Genome wide integration study indirectly helps to analyse and predict protein-protein integration in large scale analysis.^[9] Based on that an attempt was made in

this study to analyse GWI on lung cancer genome and observed some interesting facts.

In cancer differential (up/down) expression of genes is observed. This differential expression helps us to classify cancer in to different types^[10] and also used to characterize each type of cancers. Hence, an attempt was made in this study to find out differential expression of genes in lung cancer cells. Using tensor algorithm, lung cancer gene(s) are characterized in to different groups like down-regulated/up-regulated/insignificant/significant after Mgd treatment (Fig.2).

Results revealed that several gene(s) expressions are down-regulated in A549 lung cancer cells. Some of them are MMP1, ADIRF, CCND2, EGR1 (Table.2). 1) Matrix metalloproteinase-1, 2) Adipogenesis regulatory factor, 3) Cyclin D2 and 4) Early growth response 1 etc., Similarly, other researchers also reported the down-regulation of these gene(s) expression and it is highly correlated with cancer prognosis i.e. Matrix metalloproteinase in endometrial and breast cancer^[11,12]; Adipogenesis regulatory factor in cellular growth and differentiation^[13]; Cyclin D2 and Early growth response 1 are the genes involved in cell cycle regulation.^[14] Lawrence^[15] evidenced that these genes has major impact in lung cancer prognosis. This study found that Mgd treatment down-regulated (10 – 90%) these genes and the expression is lower than that of the normal (control), hence the effect of drug is not significant. In addition Mgd treatment also trigger (10 – 20%) the expression of certain genes greater than that of normal lung cell (MAP2K5, BCL9 and E2F4 indicating that Mgd acts differently in different genes and its action is not significant.

In this study several genes expressions are up-regulated in A549 lung cancer cells (SMAD3, CREBZF, RASSF7, PSMC2 etc.,) Table.3. Particularly 1) SMAD family member III, 2) Ras association (RalGDS/AF-6) domain family (N-terminal) member VII, 3) Proteasome (prosome, macropain) 26S subunit ATPase II, 4) CREB/ATF bZIP transcription factor etc., Other researchers also evidenced the overexpression of these genes, SMAD^[16], Ras^[17], Proteasome 26S subunit ATPase II^[18] and CREB/ATF bZIP^[19] and considered as cancer target for drug designing. However, Mgd treatment further up-regulated these genes that are already up-regulated in cancer condition. This result indicates that Mgd treatment targets these genes directly or indirectly, but its action is not significant. It indicates that, the use of Mgd for cancer treatment need to worked further.

However, Mgd treatment restores 4% of total genes that are either down-regulated/up-regulated. Particularly, 1) TP53BP1, 2) RBL1, 3) SKP1 etc., (Table.4) are down-regulated/up-regulated in cancer condition and their expressions were reverted back and equivalent to normal

cells. Interestingly, all these gene(s) expressions were recovered to 90% - 100%. Similar results were observed by Kiyohara^[20], he showed 80% recovery of mutant TP53BP1 in lung cancer, Modi *et. al.*,^[21] also reported the restoration of inactivated RBL1 that contributes much in lung cancer recovery. But in this study, Mgd treatment restores only 4% of genes and rest of them were deregulated. Based on this results it is inferred that Mgd treatment is not significant & it acts differently on different genes and hence further study is warranted.

CONCLUSION

This study made an attempt to characterize and predict genes that are involved in human lung cancer. It characterizes a list of genes that are up regulated, down regulated, significant and insignificant for drug treatment. It provides a list of differentially expressed genes against Mgd treatment and provides many hidden facts of cancer. Ultimately it provides a comprehensive and user-friendly platform (tool), where any user can analyse their genomic data and find out the gene(s) responsible for cancer.

REFERENCE

1. Alter, O. and G. H. Golub. "Integrative analysis of genome-scale data by using pseudoinverse projection predicts novel correlation between DNA replication and RNA transcription". Proceedings of the National Academy of Sciences of the United States of America, 2004; 101(47): 16577-16582.
2. Balmain, A., J. Gray and B. Ponder. "The genetics and genomics of cancer". Nature genetics, 2003; 33: 238-244.
3. Liu, P., Y. Wang, H. Vikis, A. Maciag, D. Wang, Y. Lu, Y. Liu and M. You. "Candidate lung tumor susceptibility genes identified through whole-genome association analyses in inbred mice". Nature genetics, 2006; 38(8): 888-895.
4. Schadt, E. E. "Molecular networks as sensors and drivers of common human diseases". Nature, 2009; 461(7261): 218-223.
5. Wang, Y., X. S. Zhang and L. Chen. "Modelling biological systems from molecules to dynamical networks". BMC Systems Biology, 2012; 6(Suppl 1): S1.
6. Nikolsky, Y., T. Nikolskaya and A. Bugrim. "Biological networks and analysis of experimental data in drug discovery". Drug Discovery Today, 2005; 10(9): 653-662.
7. Noronha, V., et al. "Epidemiology of lung cancer in India: Focus on the differences between non-smokers and smokers: A single-centre experience." *Indian journal of cancer*, 2012; 49.1: 74.
8. Johnson, David S., et al. "Genome-wide mapping of in vivo protein-DNA interactions." *Science*, 2007; 316.5830: 1497-1502.
9. Utleg, Angelita G., et al. "Proteomic analysis of human prostasomes." *The Prostate*, 2003; 56.2: 150-161.

10. Golub, Todd R., et al. "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring." *Science*, 1999; 286.5439: 531-537.
11. Lin, Chiao-Yun, et al. "A dual tyrosine kinase inhibitor lapatinib suppresses overexpression of matrix metalloproteinase 1 (MMP1) in endometrial cancer." *Journal of Molecular Medicine*, 2014; 92.9: 969-981.
12. Henckels, Eric, and Ron Prywes. "Fra-1 regulation of Matrix Metalloproteinase-1 (MMP-1) in metastatic variants of MDA-MB-231 breast cancer cells." *F1000 Research* 2 (2013).
13. Rosen, Evan D., and Bruce M. Spiegelman. "Molecular regulation of adipogenesis." *Annual review of cell and developmental biology*, 2000; 16.1: 145-171.
14. Zhu, Wuqiang, et al. "Cyclin D2-mediated cardiomyocyte cell cycle activity reverses doxorubicin-induced cardiotoxicity." *Cancer Research* 73.8 Supplement, 2013; 5602-5602.
15. Lawrence, Michael S., et al. "Discovery and saturation analysis of cancer genes across 21 tumour types." *Nature*, 2014; 505.7484: 495-501.
16. Li, Na, et al. "Altered β 1, 6-GlcNAc branched N-glycans impair TGF- β -mediated Epithelial-to-Mesenchymal Transition through Smad signalling pathway in human lung cancer." *Journal of cellular and molecular medicine*, 2014; 18.10: 1975-1991.
17. van der Drift, Miep A., et al. "Diagnosing peripheral lung cancer: the additional value of the Ras-association domain family 1A gene methylation and Kirsten rat sarcoma 2 viral oncogene homolog mutation analyses in washings in nondiagnostic bronchoscopy." *CHEST Journal*, 2012; 141.1: 169-175.
18. Millward, Michael, et al. "Phase 1 clinical trial of the novel proteasome inhibitor marizomib with the histone deacetylase inhibitor vorinostat in patients with melanoma, pancreatic and lung cancer based on in vitro assessments of the combination." *Investigational new drugs*, 2012; 30.6: 2303-2317.
19. Park, Sin-Aye, et al. "GSK-3 is a novel target of CREB and GSK-3-CREB signaling participates in cell survival in lung cancer." *Cancer Research*, 2015; 75.15 Supplement: 2094-2094.
20. Kiyohara, Chikako, et al. "Cigarette smoking, TP53 Arg72Pro, TP53BP1 Asp353Glu and the risk of lung cancer in a Japanese population." *Oncology reports*, 2010; 23.5: 1361-1368.
21. Modi, Sanjay, et al. "Protein expression of the RB-related gene family and SV40 large T antigen in mesothelioma and lung cancer." *Oncogene*, 2000; 19.40.