

# A Framework for Identifying of Gene to Gene Mutation causing Lung Cancer using SPI - Network

Praveen Tumuluru  
Research Scholar  
Department of CSE,  
GITAM University,  
Visakhapatnam, India

Bhramaramba Ravi, PhD  
Associate Professor,  
Department of IT,  
GITAM University,  
Visakhapatnam, India

## ABSTRACT

Biomedical area of research is grown dynamical for identification of various diseases and prediction of disease, among the most cancer is vital and critical disease caused from various sources of gene mutation. Now a day's Lung cancer has been malignancy and primary related cause of cancer deaths mostly worldwide. Our work is based on the observation of Inimitable mutations frequency patterns of cancer are found in lung subtypes, or in tumor, target drug that cause genetic mutations, characteristic cancer and transduction signal pathway is beneficial to the patients. Our work is based on Ampliseq Ion Torrent cancer sequence of 800 loci of 50 cancer Oncogenes and genes for identification of mutation related to genetic for 48 fixed-formalin taken from samples of Indian patients on human lung cancer disease. Genetic gene mutation of lung cancer is identified by using interaction network method.

Mutations occur frequent in PIK3CA, KRAS, TP53 AND EGFR genes, according to our observation from samples gained from lung cancer, one or more mutations occur in the gene is the basic key for finding the cancer disease. Our study narrates or gives flexibility in identifying genetic mutation for lung cancer individually through tumors.

## Keywords

lung cancer, genetic mutations, targeted sequencing, targeted therapy, algorithms, pathways, statistical significance, cancer genomes, SPI Network-Shortest Path Interactive Network.

## 1. INTRODUCTION

Biomedical research in lung cancer is caused by genetic changes that occur spontaneously with response to chewing and smoking of tobacco and pollution of bad air. Due to development of genetic underlying lung cancer disease is progressive. Smoking of cigarette may increase the risk in lung cancer development, from research it has been observed that nearly 30% of lung cancer is attacked by smoking [1].

These diseases are commonly and casual exhibited at high risk with symptoms of unknown, according to the various studies this disease is diagnosed only in advanced stage. The early diagnosis of this disease and treatment for that is not observed before for better treatment. Continuous effort is made by the researchers for identification of lung cancer at early stages for treatment, It is observed that poor diagnosis and prognosis poor has caused a failure for early detection and treatment for cancer disease. This disease has upmost rate worldwide and an overall estimated of 1.9 million cases and 1.7 million people deaths in 2013. As observed early, it was notified that nearly 25% of people deaths in India alone, among all the cancer disease, lung cancer is the leading in both women and men.

## 1.1 NSCL (Small lung cells) and LCSC(Lung cancer Small cells)

Mostly in lung cancer patient 88% of cells are NSCL, of which is categorized into 3 subtypes – large cell lung cancer, adenocarcinoma and squamous-cell carcinoma. In tobacco chewers it was observed and proved that they are associated strongly with NSCL AND LCSC [2]. From the association it was observed that nearly 65 mutagens are capable of combining and bonding chemically which modifies the DNA. The changes observed leads the change in the mutational pattern characteristics which are observed in lung cancer person with that of non lung cancer patient. As of mutation points in TP53 and KRAS patterns was observed a dynamic change with a patient history of smoking with that of non-smoking patient.

Various observations have shown that smokers have highest rate of genetic mutation occurrence compared to the non-smokers, but non-smoker genetic and heredity changes can affect the DNA structure can also caused lung cancer. Since non-smoking versus smoking patients of lung cancer have discrete patterns in mutations which has to be treated with different drug based on group [3].

The environmental factors and various factor of genetic consist for a lung cancer disease widely, and also mutation of gene profile in tumor and lung cancer is unique. As effectively observed that EGFR mutation can caused lung cancer in non-smoking women as well in advance of NSCL patients. Various methods has been used to detect mutation gene in identification lung cancer in patients. The basic and advanced clinical methods used for detect are fluorescence in situ hybridization (fish), immunohisto-chemistry (IHC) and direct polymerase chain reaction (PCR). These methods does not follow standards in diagnosis, each of them has certain pros and cons.

As and now sequencing of sanger and NGS is the popular and most used sequence in identifying genome cancer in tumors, To identify the tumors instruments are the easiest and costly method which has been used widely and practically worldwide. The next and last methods used were gene sequence identification methods and post light sequence technology methods for identification of tumor cells in gene pattern. These two methods require many algorithms for analysis and prediction of cancer cells in lung cancer based on sequence and frequency of the genome cancer [4].

Abnormal growth of cells occurred by proliferation of cells which are influence from tumor gene. In the world tumors are the second killing disease after cardiovascular disease. Researchers have defined tumor tissues in alteration with various tumor cells. Lung cancer is a subtype of Lung

adenocarcinoma and is related highly with a genetic specified background. In our study, progression of lung cancer is caused with mutation of gene and copy ov variations number in particular Oncogenes. Our observations have shown that KRAS and EGFR in a specific mutation can cause a lung cancer progression in adenocarcinoma [5,6]. Alike that , it is also observed that the alteration of genomic may contribute a tumor in lung, which leads to lung cancer and can multiple at various levels of impactions in original genomic alterations.

Chosen database related to TCGA, our study is based on various level to reach in identifying the lung cancer in patients of non-smoker and smoker using shortest path network interactions method. Our study is based on identification of lung adenocarcinoma functional level of 4 types based on RNA expression, Gene expression, mutation of somatic and DNA methylation of lung tissues, normal sample control of Cancer Genome [7].

A shortest path search processing at different levels of connectivity in a constructed protein-protein, this is used to identify the candidate cancer gene by this paths. After identifying the gene, these genes tested using permutation test for filtering, which is used for better selection of lung cancer candidate. The sample procedure is repeated in the network using shortest path network interaction algorithm [8].

The output of candidate gene identified has been differentiated into 2 levels. One level of gene of identification is number of times a mutated cancer gene identified in a level , another level of lung cancer gene identification is done , how many times a mutation cancer gene occurred in various level and how many times it has occurred. And used gene expression from pubmed [9,10], DNA methylation, Micro RNA expression, mutation somatic data of lung tissues and various patients data under category of smoker and non-smoker samples of TCGA. The level of Expression of 20,533 genes in 228 tumor cells in lung samples 44 and also normal cell samples were taken with RNA sequence. The samples are transformed into scale of log2.

## 1.2 Background

Many research studies have been done early using interactive network for disease identification in gene to gene or protien to protein interaction, Many of them have focused on . Jeong(2001), has suggest that protein mutation cause lethality, which arehighly connected among themselves, where as gene to gene mutationis non-lethal. Naher(1999) developed an algorithm for combinatorial computing, which is used to analyze and store graph results of mutation datasets.

Sneppen (2012), has developed an algorithm which supports and increase the degree of smaller nodes, and also reduces the pair connectivity between the bigger degree nodes. Buet(2013), analysed a method which identifies interaction of protien to protien and also identified the characterized proteins. Mewes(2012),analyzed data from MIPS which has genetic interaction information.

## 2. PROCEDURE SEQUENCE – A FRAME FOR IDENTIFICATION OF LUNG CANCER

**Algorithm steps involved in identification of Lung Cancer causing Mutation**

1. Identification of various types of gene and genes which involve in mutation

CpG Gene, Micro RNA, Somatic Mutation Gene, MRna Gene

2. Search method which causes gene mutation
3. Six Genes are identified which lead to lung cancer disease
4. Performance of Permutation test between the mutation genes
5. Refined genes from permutation test
6. Selection of gene mutation are done based on routing of shortest path algorithm

Sequence of RNA – 33 normal samples with micro RNA sequence 180 are transformed into scale of log2. Methylation DNA level 485,576 of which 180 samples of CPG, 21 tumor samples were measured using human infinium. Mutation of somatic data of 15,200 gene and tumor 229 samples were used with MuTect24

### 2.1 Network construction

A network model was built using the information of PPI accessed from retrieval of Interacting Gene for protein to protein interaction. The file selected contains outsized PPIs over's types of mutation organisms [6].

The file contains human PPIs total of 2,425,234 was extracted from a file by using sequential selection process of STRING to STRING interaction from the source associated with

1. Context of Genomic
2. Experimental throughput
3. Expression of conserved
4. Knowledge of history

There are various measure used for identification of protein to protein association and are used to identify problem occur in protein to protein relation (nearly 40). Protein to protein interaction and disease identification is done with the measure of strength of interaction, score and ensemble.

Interaction of protein with larger score has strong form of association. The interactive network constructed has 2,425,315 PPIs of human, 20,768 are totally proteins, of which are nodes and are connected with edges if and only if the related proteins are interacted with another. Human PPI represented edge, which indicates interaction with other edges associate with weights. The score of interaction range is 148 and 999. The value of interaction maximum is 989, which uses shortest path algorithm for low weight assignment for a strong association between each other nodes. The node weights are associated with network as defined with 1000 minus interaction with the other corresponding proteins [11,13].

### 2.2 A Key for Searching Lung Cancer Disease

It has been observed that identical proteins in PPI which has identical function and likely to have same score of interaction and share similar property among each other. As from observed of PPI based on the score, it is observed that  $p_1$ ,  $p_2$ ,  $p_3$  .... Pr are connected in a network series according to the large score in a series of proteins from which  $p_{i+1}$  to  $p_i$  are comprised of lower scores. The mapping process is constructed with the corresponding nodes based on shortest path reconstruction algorithm of network. It is also observed that proteins are encoded with dysfunctional genes which may range from  $p_2$ .....  $p_{s-1}$ . These encoded dysfunctional genes are mapping according to the shortest path network

construction [12,14]. Later then dysfunctional gene are divide into 4 levels. Which corresponding to different sets of gene patterns.

### 2.3 Our Proposed Method.

- a. Chosen groups of dysfunctional gene at various levels
- b. Search method uses for identification of candidate in an interactive network. The node indicates gene of dysfunctional at various level, and the archs indicted in dash show the shortest path between a&& i, e && h will be selected.
- c. Gene candidates used for search method
- d. Filtering using permutation test.

The proposed framework show how mutation causing gene induces lung cancer. Sample consists of 50 patients

### 2.4 Patients of Lung Cancer:

Selected nearly 50 patients of lung cancer with an average age of patients is nearly 68 yrs.

Lung Cancer patients have been divided into 3 subtypes of pathologic . n=22 for AC ,n=22 for SCC, and other n=4.

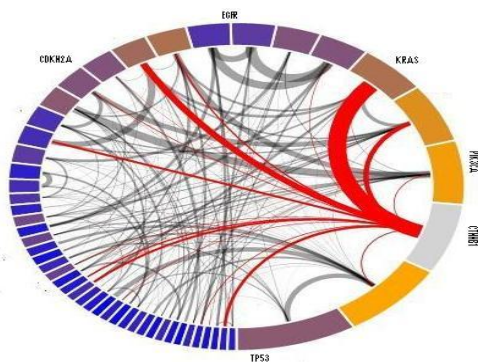


Figure 1. Shows a Shortest path between gene to gene mutation

Chosen 6 genes which are involve in lung cancer and tumors. Interaction between genes to gene is Applying a shortest path algorithm between genes and associated genes are applying , the above figure 1 shows shortest path interaction diagram between gene to gene. It has been observed that gene with highest betweenness ratio has maximum shortest connectivity path between genes and set of gene with a particular gene. And assumed a threshold value of 0.01 between the ratio of selected with a particular source gene.

$$M(g) = \min\{\max\{Q(g, x): x \in G_j\}, \max\{Q(g, x): x \in G_j\}\} \dots \dots \dots \text{Equ... (1)}$$

According to the rule of shortest path , PPIs and scores of interaction is to be built first, then two or more PPIs with similar in scores are likely to have similar function. Let us assume a group of genes with similar score compared to another with same score are likely to be bounded and will form an association. And prepare a max-min score of interaction using interaction network graph. Which is developed from Xgraph.

Likewise, a gene with property of cancer, having a strong score and related to another group of gene with same score have an identical property of cancer disease occurrence. It is

observed that almost 50% are non-smoker of which are females (55.5%) and nearly 86.4 are heavy smokers.

Table 1. Clinical trail of 50 Lung cancer patients

CHARACTERISTIC		AC n (% OF 22 SAMPLES)	SCC n (% OF 22 SAMPLES)	OTHER n (% OF 4 SAMPLES)
Age (years)		60.4 ± 9.0	64.8 ± 8.1	63.8 ± 8.3
Sex	Male	10 (45.5%)	19 (86.4%)	3 (75.0%)
	Female	12 (54.5%)	3 (13.6%)	1 (25.0%)
TNM stage	1a	6 (27.3%)	2 (9.1%)	0 (0.0%)
	1b	4 (18.2%)	3 (13.6%)	0 (0.0%)
	2a	8 (36.4%)	3 (13.6%)	1 (25.0%)
	2b	1 (4.5%)	3 (13.6%)	0 (0.0%)
	3a	3 (13.6%)	8 (36.4%)	2 (50.0%)
	3b	0 (0.0%)	3 (13.6%)	0 (0.0%)
	ND	0 (0.0%)	0 (0.0%)	1 (25.0%)
Differentiation	Low	7 (31.8%)	17 (77.3%)	2 (50.0%)
	Middle	10 (45.5%)	5 (22.7%)	0 (0.0%)
	Middle	4 (18.2%)	0 (0.0%)	0 (0.0%)
	ND	1 (4.5%)	0 (0.0%)	2 (50.0%)
Smoking history	Never	16 (72.7%)	3 (13.6%)	1 (25.0%)
	Light	4 (18.2%)	5 (22.7%)	0 (0.0%)
	Heavy	2 (9.1%)	14 (63.6%)	3 (75.0%)

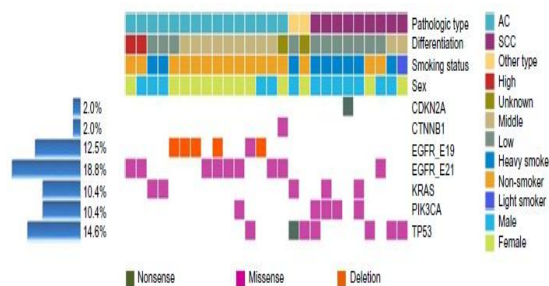


Figure 2. Shows the Summary Organization Of Lung Cancer Gene With Various Mutation Occurrence

Mutation Gene of Sub-type in Lung Cancer-In our study , and have screened 50 genes, out of which, have identified 36 mutations in TP53, PIK3CA,EGFR,KRAS,CTNNB1 and CDKN2A, there were identified in 28 out of 50 as shown in figure 2 and tables 2 and 3. The given sum of 16(68.3%) sample AC posses a least 1 mutation, 14(87.7%) AC mutations are identified in non-smokers . The P and OR values (0.116 , 0.055) in sample SSC posses 1 mutation , (0.292 ; 0.545) from non-smokers history.

Rather than this 8 of 50 samples contain nearly (15%) occurrence of mutations of combination with two genes shown in table 4. The combination mutations is found only in Sample AC of females, in which 3 samples were identified with EGFR mutation, or PIK3CA and CTNNB1 or TP53, and other 3 samples of SCC with PIK3CA or either TP53 or KRAS mutations. Mutation TP53 - various types of cancers are also developed with the changes that occur in genetic mutation through pathway signaling and hinder functions like cell growth, proliferation, survival and apoptosis. And observed from our study that disruption bath is generated which cause

suspect of tumor in various types of mutation occurrence between SCC and AC pathway for EGFR, suppressor pathway for CDKN2A and TP53 as shown in figure 3.

The mutated genes which are study currently have been classified previously as mutation drivers, these drivers act as promoters in identifying tumor growth of cells in advance in identifying the cancer and cancer type.

**Table 2. Show the Mutation Frequency of Lung cancer 50 patients based on history**

GENE	TOTAL NO OF SAMPLES WITH MUTATIONS (50 SAMPLES)	SEX		PATHOLOGIC TYPE			NEVER SMOKER SAMPLE WITH MUTATION IN 20 SAMPLES	SMOKING HISTORY	HEAVY SMOKER SAMPLE WITH MUTATION (% IN 11 SAMPLES)
		MALE	FEMALE	AC SAMPLE WITH MUTATION (% 22 SAMPLES)	SCC SAMPLE WITH MUTATION (% 22 SAMPLES)	OTHERS SAMPLES WITH MUTATION (IN 6 SAMPLES)			
CDKN2A	1(2.1%)	1(3.1%)	0(0.0%)	0(0.0%)	1(4.5%)	0(0.0%)	0(0.0%)	0(0.1%)	1(5.3%)
CTNNB1	1(2.1%)	0(0.0%)	1(6.3%)	1(4.5%)	0(0.0%)	0(0.0%)	1(100.0%)	0(0.0%)	0(0.0%)
EGFR	14(29.2%)	4(12.5%)	10(62.5%)	13(59.1%)	1(4.5%)	0(0.0%)	14(70.0%)	0(0.0%)	0(0.0%)
KRAS	5(10.4%)	5(15.6%)	0(0.0%)	2(9.1%)	2(9.1%)	1(25.0%)	0(0.05)	3(33.0%)	3(15.8%)
PIK3CA	5(10.4%)	4(12.5%)	1(6.3%)	1(4.5%)	4(18.1%)	0(0.0%)	1(5.3%)	2(22.0%)	2(10.5%)
TP53	7(14.6%)	3(9.4%)	4(25.0%)	1(4.5%)	4(18.1%)	2(50.0%)	3(15.8%)	1(11.1%)	3(15.8%)

Various driver tends to differ from patient and type of cancer, some may have one type and some may have many types of driver mutations which tentative to occur in oncogene, as and then driver mutations with tumor combination occurs rarely in oncogene.

The observation in our study 73% has 1 mutation in oncogene (KRAS,EGFR,CTNNB1 or PIK3CA) as shown in Table 3, 38% of sample has more than 2 or at least 2 mutation in Oncogenes as shown TP53(table 4).

The table 3, shows the various type of mutation causing gene and their relationship with the patient is shown based on low , middle , unknown and high , with respect to smoking history and non smoking history. An interaction diagram show in figure 3, show how mutation are occurred in gene to gene interaction and their status is represented in color pallets.

Mutations EGFR – it is the common alternations found in genetic mutation in EGFR of NSCLSCS , 35 % of lungs in ACs posses mutations of EGFR and is most common in never smokers. But from our study nearly 39-40 mutations of EGFR are common in SCC and is found to occur tumor with this mutations table 4. According to our analysis , 14 samples of AC , 5 samples of SCC were found with EGFR mutations causing tumor. It is been observed that AC (0.0002 , 31.3) rate of occurrence of lung cancer or tumor are found in non-smokers of AC sample patients.

As studied EGFR mutation in gene can increase or downstream the occurrence of angiogenesis, proliferation and metastatic in patients samples on smoker and decrease in non-smoker. Mutation TP53 – it is the basic genetic alternative identified in various lung cancers nearly 45% in SCCs and 40% in AC are found in genes.

**Table 3: Single mutation of lung cancer patient**

GENE	MUTATION	AGE	SEX	PATHOLOGIC TYPE	DIFFERENTIATION	TNM STAGE	SMOKING HISTORY
CDKN2A	E69*	62	M	SCC	Low	2a	Heavy
EGFR	L861Q	49	M	AC	High	1a	Never
EGFR	E746_A750del	48	M	AC	Middle	2a	Never
EGFR	E746_A750del	71	F	AC	Middle	2b	Never
EGFR	L858R	67	M	AC	Middle	2a	Never
EGFR	L858R	73	M	SCC	Low	3b	Never
EGFR	E746_A750del	65	F	AC	Middle	1a	Never
EGFR	L858R	66	F	AC	Middle	1a	Never
EGFR	L858R	62	F	AC	High	2a	Never
EGFR	L747_P753>S	52	F	AC	Low	3a	Never
EGFR	L858R	50	F	AC	Middle	2a	Never
KRAS	G12C	53	M	AC	Low	2a	Light
KRAS	G12A	57	M	AC	Low	2a	Heavy
PIK3CA	E542K	66	M	SCC	Low	3b	Heavy
TP53	R158L	53	M	SCC	Middle	3a	Heavy
TP53	R249S	65	F	other	Unknown	ND	Never
TP53	V157F	68	F	SCC	Middle	1b	Light
TP53	G245V	78	F	SCC	Low	2a	Never

Our observation has shown that TP53 mutation were found in 8 samples 2 in AC, 4 in SCCs and 2 in other of lung cancer patients.

Here figure 4. Show the interaction diagram of 2 mutations gene X and Y in a gene group cause lung cancer is shown.



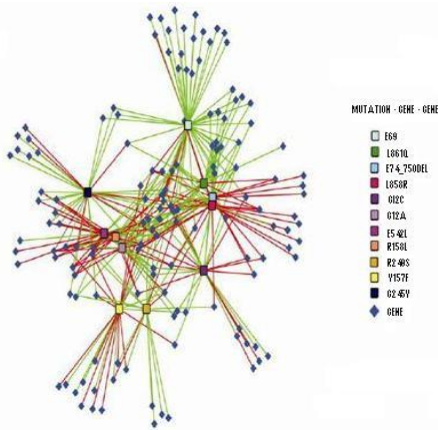


Figure 3. Shows a Gene Single gene mutation using shortest path algorithm

Table 4. Combination mutation of Lung cancer

GENE 1	MUTATION 1	GENE 2	MUTATION 2	GENE 3	MUTATION 3	AGE	SEX	PATHOLOGIC TYPE	DIFFERENTIATION	TNM	SMOKING HISTORY
CTNNB1	S33F	EGFR	L858R	-	-	72	F	AC	ND	1b	Never
EGFR	E746_A750del	EGFR	L858R	-	-	56	F	AC	Middle	3a	Never
PIK3CA	R88Q	KRAS	G12C	-	-	69	M	SCC	Low	3a	Light
PIK3CA	R88Q	KRAS	G12C	-	-	56	M	SCC	Low	3b	Light
PIK3CA	E545K	TP53	E285K	-	-	68	M	SCC	Low	2b	Heavy
PIK3CA	H1047L	EGFR	L858R	-	-	73	F	AC	Middle	3a	Never
TP53	R306*	KRAS	G12C	-	-	58	M	other	Low	3a	Heavy
EGFR	L747_A750>P	EGFR	A750P	TP53	R248W	59	F	AC	Middle	1b	Never

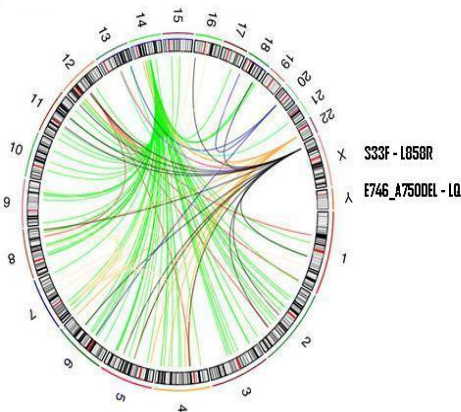


Figure 4 Shown a interactive network between two X and Y mutation in a Gene

It has been observed that TP53 mutation is found in tobacco associate sample of lung cancer patients, which increase the higher rate of cancer detection of 60-80% in tobacco user and 30-45% of non-smokers particularly [11]. From our datasets it has shown AC sample 30% and SCCs of 45% occurrence of lung cancer is identified with higher rate. At various stages mutations of TP53 can lead to abnormal growth of tumor cell in protein at higher rate and can caused to death immediately.

Mutation KRAS – This is a lung cancer causing mutation, and is normal identified in AC patients sample of 20% and SCC of 30%. Our study has shown that majority this mutation occurs in patients of smoking history in males compared to female. Mutation PIK3CA – This mutation is fast growing in gene which reduces the survival of patient of lung cancer, PIK3CA is a rear mutation which changes the structure of condon in nucleus and reduces the life sample of cell and causes the patient to sudden death. The sample data sets of nearly 30% patients with tabocca associate were detected with this. In TP53 , it has been observed ACs and SCCs mutations 8%

and 9% of CDKN2A, which was lesser compared to the actual percentage in smokers than non-smokers. CTNNB1 mutation is occurred in sample of AC, with a semi-occurring at EGFR.

Table 5. Multiple Mutations In Lung Cancer Patients

MUTATION	SUBSTITUTION	MUTATION TYPE	LUNG CANCER TYPE	SMOKING HISTORY
V157F	G>T	transversion	SCC	Light
R158L	G>T	transversion	SCC	Heavy
G245V	G>T	transversion	SCC	Never
R248W	C>T	transition	AC	Never
R249S	G>T	transversion	other	Never
E285K	G>A	transition	SCC	Heavy
R306*	C>T	transition	other	Heavy

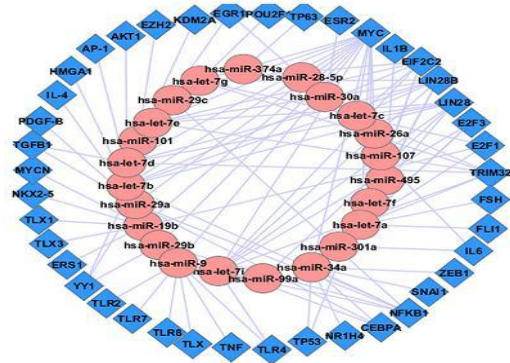


Figure 5. Shows multiple mutation occurs in Gene to Gene

Nearly 3% and 1% is notified in AC and SCC mutations in gene, which is very rare in occurrence of lung cancer at this point. At certain point of time, due to DNA structure change, CTNNB1 will encode a  $\beta$ -catenin, Protein which plays a critical path between intracellular protein growth can cause a growth of tumor. From table 5, one can observe that 2 or more mutation occurred in gene to gene interaction can cause cancer in patients. The short path interaction from figure 1 gives us that the number of scores obtained from association between group of gene of same behavior. From 50 patients datasets of gene and gene mutations datasets obtained from pub med, an interactive network diagram shows , how multiple mutation occurs in gene and their relationship are shown below.

### 3. CONCLUSION

Our study of work uses 800 Ampliseq sequence panel of cancer from 50 related gene of cancer, from which mainly are tumor and Oncogenes in 50 patient samples of lung cancer pathologic type. In our work, recognized occurrence of mutations frequent in PIK3CA, KRAS, TP53 and EGFR and low frequent mutations in CTNNB1 and CDKN2A, these pattern are observed in SCC samples and AC . our proposed algorithm for identification of lung cancer using shortest path algorithm gives a solution for detection of mutation gene occurs in cause of single mutation , two mutation and multiple mutations applied on 50 lung cancer patients and empirical results are shown based on the type of mutations occurred between them. Future work can be extended for identification of various types of cancer using the shortest path mutation algorithm more effectively.

### 4. ACKNOWLEDGEMENTS

The authors published this paper under their Ph.D work. The authors wish to thank the University grants commission(UGC) for extending financial support for this study, under the project — Development of a software tool to

identify Lung-cancer related genes using Protein Protein interaction Networks with sanctioned Proposal number: 137, F.NO:4-4/2014-

## 5. REFERENCES

- [1] Liu YY, Slotine JJ, Barabási AL (2013) Controllability of complex networks. *Nature* 473(7346):167–173.
- [2] Nacher J, Akutsu T (2012) Dominating scale-free networks with variable scaling exponent: Heterogeneous networks are not difficult to control. *New J Phys* 14(7):073005–073028.
- [3] Jeong H, Mason SP, Barabási AL, Oltvai ZN (2011) Lethality and centrality in protein networks. *Nature* 411(6833):41–42.
- [4] Yu H, et al. (2012) High-quality binary protein interaction map of the yeast interactome network. *Science* 322(5898):104–110.
- [5] Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011; 12: 56–68.
- [6] Rual JF, Venkatesan K, Hao T, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature.* 2005; 437: 1173–8.
- [7] Stelzl U, Worm U, Lalowski M, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell.* 2015; 122: 957–68.
- [8] R. Siegel, D. Naishadham, and A. Jemal, —Cancer statistics,||*A Cancer Journal for Clinicians*, vol. 62, pp. 10–29, 2012.
- [9] J. Trepel, M. Mollapour, G. Giaccone, and L. Neckers, —Targeting the dynamic HSP90 complex in cancer,||*Nature Reviews Cancer*, vol. 10, no. 8, pp. 537–549, 2010.
- [10] Y. P. Choi, B. G. Kim, M. Q. Gao, S. Kang, and N. H. Cho, —Targeting ILK and 4 integrin abrogates the invasive potential of ovarian cancer,||*Biochemical and Biophysical Research Communications*, vol. 427, no. 3, pp. 642–648, 2012.
- [11] L. P. Chen, Y. D. Lai, D. C. Li et al., — 4 is highly expressed in carcinogen-transformed human cells and primary human cancers,||*Oncogene*, vol. 30, no. 26, pp. 2943–2953, 2011
- [12] S. Sakashita, D. Li, N. Nashima et al., —Overexpression of immunoglobulin (CD79a) binding protein1 (IGBP-1) in small lung adenocarcinomas and its clinicopathological significance,||*Pathology International*, vol. 61, no. 3, pp. 130–137, 2011.
- [13] Bi-Qing Li, Jin You, Lei Chen, Jain Zhang..|| —Identification of Lung cancer related genes with shortest path approach in a protein-protein interactin network|| *BioMed Research International, Hindawi*, 1-8, 2013
- [14] D. Li, S. Sakashita, Y. Morishita et al., —Binding of lactoferrin to IGBP1 triggers apoptosis in a lung adenocarcinoma cell line,||*Anticancer Research*, vol. 31, no. 2, pp. 529–534, 2011.

## 6. AUTHORS PROFILE

**Mr.T.Praveen**, received the M.Tech degree in Computer Science and Engineering from Koneru Lakshmaiah College of Engineering, Acharya Nagarjuna University, in 2008. He is a research Scholar in GITAM University working in Data Mining Techniques for Bioinformatics Protein-Protein Interaction. He has been working as Assistant Professor, in the Department of Electronics & Computer Engineering, Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, since 2008.

**Dr. Bhramaramba Ravi**, presently working as Associate Professor in GITAM University. She has a total of 13 years of rich experience and 16 years of teaching. She received her Ph.D from Jawaharlal Nehru Technological University, in 2011 and MS degree in Software Systems from BITS, in 1999. She has published 14 papers in various National and International Journals/ Conferences. Her current research interests are in the areas of Data Mining Techniques for Bioinformatics.