Analysis of Protein Kinase Domain and Tyrosine Kinase Or serine/Threonine Kinase Signatures Involved In Lung Cancer

Lung cancer results when normal check and balance system of cell division is disrupted and ultimately the cells divide and proliferate in an uncontrollable manner forming a mass of cells in our body, known as tumor. Frequent mutations in Protein Kinase Domain alter the process of phosphorylation which results in abnormality in regulations of cell apoptosis and differentiation. Tyrosine Protein kinases and Serine/Threonine Protein Kinases are the two broad classes of protein kinases in accordance to their substrate specificity. The study of Tyrosine protein kinase and serine Kinase coding regions have the importance of sequence and structure determinants of cancer-causing mutations from mutation-dependent activation process. In the present study, we analyzed huge amounts of data extracted from various biological databases and NCBI. Out of the 534 proteins that may play a role in lung cancer, 71 proteins were selected that are likely to be actively involved in lung cancer. These proteins were evaluated by employing Multiple Sequence Alignment and a Phylogenetic tree was constructed using Neighbor-Joining Algorithm. From the constructed phylogenetic tree, protein kinase domain and motif study was performed. The results of this study revealed that the presence of Protein Kinase Domain and Tyrosine or Serine/Threonine Kinase signatures in some of the proteins are mutated, which play a dominant role in the pathogenesis of Lung Cancer and these may be addressed with the help of inhibitors to develop an efficient anticancer drugs. Furthermore, the present study contributes to the possibility that genetic components are more important in Lung Cancer as compared to environmental and smoking(carcinogens) factors.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.414v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

Analysis of Protein Kinase Domain and Tyrosine Kinase Or serine/Threonine Kinase Signatures Involved In Lung Cancer

Manas Ranjan Barik[1], Tripti Kaur Bagga[2], Anupam Bhattacharya[3], Bhagath Kumar Palaka[1], Viswanath Kasi Kotapati [1], Dinakara Rao Ampasala *[1]

1 .Centre for Bioinformatics, School of Life sciences, Pondicherry University, Pondicherry-605014, India

2. Department of Biotechnology, Amity University,

3. Bioinformatics, Jaivik Data Consulting, New Delhi, Delhi, India.

*Mail Correspondance:

Centre for Bioinformatics

School of Life Sciences

Pondicherry University

Puducherry-605014

Email: ampasaladr@bicpu.edu.in

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.414v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

ABSTRACT

Lung cancer results when normal check and balance system of cell division is disrupted and ultimately the cells divide and proliferate in an uncontrollable manner forming a mass of cells in our body, known as tumor. Frequent mutations in Protein Kinase Domain alter the process of phosphorylation which results in abnormality in regulations of cell apoptosis and differentiation. Tyrosine Protein kinases and Serine/Threonine Protein Kinases are the two broad classes of protein kinases in accordance to their substrate specificity. The study of Tyrosine protein kinase and serine Kinase coding regions have the importance of sequence and structure determinants of cancer-causing mutations from mutation-dependent activation process.

In the present study, we analyzed huge amounts of data extracted from various biological databases and NCBI. Out of the 534 proteins that may play a role in lung cancer, 71 proteins were selected that are likely to be actively involved in lung cancer. These proteins were evaluated by employing Multiple Sequence Alignment and a Phylogenetic tree was constructed using Neighbor-Joining Algorithm. From the constructed phylogenetic tree, protein kinase domain and motif study was performed. The results of this study revealed that the presence of Protein Kinase Domain and Tyrosine or Serine/Threonine Kinase signatures in some of the proteins are mutated, which play a dominant role in the pathogenesis of Lung Cancer and these may be addressed with the help of inhibitors to develop an efficient anticancer drugs. Furthermore, the present study contributes to the possibility that genetic components are more important in Lung Cancer as compared to environmental and smoking(carcinogens) factors.

Keywords: Lung Cancer; Protein Kinase Domain; Tyrosine Kinase; Phylogenetic Tree; Motif

INTRODUCTION

Lung cancer results when the body's system of checks and balances on the growth of cells is disrupted. This disruption results in uncontrolled division and proliferation of cells that ultimately form a mass of cells known as tumor. Cancers generally refer to those tumors which are malignant. The Protein Kinase family is the most frequently mutated gene family found in human Lung Cancers and kinases enzymes are widely targeted for the designing of anticancer drugs (Futreal et al., 2005). Protein Kinase domain is an evolutionarily conserved domain comprising of Protein Kinases enzymes. These enzymes function efficiently in the process of Phosphorylation by transferring phosphate groups from nucleoside triphosphate donor to acceptor molecules. Cancer is the result of abnormal phosphorylation. Tyrosine kinases and Serine/Threonine Kinases constitute to form 2 broad classes of protein kinases differing with respect to their substrate specificity. A substantial percentage of Lung Cancers express Epidermal Growth Factor Receptors (EGFRs) (Sequist et al., 2007); a cell surface receptor and several clinical researches efforts are performed towards the development of inhibitors of EGFR. Clinical responses in non-small cell lung cancer patients with EGFR Tyrosine Kinase inhibitor, Gefitinib were found to correlate with the presence of somatic EGFR kinase domain mutations in tumors (Daub et al., 2004, Lynch et al., 2004, Paez et al., 2004). Several other proteins that are considered to possess Protein Kinase Domain and are effectively responsible for the pathogenesis of Lung cancer include: BRAF, ERBB2,PTK2, TGFBR2, KDR, MET, IGF1R.MAPK1 and ALK. These proteins contain Protein Kinases ATP binding signatures and Tyrosine kinases specific active site signatures or Serine/Threonine Kinases specific active site signatures [Table 2].In the present study, we focused on the proteins that are believed to have

major role in pathogenesis of Lung Cancer by performing Domain and Motif study using Bioinformatics approach.

MATERIALS AND METHODS

2.1Sequence Retrieval from NCBI

We collected 71 known proteins that are actively involved in the pathogenesis of Lung cancer. The functional protein sequences in FASTA format for these proteins were retrieved from NCBI (National Center for Biotechnology Information: (http://www.ncbi.nlm.nih.gov) including their function, domain structure. Cross references with other databases modemize NCBI entries to hold details expertise.

2.2 Multiple Sequence Alignment

ClustalW is a widely used multiple sequence alignment (MSA) computer program. ClustalW is a general rationale multiplesequence alignment program for DNA or proteins (Chenna et al., 2003). It finds out the most outstanding match for the selected sequences and aligns them up so that the identities, similarities and differences can be seen. These sequences were employed to ClustalW (http://www.ebi.ac.uk/clustalw) for multiple sequence alignment which calculates the best match for selected sequences and lines them up so that the identities, similarities and differences can be seen.

2.3 Development of Phylogenetic tree

The sequence analysis score shows a high degree of evolutionary conservation among the sequences of the proteins. The phylogenetic profile of the proteins depicts low substitution rate & less gap penalty which indicates that they belong to same protein family. The evolutionary tree was drawn using NEIGHBOR. A rooted Phylogenetic tree with a unique node corresponding to the most recent common ancestor was found using the evolutionary analysis study. Based on these results, the score table and Phylogenetic tree that shows the distance between the protein sequences were constructed [Fig.1].

2.4 Identification of Domain and motif

Multiple Sequence Alignment was succeeded by Domain and Motif Study which revealed 10 proteins containing Protein Kinase domain and different Protein Kinases ATP binding signatures and Tyrosine Kinases specific active site signatures or Serine/Threonine Kinases specific active site signatures.

RESULTS AND DISCUSSION

The Bioinformatics analysis revealed 10 proteins out of 71 proteins that are key pathological proteins in the evolution of Lung Cancer. From the phylogenetic study we found that these 10 proteins are evolutionarily related. The present domain and motif study revealed that 10 proteins

were found to contain Protein Kinase domain and Tyrosine Kinase specific or Serine/Threonine Kinase specific active-site signatures which include ALK, BRAF, EGFR, ERBB2,IGF1R, KDR, MAPK1, MET, PTK2 and TGFBR2. From Table -2 these proteins are mainly responsible for the lung cancer. From the above domain study, we found that these proteins have conserved domains. These domains have some conserved patterns and the active sites those are responsible for the proliferation of the lung cancer.

3.1 EGFR Mutations and Lung Cancer

In tumors from patients with NSCLC responsive to the tyrosine kinase inhibitor gefitinib, (Lynch et al., 2004) and (Paez et al., 2004) identified mutations in the EGFR gene. Paez et al., (2004) found somatic mutations in EGFR in 15 of 58 unselected NSCLC tumors from Japan and 1 of 61 from the United States. Pao et al., (2004) found that in-frame deletions in exon 19 of the EGFR gene and somatic point mutations in codon 858 (exon 21) were common particularly in lung cancers from 'never smokers' and were associated, as found by others, with sensitivity to the tyrosine kinase inhibitors gefitinib and erlotinib.

3.2 BRAF Mutations and Lung Cancer

Mutations of the BRAF protein serine/threonine kinase gene have been identified in a variety of human cancers, most notably melanomas. (Naoki et al., 2002) analyzed the BRAF sequence in 127 primary human lung adenocarcinomas and found mutations in 2 tumor specimens, one in exon 11 and another in exon 15. The specimens belonged to the same adenocarcinoma subgroup as defined by clustering of gene expression data. The authors proposed that BRAF may provide a target for anticancer chemotherapy in a subset of lung adenocarcinoma patients.

3.3 .ERBB2 Mutations and Lung Cancer

The Cancer Genome Project and Collaborative Group (2004) sequenced the ERBB2 gene from 120 primary lung tumors and identified 4% that had mutations within the kinase domain; in the adenocarcinoma subtype of lung cancer, 10% of cases had mutations. In-frame deletions within the kinase domain of EGFR are associated with lung tumors that respond to therapy with gefitinib, an EGFR inhibitor. The Cancer Genome Project and Collaborative Group (2004) suggested that ERBB2 inhibitors, which had to that time proved to be ineffective in treating lung cancer, should be clinically reevaluated in the specific subset of patients with lung cancer whose tumors carry ERBB2 mutations.

3.4 ALK Mutations and Lung Cancer

Soda et al., (2007) identified a fusion gene, ALK/EML4, that was present in 5 of 75 Japanese non-small cell lung cancer patients examined. None of these patients had mutations in EGFR.

CONCLUSION

As research related to currently most advanced Lung cancer, reviewed the focuses on progress in the development and improve the diagnosis and treatment of Lungs cancer. It is evident from the discussion that ALK, BRAF, EGFR, ERBB2, IGF1R, KDR, MAPK1, MET, PTK2, TGFBR2 proteins play a significant role in the pathogenesis of lung cancer. Mutation in this particular EGFR, ERBB2, BRAF and ALK genes are responsible for the disease. From the domain study from Table2, we found that these proteins are having common Tyrosine protein kinases domain.

This Tyrosine protein kinase domain is responsible for the disease prognosis. In tumors from patients with NSCLC responsive to the Tyrosine Kinase inhibitor Gefitinib, (Lynch et al., 2004) and (Paez et al., 2004) identified mutations in the EGFR gene. The present study resultswill be helpful for the identification and designing the drugs for the lung cancer.

Although clinical responses to some Tyrosine protein kinases inhibitors as a single agent have been observed in several clinical trials, there is evidence that some of these inhibitors are more effective if given concurrently or sequentially with chemotherapy such as taxol, carboplatin, cisplatin, and etoposide(Woodburn et al., 2000)(Lydon et al., 1998)(Baselga et al., 1998)(Pietras et al., 1998)(Pegram et al., 1999)(Demidem et al., 1997)(Ciardiello et al., 2000). It is likely that several clinically useful Tyrosine protein kinases inhibitors will be in the market within the next decade, and that these new therapies will be more effective and with fewer side effects on patients with lung cancer.

ACKNOWLEDGEMENT

Authors are grateful to the UGC-SAP, New Delhi, for financial support. Special thanks to DBT and DIT for providing computational facilities. Manas Ranjan Barik is grateful to Centre for Bioinformatics, Pondicherry University.

REFERENCES

BASELGA, J., NORTON, L., ALBANELL, J., KIM, Y. M. & MENDELSOHN, J. 1998. Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and doxorubicin against HER2/neu overexpressing human breast cancer xenografts. *Cancer Res,* 58, 2825-31.

CHENNA, R., SUGAWARA, H., KOIKE, T., LOPEZ, R., GIBSON, T. J., HIGGINS, D. G. & THOMPSON, J. D. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res,* 31, 3497-500.

CIARDIELLO, F., CAPUTO, R., BIANCO, R., DAMIANO, V., POMATICO, G., DE PLACIDO, S., BIANCO, A. R. & TORTORA, G. 2000. Antitumor effect and potentiation of cytotoxic drugs activity in human cancer cells by ZD-1839 (Iressa), an epidermal growth factor receptor-selective tyrosine kinase inhibitor. *Clin Cancer Res,* 6, 2053-63.

DAUB, H., SPECHT, K. & ULLRICH, A. 2004. Strategies to overcome resistance to targeted protein kinase inhibitors. *Nat Rev Drug Discov,* 3, 1001-10.

DEMIDEM, A., LAM, T., ALAS, S., HARIHARAN, K., HANNA, N. & BONAVIDA, B. 1997. Chimeric anti-CD20 (IDEC-C2B8) monoclonal antibody sensitizes a B cell lymphoma cell line to cell killing by cytotoxic drugs. *Cancer Biother Radiopharm,* 12, 177-86.

FUTREAL, P. A., WOOSTER, R. & STRATTON, M. R. 2005. Somatic mutations in human cancer: insights from resequencing the protein kinase gene family. *Cold Spring Harb Symp Quant Biol,* 70, 43-9.

LYDON, N. B., METT, H., MUELLER, M., BECKER, M., COZENS, R. M., STOVER, D., DANIELS, D., TRAXLER, P. & BUCHDUNGER, E. 1998. A potent protein-tyrosine kinase inhibitor which selectively blocks proliferation of epidermal growth factor receptor-expressing tumor cells in vitro and in vivo. *Int J Cancer,* 76, 154-63.

LYNCH, T. J., BELL, D. W., SORDELLA, R., GURUBHAGAVATULA, S., OKIMOTO, R. A., BRANNIGAN, B. W., HARRIS, P. L., HASERLAT, S. M., SUPKO, J. G., HALUSKA, F. G., LOUIS, D. N., CHRISTIANI, D. C., SETTLEMAN, J. & HABER, D. A. 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med,* 350, 2129-39.

NAOKI, K., CHEN, T. H., RICHARDS, W. G., SUGARBAKER, D. J. & MEYERSON, M. 2002. Missense mutations of the BRAF gene in human lung adenocarcinoma. *Cancer Res,* 62, 7001-3.

PAEZ, J. G., JANNE, P. A., LEE, J. C., TRACY, S., GREULICH, H., GABRIEL, S., HERMAN, P., KAYE, F. J., LINDEMAN, N., BOGGON, T. J., NAOKI, K., SASAKI, H., FUJII, Y., ECK, M. J., SELLERS, W. R., JOHNSON, B. E. & MEYERSON, M. 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science,* 304, 1497-500.

PAO, W., MILLER, V., ZAKOWSKI, M., DOHERTY, J., POLITI, K., SARKARIA, I., SINGH, B., HEELAN, R., RUSCH, V., FULTON, L., MARDIS, E., KUPFER, D., WILSON, R., KRIS, M. & VARMUS, H. 2004. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U S A,* 101, 13306-11.

PEGRAM, M., HSU, S., LEWIS, G., PIETRAS, R., BERYT, M., SLIWKOWSKI, M., COOMBS, D., BALY, D., KABBINAVAR, F. & SLAMON, D. 1999. Inhibitory effects of combinations of HER-2/neu antibody and chemotherapeutic agents used for treatment of human breast cancers. *Oncogene,* 18, 2241-51.

PIETRAS, R. J., PEGRAM, M. D., FINN, R. S., MANEVAL, D. A. & SLAMON, D. J. 1998. Remission of human breast cancer xenografts on therapy with humanized monoclonal antibody to HER-2 receptor and DNA-reactive drugs. *Oncogene,* 17, 2235-49.

SEQUIST, L. V., BELL, D. W., LYNCH, T. J. & HABER, D. A. 2007. Molecular predictors of response to epidermal growth factor receptor antagonists in non-small-cell lung cancer. *J Clin Oncol,* 25, 587-95.

SODA, M., CHOI, Y. L., ENOMOTO, M., TAKADA, S., YAMASHITA, Y., ISHIKAWA, S., FUJIWARA, S., WATANABE, H., KURASHINA, K., HATANAKA, H., BANDO, M., OHNO, S., ISHIKAWA, Y., ABURATANI, H., NIKI, T., SOHARA, Y., SUGIYAMA, Y. & MANO, H. 2007. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature,* 448, 561-6.

WOODBURN, J., KENDREW, J., FENNELL, M. & WAKELING, A. ZD1839 ('Iressa') a selective epidermal growth factor receptor tyrosine kinase inhibitor (EGFR-TKI): inhibition of c-fos mRNA, an intermediate marker of EGFR activation, correlates with tumor growth inhibition. Proc Am Assoc Cancer Res, 2000. A2552.

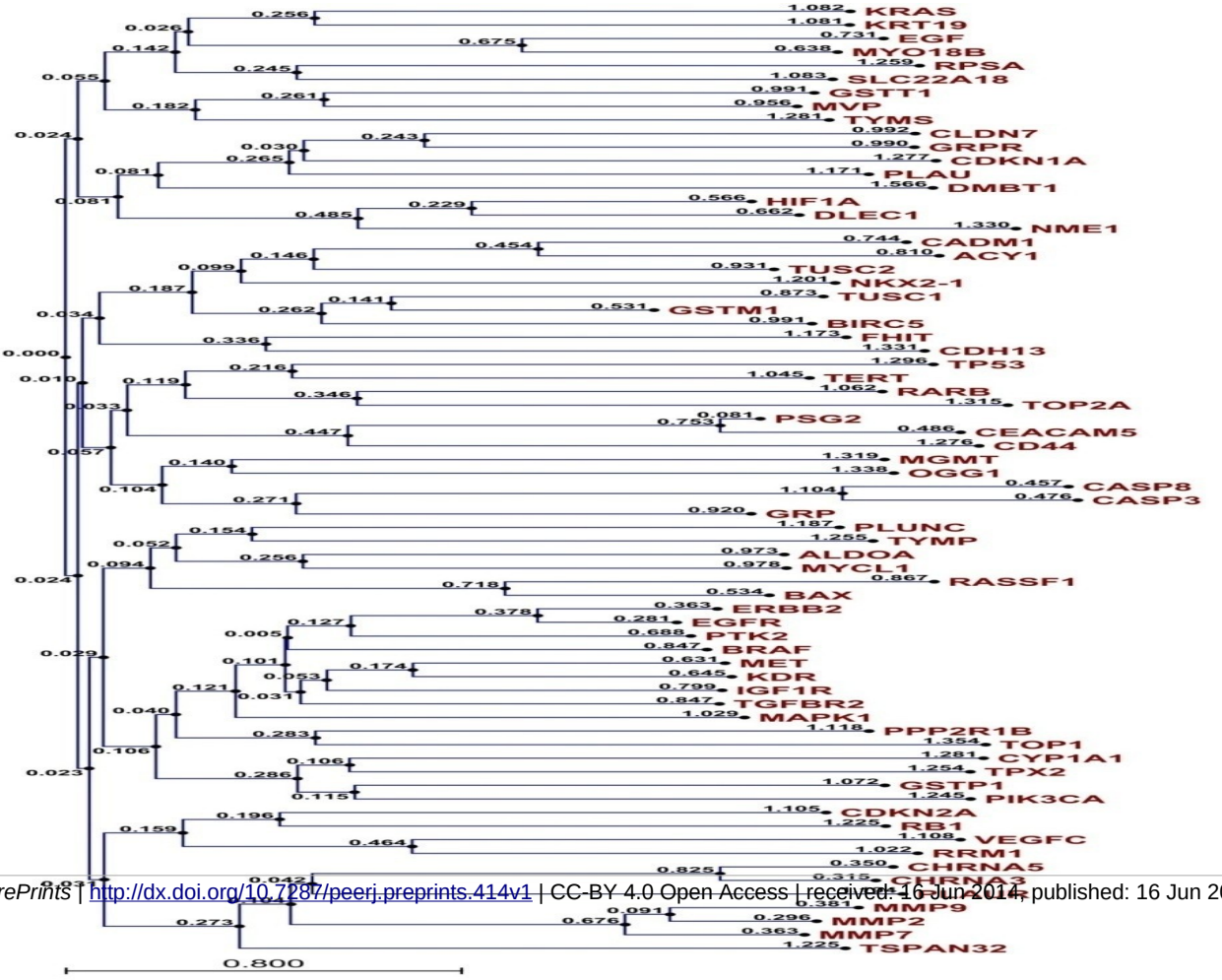

Figure1.The Phylogenetic tree that was constructed based on the alignment score of all the protein sequences involved in Lung Cancer. A high degree of homology was noticed between the proteins containing Protein Kinase Domain.

S.No	Protein Name	Swiss Prot ID	Length(Amino acids)	Domain Present
1	ALK	Q9UM73	1620	PROTEIN KINASE
2	ACY1	Q03154	408	ARGE_DAPE_CPG2_1
3	ALDOA	P04075	364	ALDOLASE _CLASS_1
4	BAX	Q07812	192	BCL2 like apoptosis inhibitors family profile
5	BIRC5	O15392	142	BIR-REPEAT-2
6	BRAF	P15056	766	PROTEIN KINASE
7	CADM1	Q9BY67	442	Ig-like domain profile
8	CASP3	P42574	277	Caspase family p20 domain profile
9	CASP8	Q14790	479	Death effector domain(DED)profile
10	CD44	P16070	742	LINK_2 Link domain profile
11	CDH13	P55290	713	CADHERIN_2 cadherins domain profile
12	CDKN1A	P38936	164	No significant domain found
13	CDKN2A	P42771	156	ANK_REP_REGION

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.414v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014

14	CEACAM5	P06731	702	Ig-like domain profile
15	CHRNA3	P32297	505	Neurotransmitter-gated ion-channels signature
16	CHRNA5	P30532	468	Neurotransmitter-gated ion-channels signature
17	CLDN7	O95471	211	Claudin family signature
18	CYP1A1	P04798	512	CYTOCHROME_P450
19	DLEC1	Q9Y238	1755	No significant domain found
20	DMBT1	Q9UGM3	2413	SRCR domain
21	EGF	P01133	1207	EGF-like domain profile
22	EGFR	P00533	1210	PROTEIN KINASE
23	ERBB2	P04626	1255	PROTEIN KINASE
24	FHIT	P49789	147	HIT-2 HIT domain profile
25	GRP	P07492	148	BOMBESIN Bombesin-like peptides family signature
26	GRPR	P30550	384	G_PROTEIN_RECEP_F1_2
27	GSTM1	P09488	218	GST_NTER,GST_CTER
28	GSTP1	P09211	210	GST_NTER,GST_CTER
29	GSTT1	O15392	240	GST_NTER,GST_CTER
30	H1F1A	Q16665	826	HLH"helix-loop-helix" domain profile
31	IGF1R	P08069	1367	PROTEIN KINASE
32	KDR	P35968	1356	PROTEIN KINASE
33	KRAS	P01116	189	RAS Small GTPaseRas family profile
34	KRT19	P08727	400	IF Intermediate Filaments signature
35	MAPK1	P28482	360	PROTEIN KINASE
36	MET	P08581	1390	PROTEIN KINASE
37	MGMT	P16455	207	THIOL_PROTEASE_HIS
38	MMP2	P08253	660	FN2_2,Fibronectin type-2

				collagen binding domain profile
39	MMP7	P09237	267	CYSTEINE_SWITCH
40	MMP9	P14780	707	FN2_2
41	MVP	Q14764	893	MVP;MVP(Vault)domain profile
42	MYCL1	P12524	364	HLH ”helix-loop-helix”domain profile
43	MYO18B	Q8IUG5	2567	IQ Motif Profile
44	NKX2-1	P43699	371	HOMEOBOX-2
45	NME1	P15531	152	Nucleoside diphosphate kinases active site
46	OGG1	O15527	345	No significant domain found
47	PIK3CA	P42336	1068	PI3_4_KINASE_3
48	PLAU	P00749	431	EGF-Like domain profile
49	PLAUR	Q03405	335	No significant domain found
50	PLUNC	Q9NP55	256	No significant domain found
51	PPP2R1B	P30154	601	HEAT_REPEAT
52	PSG2	P11465	335	Ig-like domain profile
53	PTK2	Q05397	1052	PROTEIN KINASE
54	RARB	P10826	455	NUCLEAR_REC_DBD_2
55	RASSF1	Q9NS23	344	Zinc finger phorbol -ester/DAG-Type profile
56	RB1	P06400	928	No significant domain
57	RPSA	P08865	295	RIBOSOMAL_S2_1
58	RRM1	P23921	792	ATP_CONE
59	SLC22A18	Q96BI1	424	Major Facilitator Superfamily(MFS)profile
60	TERT	O14746	1132	Reverse Transcriptase(RT) Catalytic domain profile
61	TGFBR2	P37173	567	PROTEIN KINASE
62	TOP1	P11387	765	TOPOISOMERASE_I_EUK
63	TOP2A	P11388	1531	TOPOISOMERASE_2

64	TP53	P04637	393	p53 family signature
65	TPX2	Q9ULW0	747	No significant domain found
66	TSPAN32	Q96QS1	320	No significant domain found
67	TUSC1	Q2TAM9	212	No significant domain found
68	TUSC2	O75896	110	No significant domain found
69	TYMP	P19971	482	THYMID_PHOSPHORYLASE
70	TYMS	P04818	313	THYMIDYLATE_SYNTHASE
71	VEGFC	P49767	419	PDGF_1

Table1: Table showing proteins that have been studied in the present study, which are believed to be actively involved in the disease.

S.NO	PROTEIN NAME	SIGNATURE NAME	SEQUENCE
1	ALK	Tyrosine protein kinases	FIHrDIAARNCLL
2	BRAF	Serine/Threonine protein kinases	IiHrDLKsnNIFL
3	EGFR	Tyrosine protein kinases	LVHrDLAARNVLV
4	ERBB2	Tyrosine protein kinases	LVHrDLAARNVLV
5	IGF1R	Tyrosine protein kinases	FVHrDLAARNCMV

6	KDR	Tyrosine protein kinases	CIHrDLAARNILL
7	MAPK1	Serine/Threonine protein kinases	VlHrDLKpsNLLL
8	MET	Tyrosine protein kinases	FVHrDLAARNCML
9	PTK2	Tyrosine protein kinases	FVHrDIAARNVLV
10	TGFBR2	Serine/Threonine protein kinases	IvHrDLKssNILV

212 Table2: Table showing 10 proteins possessing Protein Kinase Domain and Tyrosine Kinase or
213 Serine/Threonine Kinase specific active-site signature.

214 Figure 2: 3-D structure of Protein Kinase Domain

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.414v1 | CC-BY 4.0 Open Access | received: 16 Jun 2014, published: 16 Jun 2014