Clinical and *in vitro* studies on a Breast Tumour Kinase splice variant, λm5: A case study of tyrosine kinome diversity due to exon skipping

Thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

by

VAMSHI KRISHNA IRLAPATI

UOH Enrolment ID: 09LTPH15

PhD Advisory committee
Dr. Kiranam Chatti (Guide)
Dr. Niyaz Ahmed (Co-Guide)
Dr. Prasenjit Mitra

Department of Biology
Dr. Reddy's Institute of Life Sciences
University of Hyderabad Campus
Hyderabad, Telangana, INDIA.

Department of Biotechnology & Bioinformatics

University of Hyderabad

May 2016

DECLARATION

I hereby declare that the work presented in this thesis entitled "**Clinical and in vitro studies on a Breast Tumour Kinase splice variant, λm5: The significance of tyrosine kinome diversity due to exon skipping"** has been carried out by me under the supervision of **Dr. Kiranam Chatti,** Principal Research Scientist and Head, Department of Biology at Dr. Reddy's Institute of Life Sciences, University of Hyderabad. I am registered for Ph.D. degree in the department of Biotechnology, School of Life Sciences, University of Hyderabad and this thesis has not been submitted for any degree or diploma of any other university.

Vamshi Krishna Irlapati

DECLARATION

I hereby declare that the work presented in this thesis entitled "**Clinical and in vitro studies on a Breast Tumour Kinase splice variant, λm5: The significance of tyrosine kinome diversity due to exon skipping"**, carried out by me under the supervision of **Dr. Kiranam Chatti,** Principal Research Scientist and Head, Department of Biology at Dr. Reddy's Institute of Life Sciences, University of Hyderabad is a bonafide research work and free from plagiarism. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma. I hereby agree that my thesis can be deposited in Shodhganga/INFLIBNET. A report on plagiarism statistics from the university librarian is enclosed.

Date: Vamshi Krishna Irlapati

Reg. No: 09LTPH15

Signature of Supervisor Signature of Student

University of Hyderabad
(Central University established in 1974
by Act of Parliament)
School of Life Sciences
Department of Biotechnology and Bioinformatics
Hyderabad – 500 046. India.

Dr. Reddy's Institute of Life
Sciences
Department of Biology
University of Hyderabad Campus
Hyderabad – 500 046.
India.

CERTIFICATE

This is to certify that the research work embodied in this thesis entitled **"Clinical and *in vitro* studies on a Breast Tumour Kinase splice variant, λm5: A case study of tyrosine kinome diversity due to exon skipping"** has been carried out by **Vamshi Krishna Irlapati**, under the guidance of **Dr. Kiranam Chatti** at Dr. Reddy's Institute of Life Sciences, Hyderabad and is being submitted to Department of Biotechnology and Bioinformatics, School of Life Sciences, University of Hyderabad, for a full period prescribed under the PhD ordinance of this University. We recommend this thesis for submission for the award of the degree of Doctor of Philosophy of this University.

Dr. Kiranam Chatti (Supervisor)
Principal Research Scientist & Head,
Dr. Reddy's Institute of Life Sciences
University of Hyderabad Campus,
Hyderabad-500 046

Head of Department
Department of Biotechnology
& Bioinformatics
University of Hyderabad
Hyderabad-500 046

Dean of the School
School of Life Sciences
University of Hyderabad
Hyderabad – 500 046.

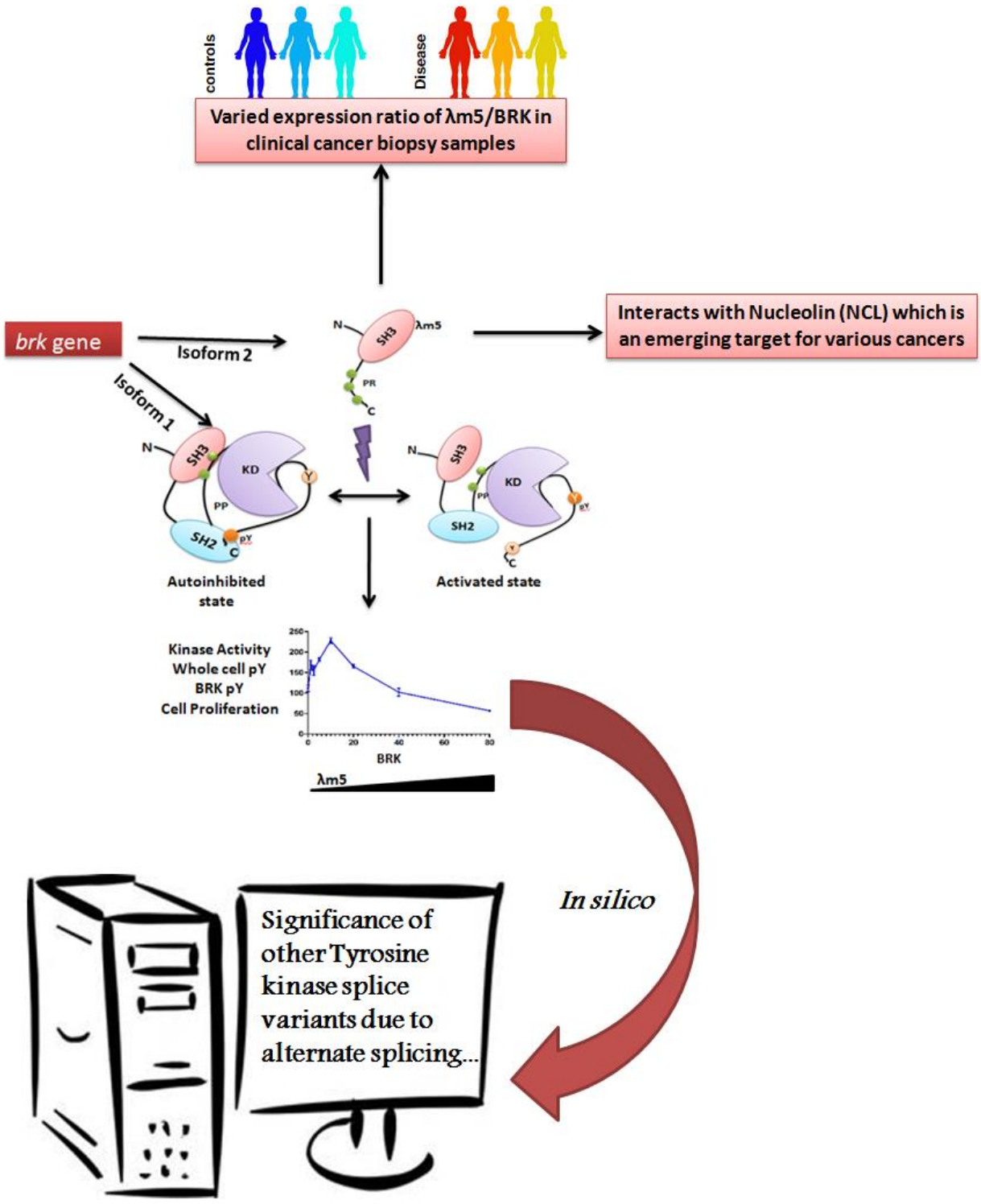
ABSTRACT

Breast Tumour Kinase (BRK) or Protein Tyrosine Kinase 6 (PTK6) is a potential biomarker and target for cancer therapy, and dysregulation of its activity and expression result in oncogenesis. This work explored the function of a non-catalytic protein variant of BRK (dubbed λm5). λm5 is produced as a result of alternate splicing of the brk gene, where the second exon is skipped. This results in a frameshift mutation and premature termination. *In vitro* and in cell-based studies, we found that λm5 affected BRK activity and cell proliferation in a biphasic dose-dependent manner. Quantitative expression profiling of cancer cell lines and patient biopsies revealed a variable expression ratio of BRK and λm5. Our work indicates that the expression ratio of BRK and λm5 is relevant in cancer, and suggests that the biphasic effect of λm5 expression level is an underlying variable. We also sought to find potential cellular interactions of λm5 and identified Nucleolin, a known oncoprotein as a common interacting partner for BRK and λm5.

We then used a computational strategy to expand our study to other non-catalytic splice variants of the human tyrosine kinome by manual curation of available RNA sequencing data. We found that the 90 human tyrosine kinase genes produce 835 splice variants, of which almost half (52%) contain open reading frames and can potentially produce proteins. When we analyzed the sequences of the predicted tyrosine kinase splice variant proteins, we found that almost half of them are non-catalytic in nature (47%). An algorithm was developed to predict and catalog all tyrosine kinase variants resulting from exon skipping, including their sequences and biophysical properties. The resulting database is hosted and is publicly accessible at www.kinexon.in. A potential application of the database is to aid in curation of unassigned or orphan protein sequences obtained in proteomics studies.

Collectively, this work explores the function and expression of a non-catalytic BRK splice variant in a sample of cancer patients from India, and highlights the potential significance of tyrosine kinase splice variants and kinome diversity in cancer research.

GRAPHICAL ABSTRACT



DEDICATION

Dedicated to….

My loving parents who struggled for my life, a success…

My wife for her untiring support being at professional equality…

My daughter for erasing all the stress with her smile…

My guide, an ideal teacher…

ACKNOWLEDGEMENTS

Its July 12th, 2009 morning 10:30 when three eminent scientists (Dr. Kiranam Chatti, Dr. Kishore Parsa and Dr. Devyani Haldar) took my interview at DRILS and gave a green signal to a completely new career of full- fledged research. The career seed was sown by a special friend who referred my CV to the then HOD. The HOD was kind enough to give an opportunity to work under a newly appointed scientist in the institute. The HOD from then deliberately improved my personality development too. The new scientist laid the strong foundation stone of research by teaching me all the fundamentals of bacterial and cell culture work. It was only 7 months that I worked under the new scientist while I got through the national entrance ICMR-JRF and shifted to another senior research scientist who agreed for being my Ph.D supervisor. The friend who referred my CV to the institute was also instrumental in finding the Ph.D. guide for me. Taken all together, my special friend, the then HOD, the new scientist and the Ph.D guide (names would be revealed at the end) have influenced my life so much for whatever I am capable of today…. Even if I forget to thank them while writing this long story short, the irrevocable fact is that those people will remain in my mind till I forget my parents…..

Research is never a one man show. I realized that the working environment matters a lot in easing our success. I got good friends in perpetuating my research career. My sincere thanks to all of them. From then till now, the names are as follows…

Vanaja G R, S. TejaSudha, Srinadh M, G Suresh Babu, Yakub, Keerthana Sarma, Tulasi Kirla, Rakesh Bonate, Dr. Aleem Pasha, Amrita Sengupta, Lahari, Neelam, Khubaib, Bandish K, Govardhan KS, Dr. Ritta Mathew, Neeraja AP, Malathi T, G Sravan Babu, Ramya Sri K, Suman, Soma B, Tapan Kumar Singh Nayak, Ansu Mathew, Maithreyi, Ramudu B, Yaseen, Raghavendra M, Swapna, Raghu CDFD, Shilpak and thanks to all other DRILS friends who made my life easy.

My heartful thanks to Sravan and Suman who helped me a lot with wet lab work during my last days of Ph.D.

My sincere thanks to the Director, Management, Teaching and non-teaching staff of Dr. Reddy's Institute of Life Sciences (DRILS) for the wonderful facility, infra structure that any researcher would dream of, and for the working time I enjoyed in the institute.

My special thanks to each and every faculty for all the knowledge they have spread with their valuable talks and inputs during my Ph.D career.

My sincere thanks to Dr. Prasenjit Mitra (Principal Research Scientist, DRILS) who took deliberate care in my Ph.D work.

My thanks to Dr. Niyaz Ahmed (HOD, Biotechnology Dept. University of Hyderabad) for being my co-guide and helping me with the official formalities of the department.

Heartfelt thanks to Aditya, student of Bioinformatics who gave his full efforts helping me with the *in silico* work and KINEXON web site hosting.

Few lines from the heart for 3 special people – Dr. Kirmani Natukula, Dr. Krishna Veni and Dr. Ashraf Ashfaq -

I learnt that seniority; status, scientific caliber, authority, achievements, ego and attitude should never stop us from being kind and helpful to others. They don't speak much but we can learn everything from them. They never take the merit of inspiring others with their down to earth nature, politeness, kindness, helping nature and ready to teach knowledge. I am very happy to be associated with these 3 knowledge hubs during my Ph.D. who were at my back then and even now, at all the times for any problem. I thank god for giving me an opportunity to work and learn from these seniors during my research career.

Lines of love for my wife-

Seeing my wife I often feel it's true that woman is born for multitasking. Being at professional equality with me, she never got tired in managing my family and the new born we are blessed with. I am a fan of her management skills and hard-working nature. Thanks to her indispensable role in my Ph.D. I keeping loving her and learning from her…..

My boss Dr. Kiranam Chatti-

He talks less… but teaches more…

He seems gentle… but solid scientist…

He remains reserved… but deserves all the respect…

He never scolds… but moulds the student to the right path…

He may not need science… but science and students need him…

He is not my family member… but gives the same support.

He is not only an ideal human being… but also an ideal teacher…

He not only teaches us… but also inspires us how to teach others…

When he comes into my mind… My eyes talk more about him than my lips…

I DON'T WANT TO SAY JUST 'THANKS', BUT THANK YOU SIR……

The special friend of my Ph.D was Vanaja G R who is my wife now…

The then HOD was Prof. Parimal Misra…

The new scientist was Dr. Kishore Parsa…

The Ph.D guide…sorry, Ph.D God was Dr. Kiranam Chatti…

My heartfelt thanks to Indian Council of Medical Research (ICMR) for the timely disbursement of fellowship to me which made my life comfortable during the Ph.D.

TABLE OF CONTENTS

ABBREVIATIONS

BRK/PTK6: Breast Tumor kinase/Protein tyrosine kinase 6

λm5: Alternative splice variant of BRK

TK: Tyrosine kinase/Tyrosine kinome

RTKs: Receptor Tyrosine Kinases

NRTKs: Non-Receptor Tyrosine Kinases

RNA: Ribonucleic acid

NCBI: National Centre for Biotechnology Information

ORF: Open Reading Frame

ATP: Adenosine triphosphate

NGFR: Nerve growth factor receptor

TrkA: Tropomysin receptor kinase A

KIT: A Cytokine receptor belonging to RTK family

SCF: Stem cell factor

FGFR: Fibroblast growth factor receptor

EGFR/ErbB: Epidermal growth factor receptor

c-Src: Cellular Src kinase

c-Abl: Cellular Abelson murine leukemia kinase

JAK: Janus Kinase

Y527: Tyrosine residue at 527 position of protein

Y530: Tyrosine residue at 530 position of protein

pY527: Phosphorylated tyrosine at 527 position of protein

SH3: Src homology like domain

SH2: Src homology like domain

Bcr-Abl: Break point cluster-Abelson leukemia kinase

STATs: Signal Transducer and Activators of transcription

SOCS: Suppressors of cytokine signaling

mRNA: messenger RNA

cDNA: complementary DNA

T47D: Metastatic breast cancer cell line derived from ductal carcinoma

MCF-7: Metastatic breast cancer cell line derived from adenocarcinoma

K219M: Lysine at position 219 mutated to methionine

Y447F: Tyrosine at position 447 mutated to phenylalanine

W44A: Typrophan at position 44 mutated to alanine

p190RhoGAP: Rho family specific GTPase activating protein

Akt: Protein kinase B (serine-threonine kinase)

IRS-4: Insulin receptor substrate 4

BKS/STAP-2: Signal-transducing adaptor protein 2

KAP3A: Keratin-associated protein 3A

KLF9: Krueppel-like factor 9

NFκB: Nuclear factor kappa-light-chain-enhancer of activated B cells

EGF: Epidermal growth factor

HRG: Histidine rich glycoprotein

OPN: Osteopontin

IGF-1: Insulin growth factor 1

SOCS3: Suppressor of cytokine signaling 3

Myc: human myelocytomatosis viral oncoprotein

CycD1: Cyclin D1

Ras: Small GTPase protein family member

MAPK: Mitogen-activated protein kinase

ATF-4: Activating transcription factor 4

CrkII: Human sarcoma viral proto-oncogene

RacI: Rho related protein

VEGF: Vascular endothelial growth factor

PSF: Polypyrimidine tract-binding *protein* (PTB)-associated splicing factor

Sam68: Src-Associated substrate in Mitosis of 68 kDa

SLM1: Sam68 like mammalian protein 1

SLM2: Sam68 like mammalian protein 2

HNSCC: Head and neck squamous cell carcinoma

NSCLC: Non-small cell lung cancer

IC50: Inhibitory concentration 50

HER2: Human epidermal growth factor receptor 2

PP: Poly-proline motif

PR: Proline rich region

KD: Kinase domain

α4G10: phosphotyrosine specific antibody

αc-myc: c-myc specific antibody

LIST OF FIGURES

Figure **Page**

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

LIST OF TABLES

Table **Page**

Chapter 3

Chapter 4

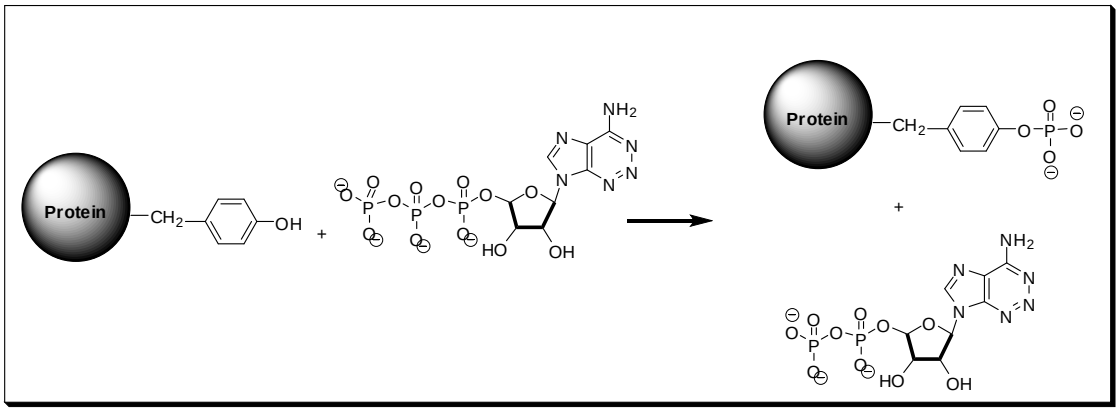
Chapter 5

Chapter 1

Introduction and review of literature.

1.1 Human Tyrosine Kinome: Key player in cellular signaling.

The tyrosine kinome represents the complement of protein tyrosine kinase enzymes encoded by the human genome. Tyrosine kinases are enzymes which catalytically transfer the γ-phosphate group of the nucleoside triphosphate donor ATP to tyrosine residues on proteins, as shown in the figure below.



The history of tyrosine phosphorylation dates back to 1979 when it was first detected in hydrolysates of viral transforming proteins labeled by incubation of immunoprecipitates with radioactive ATP (Eckhart, Hutchinson et al. 1979; Hunter and Sefton 1980; Witte, Dasgupta et al. 1980). Six years after the discovery of tyrosine phosphorylation, the first comprehensive review of tyrosine kinases by Hunter and Cooper highlighted the protein tyrosine kinase activity of growth factor receptors as pertinent to intercellular signaling and thereby the functioning of multicellular organisms (Cooper and Hunter 1985; Hunter and Cooper 1985). Since then, tyrosine kinases have emerged as clinically useful drug targets for treating many human cancers. Tyrosine phosphorylation directs the activity, localization and overall function of many proteins and serves to orchestrate the signaling activity of almost all cellular processes.

1.1.1 Classification of Human Tyrosine kinases.

Tyrosine kinases (TKs) mark their existence in most metazoan phyla like poriferans, cnidarians, nematodes, annelids, arthropods, echinoderms, and chordates including the experimental model organisms *Saccharomyces cerevesiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster* (Hunter and Plowman 1997; Sakuma, Onodera et al. 1997; Lucini, Castaldo et al. 1999; Muller, Kruse et al. 1999; Plowman, Sudarsanam et al. 1999; Popovici, Roubin et al. 1999; Suga, Hoshiyama et al. 1999; Miller, Malik et al. 2000; Morrison, Murakami et al. 2000; Rikke, Murakami et al. 2000). Tyrosine kinases were differentiated from other protein kinases based on the unique conserved sub-domain and motif sequences (Hanks and Lindberg 1991; Hanks and Quinn 1991). Thorough manual curation of human genome sequence data has allowed the identification of 90 unique tyrosine kinase genes (Robinson, Wu et al. 2000). Based on the phylogenetic analysis of amino acid sequences of kinase domains, this class of TKs was broadly sub-classified into receptor tyrosine kinases (RTKs) and non-receptor tyrosine kinases (NRTKs) that can be further grouped into 20 families and 10 families respectively.

1.1.2 Receptor tyrosine kinases (RTKs).

TKs that mostly remain anchored to the cell membrane and display their protein complements on the cell surface to receive external stimuli are termed as receptor tyrosine kinases. The human genome encodes 58 known RTKs which fall into 20 subfamilies (Figure 1). All RTKs are known to have a similar molecular architecture, with ligand-binding domains in the extracellular region, a single transmembrane helix, and a cytoplasmic region that contains the protein tyrosine kinase (TK) domain along with additional carboxy (C-) terminal and juxtamembrane regulatory regions.

1.1.3 Non-Receptor tyrosine kinases (NRTKs).

TKs that mostly lack membrane localization signals and involved in cytoplasmic phosphotransfer-directed functions are termed as non-receptor tyrosine kinases. The human genome encodes 32 known NRTKs which fall into 10 protein subfamilies based on initial studies carried out on their viral counterparts (Figure 2). NRTKs lack receptor-like features

such as extracellular ligand-binding domains and membrane-spanning regions. The typical structure of NRTKs consists of protein domains for substrate recognition linked to a kinase domain which possesses catalytic activity with the ability of binding and utilizing ATP for phosphotransfer reactions.

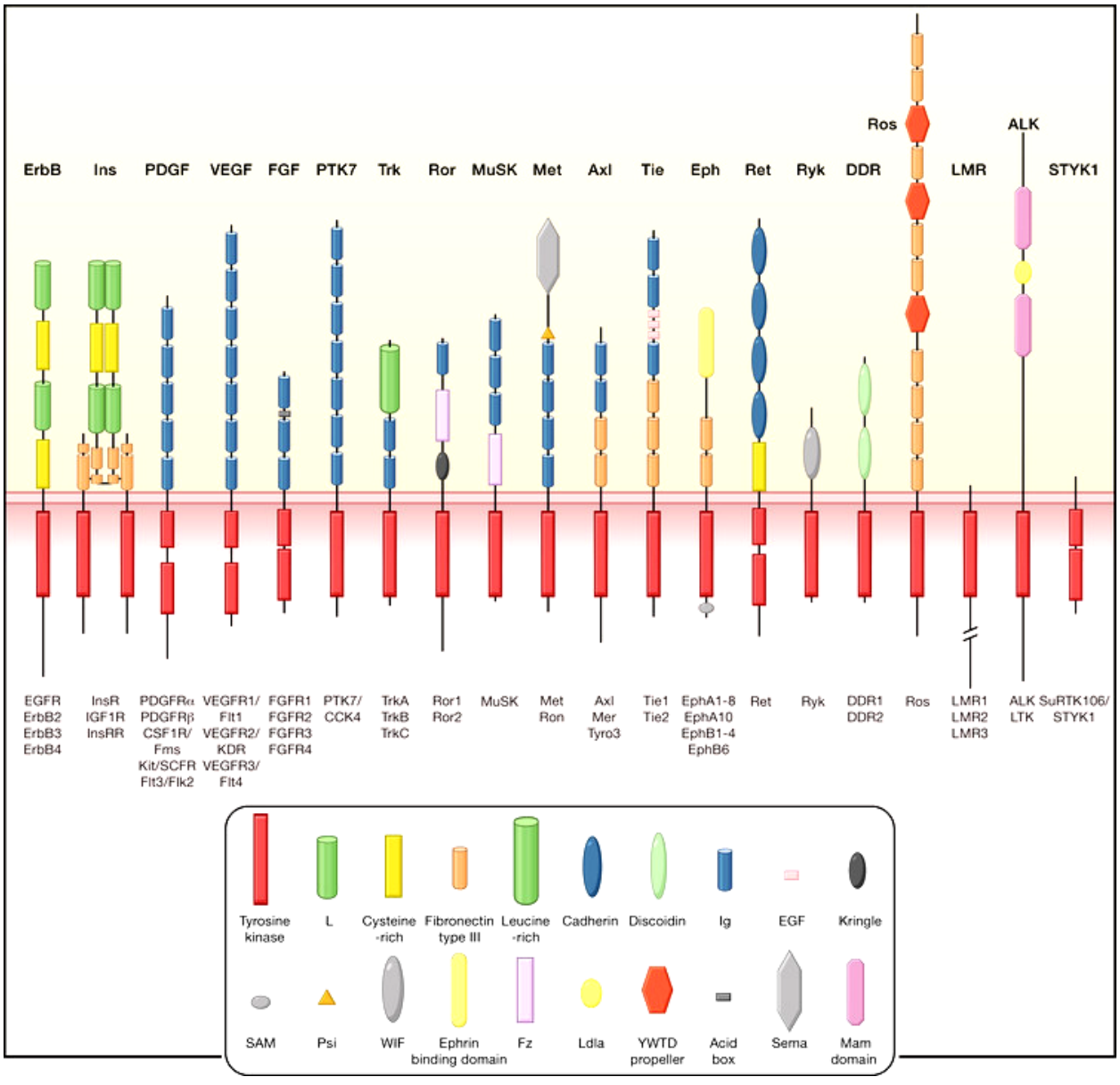


Figure 1. Depiction of receptor tyrosine kinase families.
Human receptor tyrosine kinases (RTKs) contain 20 subfamilies, shown here schematically with the family members listed beneath each receptor. Structural domains in the extracellular

regions, identified by structure determination or sequence analysis, are marked according to the key. The intracellular domains are shown as red rectangles
[Lemmon and Schlessinger. Cell (2010) 141(7): 1117-34].

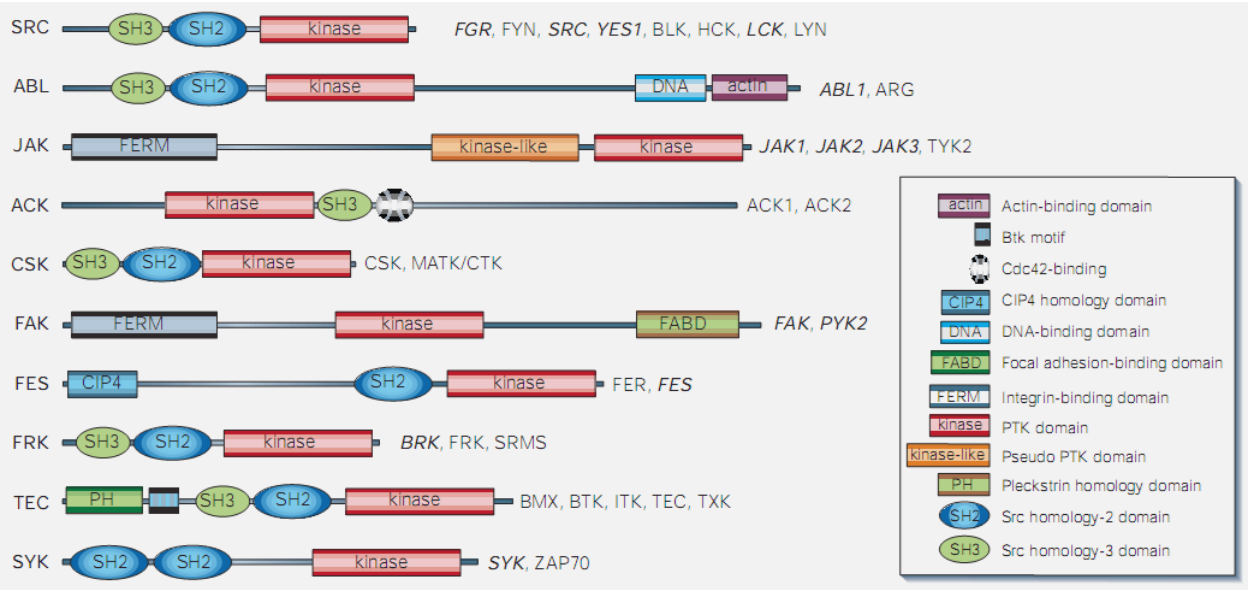


Figure 2. Depiction of non-receptor tyrosine kinase families.
The family members are indicated to the right and the family name to the left of each PTK
[Blume-Jensen and Hunter. Nature (2001) 411(6835): 355-65].

1.2 Regulation of tyrosine kinases.

The common theme in RTKs and NRTKs is tyrosine phosphorylation. Tyrosine phosphorylation of protein substrates triggers cellular signaling pathways. Autophosphorylation of TKs is not uncommon and often required for kinase functionalities. The homeostasis of phosphorylation is maintained by protein tyrosine phosphatases. In general, tyrosine phosphorylation on tyrosine kinases has a regulatory effect, both positive and negative and RTKs and NRTKs have been observed to follow specific trends of regulation.

1.2.1 Regulation of RTKs.

In general, ligand (growth factor) binding activates RTKs by inducing receptor dimerization or oligomerization (Ullrich and Schlessinger 1990). A ligand interacts with receptor subunits on the cell surface and drives formation of dimeric or multimeric complexes. Based on this

concept, the regulation of RTKs was observed to fall under four types attributed to the specific participation of receptor components(Lemmon and Schlessinger) (Figure 3).

1) ***Ligand-mediated receptor dimerization*** where the receptors make no direct contact till ligand stimulation. Example: Nerve growth factor receptor (NGFR) called TrkA (Wehrman, He et al. 2007).

2) ***Ligand-mediated dimer with receptor contacts*** where the ligand pair helps crosslink the primary receptor domains and initiates receptor activation that triggers other domains of the receptor to reorient and interact at the dimer interface. Example: Activation of KIT receptor tyrosine kinase by stem cell factor dimer (SCF) (Yuzawa, Opatowsky et al. 2007).

3) ***Multiple contacts (combination of ligand binding, receptor-receptor contacts and involvement of accessory molecules).*** Example: Dimerization of Fibroblast growth factor receptor (FGFR) by its ligands in the presence of heparin as accessory molecule (Plotnikov, Hubbard et al. 2000; Stauber, DiGabriele et al. 2000).

4) ***Receptor-mediated dimerization*** where the activating ligand does not make a direct contribution to the dimerization interface. The classical example is the EGFR/ErbB family of RTKs (Garrett, McKern et al. 2002; Ogiso, Ishitani et al. 2002). Here, the bivalent ligand contacts two distinct sites within the single receptor molecule and unlocks the auto-inhibited tethered conformation substantially making the individual receptors compatible for dimerization (Cho and Leahy 2002; Burgess, Cho et al. 2003; Ferguson, Berger et al. 2003; Bouyain, Longo et al. 2005).

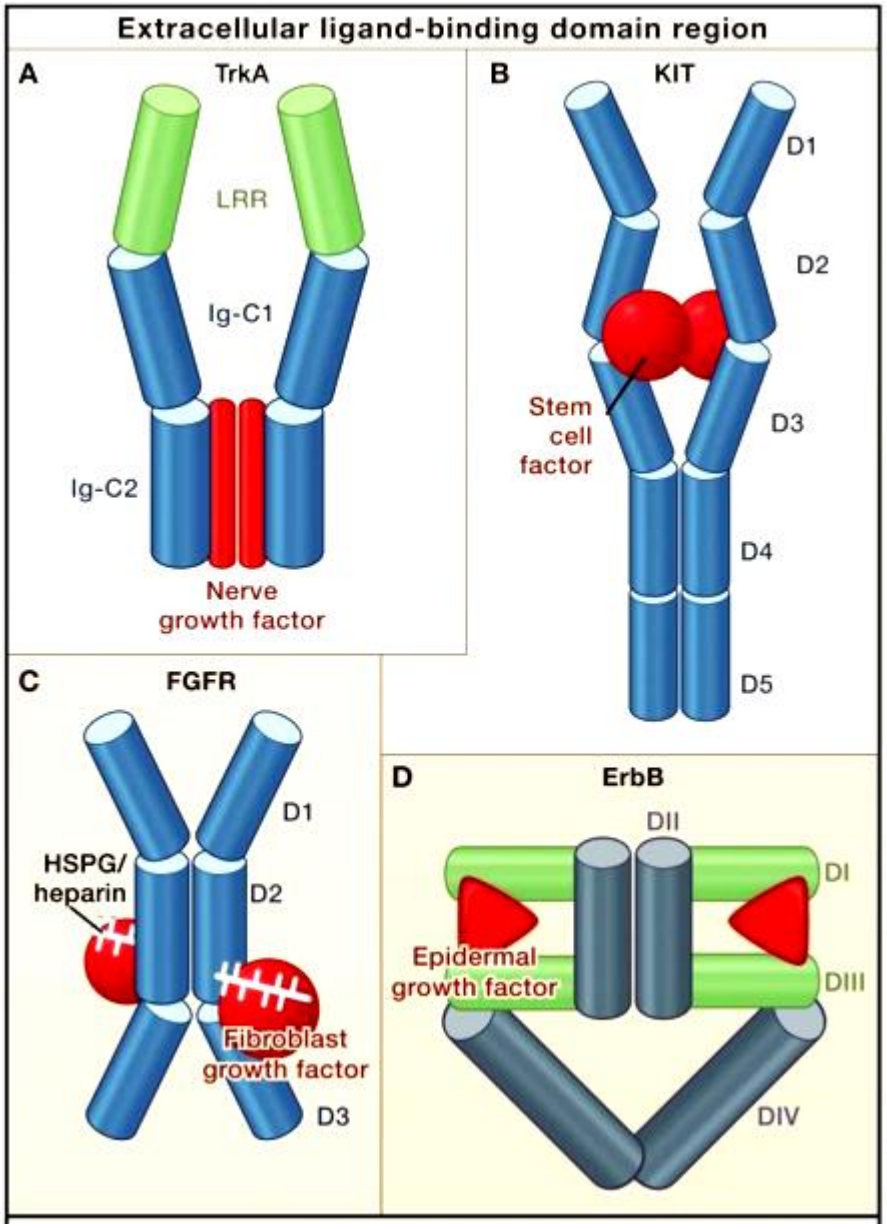


Figure 3. Regulation of RTKs.

[Lemmon and Schlessinger. Cell (2010) 141(7): 1117-34]

(A) A nerve growth factor dimer (red) crosslinks two TrkA molecules without any direct contact between the two receptors.

(B) A stem cell factor dimer (red) also crosslinks two KIT molecules. In addition, two Ig-like domains (D4 and D5), which reorient upon receptor activation, interact across the dimer interface (Yuzawa et al., 2007). Thus, KIT combines ligand-mediated and receptor-mediated dimerization modes.

(C) Two fibroblast growth factor receptor (FGFR) molecules contact one another through the Ig-like domain D2, and the accessory molecule heparin or heparin sulfate proteoglycans

(white sticks) also contact this domain (Schlessinger et al., 2000). In addition, each fibroblast growth factor molecule (red) contacts Ig-like domains D2 and D3 of both FGFR molecules. **(D)** Dimerization of ErbB receptors is mediated entirely by the receptor. Binding simultaneously to two sites (DI and DIII) within the same receptor molecule, the ligand drives conformational changes in epidermal growth factor receptor (EGFR) that expose a previously occluded dimerization site in domain II. Bottom: Dimerization of the extracellular regions of RTKs activates the intracellular tyrosine kinase domains (TKDs), which contain a C-lobe (light purple), N-lobe (dark purple or yellow in the inactive and active states, respectively), and an activation loop (purple or yellow in the inactive and active states, respectively). Although the crystal structures of the activated TKDs are very similar (Huse and Kuriyan, 2002), structures of inactive TKDs differ substantially among the receptors (top row), refecting the diversity in their regulatory mechanisms. However, many receptors are inhibited by a set of intramolecular (or cis) interactions.

1.2.2 Regulation of NRTKs.

NRTKs are critical conveyers of extracellular signals and determine the cellular functional outputs. A major part of regulation in NRTKs happens through intramolecular interactions. The most extensively studied NRTK families are c-SRC, c-ABL and JAK kinases(Blume-Jensen and Hunter 2001).

***c-SRC family*:** The conserved domain structure of SRC consists of a myristoylated N-terminal segment, SH3, SH2, kinase domains and a C-terminal segment. Under inactive conditions, the C-terminal tyrosine residue (Y527 in mouse/Y530 in humans) of SRC is phosphorylated and interacts with the SH2 domain, while the SH3 domain interacts with linker region between the SH2 domain and the N-terminal kinase domain (Figure 4).

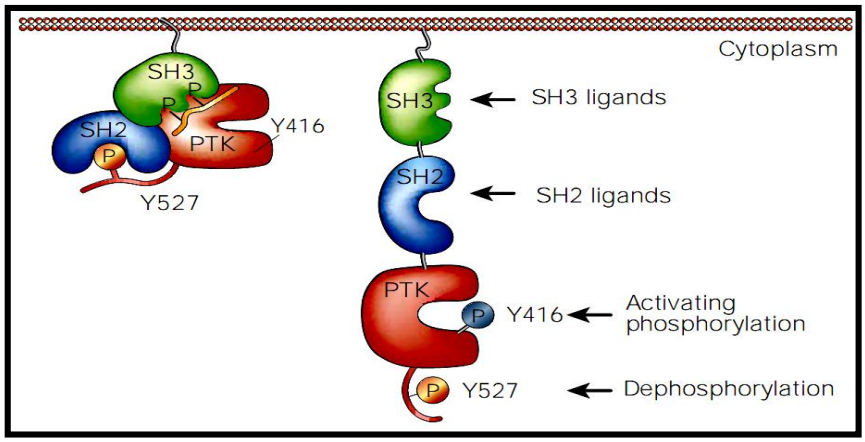


Figure 4. Regulation of c-SRC kinase.
Repression (left) and activation (right) modes of c-SRC kinase. [Blume-Jensen and Hunter. Nature (2001) 411(6835): 355-65].

The intramolecular interactions of SH3 and SH2 domains regulate tyrosine kinase activity by displacing the C-alpha helix in the N-terminal with changed orientation of activation loop to block access to the active site. c-SRC is activated upon dephosphorylation of the phosphorylated c-terminal tyrosine (pY527) residue and by binding through its SH2 domain to specific tyrosine autophosphorylation sites in ligand-stimulated RTKs, resulting in displacement of SH2 domain from pY527. Binding of the SH3 domain to Pro-X-X-Pro motifs in target proteins (Sicheri and Kuriyan 1997) is also a mechanism of c-SRC activation.

c-ABL family: Structurally, c-ABL kinase consists of the SH3, SH2, kinase, DNA-binding and actin binding domains. c-ABL kinase is auto-inhibited by a set of intramolecular interactions which were found similar to c-SRC kinase but with some unique differences. A high level of interest rests on BCR-ABL fusion kinase which is a result of a chromosomal translocation that codes for a hyperactive form of the kinase leading to human leukemic malignancies. The BCR-ABL tyrosine kinase domain is activated by formation of homo-oligomeric complexes mediated by the BCR coiled-coil domain, therefore enhancing trans-autophosphorylation (Hantschel and Superti-Furga 2004).

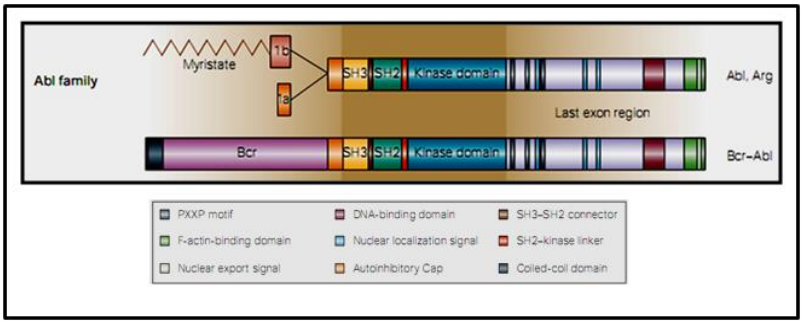


Figure 5. Domain structures of ABL family kinases
[Hantschel et al. Nat Rev Mol Cell Biol (2014) 5(1): 33-44].

JAK *family:* The Janus kinase family along with STATs (Signal Transducer and Activators of Transcription) forms a well characterized signaling cascade activated by cytokines or growth factors. Receptor engagement triggers Janus kinases to get into close proximity to one another and the kinases get activated by trans-phosphorylation. JAK proteins in turn phosphorylate specific tyrosine residues on the C-terminal end of STATs which orient them to dimerize and relocate to the nucleus to either activate or repress specific gene transcripts. The factor that negatively regulate JAK-STAT pathway are down-regulation of receptors, dissociation of JAKs from their receptors, suppressors of cytokine signaling (SOCS), protein inhibitors of STATs and protein tyrosine phosphatases (Babon, Lucet et al.).

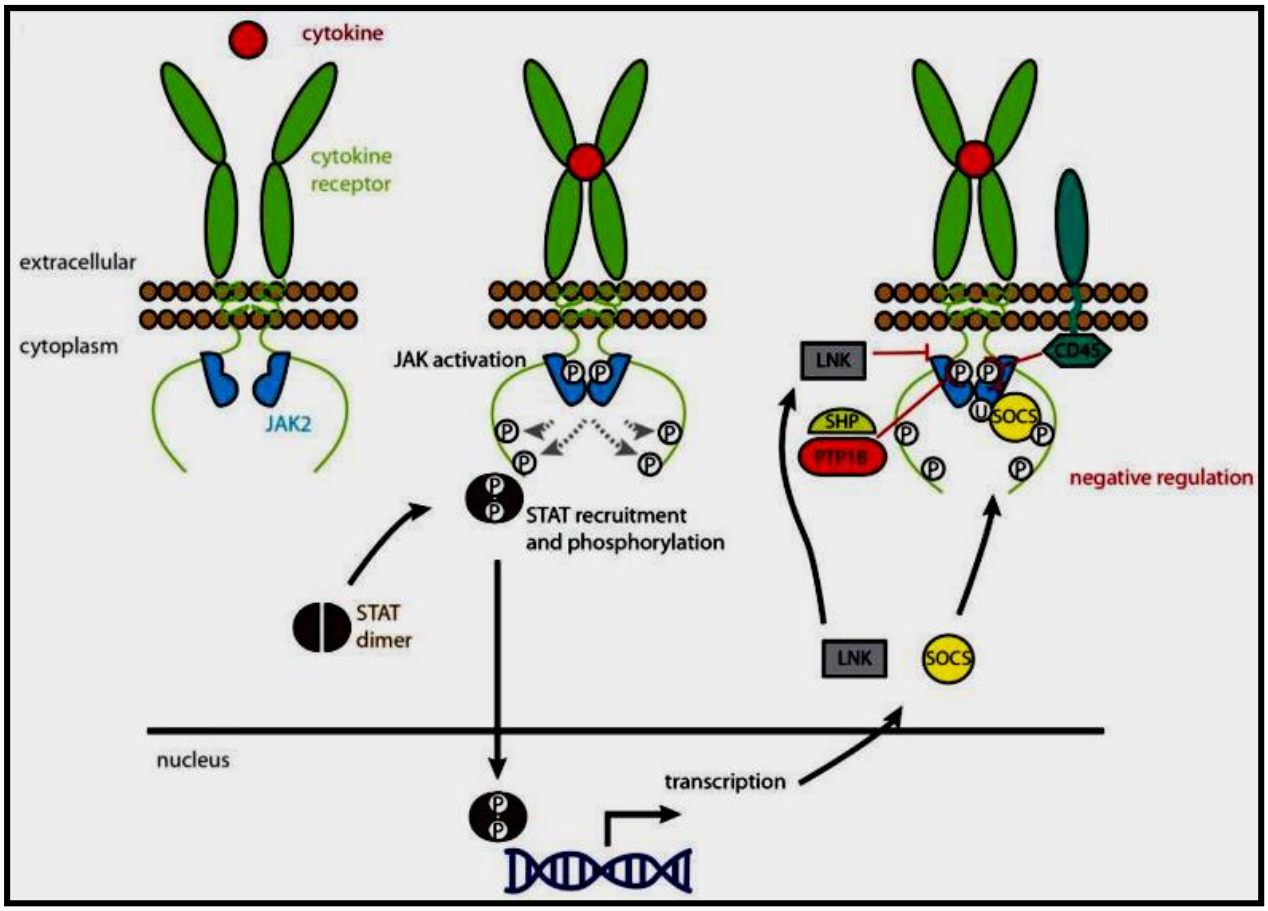


Figure 6. Overview of JAK activation and regulation
[Babon, Lucet et al. Biochem. J (2014) 462(1): 1-13].

1.3 BRK/PTK6 – Breast tumor kinase / Protein tyrosine kinase 6.

BRK (PTK6) is a human non-receptor protein tyrosine kinase (NRTK) previously classified under the FRK family of kinases (Robinson, Wu et al. 2000) but later on known to form a distinct family of intracellular SRC-like tyrosine kinases based on highly conserved structural and regulatory features (Serfas and Tyner 2003). BRK was first identified during an extensive survey of protein tyrosine kinase mRNAs expressed in normal human melanocytes (Lee, Strunk et al. 1993). It consists of SH3, SH2 and kinase domains, with a proline-rich linker between SH2 and the kinase domains.

1.3.1 Identification and characterization of BRK.

As mentioned above, breast tumor kinase was first identified during an extensive survey of protein tyrosine kinase mRNAs expressed in normal human melanocytes (Lee, Strunk et al. 1993), when it was referred to as PTK6. Subsequently, an identical *brk* gene was isolated and its cDNA was characterized from human metastatic breast tumour using a polymerase chain reaction based differential screening approach. BRK was found to have the capability of autophosphorylation on tyrosine residues when synthesized in baculovirus and bacterial expression systems. Sequencing of the brk cDNA from the human breast cancer cell lines T47D and MCF-7 revealed features of a cytoplasmic tyrosine kinase (Mitchell, Barker et al. 1994).

1.3.2 Gene and protein domain structure, and activity regulation.

Human BRK maps to chromosome 20q13.3 which is a known cancer amplicon. The coding complement of the *brk* gene is composed of 8 exons that encode a catalytically active protein product of 451 amino acids (Mitchell, Barker et al. 1997).

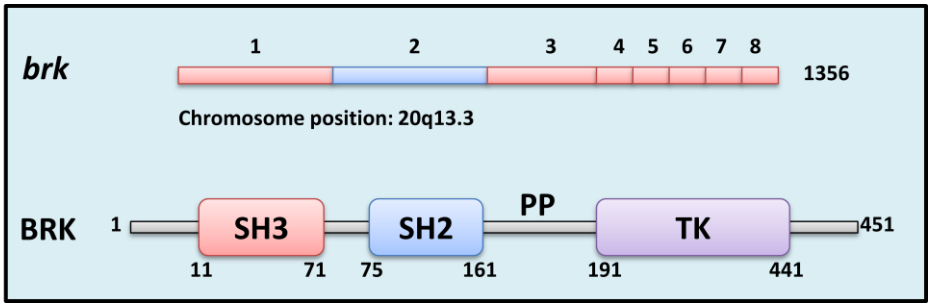


Figure 7. Schematic of gene structure and protein domain structure of BRK.
SH3:SRC homology 3 domain, SH2-SRC homology 2 domain, PP-Polyproline motif, TK-Tyrosine kinase domain.

The protein domain structure of BRK includes SH3 (SRC Homology 3 domain), SH2 (SRC Homology 2 domain) and tyrosine kinase (TK) domains. SH2 and TK domains are connected by a linker rich in proline residues, termed as polyproline (PP) motif. The regulation of BRK was found to be similar to that of SRC kinase attributed to their common structural features with some unique differences. Mutation studies revealed key residues in the BRK protein

which can regulate its activity by modulating the intramolecular interactions. Phosphorylation of C-terminal tyrosine 447 negatively regulates BRK activity by interacting with the SH2 domain, while phosphorylation of tyrosine 342 present in the catalytic domain exhibits positive regulation (Qiu and Miller 2002). Other mutations such as the K219M (Lysine to Methionine) and Y447F (Tyrosine to Phenylalanine) make the kinase inactive and constitutively active respectively (Kamalati, Jolin et al. 1996; Derry, Richard et al. 2000). In the SH3 domain, tryptophan 44 is known to be involved in auto-inhibitory interaction with the proline residues of the linker present between SH2 and catalytic domains. A W44A (Tryptophan to Alanine) mutation breaks the interaction between SH3 domain and proline residues of the linker thus making the kinase active (Qiu and Miller 2004).

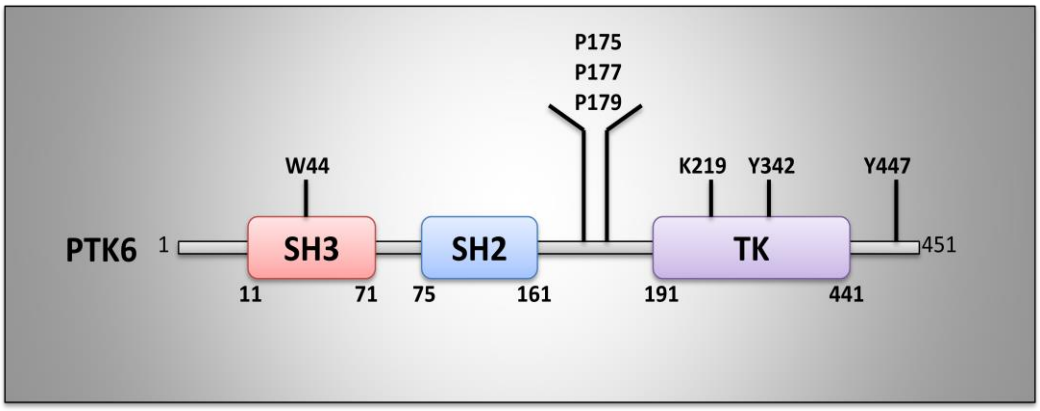


Figure 8: Schematic of important residues of BRK kinase.

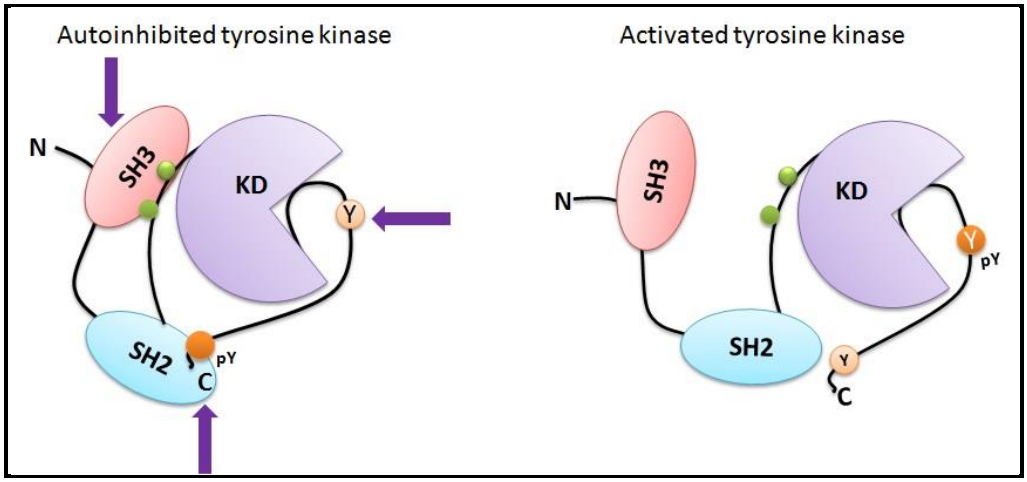


Figure 9. Schematic of the current model of tyrosine kinase regulation applicable to BRK. Arrows indicate the regulatory features.

1.3.3 BRK Signaling.

Potential BRK substrates include the RNA-binding proteins Sam68, SLM-1, SLM-2, and PSF (Coyle, Guzik et al. 2003; Haegebarth, Heap et al. 2004; Lukong, Larocque et al. 2005; Lukong, Huot et al. 2009), transcription factors STAT3 and STAT5a/b (Liu, Gao et al. 2006; Weaver and Silva 2007) and a variety of signaling molecules p190RhoGAP, paxillin, Akt, IRS-4, BKS/STAP-2, and KAP3A (Chen, Shen et al. 2004; Qiu, Zappacosta et al. 2005; Zhang, Ostrander et al. 2005; Lukong and Richard 2008; Shen 2008; Ikeda, Miyasaka et al. 2009). Together, accumulating literature suggests that BRK plays more potential roles in oncogenic signaling than non-oncogenic signaling thus giving room to clinical implications.

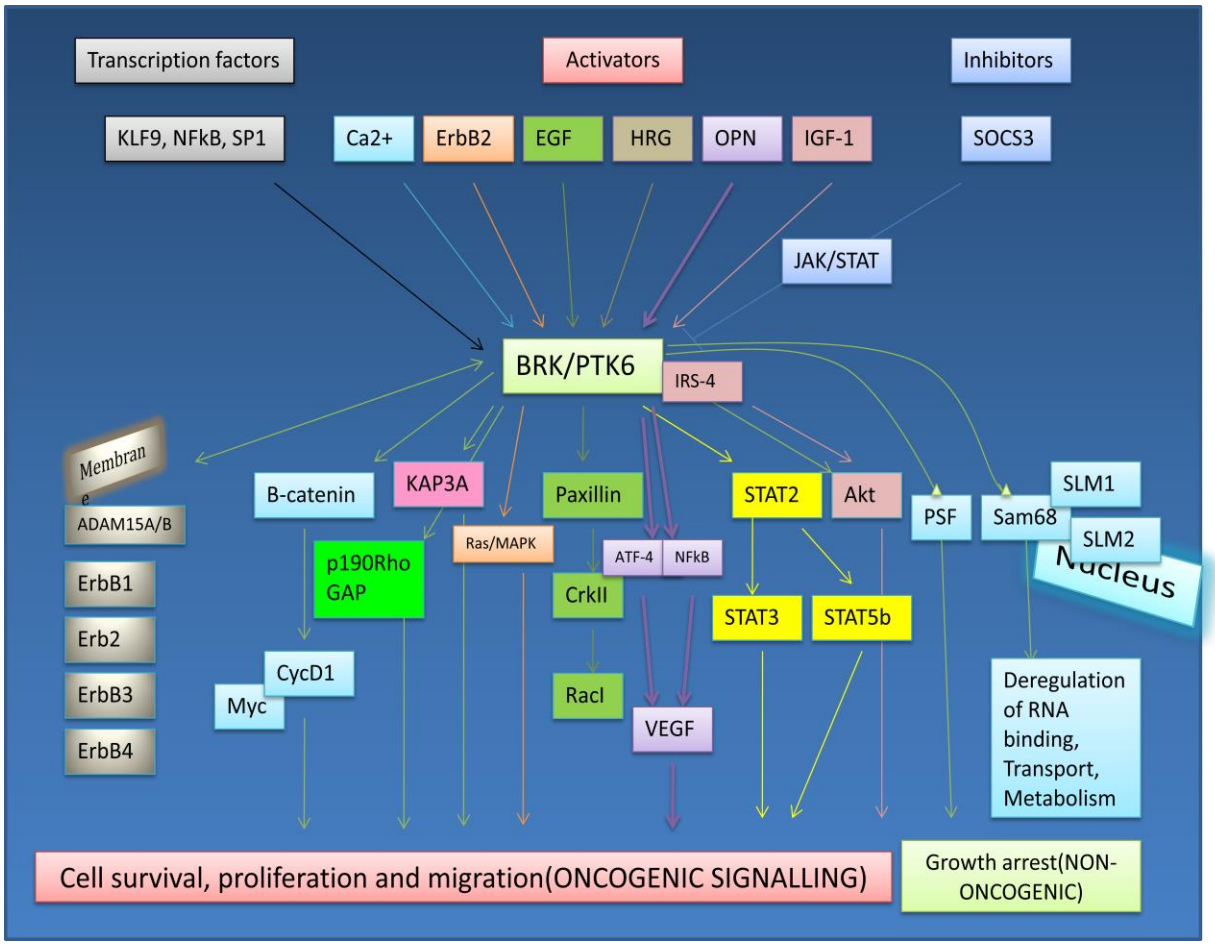


Figure 10. Schematic of BRK/PTK6 signaling pathways.
The signals received and the end effectors are indicated with common colors.

1.3.4 Clinical significance of BRK.

BRK has been gaining clinical significance implicated by its expression patterns in different cancers. 60-80% of human breast carcinomas overexpress BRK unlike the normal breast mammary epithelium (Mitchell, Barker et al. 1994; Barker, Jackson et al. 1997; Aubele, Auer et al. 2007; Ostrander, Daniel et al. 2007; Aubele, Spears et al. 2010). Notable overexpression patterns of BRK were also demonstrated in other major cancers which include mucinous epithelial ovarian cancer and high grade serous carcinoma of the ovary (Heinzelmann-Schwarz, Gardiner-Garden et al. 2006; Schmandt, Bennett et al. 2006), head and neck squamous cell carcinoma (HNSCC) (Lin, Berry et al. 2004; Petro, Tan et al. 2004; Liu, Huang et al. 2013; Liu, Zhang et al. 2013), prostate cancer (Derry, Prins et al. 2003; Chang, Kung et al. 2007; Zheng and Tyner 2013; Zheng, Wang et al. 2013), lymphomas (Kasprzycka, Majewski et al. 2006), colon cancer (Llor, Serfas et al. 1999), non-small cell lung cancer (NSCLC) (Zhao, Chen et al. 2013), metastatic melanomas (David J. Easty1 1997), and pancreatic cancer (Ono, Basson et al. 2014).

1.3.5 BRK inhibitors.

Following an appreciation of the potential clinical importance of BRK in cancer, small molecule inhibitors have been designed by several research groups including those at pharmaceutical companies. A series of substituted imidazo[1,2-a]pyrazin-8-amines were discovered as novel breast tumor kinase (BRK)/protein tyrosine kinase 6 (PTK6) inhibitors with IC50 values ranging between 500nM-7 nM (Zeng, et al. 2011).

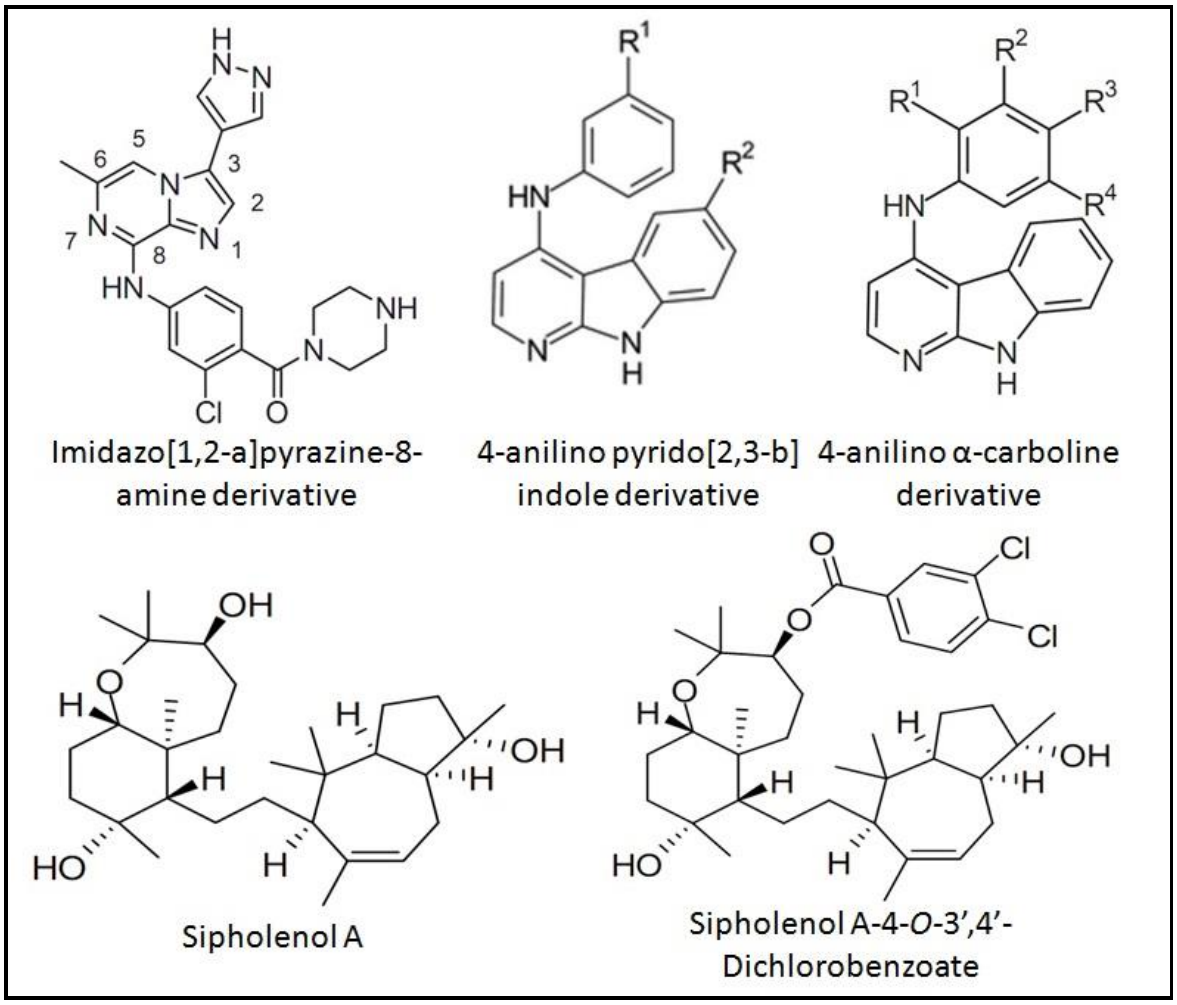


Figure 11. Structural details of BRK inhibitors.

Sipholenol A, a sipholane triterpene isolated from the red sea sponge *Callyspongia siphonella*, and semi-synthetically generated sipholenol derivatives were found to act as potent inhibitors of breast cancer migration and invasion with IC_{50} values of 5.3 and 5.9 μM. A kinome scan of 19,20-anhydrosipholenol A 4β-benzoate ester against 451 human protein kinases identified BRK as a potential target (Foudah AI 2013). Subsequently, an extended study of the Sipholenol A derivative, sipholenol A-4-O-3',4'-dichlorobenzoate revealed the inhibition of breast cancer growth and motility under *in vitro* and *in vivo* conditions through the suppression of Brk and FAK signaling (Mohamed R. Akl 2014). Novel 4-anilino pyrido[2,3-*b*]indoles have been discovered as dual inhibitors of the breast cancer relevant protein kinases Brk and HER2 (Kazem Ahmed Mahmoud 2014). 4-anilino α-carbolines have also been profiled for kinase

activity inhibition and antiproliferative activity confirming a novel group of BRK inhibitors (Mahmoud, et al. 2014). However, none of the small molecules have progressed to clinical trials which underline the need for more research in developing better BRK inhibitors with desired pharmacological profiles.

1.3.6 The BRK splice variant λm5.

Alternate splicing of the primary *brk* transcript due to a skipped second exon (exon 2) and a resultant frameshift mutation generates a distinct mRNA that codes for a truncated protein product of 134 amino acids composed of N-terminal SH3 domain and a unique proline rich C-terminal region (Mitchell, Barker et al. 1997). The regulatory and functional aspects of this splice variant which we refer to here as λm5 are largely unknown. Recently, it was demonstrated that λm5 negatively regulates growth and enhances PTK6-mediated inhibition of β-Catenin signaling (Brauer, Zheng et al. 2011).

A comparison of the protein products BRK and λm5 reveal the intriguing features of the splice variant, *viz.* the SH3 domain identical to full-length BRK and the unique proline-rich region. The characteristic feature of SH3 domains in general is to bind to proline-rich substrates. Accumulating evidence on BRK expression reveals the possibility that the kinase has a role in disease pathology over a range of cancers. This work sought to understand the functional effect of λm5 *in vitro* and to understand its clinical relevance. The results shown here involve studies on the effect of λm5 on the kinase activity of BRK, its effect on cells, and its relative expression levels in different cancer biopsy samples and cancer cell lines. One of the significant efforts of this work was to access and study cancer samples from patients in India.

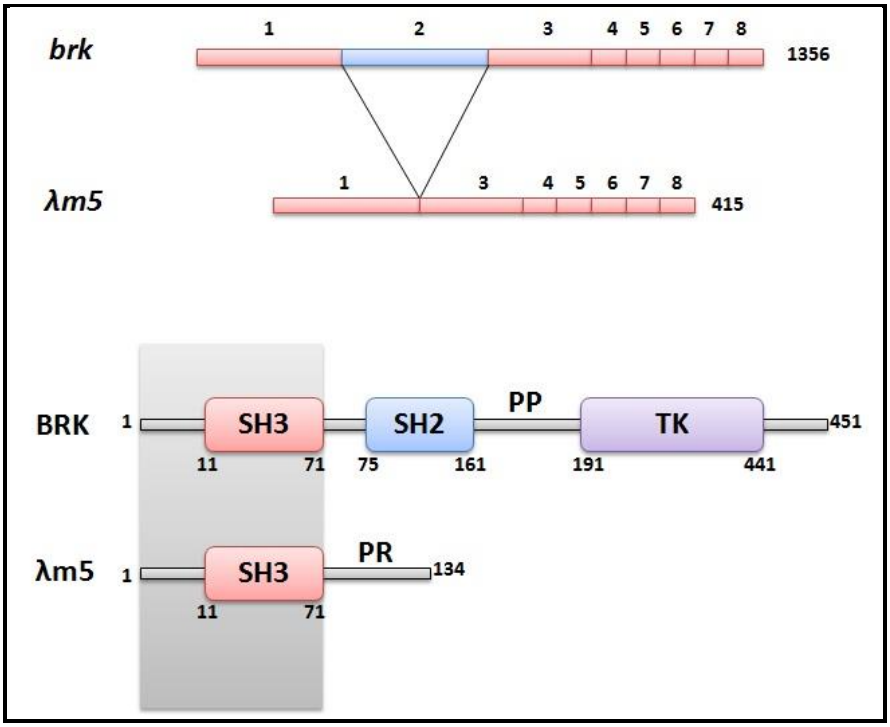


Figure 12. Schematic comparison of brk and λm5 cDNAs and protein domain structures. Upper: Skipping of exon 2 in brk results in generation of λm5. **Lower:** The shifted reading frame in exon 3 codes for a unique proline-rich (PR) region in λm5 with an unchanged SH3 domain.

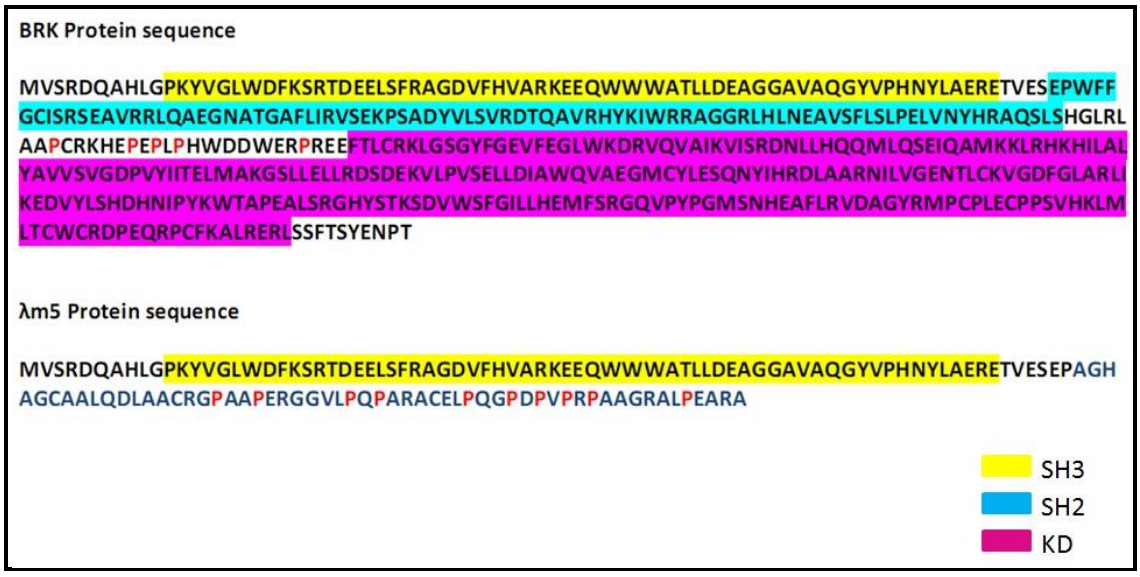


Figure 13. Protein sequences of BRK and λm5. Prolines are indicated in red color**.**

Chapter 2

λm5 regulates the kinase activity of BRK in a biphasic manner and modulates BRK-induced whole cell tyrosine phosphorylation and cell proliferation.

2.1 Cloning and expression of human brk/ptk6 and λm5 cDNAs.

The human *brk* cDNA was constructed by mRNA isolation and reverse transcriptase polymerase chain reaction from the total RNA isolated from a breast cancer cell line T47D(Mitchell, Barker et al. 1994). The human brk cDNA clone we used was a gift from Prof. W. Todd Miller, State University of New York at Stony Brook, NY, USA. The T47D cell line was shown to contain both the full length *brk* and its alternately spliced transcript *λm5* (chapter 1). Based on the reported cDNA sequence of *λm5,* we sought to identify any available sources of the the cDNA clone of *λm5*. We were able to obtain it from NITE (National Institute of Technology and Evaluation) Biological Resource Center (NBRC), Japan. NBRC contains a unique collection of human cDNA clones (14,490) library constructed by oligo-capping method, which included splice variants also (Ota, Suzuki et al. 2004).

2.1.1 Subcloning into bacterial and mammalian expression plasmids.

Both *brk/ptk6* and *λm5* cDNAs were subcloned into the bacterial expression vector pET28a(+) for protein expression and purification, and used in subsequent biochemical experiments including *in vitro* kinase assays. The *ptk6* and *λm5* cDNAs were also cloned into the mammalian expression vectors pCMV-HA and pcDNA3.1/myc-His(-)A respectively. The BRK-K219M kinase-inactive mutant clone was generated from pCMV-HA-PTK6 clone using site-directed mutagenesis, and was also a gift from Prof. W. Todd Miller. The pEGFP-C1-BRK-Y447F (Lukong and Richard 2008) constitutively active kinase mutant plasmid was a gift from Dr. Kiven Erique Lukong, University of Saskatchewan, Canada. All the mammalian expression constructs were used for ectopic expression in cell lines to study the cellular function of λm5.

2.2 λm5 regulates BRK activity in a biphasic manner.

An *in vitro* kinase assay was performed using BRK and λm5 purified using bacterial recombinant expression and Ni-NTA affinity chromatography. The concentration of BRK was maintained at a constant 1μM, with increasing concentrations of λm5 ranging from 1.25μM to 80 μM.

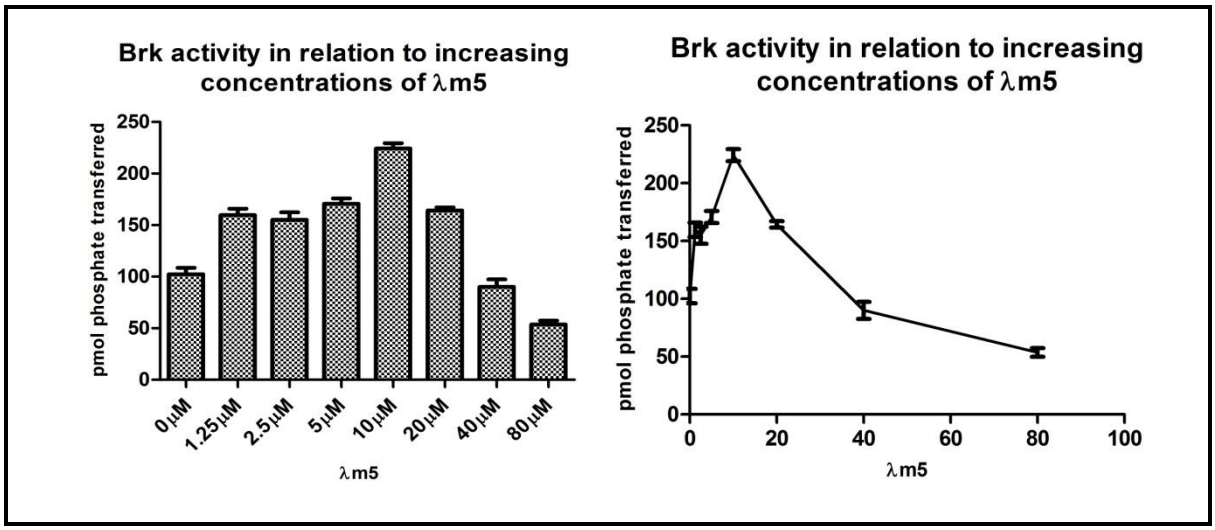


Figure 14. Dose dependent regulation of BRK activity by λm5.

Left: Bar graph indicating the change in BRK activity at increasing concentrations of λm5. Right: Line graph of the same data highlighting the biphasic response of BRK activity with increasing λm5.

Maximum activity of BRK was observed at 10μM concentration of λm5 which was >2 fold compared to BRK alone. The activity declined to <0.5 fold at 80μM concentration of λm5. The results were reproducible and performed with purified protein and appropriate controls. The data led us to conclude that λm5 has a concentration-dependent biphasic response on BRK kinase activity, with increased activity at a concentration up to 10μM and suppressed activity at >10μM λm5.

2.4 λm5 modulates BRK-induced cell proliferation.

Ectopic expression of a constant amount of the constitutively active mutant Brk-Y447F (1μg cDNA) with increasing concentrations of λm5 (0.4μg, 0.8μg, 1.2μg, 1.6μg and 2.0μg cDNAs) in non-cancerous HEK-293T cells revealed a biphasic manner of cell proliferation showing a

maximum effect at 1.2μg cDNA consistent with the data of *in vitro* kinase assay. A similar experiment was performed with the inactive kinase mutant BRK-K219M to confirm the specific role of λm5 in modulating BRK-Y447F induced cell proliferation. It was observed that cell proliferation is unaffected upon ectopic expression of λm5 and BRK-K219M. Taken together, the data led us to conclude that λm5 modulates BRK-induced cell proliferation in a biphasic manner and that this effect is dependent on the tyrosine kinase activity of BRK.

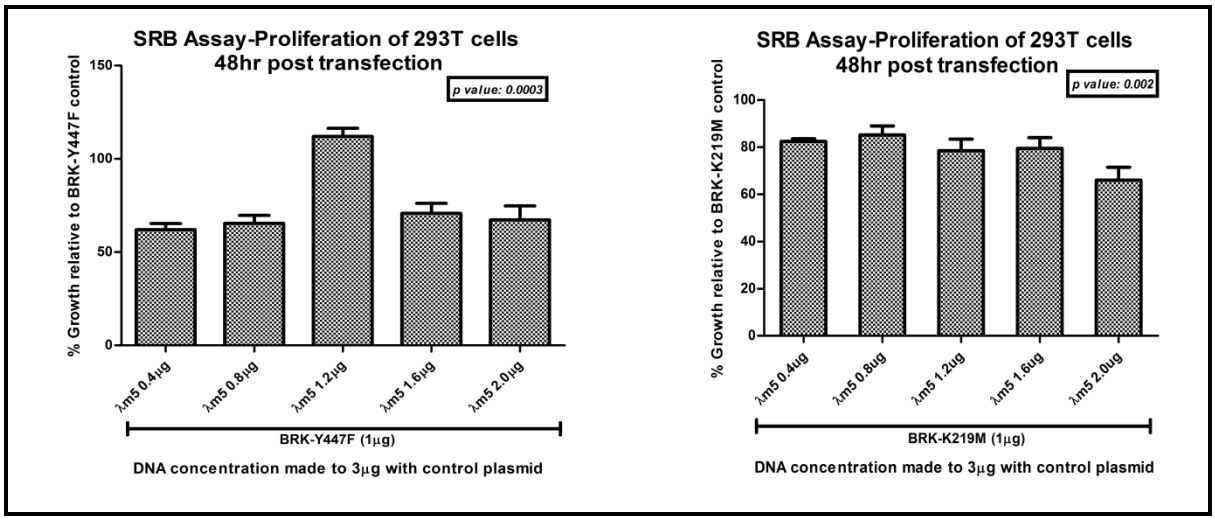


Figure 15. λm5 modulates BRK-induced cell proliferation in a biphasic manner.
Left: Cell proliferation in the presence of constitutively active BRK mutant and increasing amounts of λm5. Right: The same experiment with inactive BRK mutant.

2.3 λm5 modulates BRK-induced whole cell tyrosine phosphorylation

A constant amount of constitutively active mutant BRK-Y447F cDNA (1μg) with increasing amounts of λm5 cDNA (0.4μg, 0.8μg, 1.2μg, 1.6μg and 2.0μg) were used to transfect 293T cells. Western blotting of whole cell lysates of the samples at 48h post-transfection was done using an anti-phospho-tyrosine antibody to detect tyrosine phosphorylation status of cellular proteins. The results revealed a biphasic response with respect to whole cell tyrosine phosphorylation. Plots of densitometric scans of western blots show an approximately four-fold increase of whole cell tyrosine phosphorylation at 1.2μg of λm5 compared to 0μg. This biphasic response was clearly consistent with the cell proliferation data, which also showed

maximal cell proliferation at an identical amount of λm5. The biphasic response effect of λm5 on the constitutively active kinase mutant BRK-Y447F was absent when a similar experiment was performed with the inactive kinase mutant BRK-K219M. The results led us to conclude that the effect of λm5 on cellular tyrosine phosphorylation is dependent on BRK activity.

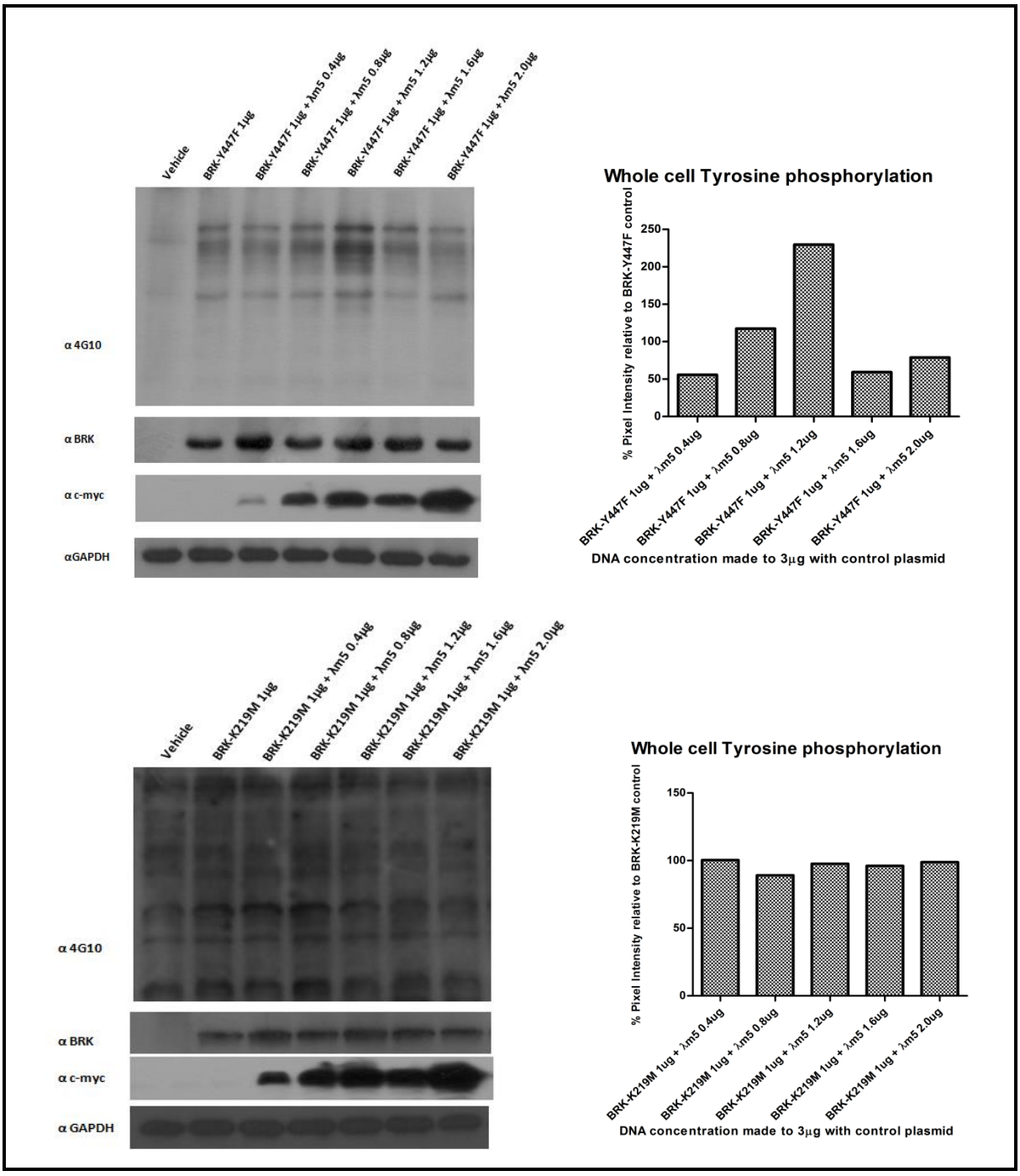

Vehicle
BRK-Y447F 1μg
BRK-Y447F 1μg + λm5 0.4μg
BRK-Y447F 1μg + λm5 0.8μg
BRK-Y447F 1μg + λm5 1.2μg
BRK-Y447F 1μg + λm5 1.6μg
BRK-Y447F 1μg + λm5 2.0μg
α 4G10
α BRK
α c-myc
αGAPDH
Whole cell Tyrosine phosphorylation
% Pixel Intensity relative to BRK-Y447F control
250
200
150
100
50
0
BRK-Y447F 1ug + λm5 0.4ug
BRK-Y447F 1ug + λm5 0.8ug
BRK-Y447F 1ug + λm5 1.2ug
BRK-Y447F 1ug + λm5 1.6ug
BRK-Y447F 1ug + λm5 2.0ug
DNA concentration made to 3μg with control plasmid
Vehicle
BRK-K219M 1μg
BRK-K219M 1μg + λm5 0.4μg
BRK-K219M 1μg + λm5 0.8μg
BRK-K219M 1μg + λm5 1.2μg
BRK-K219M 1μg + λm5 1.6μg
BRK-K219M 1μg + λm5 2.0μg
α 4G10
α BRK
α c-myc
α GAPDH
Whole cell Tyrosine phosphorylation
% Pixel Intensity relative to BRK-K219M control
150
100
50
0
BRK-K219M 1ug + λm5 0.4ug
BRK-K219M 1ug + λm5 0.8ug
BRK-K219M 1ug + λm5 1.2ug
BRK-K219M 1ug + λm5 1.6ug
BRK-K219M 1ug + λm5 2.0ug
DNA concentration made to 3μg with control plasmid

Figure 16 (previous page): λm5 modulates BRK-induced tyrosine phosphorylation in a biphasic manner. Top: Phosphotyrosine western blot with active BRK and Densitometry. Bottom: Phosphotyrosine western blot with inactive BRK and Densitometry. α 4G10 is phosphotyrosine-specific antibody; α c-myc represents expression of λm5 since it was myc-tagged; An λm5 antibody is currently not available.

2.5 Summary

Brk and λm5 were cloned into bacterial expression vectors and the expressed proteins purified using Ni-NTA affinity chromatography. The proteins were used to perform *in vitro* kinase assays with a constant concentration of BRK and increasing concentrations of λm5. The assay was performed to test the effect of λm5 on BRK kinase activity. We found that λm5 regulates the kinase activity of BRK in a dose-dependent biphasic manner, with BRK activity varying between ~2-fold above baseline at 10μM λm5, and ~0.5-fold below baseline at 80μM λm5. A similar phenomenon was observed under cellular conditions, where a constant amount of active Brk DNA was transfected with increasing doses of λm5 DNA into non-cancerous human embryonic kidney fibroblast cells (HEK-293T cells). We found that at 48h post-transfection, cell proliferation and whole cell tyrosine phosphorylation of HEK-293T cells followed a biphasic response, with ~2-fold effect above baseline at 1.2μg of λm5 cDNA, dropping to baseline at higher amounts of λm5 cDNA. This effect was reproducible and consistent with the *in vitro* kinase assay data. The specific requirement of BRK activity in this phenomenon was confirmed using an inactive BRK mutant under similar experimental conditions, where no significant effect was noticed.

The above results confirmed that λm5 regulates the activity of BRK *in vitro* and in cultured cells in a biphasic manner. Further, λm5 also affects cell proliferation similarly, indicating that its expression is relevant in the cellular context.

Chapter 3

Cancer-specific expression profiles of brk and λm5.

The dose-dependent regulatory effect of λm5 on *in vitro* BRK activity and its cellular effects led us to investigate the expression pattern of λm5 and its levels relative to BRK. We conducted our studies using cancer cell lines, cancer tissue mRNA arrays and finally using cancer patient tissue biopsies from a local hospital in Hyderabad, India (The Basavatarakam Indo-American Cancer Hospital and Research Institute). Selective antibodies that recognize BRK and λm5 are unavailable. We therefore chose to use mRNA detection to selectively detect and quantify BRK and λm5 levels. In order to carry out this study, it was necessary to first design primers able to differentiate between brk and λm5 mRNA transcripts.

3.1 Differentiation of *brk* and *λm5* transcripts.

The major difference between *brk* and *λm5* is the absence of exon 2 in *λm5*. The strategy we adopted was to design primers binding within exon 2 for brk, and those binding the junction of exon 1 and 3 for λm5 (since this junction is absent in the brk transcript). Initially, 3 sets of semi-quantitative primers were designed:

Set 1 which can amplify both *brk* and *λm5* transcripts.

Set 2 which can amplify brk selectively.

Set 3 which can amplify *λm5* selectively. Set 3 has two sub-sets of primers

Sequences of brk and λm5 cDNAs:

> [brk cDNA] - 1356 bp (Highlighted background colour indicate alternate exons)

ATGGTGTCCCGGGACCAGGCTCACCTGGGCCCCAAGTATGTGGGCCTCTGGGACTTCA
AGTCCCGGACGGACGAGGAGCTGAGCTTCCGCGCGGGGGACGTCTTCCACGTGGCCA
GGAAGGAGGAGCAGTGGTGGTGGGCCACGCTGCTGGACGAGGCGGGTGGGGCCGT
GGCCCAGGGCTATGTGCCCCACAACTACCTGGCCGAGAGGGAGACGGTGGAGTCGGA
AC**C**GTGGTTCTTTGGCTGCATCTCCCGCTCGGAAGCTGTGCGTCGGCTGCAGGCCGAG

GGCAACGCCACGGGCGCCTTCCTGATCAGGGTCAGCGAGAAGCCGAGTGCCGACTAC
GTCCTGTCGG**T**GCGGGACACGCAGGCTGTGCGGCACTACAAGATCTGGCGGCGTGCC
GGGGGCCGGCTGCACCTGAACGAGGCGGTGTCCTTCCTCAGCCTGCCCGAGCTTGTGA
ACTACCACAGGGCCCAGAGCCTGTCCCACGGCCTGCGGCTGGCCGCGCCCTGCCGGAA
GCACGAGCCTGAGCCCCTGCCCCATTGGGATGACTGGGAGAGGCCGAGGGAGGAGTT
CACGCTCTGCAGGAAGCTGGGGTCCGGCTACTTTGGGGAGGTCTTCGAGGGGCTCTG
GAAAGACCGGGTCCAGGTGGCCATTAAGGTGATTTCTCGAGACAACCTCCTGCACCAG
CAGATGCTGCAGTCGGAGATCCAGGCCATGAAGAAGCTGCGGCACAAACACATCCTGG
CGCTGTACGCCGTGGTGTCCGTGGGGGACCCCGTGTACATCATCACGGAGCTCATGGC
CAAGGGCAGCCTGCTGGAGCTGCTCCGCGACTCTGATGAGAAAGTCCTGCCCGTTTCG
GAGCTGCTGGACATCGCCTGGCAGGTGGCTGAGGGCATGTGTTACCTGGAGTCGCAG
AATTACATCCACCGGGACCTGGCCGCCAGGAACATCCTCGTCGGGGAAAACACCCTCTG
CAAAGTTGGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGACGTCTACCTCTCCCAT
GACCACAATATCCCC
TACAAGTGGACGGCCCCTGAAGCGCTCTCCCGAGGCCATTACTCCACCAAATCCGACGT
CTGGTCCTTTGGGATTCTCCTGCATGAGATGTTCAGCAGGGGTCAGGTGCCCTACCCA
GGCATGTCCAACCATGAGGCCTTCCTGAGGGTGGACGCCGGCTACCGCATGCCCTGCC
CTCTGGAGTGCCCGCCCAGCGTGCACAAGCTGATGCTGACATGCTGGTGCAGGGACCC
CGAGCAGAGACCCTGCTTCAAGGCCCTGCGGGAGAGGCTCTCCAGCTTCACCAGCTAC
GAGAACCCGACCTGA

> [λm5 cDNA] - 1234 bp (With skipped exon 2 and the merge point indicated in bold and increased font size)

ATGGTGTCCCGGGACCAGGCTCACCTGGGCCCCAAGTATGTGGGCCTCTGGGACTTC
AAGTCCCGGACGGACGAGGAGCTGAGCTTCCGCGCGGGGGACGTCTTCCACGTGGC
CAGGAAGGAGGAGCAGTGGTGGTGGGCCACGCTGCTGGACGAGGCGGGTGGGGC
CGTGGCCCAGGGCTATGTGCCCCACAACTACCTGGCCGAGAGGGAGACGGTGGAGT
CGGAAC**CT**GCGGGACACGCAGGCTGTGCGGCACTACAAGATCTGGCGGCGTGCCG
GGGGCCGGCTGCACCTGAACGAGGCGGTGTCCTTCCTCAGCCTGCCCGAGCTTGTG
AACTACCACAGGGCCCAGAGCCTGTCCCACGGCCTGCGGCTGGCCGCGCCCTGCCGG
AAGCACGAGCCTGAGCCCCTGCCCCATTGGGATGACTGGGAGAGGCCGAGGGAGG
AGTTCACGCTCTGCAGGAAGCTGGGGTCCGGCTACTTTGGGGAGGTCTTCGAGGGG
CTCTGGAAAGACCGGGTCCAGGTGGCCATTAAGGTGATTTCTCGAGACAACCTCCTG
CACCAGCAGATGCTGCAGTCGGAGATCCAGGCCATGAAGAAGCTGCGGCACAAACA
CATCCTGGCGCTGTACGCCGTGGTGTCCGTGGGGGACCCCGTGTACATCATCACGGA

GCTCATGGCCAAGGGCAGCCTGCTGGAGCTGCTCCGCGACTCTGATGAGAAAGTCCT
GCCCGTTTCGGAGCTGCTGGACATCGCCTGGCAGGTGGCTGAGGGCATGTGTTACC
TGGAGTCGCAGAATTACATCCACCGGGACCTGGCCGCCAGGAACATCCTCGTCGGG
GAAAACACCCTCTGCAAAGTTGGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGAC
GTCTACCTCTCCCATGACCACAATATCCCCTACAAGTGGACGGCCCCTGAAGCGCTCT
CCCGAGGCCATTACTCCACCAAATCCGACGTCTGGTCCTTTGGGATTCTCCTGCATGA
GATGTTCAGCAGGGGTCAGGTGCCCTACCCAGGCATGTCCAACCATGAGGCCTTCCT
GAGGGTGGACGCCGGCTACCGCATGCCCTGCCCTCTGGAGTGCCCGCCCAGCGTGC
ACAAGCTGATGCTGACATGCTGGTGCAGGGACCCCGAGCAGAGACCCTGCTTCAAG
GCCCTGCGGGAGAGGCTCTCCAGCTTCACCAGCTACGAGAACCCGACCTGA

Primer	Sequence (5'- 3')	Feature	Amplicon size (bp)
Set 1	CCCGCCCGCCATGGTGTC (for)	Amplifies both brk and λm5	525
	TCAGGCTCGTGCTTCCGGCA (rev)		405
Set 2	GTGGTTCTTTGGCTGCATCTCCCG (for)	Amplifies only brk	1125
	TCAGGTCGGGTTCTCGTAGCTGG TG (rev)		
Set 3	AAGGAGGAGCAGTGGTG (for)	Amplifies only λm5	120
	TCCCGCA***GGTTCCGA*** (rev)		

Table 1. Table showing the list of semi-quantitative PCR primer sets used for differentiation of brk and λm5. (for) and (rev) represent forward and reverse primer respectively. Italicized bold region of Set 3 reverse primers represent the junction primers spanning the junction of exon 1 and 3 in λm5.

The primers were used in multiple rounds of PCR reactions to optimize the conditions for the required amplification. The reactions were optimized and we were successful in achieving desired amplifications as shown in figures 17 and 18.

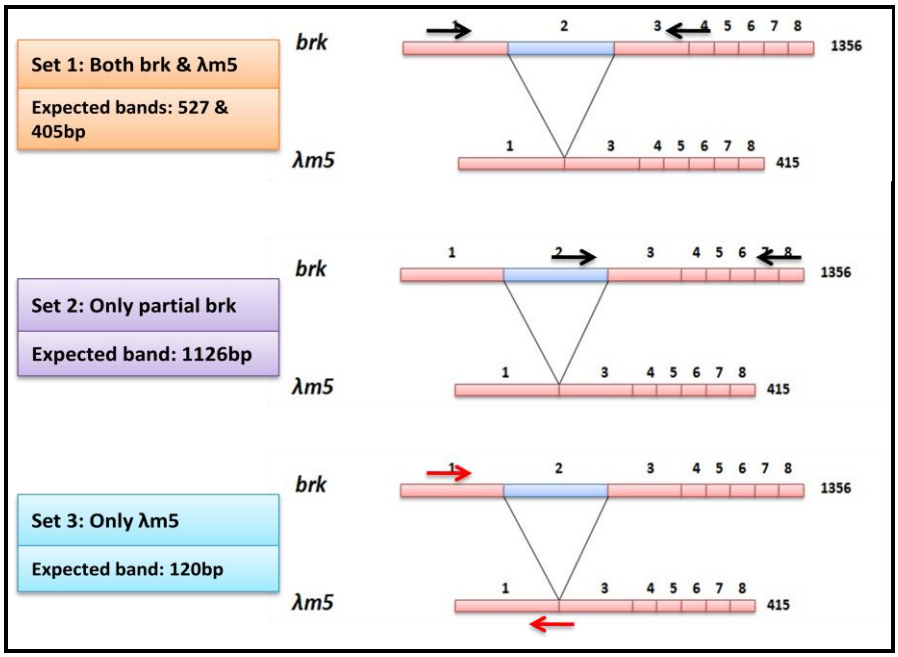


Figure 17. Schematic for designing PCR primers for brk and λm5.

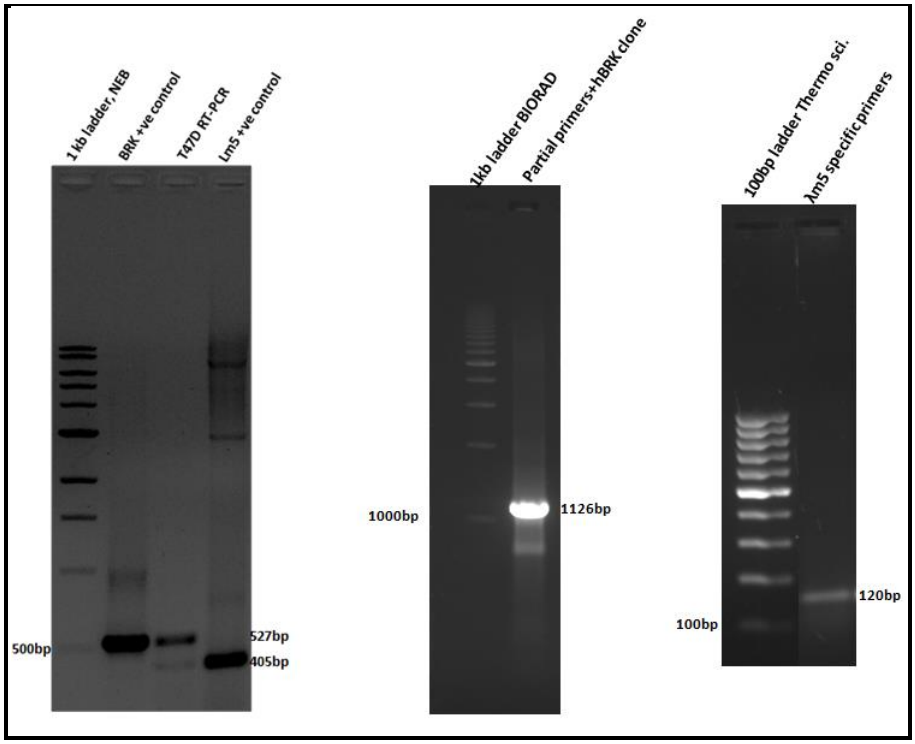


Figure 18. Semi-quantitative PCR of brk and λm5.
Images are results from Set 1, Set 2 and Set 3 (from left to right).

Based on the above primers, PCR primers were designed for use in real-time quantitative PCR (qPCR) reactions. Brk-selective primers for qPCR were designed to bind within exon 2. The set 3 primers used above for semi-quantitative PCR were used as such for qPCR since they were selective to λm5 and met the essential amplicon size requirement for qPCR analysis.

Primer	Sequence	Feature	Amplicon size (bp)
BRK exon 2	TGGTTCTTTGGCTGCATCT (for)	brk/ptk6 specific	117
	CAGGACGTAGTCGGCACTC (rev)		
λm5 qPCR primers	AAGGAGGAGCAGTGGTG (for)	λm5 specific	120
	TCCCGCA***GGTTCCGA*** (rev)		

Table 2: Table showing the list of quantitative PCR (qPCR) primers selective to brk and λm5.

3.2 Expression profiles of brk and λm5 in different cancer cell lines.

qPCR primers were used in test reactions to optimize the reaction conditions. The optimized qPCR primers were used to check the expression levels of brk and λm5 in six different cancer cell lines. The non-cancerous cell line HEK-293T (kidney epithelial cell line) was used as a control for data normalization. The cell lines represented different cancer types which included A549 (lung carcinoma), K562 (chronic myelogenous leukemia (CML), T47D (breast carcinoma), HepG2 (hepatocyte carcinoma), TZM-BL (HeLa human cervical carcinoma) and CAL-27 (tongue adenosquamous cell carcinoma).

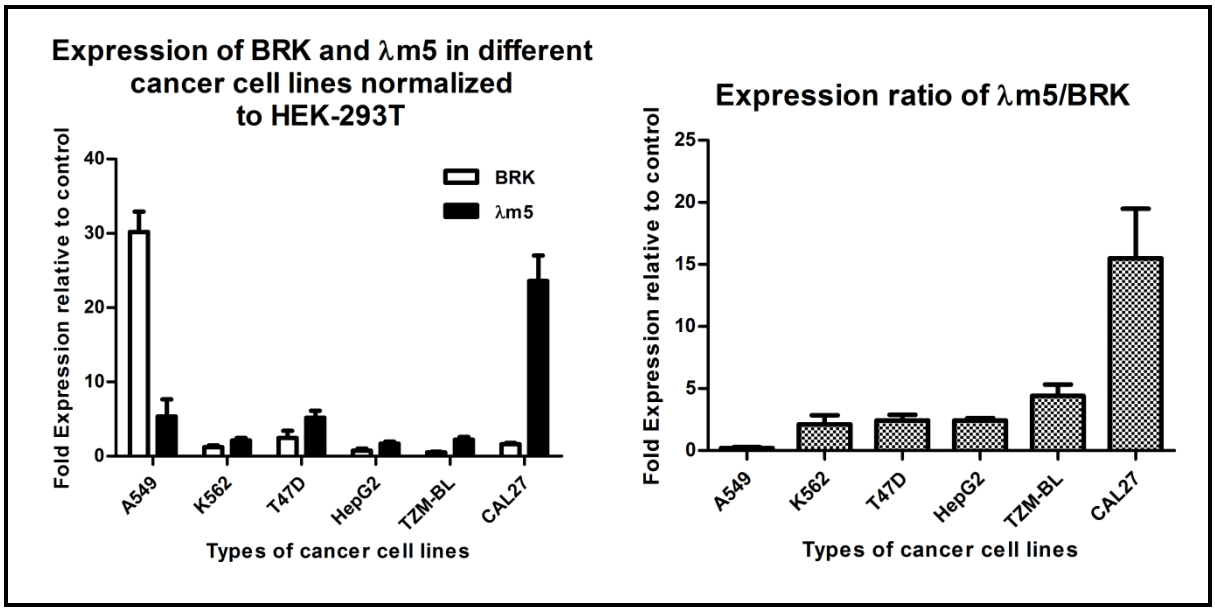


Figure 19. Relative expression profiles of brk and λm5

Left: Relative expression of brk and λm5 in cancer cell lines. Right: Expression ratio of λm5/brk in cancer cell lines.

Cell line	Brk	λm5	Fold expression of λm5 relative to Brk
A549	30.17611	5.346938	0.177191096
K562	1.226416	2.130585	1.737244948
T47D	2.451748	5.167751	2.107782284
HepG2	0.753342	1.683092	2.234167218
TZM-BL	0.528046	2.227925	4.219187344
CAL27	1.6071	23.58021	14.67252193

Table 3. Expression study with cancer cell lines.

Among the cancer cell lines, the expression of λm5 was higher than brk in all cases except the A549 (lung carcinoma) cell line, where with λm5 expression was $1/10^{th}$ (0.1 fold) that of brk. In K562 (chronic myelogenous leukemia), T47D (breast ductal carcinoma), HepG2 (hepatocyte carcinoma), and TZM-BL (HeLa human cervical carcinoma), λm5 expression was

higher than brk with a higher a fold difference range of approximately 2-5 times. We found the highest relative expression of λm5 in the CAL27 (tongue adenosquamous carcinoma) cell line, with a fold difference of approximately 15 times.

3.3 Expression profiles of brk and λm5 in Indian cancer patient biopsy samples.

We next focused our attention on the expression levels of brk and λm5 in cancer patients. Adhering to all prescribed institutional and hospital guidelines for human ethics, we sought and acquired access to a collection of 14 pairs of matched (cancer vs. normal) patient samples from a local hospital (Basavatarakam Indo-American Cancer Hospital & Research Institute (BIACHRI)). The collection of samples was from various sites of the human body, depending on sample availability and suitability for study. Such a collection also allowed us to assess λm5 and brk expression among various anatomical sites. The hospital guidelines allowed us access to patient age, gender, diagnosis and medical record number, which we used in our analysis.

Cancer biopsy sample processing.

Cancer tissue biopsy samples were collected and put in RNA later solution and stored at 4^0C. After 24 hours, the RNA later solution was drained out completely and the samples were stored frozen at -80^0C. During the processing step, the samples were taken out of the freezer and immediately cut into small bits with a surgical blade and suspended in 0.75ml of TRIzol (Life technologies lot#47114). The cut bits were put in 1.5ml microcentrifuge tubes along with a glass bead and lysed in a tissue disruptor (TissueLyser) for 5 min. The lysates were collected and processed for total RNA isolation as per the kit manufacturer's instructions. Total RNA isolated was stored at -80^0C for subsequent usage. The processing till the biopsy tissue lysis was carried out at the collection site (BIACHRI) following good clinical practice (GCP) guidelines. The total RNA was used for first strand cDNA synthesis and the samples were analyzed using quantitative PCR.

The results from the qPCR analysis of cancer biopsy samples showed that ten out of fourteen samples had a higher expression of λm5 compared to brk. Cancer samples derived from buccal mucosa, neck, colon and leukocytes showed lower expression of λm5 compared to

BRK, ranging from 1/20th to negligible expression. The anatomical sites maxillary sinus, tonsil (IA), gastroesophageal (GE) junction, breast, prostate, rectum, cervix, lymphocytes and bone marrow expressed a higher level of λm5 than brk with fold difference range of 2-13 times. Interestingly, one sample derived from a T-lymphoblastic lymphoma/leukemia metastasized to the bone marrow showed 75-fold higher expression of λm5 compared to brk. The data also show that 2 out of 3 patients under 20 years of age showed higher λm5 expression than brk. Overall, most cancer types in this work revealed a relatively higher expression of λm5 than brk.

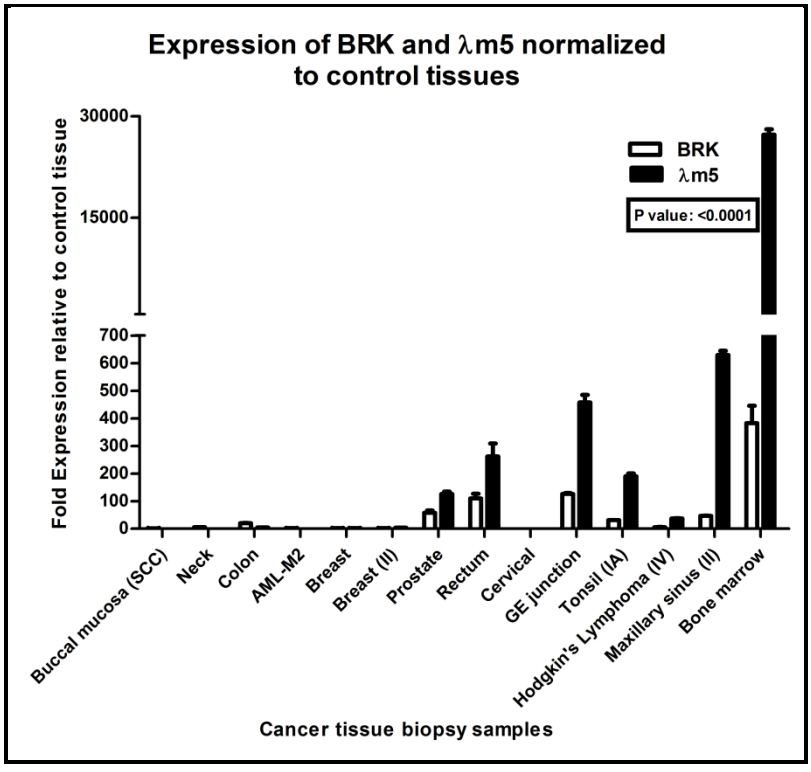


Figure 20. Expression profiles of brk and λm5 in clinical samples.

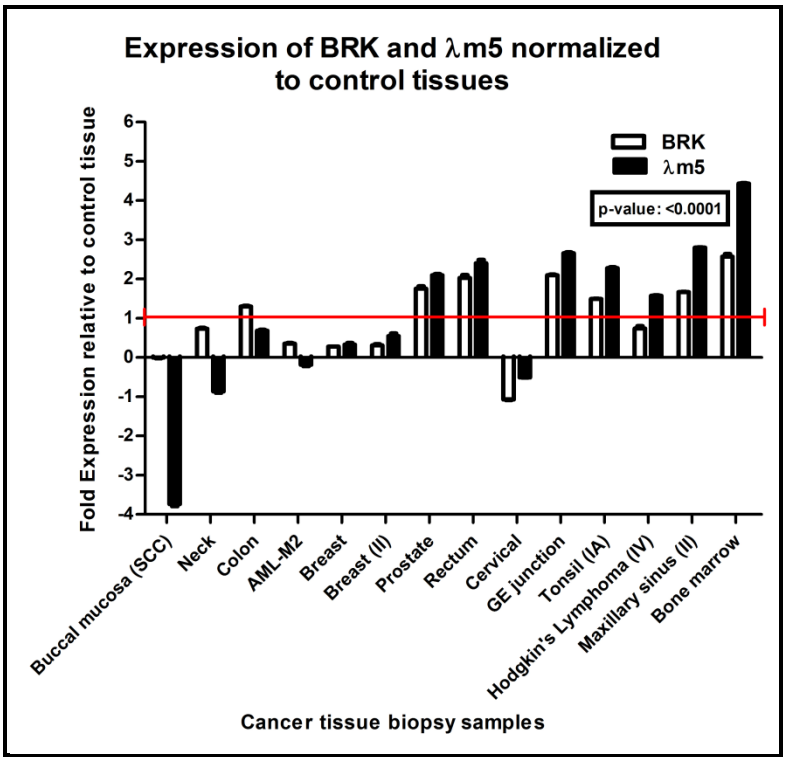


Figure 21. Expression profiles of brk and λm5 in clinical samples (log scale).

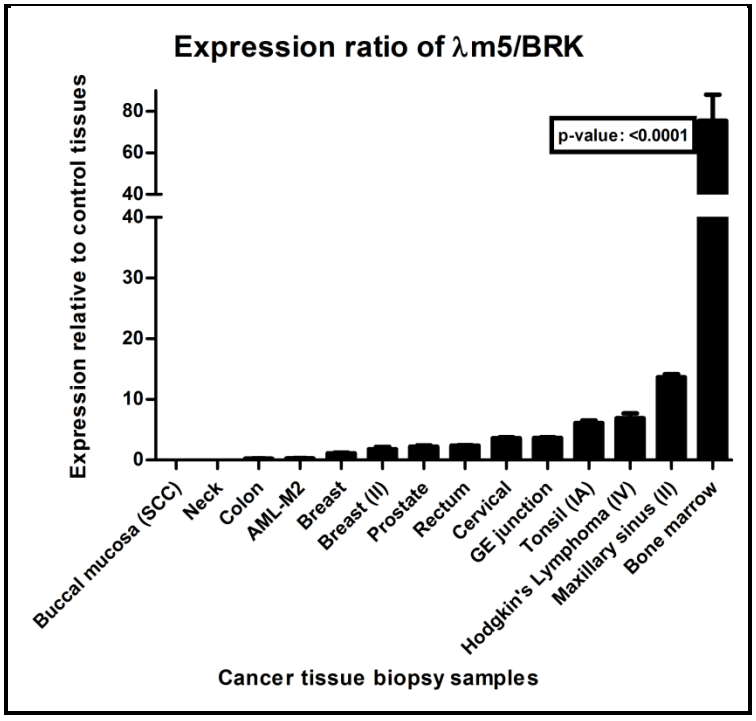


Figure 22. Expression ratio of λm5 and brk in clinical samples.

MR No	Age	Gender	Diagnosis	Fold expression of λm5 relative to brk
82643	54	M	Incisional biopsy from right buccal mucosa	0.00019
80144	52	M	MUO Neck, Post RT with residual disease.	0.025516
75726	40	F	Ca. Ascending Colon (p T3, N1, M0)	0.24551
80699	17	M	Refractory AML-M2	0.290895
82014	60	F	Ca. Left Breast, Ca. Right Ovary	1.137575
82338	67	F	Ca. Right Breast, Triple negative	1.807625
82579	75	M	Ca. Prostate	2.219989
81924	43	M	Rec. Ca. Caecum	2.366084
56566	40	F	Papsmear (post RT) Cervix	3.628082
82378	58	M	Ca. GE junction	3.637561
80732	58	F	Ca. Right Tonsil (Stage 1A)	6.116397
80865	18	M	Hodgkins's disease, Stage IV	6.880068
82137	40	M	Ca. Maxilla Right - Squamous cell carcinoma	13.63854
74839	17	M	T - Lymphoblastic lymphoma with BM involvement/Leukemia	75.45622

Table 4. Details of clinical samples along with expression details of λm5.

MR-Medical Record; MUO-Metastasis from unknown origin; RT-Radio-Therapy; Ca.-Carcinoma; AML-M2-Acute Myeloid Leukemia-Maturity level 2; GE-Gastro-Esophageal; T.-T- cell; BM-Bone Marrow.

3.4 Expression study of brk and λm5 with cancer tissue cDNA array.

In order to further confirm our findings with cancer-specific λm5 and brk expression, we sought to extend our study using other sources of cancer tissue. We procured a Tissuescan Starter qPCR array kit from Origene technologies (cat#TSRT101). The kit contained first-strand cDNAs prepared from human cancer and normal tissues covering 4 cancer types: breast, kidney, lung and ovary. The cDNAs were arrayed onto a PCR plate as a part of the kit. We followed the kit manufacturer's instructions and performed PCR using selective primers for λm5 and Brk. We then analyzed the fold-expression data with cancer type and stage. We were able to obtain and analyze data from 19 cancer tissues of different stages.

Our results showed variable λm5 and brk expression ratios depending on cancer type and stage. Overall, 8 out of 19 total samples expressed higher λm5 than Brk. Ten samples showed higher brk expression and one sample had equal expression of both. At least one sample in each type had higher λm5 than Brk: 3 of 4 breast cancer samples (stages IIA, IIB and IIIC), 2 of 5 kidney cancer samples (stage II), 1 of 5 lung cancer samples (stage IIB) and 1 of 5 ovarian cancer samples (stage IIIC). A trend of higher λm5 in advanced cancer stages with lower λm5 in early cancer stages is apparent from our results. None of the samples classified as stage I showed higher λm5, whereas all of them showed higher Brk. Analysis of the expression ratio revealed that the fold difference of λm5 expression over brk ranged from 2-6 fold. In samples showing lower λm5 than Brk, the fold difference ranged from 0.08-0.6 fold. Samples showing higher Brk relative to λm5 did not show a trend with reference to cancer type or stage.

The results from this work indicate that the expression levels of λm5 and Brk are potentially discriminatory. If the expression and expression ratios were largely similar across cancer types and stages, the power of discrimination and stratification in clinical diagnosis is lost. Our work supports the feasibility of using λm5 and Brk expression and expression ratio as a diagnostic and/or prognostic marker. This work, though limited in scope, provides a sound rational basis for follow-up studies.

Follow-up studies should ideally involve the use of follow-up patient data (such as disease progression and mortality, biochemistry, hematology and pathology data), relatively larger sample number and use of a multiple detection methods (such as protein detection in addition to mRNA).

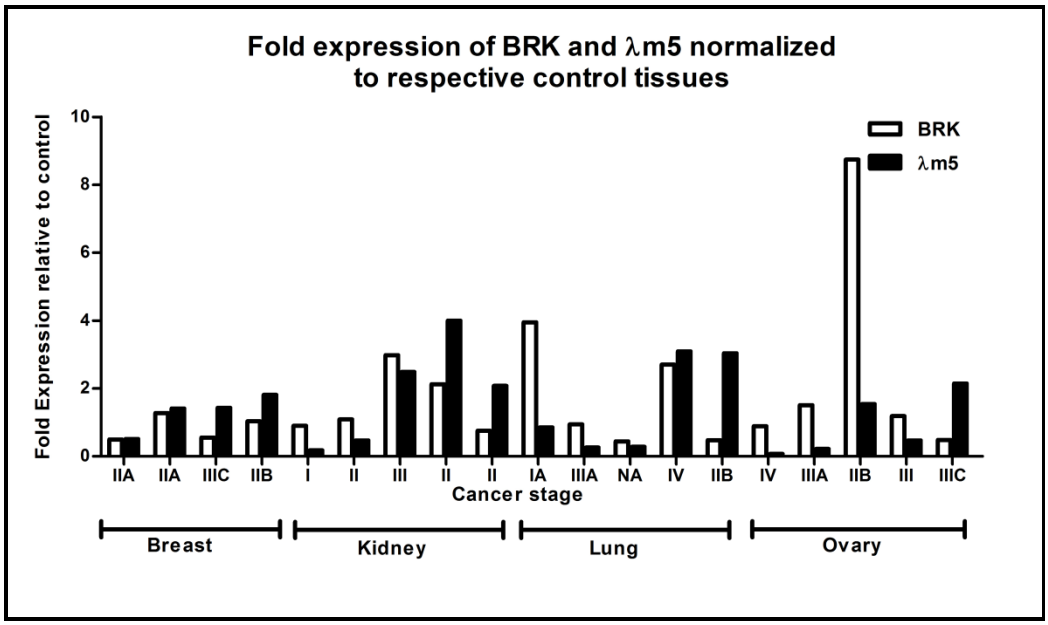


Figure 23. Quantitative expression profiles of brk and λm5 using a cancer qPCR array kit. Stages are defined differently as per the cancer type but follow the same meaning of increasing risk with increasing roman number indication.

Cancer type	Cancer grade	Fold expression of λm5 relative to brk
Breast	IIA	1.021
Breast	IIA	1.10151
Breast	IIIC	1.75139
Breast	IIB	2.59657
Kidney	I	0.195543
Kidney	II	0.431921
Kidney	III	0.835678
Kidney	II	1.88456
Kidney	II	2.75347
Lung	IA	0.213977
Lung	IIIA	0.276787
Lung	NA	0.629218
Lung	IV	1.14227
Lung	IIB	6.37018
Ovary	IV	0.0826656
Ovary	IIIA	0.144283
Ovary	IIB	0.17643
Ovary	III	0.395507
Ovary	IIIC	4.45328

Table 5. Details of samples in the cancer tissue cDNA array along with expression details of λm5.

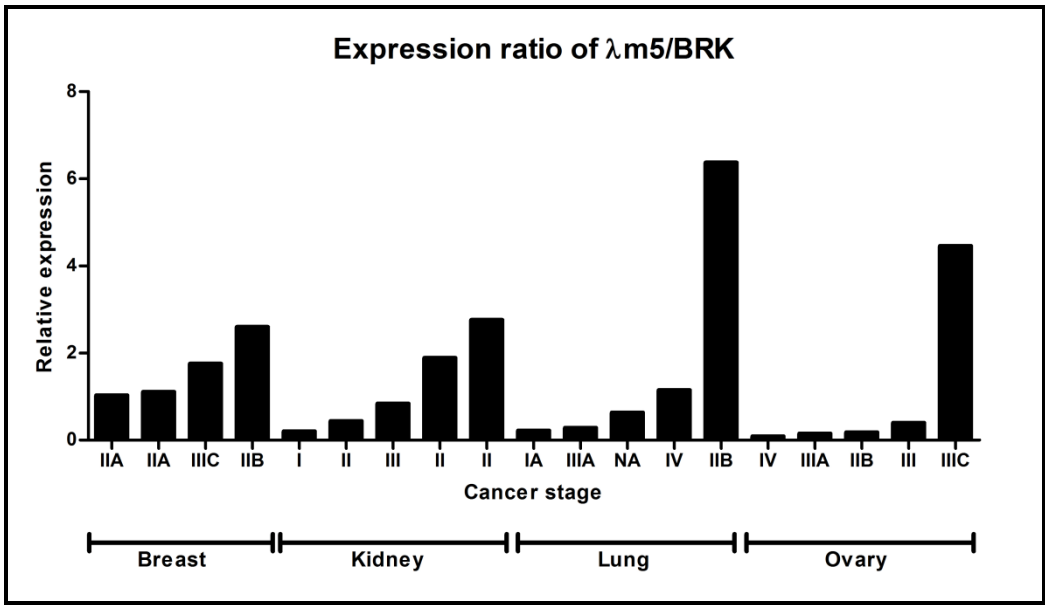


Figure 24. Relative expression ratio of λm5 and brk in cancer tissue qPCR array. Stages are defined differently as per the cancer type but follow the same meaning of increasing risk with increasing roman number indication.

3.5 Summary

The primers for quantitative expression analysis of Brk and λm5 were designed based on the differential features present in the transcripts and were well-standardized. The expression analysis was initially carried out with a set of cancer cell lines we could obtain and culture. These included 6 types of cell lines which represent cancers of different anatomical origins. It was clear from the study that the splice variant λm5 was found to be higher in expression levels than Brk in all except A549 (lung cancer cell line) with a fold difference range of 1-14 times. This led us to extend the study with a collection of 14 matched pairs (cancer vs. normal) of cancer patient biopsy samples which represented cancers of various anatomical origins, as well as a cancer tissue cDNA qPCR array kit. The patient samples were acquired from a local hospital, the Basavatarakam Indo-American Cancer Hospital and Research Institute (BIACHRI) with necessary institutional and hospital ethics committee approvals. The study indicated that expression of λm5 is higher than Brk in most cases with an interesting trend towards metastasis. The fold difference in expression (λm5/Brk) was in the range of 2-75 times. A mixed pattern of λm5/Brk expression was also noticed across cancer types in a

tissue cDNA qPCR array kit which represented 4 different cancers viz., breast, kidney, lung and ovary, with a trend of higher λm5 in advanced stages and higher Brk in earlier stages.

This work establishes, for the first time, the methods and conditions for successful selective differentiation of brk transcripts and their expression profiles from cancer cells and tissues. It also demonstrates that a non-catalytic variant of a tyrosine kinase has functional effects, and that its expression in cancer is relevant. Further, it highlights the importance of the expression ratio of a tyrosine kinase and its variant protein as a potentially useful disease marker. This is the first study of its kind in India, and only the second globally. Our findings are novel and add considerable new knowledge to the area.

Chapter 4

Cellular interacting partners of λm5.

4.1 Immunoprecipitation (IP) and Mass spectrometric (LC-MS/MS) analysis.

In order to further understand the potential cellular function(s) of λm5, we sought to find interacting partners of λm5. We utilized a pcDNA3.1/myc-His(-)A-λm5 construct to ectopically express λm5 in non-cancerous human embryonic kidney fibroblasts cell line HEK-293T. 48 hours post-transfection, the cells were lysed in RIPA buffer (Sigma) and quantified. 2µg of anti-Myc antibody was added to 500µg of total cell protein lysate making a final volume of 600µl with the RIPA/Wash buffer and incubated overnight at 4^0C with rocking. An isotypic control was also maintained under similar conditions. The next day, the immune complex mixture was added to pre-washed Protein A/G agarose beads (20µl bed volume) and mixed for 1 hour. Then, the IP mix was washed three times with 200µl of wash buffer with gentle resuspension and centrifugation steps. The final immunoprecipitated complex was eluted in sample buffer and resolved on SDS-PAGE. The experiment was performed three times. Bands appearing consistently were excised and dispatched for mass spectrometric identification analysis.

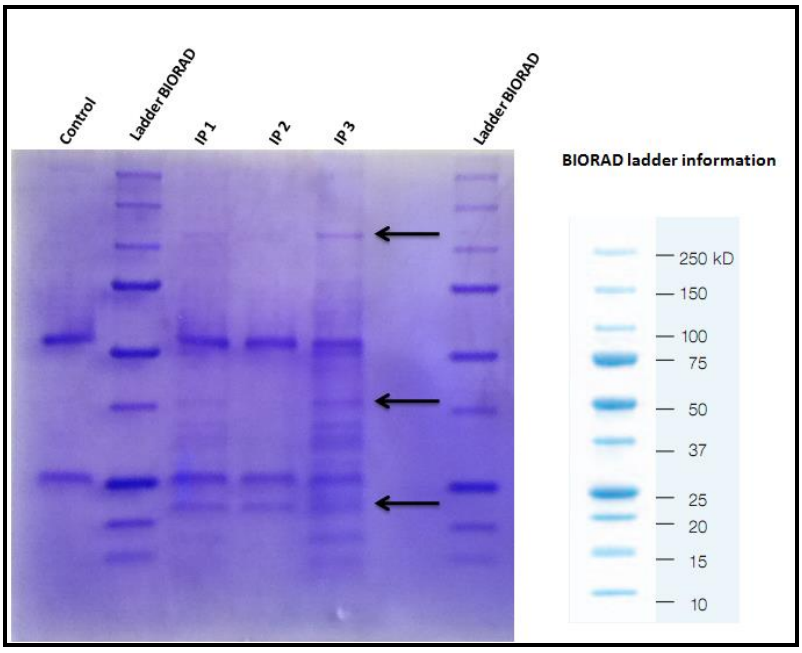


Figure 25 (previous page). Coomassie-stained 10% Tricine SDS gel of Anti-Myc immunoprecipitated cell lysates of HEK 293T (Human Embryonic Kidney cells) expressing the myc-tagged protein λm5 (~17kDa). IP1, IP2 and IP3 indicate 3 different Immunoprecipitation experiments; Arrows indicate the cut bands sent to MS analysis.

The curated list of identified proteins from all the samples submitted is given below.

Identified Protein	Score
Nucleolin OS=Homo sapiens GN=NCL PE=1 SV=3 – [NUCL_HUMAN]	1262.62
Putative ribosomal RNA methyltransferase NOP2 OS=Homo sapiens GN=NOP2 PE=1 SV=2 – [NOP2_HUMAN]	608.88
Staphylococcal nuclease domain-containing protein 1 OS=Homo sapiens GN=SND1 PE=1 SV=1 – [SND1_HUMAN]	600.31
Putative helicase MOV-10 OS=Homo sapiens GN=MOV10 PE=1 SV=2 – [MOV10_HUMAN]	363.89
116 kDa U5 small nuclear ribonucleoprotein component OS=Homo sapiens GN=EFTUD2 PE=1 SV=1 – [U5S1_HUMAN]	302.20
5'-3' exoribonuclease 2 OS=Homo sapiens GN=XRN2 PE=1 SV=1 – [XRN2_HUMAN]	292.65
Heterogeneous nuclear ribonucleoprotein U OS=Homo sapiens GN=HNRNPU PE=1 SV=6 – [HNRPU_HUMAN]	135.87
Zinc finger RNA-binding protein OS=Homo sapiens GN=ZFR PE=1 SV=2 – [ZFR_HUMAN]	120.40
Annexin A2 OS=Homo sapiens GN=ANXA2 PE=1 SV=2 – [ANXA2_HUMAN]	101.65
Cyclin-dependent kinase 11A OS=Homo sapiens GN=CDK11A PE=1 SV=4 – [CD11A_HUMAN]	93.70
Cyclin-dependent kinase 11B OS=Homo sapiens GN=CDK11B PE=1 SV=3 – [CD11B_HUMAN]	93.70
Replication factor C subunit 1 OS=Homo sapiens GN=RFC1 PE=1 SV=4 – [RFC1_HUMAN]	87.89
ATP-binding cassette sub-family F member 1 OS=Homo sapiens GN=ABCF1 PE=1 SV=2 – [ABCF1_HUMAN]	79.69

Table 6 (previous page). Curated list of identified proteins from IP & LC-MS/MS of myc-tagged λm5 in HEK-293T cells.

The list revealed the identification of Nucleolin (NCL) with the highest score. Other proteins included RNA methyltranferase, predominant components of ribonucleoprotein complexes, Annexin A2 and cyclin dependent kinases that play significant roles in cell survival.

4.2 Comparative analysis with BRK-interacting protein partners.

BRK-interacting protein partners have been reported earlier. We also carried out immunoprecipitation and mass spectrometric analysis with anti-BRK antibodies in HEK-293T cells. Interestly, we found Nucleolin as one of the interacting partners of BRK-directed immunoprecipitation as well. This indicated that Nucleolin was interacting with SH3 domain which is present in both BRK and λm5.

Identified Protein	Score
ATP5A1 protein [Homo sapiens]	10
G22P1 [Homo sapiens]	18
epithelial microtubule-associated protein [Homo sapiens]	22
heat shock 70kDa protein 8 isoform 2 [Homo	26
dnaK-type molecular chaperone HSP70-Hom - human	27
heat shock protein [Homo sapiens]	28
aralar2 [Homo sapiens]	31
tubulin beta chain - human	36
Calcium-binding mitochondrial carrier protein Aralar1	37
ribosomal protein S18 [Rattus norvegicus]	44
solute carrier family 25, member A6 [Homo	47
plectin 1 isoform 7 [Homo sapiens]	49
P43 [Homo sapiens]	50
Nucleolin - human	55
solute carrier family 25 member 3 isoform a	56

1 ribosomal protein L4 - human	57
ATPase family, AAA domain containing 3A	59
probable ataxia-telangiectasia group D protein -	61

Table 7. Curated list of identified proteins from IP & LC-MS/MS of BRK-SH3 in HEK-293T cells.

Nucleolin (NCL) is an abundant protein of the nucleolus. Nucleolar proteins structurally related to nucleolin are found to be conserved in organisms like yeast, plants and mammals. The structure of Nucleolin consists of an N-terminal domain which is made up of highly acidic regions interspersed with basic sequences and containing multiple phosphorylation sites. The central domain contains four RNA-binding domains called RBD. The C-terminal domain called GAR or RGG domain is rich in glycine, arginine and phenylalanine residues (Ginisty, Sicard et al. 1999).

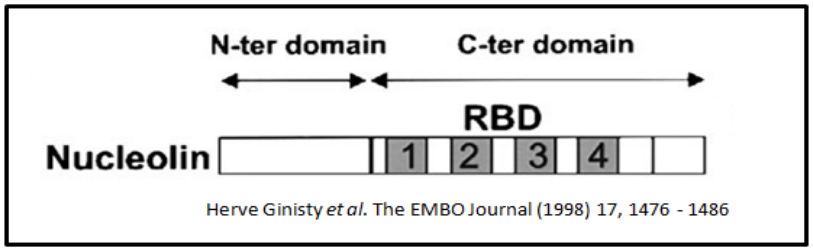


Figure 26. Schematic of Nucleolin structure.

Nucleolin has been implicated in chromatin structure, rDNA transcription, rRNA maturation, ribosome assembly, microRNA biogenesis and nucleo-cytoplasmic transport due to the evolutionarily conserved domain structure. Numerous reports have implicated the involvement of nucleolin either directly or indirectly in the regulation of cell proliferation and growth, cytokinesis, replication, embryogenesis, and nucleogenesis (Srivastava and Pollard 1999). It was found that apoptosis in leukemia cells is accompanied by alterations in the levels and localization of NCL (Mi, Thomas et al. 2003). Moreover, NCL was proved to be a negative regulator of hydrogen peroxide induced apoptosis in human umbilical vascular

endothelial cells (HUVECs) (Zhang, Wang et al.). We performed cell proliferation assays to address the effect of λm5 on NCL.

4.3 λm5 reduces Nucleolin-induced cell proliferation.

Sulforhodamine B assay (SRB assay), a cell proliferation assay standardized by the National Cancer Institute (NCI) was used to understand the role of λm5. Assays conducted 48 hours post-transfection with NCL alone or NCL+λm5 revealed that λm5 can significantly abrogate NCL-induced cell proliferation of HEK-293T cells. The figure below depicts λm5-induced decrease of cell proliferation by 43.2% (***p value < 0.0001) compared to NCL.

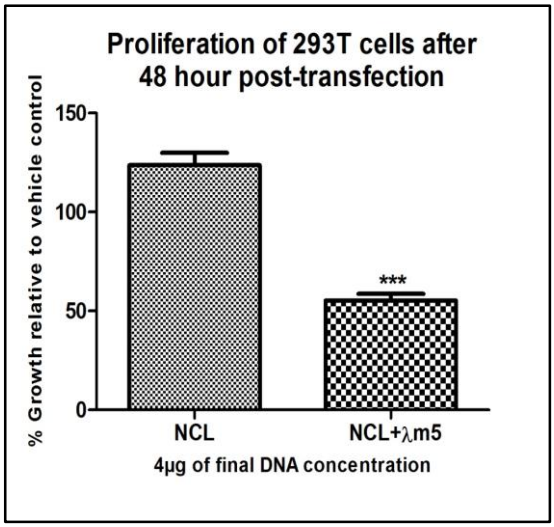


Figure 27. Anti-proliferative role on NCL-induced cell proliferation in HEK-293T cells.

4.4 Summary and future directions.

It was demonstrated that Nucleolin is a potentially significant interacting partner of λm5 and BRK, possibly via the SH3 domain which is common to both proteins. λm5 abrogated the proliferative function of over expressed NCL in HEK cells. It has been known that over-expression of NCL in chronic lymphocytic leukemia cells induces stabilization of bcl-2 mRNA (Otake, Soundararajan et al. 2007). Recently, it was demonstrated that NCL suppresses Fas-mediated apoptosis in vivo via a surface Nucleolin-Fas complex (Wise, Berkova et al. 2013).

Further, a tumor-homing peptide F3 was shown to target Nucleolin. Interestingly, a proline-rich sequence of F3 peptide shares 56.2% of similarity with the C-terminal unique proline-rich region of λm5.

```
F3              12 ARLSAKPAPPKPEPKP     27
                   ||....|..|:|.|:|
Lm5              8 ARACELPQGPEPVPRP     23
```

Figure 28. Sequence alignment of tumor homing peptide F3 and λm5.
Similarity was 56.2% as per Serial cloner software.

A DNA aptamer targeting Nucleolin, AS1411 (Antisoma Therapeutics), has been developed as a therapeutic and diagnostic agent and has progressed to Phase II clinical trials in the U.S. Given the clinical importance of Nucleolin, this represents a strong direction for future studies with λm5 and BRK. Future directions would include evaluation of λm5 and NCL expression levels in cancer tumor samples, understanding the protein-protein interactions involved and deciphering possible mechanistic signaling insights in cancer. Peptide sequences from λm5 could be potential binders to Nucleolin, and such peptides would have diagnostic and/or therapeutic potential. Thus, this work leads to a clear, testable and clinically relevant hypothesis.

Chapter 5

Characterization of tyrosine kinome proteome diversity due to alternate splicing: *in silico* approach.

The biochemical, cellular and clinical studies described thus far formed the basis to explore the larger question of the origin and occurrence of other TK splice variants. While we were able to perform a set of feasible laboratory experiments with λm5 (a non-catalytic splice variant of BRK), a significant question of the broader relevance of such splice variants remained. We addressed this question using computational methods and available sequence databases.

5.1 Tyrosine kinase (TK) isoforms.

Approximately 1.7% of the human genome codes for protein kinases (Manning, Whyte et al. 2002). Tyrosine kinases in particular constitute only ~0.3% of the genome but are involved in a diverse spectrum of cellular processes. Alternate splicing has the potential of generating a varied number of distinct protein isoforms from the primary transcript and the produced protein isoforms may be structurally and functionally different from one another (Ghigna, Valacca et al. 2008). Protein kinase isoforms can be broadly distinguished into two types viz; catalytic splice variants that retain the kinase domain and non-catalytic splice variants that have lost the kinase domain after alternate splicing. These isoforms add additional complexity to cellular signaling by altering post-translational modifications such as phosphorylation (Papin, Denouel-Galy et al. 1998; Hmitou, Druillennec et al. 2007), modifying intramolecular interactions due to novel structural constraints (Brignatz, Paronetto et al. 2009) and causing changes in the set of cellular interacting partners which may be dependent or independent of the above two previous reasons (Lorenzo, Gish et al. 1997; Rossel, Pasini et al. 1997; Ishiguro, Iwashita et al. 1999; Scott, Eketjall et al. 2005).

From the plethora of TK splice variants encoded by human genome, it is important to know the role of protein isoforms resulting from kinases that are well known for their oncogenic transformations. Few notable examples are listed in table 7 which explains the physiological

consequences of alternate splicing as studied in mouse models (Druillennec, Dorard et al.). In general, information on splice variants is currently available on a few web-based sequence databases. We chose to use the comprehensive Ensemble genome database which generates transcript data using RNA sequencing technologies. We collected the sequence data of TK isoforms from Ensemble and performed manual curation to identify their protein-coding potential using the translation tool of Expasy. We subsequently arrived at their domain structures using SMART (Simple Modular Architecture Research Tool) in order to identify the percentage of non-catalytic splice variants that constitute for human tyrosine kinome.

We found that the 90 Tyrosine kinase (TK) genes of the human genome produce 835 splice variants among which 438(52.46%) can produce proteins due to the presence of ORFs (Open Reading Frames). Interestingly, 47.76% of Receptor TK splice variants and 47.64% of Non-receptor TK splice variants are non-catalytic in nature, like the BRK splice variant λm5.

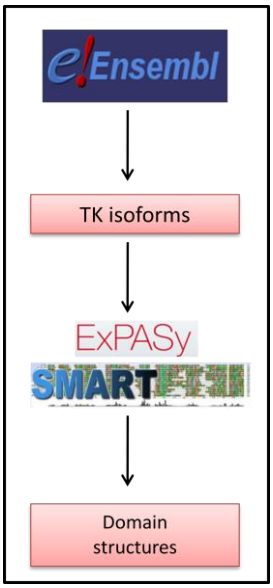


Figure 29. Schematic showing the protocol for manual curation of TK isoforms.

Splicing resulting in catalytic splice variants			
Kinase name	**Kinase type**	**Splicing**	**Regulation type**
Fyn	Cytosolic tyrosine kinase	Alternative use of exon 7a or 7b upstream of the Kinase domain	Kinase activity modulation by interfering with auto-inhibition
Fak	Focal adhesion tyrosine kinase	Multiple alternative splicing upstream of the kinase domain	Kinase activity modulation by interfering with autophosphorylation
Ret	Membrane-bound tyrosine kinase receptor	C-terminal alternative splicing generating three isoforms	Modulation of signaling partners binding
ErbB4	Membrane-bound tyrosine kinase receptor	N- and C-terminal alternative splicing generating four isoforms	Modulation of partners binding, cleavage, and subcellular localization
FGFR1, 2 & 3	Membrane-bound tyrosine kinase receptors	Alternative use of exon 8 or 9 generating distinct extracellular immunoglobulin-like domain III	Modified FGF binding specificity
Splicing resulting in Non-Catalytic splice variants			
TrkB; TrkC	Membrane-bound tyrosine kinase receptors	C-terminal alternative splicing replacing kinase domain by short amino acid sequences	Ligand sequestering, dominant negative and/or specific signaling functions
VEGFR1; VEGFR2	Membrane-bound tyrosine kinase receptors	C-terminal alternative splicing eliminating the kinase and transmembrane domains	Synthesis of secreted/soluble extracellular ligand-binding domains

Table 8 (previous page). Regulatory mechanisms of oncogenic kinases activity by alternative splicing resulting in catalytic and non-catalytic splice variants respectively.

5.1.1 Curated information on TK isoforms.

Transcript sequence information of TK genes was obtained from the Ensemble genome browser and curated manually to understand their protein coding potential and to compare the variant protein sequences with the full-length or main transcript-encoded protein sequences. 52.46% of splice variants were found to potentially produce proteins (analyzed using Expasy online tool). Unique sequences were identified by global sequence alignments and comparisons with the main transcript-encoded protein sequences. When all the unique sequences were aligned, we observed that unique sequences were conserved only within the respective genes but not across tyrosine kinase genes. We also observed that some unique sequences contained proline-rich motifs which may be of interest in the context of our studies on λm5. It is important to note that proline-rich peptides and sequences are known to be potential interacting partners for SH3 domain containing kinases in the cell (Alexandropoulos, Cheng et al. 1995). Tables 8-11 show tyrosine kinase gene-specific information regarding the coding potential of splice variants and the unique sequences that can be produced, potentially adding to human proteome diversity.

RECEPTOR TYROSINE KINASES (58)			
Gene	**No. of Transcripts**	**No. of splice variants relative to main transcript**	**No. of protein coding splice variants**
ALK	4	3	2
LTK	6	5	3
AXL	5	4	2
MER	6	5	3
TYRO3	9	8	2
DDR1	59	58	43
DDR2	6	5	5

EGFR	11	10	7
ERBB2	24	23	13
ERBB3	22	21	9
ERBB4	9	8	4
EPHA1	7	6	0
EPHA2	4	3	0
EPHA3	3	2	2
EPHA4	15	14	8
EPHA5	4	3	3
EPHA6	11	10	6
EPHA7	2	1	1
EPHA8	3	2	2
EPHB1	12	11	5
EPHB2	7	6	4
EPHB3	3	2	0
EPHB4	9	8	1
EPHB6	10	9	2
EPHX	DATA NOT AVAILABLE		
FGFR1	40	39	18
FGFR2	22	11	15
FGFR3	9	8	6
FGFR4	17	16	9
IGF1R	17	16	4
INSR	7	6	2
INSRR	1	0	0
MET	8	7	6
RON	12	11	3
MUSK	6	5	5
CSF1R	9	8	2

FLT3	5	4	2
KIT	4	3	1
PDGFRA	10	9	5
PDGFRB	11	10	2
PTK7	20	19	9
RET	4	3	2
ROR1	4	3	2
ROR2	8	7	1
ROS1	3	2	2
RYK	10	9	3
TEK	4	3	3
TIE	12	11	3
NTRK1	10	9	3
NTRK2	9	8	8
NTRK3	20	19	11
VEGFR1	5	4	4
VEGFR2	3	2	0
VEGFR3	11	10	3
AATYK	10	9	2
AATYK2	2	1	0
AATYK3	DATA NOT AVAILABLE		
DKFZp761P1010	6	5	5

Table 9. Curated information on protein coding potential of RTKs.

NON-RECEPTOR TYROSINE KINASES (32)			
Gene	**No. of Transcripts**	**No. of splice variants relative to main transcript**	**No. of protein coding splice variants**
ABL1	3	2	2
ARG	12	11	8

ACK1	24	23	10
TNK1	9	8	3
CSK	15	14	6
MATK	14	13	10
FAK	59	58	34
PYK2	19	18	11
FER	9	8	2
FES	18	17	9
BRK	2	1	1
FRK	2	1	1
SRMS	1	0	0
JAK1	6	5	0
JAK2	5	4	2
JAK3	7	6	2
TYK2	21	20	10
FGR	8	7	5
FYN	31	30	20
SRC	13	12	5
YES1	5	4	2
BLK	6	5	1
HCK	9	8	5
LCK	13	12	8
LYN	4	3	2
BMX	5	4	2
BTK	8	7	1
ITK	10	9	1
TEC	5	4	0
TXK	6	5	2
SYK	5	4	3

ZAP70	9	8	2

Table 10. Curated information on protein coding potential of NRTKs.

UNIQUE PROTEIN SEQUENCES OF RTK SPLICE VARIANTS	
RTK splice variant Ensemble name	**Unique sequence information**
AATK-006	MLTLCYLIARCIGRTRLAPGQQ
ALK-201	KERSPAAPPPLPTTSSGKAAKKPTAAEISVRV
ALK-002	LSWKQLFCFTFCLRPSWEKSTMNSSTQSTHQYLRKTLSCPLRHF LTQVPEQHPHSSSHSPANGTPVEPAVVLPTCSPARFWV
CSF1R-201	GTPSPSLCPA
DDR1-020	MSLPRCCPHPLRPEGSGA
DDR1-027	MPHLASLFTASSRCCPHPLRPEGSGA
DDR1-206	GWMTLGRVRSCGSGQAMTMWDGATTASPVAMWRWSLSLTG
DDR1-203	KVLESHPRTRSPGLVGIRPTPLPVSPMALRCCSPIQPTASFWPLTP VPLEARA
DDR1-007	KVLESHPRTRSPGLVGIRPTPLPVSPMALRCCSPIQPTASFWPLTP VPLEARA
DDR1-051	SFSLFS
DDR1-205	SFSLFS
DDR1-018	SFSLFS
DDR1-203	FWGERHRFLH
DDR1-009	RDQGRQVRVEEREDGSEAGDRRGRVVILETKEYLFPDSHPHC HNAGVPVP AACLPAGPI
DDR2-005	EDDVVGRVSREKPQQASGRYDTPAPSSLHLFLVSANLLRSPHSL SRFGLPLLLA
EGFR-201	SYIVSHFPRSFYKMSVH
EGFR-002	PGNESLKAMLFCLFKLSSCNQSNDGSVSHQSGSPAAQESCLGW

	IPSLLPSEFQLGWGGCSHLHAWPSASVIITASSCH
EGFR-004	GEWLVWKQSCSSTSSTHSAAASLQCPSQVLPPASPEGETVADLQTQ
EPHA3-003	CMYYFNAV
EPHA3-002	VT
EPHA4-202	MKWEEVSI
EPHA4-006	FHGSLSSTHFT
EPHA4-012	MAGIFYFALFSCLFGICDAVTGSRVYPANEG
EPHA5-003	SV
EPHA5-004	SV
EPHA6-202	GASKGRCF
EPHA6-005	GASKGRCF
EPHA6-002	APSLIGVVRKDWASQNSIALSWQAPAFSNGAILDYEIKYYEKEHEQLTYSSTRSKAPSVIITGLKPATKYVFHIRVRTATGYSGYSQKFEFETGDETSD
EPHA6-001	CSLKTDP
EPHA6-007	MKDSPFQVTKLYWLNEKWDFIASASD
EPHA6-006	MKDSPFQVTKLYWLNEKWDFIASASD
EPHA6-201	LYPKAEYNGKQCSRYSI
EPHA6-002	GGKIPIRWTAPEAIAYRKFSSASDAWSYGIVMWEVMSYGERPYWEMSNQDVILSIEEGYRLPAPMGCPASLHQLMLHCWQKERNHRPKFTDIVSFLDKLIRNPSALHTLVEDILVMPESPGEVPEYPLFVTVGDWLDSIKMGQYKNNFVAAGFTTFDLISRMSIDDIRRIGVILIGHQRRIVSSIQTLRLHMMHIQEKGFHV
EPHA6-007	GTGLVLMWKRNRRAMGASGQTRKQCDKRDNPPTD
EPHA8-002	GRRRNSVPQRPGPPASPASDPSRDQSSAGDVLWAFRQVPLWPCAPHQDPELEALHCL
EPHA8-201	GRRRNSVPQRPGPPASPASDPSRDQSSAGDVLWAFRQVPLWPCAPHQDPELEALHCL
EPHB2-004	MWVPVLALPVCTYA

EPHB2-201	QRS
EPHB6-202	MATEGAAQLGNRVAGMVS
ERBB2-005	MELAALCRWGLLLALLPPGAASTQVCTGTD
ERBB2-204	MPRGSWKPQVCTGTD
ERBB2-008	MELAALCRWGLLLALLPPGAASTQVCTGTD
ERBB2-004	MELAALCRWGLLLALLPPGAASTQD
ERBB2-015	YRDP
ERBB2-005	NM
ERBB3-006	TGQFPMVPSGLTPQPAQDWYLLDDDPRLLTLSASSKVPVTLAA V
ERBB3-004	MPMSTILPVTSPCLPA
ERBB4-004	IGSSIEDCIGLMD
ERBB4-201	IGSSIEDCIGLMD
FGFR1-003	MAAVTRDFGEMLLHSGRV
FGFR1-202	MEARVSLKRRIELTVEYPWRCGALSPTSNCRTG
FGFR1-201	VPI
FGFR1-002	RM
FGFR1-203	RM
FGFR1-036	RM
FGFR1-003	RM
FGFR1-004	RM
FGFR1-037	RM
FGFR1-201	HSGINSSDA
FGFR1-201	TLFNVTEAQSGEYVCKVSNYIGEANQ
FGFR1-201	TRPV
FGFR1-002	EA
FGFR1-203	EA
FGFR1-011	TLFNVTEAQSGEYVCKVSNYIGEANQ
FGFR1-011	TRPVAK

FGFR2-008	MCLNKTKQLFGVAKRLPFWRKER
FGFR2-016	GSQGL
FGFR2-007	HSGINSSNAE
FGFR2-004	HSGINSSNAE
FGFR2-006	HSGINSSNAE
FGFR2-010	HSGINSSNAE
FGFR2-011	HSGINSSNAE
FGFR2-007	KQQ
FGFR2-004	KQQ
FGFR2-006	KQQ
FGFR2-010	KQQ
FGFR2-011	KQQ
FGFR2-021	NTWTSANLSNSIH-LVTLTQEVLVLQEMILFFLQTPCLTNHAFLSIHT
FGFR2-202	KVSGAVDCHKPPCNPSHLPCVLAVDQ
FGFR3-201	SWISESVEADVRLRLANVSERDGGEYLCRATNFIGVAEKAFWL SVHGPR
FGFR3-006	WGPALGDLHAGGLPVPRHPCGGALQAAEGGPPHGQARQLH TRPVHDHAGVLACRALPEAHLQAAGGGPGPCPYRDVHRRVP GPVGAFRAVLPGWPGHPQLQLLRGRLRVCPRPAAPGPTQQWG LADVKGHWSPTM
FGFR4-201	TCDSLTPAGRTKSPTL
FGFR4-002	TCDSLTPAGRTKSPTL
FGFR4-003	SSPWSQALPASQAHPWYEACVSPPAAPPCSPAS
FLT3-201	VRLA
IGF1R-004	GAIAGCGVPRPANTGHH
INSR-004	RVQSSAPL
LTK-005	MGCWGQLLVWFGAA
LTK-005	GTKRLAGTVDSRLLLSM

LTK-004	GGAWPGPVLASATRCHRGFPSQCYSAQ
MERTK-005	ITHLAILARSPASCSYCR
MET-005	MKSKSKSLAECFPYDKPLI
MET-201	IQR
MET-004	VGFLHSSHDVNKEASVIMLFSGLK
MUSK-201	LSCPQI
MUSK-003	LSCPQI
MUSK-004	MAGNGREDPQRTLQIRDAFAVRARMQQASQHALGPHGLCQT ATS
MUSK-004	MKIQVRICISSENRGFPSFSPLVRAFSFSP
NTRK1-004	MKEAALICLAPSVPPILTVKSWDTMQLRAARSRCTNLLAAS
NTRK1-007	VSFSPV
NTRK1-006	SPSG
NTRK2-204	GFVLFHKIPLDG
NTRK2-005	GFVLFHKIPLDG
NTRK2-203	GFVLFHKIPLDG
NTRK2-202	DFSWFGFGKVKSRQGV
NTRK2-004	DFSWFGFGKVKSRQGV
NTRK2-003	DFSWFGFGKVKSRQGV
NTRK2-003	KPDTWPRGSPKTA
NTRK2-002	KPDTWPRGSPKTA
NTRK3-010	VGMSNASVALTVY
NTRK3-010	DGLELLAVLLKCDSKQ
NTRK3-201	NLKDNRDHLVPSTHYIYEEPEVQSGEVSYPRSHGFREIMLNPIS LPGHSKPLNHGIYVEDVNVYFSKGRHGF
NTRK3-002	NLKDNRDHLVPSTHYIYEEPEVQSGEVSYPRSHGFREIMLNPIS LPGHSKPLNHGIYVEDVNVYFSKGRHGF
NTRK3-007	YTSAKALEECSDLSDRENTVKDSLYELIPPDRLCRHED
NTRK3-003	VVQA

NTRK3-015	VK
NTRK3-203	LFNPSGNDFCIWCE
NTRK3-204	LFNPSGNDFCIWCE
PDGFRA-009	MLPRLVLNSWAQMICLPQLPKFPRA
PDGFRA-002	GTCIISFLL
PTK7-013	GSPAR
PTK7-003	GSPAR
PTK7-009	MGSFLSGEKRPSAPTVGSAMEKKEFPTPPGRVGP
PTK7-013	CESQGGCAQSPCQTLND
PTK7-015	SEHLCPEGQGEVEGNTGLGVMDRGFPA
PTK7-003	SEHLCPEGQGEVEGNTGLGVMDRGFPGTHLRSSQFWALQAW ESVHYWESV
PTK7-004	AN
PTK7-020	MVARVSGLIVTLLSSSFP
RET-003	MAKATSGAAGLRLLLLLLLPLLGKD
RET-002	RISHAFTRF
ROR2-002	VRMRGSGRQCWMTETPNALLWHNVSGVGSCRWVRGLADFK DEATDFRVMPRDPSLHCHVTSTA
ROS1-002	NDTYATVCER
RYK-201	AWAARLGRPGRSCLPGARG
RYK-202	SSLG
RYK-007	APPCGAALFGVS
TEK-003	MDSLASLVLCGVSLLL
TEK-004	MDSLASLVLCGVSLLL
TEK-005	WT
TEK-003	WT
TIE1-203	GWRDWVDTSTEKQNTDEGRFGGHVSAPVGAPG
TIE1-202	VHQGHCGAREDHS
TYRO3-010	LTLPLHPSQGLPRPSDNLL

FLT1-201	LPPANSSFMLPPTSFSSNYFHFLP
FLT1-203	SSPLSSSSSSSSSSS
FLT1-204	MNSDLLV
FLT4-003	RGTRRTRGA

Table 11. Curated information on unique protein sequences of RTKs. Proline-rich motifs are indicated in red colour.

UNIQUE PROTEIN SEQUENCES OF NRTK SPLICE VARIANTS	
NRTK splice variant Ensemble name	**Unique sequence information**
ABL1-003	MGQQPGKVLGDQRRPSLPALHFIKGAGKKESSRHGGPHCNV FVEH
ABL1-002	MGQQPGKVLGDQRRPSLPALHFIKGAGKKESSRHGGPHCNV FVEH
ABL2-006	MVLGTVLLPPNSYGRDQDTSLCCLCTEASESALPDLT
ABL2-002	MVLGTVLLPPNSYGRDQDTSLCCLCTEASESALPDLTDHF
ABL2-005	MVLGTVLLPPNSYGRDQDTSLCCLCTEASESALPDLTDHF
ABL2-201	MVLGTVLLPPNSYGRDQDTSLCCLCTEASESALPDLT
ABL2-007	VLLHCANQTCITL
TNK2-012	MRSGALGPRGRGRASALGLPVPAMASRQCWARHPRVGLCWA AFEQESGPSCHRSEVKERLGGGR
TNK2-019	VGPSRWFPVAAAGGALGAPSPWATSPGLEPQHLPSVPGAAQE SLNISMFVWLPPSPCLQADGSV
TNK2-007	MPAARRFPGLELSFPLLARLRRRLYTRLGGGR
TNK2-005	MLEARPPRTQGSDAAGAAAGRGLRALLLSLTAAAGIWGSMGE RSAYQRLAGGEEGPQRLGGGR
TNK2-005	REPPPRPPQPAFFTQ
TNK2-020	XLWGLLPPGPPALAWNHSTCPPCLELLR

TNK2-201	CPFSAFSPGHPPAETCGQVLWTGRREACASDPRLHPVSSRTKGL
TNK2-010	TQ
PTK2-010	YDRYLASSKI
PTK2-042	MLLELAGQ
PTK2-040	MLLELAGQEALKPAG
PTK2-008	MLLELAGQEALKPAGAIYMEKSGCSPFPVCWAKEYDRYLASSKI
PTK2-007	MLLELAGQEALKPAGAIYMEKSGCSPFPVCWAKEYDRYLASSKI
PTK2-005	MKYQEVRCLTSFN
PTK2-009	MRTHAVSVSGVSHCQHKVKKARRFLPLVFCSHDPPS
PTK2-053	KSYGIDEA
PTK2-203	KSYGIDEA
PTK2-002	KSYGIDEA
PTK2-054	CKNCTSDSVREKFLQEASEVIFASKEIQFGSSIFDPVCLSA
PTK2-201	GKEEKNWA
PTK2-011	GKEEKNWA
PTK2-015	MRLRFC
PTK2-014	MRLRFC
PTK2-009	PWR
PTK2-201	ER
PTK2-201	PNQ
PTK2-011	ER
PTK2-011	PWR
PTK2-015	ISSQEYNINNAESFSASYPETYDVQGMFQLTNWTLTSEGRLDTTPVGDS
PTK2-014	ISSQEYNINNAESFSASYPETYDVQGMFQLTNWTLTSEGRLDT

	TPVGDS
PTK2-014	FAQAVLPRQPPE
PTK6-201 (λm5)	AGHAGCAALQDLAACRGPAAPERGGVLPQPARACELPQGPEPVPRPAAGRALPEARA
FER-008	GRKLKRPRNDTTKPQ
FER-202	KERRGYPNLNLFVIQLLE
FRK-201	MDSTSLLPNPWIR
FYN-010	TLF
FYN-002	NLTVIASSCTPQTSGLAKDAWEVARRSLCLEKKLGQGCFAEVWL
FYN-017	NLTVIASSCTPQTSGLAKDAWEVARRSLCLEKKLGQGCFAEVWL
FYN-203	NLTVIASSCTPQTSGLAKDAWEVARRSLCLEKKLGQGCFAEVWL
FYN-201	NLTVIASSCTPQTSGLAKDAWEVARRSLCLEKKLGQGCFAEVWL
JAK3-201	HELMKLCWAPSPQDRPSFSALGPQLDMLWSGSRGCETHAFTAHPEGKHHSLSFS
JAK3-005	HELMKLCWAPSPQDRPSFSALGPQLDMLWSGSRGCETHAFTAHPEGKHHSLSFS
LCK-010	MGIPGSHNLRYFWNFPGQGPIISDVGGADLGGAPSAPSSIPSGT
LCK-011	MGIPGSHNLRYFWNFPGQGPIISDVGGADLGGAPSAPSSIPSGT
LCK-011	GDPRQQGLKDKACGSLAVGFHLSPTYFLPGLAFLVPHPVTPGFLPIPARFSLMPLVFT
LCK-201	GDPRQQGLKDKACGSLAVGFHLSPTYFLPGLAFLVPHPVTPGFLPIPARFSLMPLVFT
LCK-002	GDPRQQGLKDKACGSLAVGFHLSPTYFLPGLAFLVPHPVTPG

	FLPIPARFSLMPLVFT
LCK-012	GDPRQQGLKDKACGSLAVGFHLSPTYFLPGLAFLVPHPVTPG FLPIPARFSLMPLVFT
LCK-004	DTLLDSQLEEKGLGASPWGNLGQQLLLLPT
MATK-011	MQGHFPAERREGRPRRGTRGQQQLLKTGRTRLGPVAPS
MATK-007	MAGRGSLVSWRAF
MATK-006	MAGRGSLVSWRAFHGCDSAEELPR
MATK-004	MAGRGSLVSWRAFHGCDSAEELPR
MATK-003	MAGRGSLVSWRAFHGCDSAEELPR
MATK-009	ERPPGNGTGCWGSPPWGWAHGCPTILQ
MATK-010	ERPPGNGTGCWGSPPWGWAHGCPTILQ
MATK-009	GPQLRHSSVTIIHHLSQKPGPEPPPTPSSSIPSLARFPRP
PTK2B-003	MGRTALEQQSTLLLFQ
PTK2B-003	RECSRPSSGGFLLAPCAEPPSLPTAQGREQRSLSAFPAASAMI
SRC-005	RKVDVR
SRC-202	RKVDVR
TNK1-007	TPRKHRWRQKEGKSLGCAPSTGPEEEHAPGEDE
TYK2-007	MYPALWPETSSAIGSILAAREAGKCS
TYK2-008	VETGFHHVSQDGLDLLTSWSAHLSLPKCWDYRREPLRPADRY HSSLLQSLCPLRCSGPSLVAPKPHPRDP
TYK2-002	DPWALPAVGAPPRNPEDRIRAQSCIPSGLVTSSLSGLPAGGPLA GAQCSPSLFRSLWKSLVPGTQKASKRLDFSFRQSDQGSDGRRP GCRVGKKG
YES1-006	MLDLI
ZAP70-201	MRLGPRWK

Table 12. Curated information on unique protein sequences of NRTKs. Proline-rich sequences along with details of λm5 are indicated in red colour.

5.1.2 Latest evidence for isoform occurrence

Recently, two research groups have published the first draft of the human proteome, both of which discussed the occurrences of known and unknown splice variants from the human proteome. One group stated that Isoform-specific peptides for 2,861 protein isoforms derived from 2,450 genes were identified (Hong, Shin et al. 2001). The other group which reported that they covered 80% of the human proteome, identified λm5 in 16 of their projects which involved diseased samples (Wilhelm, Schlegl et al.). Together, both datasets highlighted the occurrence of splice variants in general and of λm5 in normal and diseased samples. This provides validation and support to this thesis work.

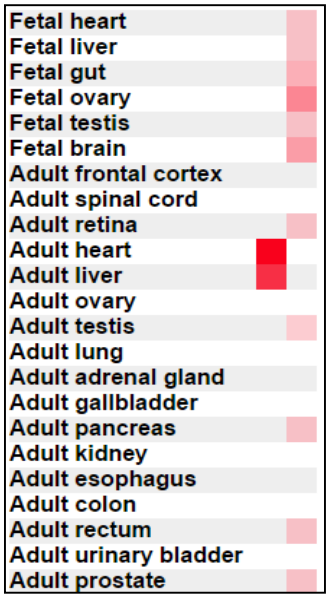


Figure 30. Evidence for λm5 expression studied in different anatomical sites of human body. Intensity of red colour represents direct proportion of expression level.
[Kim, Pinto et al. *Nature (2014) 509, 575–581*].

Uniprot Accession	Uniprot Identifier	I	Description	Length	E	L	L	S	Sequence Coverage [%]		Projects	
PTK6_HUMAN	Q13882	s	Protein-tyrosine kinase 6	451	is	4	8	1		87		41
PTK6_HUMAN	Q13882-2	s	Isoform 2 of Protein-tyrosine kinase 6	134	p	8	1	0		52		16

Figure 31. Representation of λm5 by number of projects and sequence coverage in the human proteome map.
[Schlegl et al. *Nature (2014) 509(7502): 582-7*].

Data reported by the human proteome studies also highlight the growing need for curated information portals and analysis tools in the field of TK splice variants in order to understand their origin, occurrence and complex role in fine-tuning kinase activity, cell signaling, development and disease.

5.2 Proteome diversity due to alternative splicing.

mRNA splicing can be found in a range of species from bacteria to humans. Alternative splicing (AS) is the major contributor for increased transcriptome and proteome diversity in mammals. Of all the genes encoded by human genome, >60% are believed to produce transcripts that are alternatively spliced (Fox-Walsh, Dou et al. 2005). The splicing machinery is a conserved mechanism that works mainly via the spliceosome complex. The spliceosome is a ribonucleoprotein complex, composed of five small nuclear ribonucleoproteins (snRNPs) and other non-snRNPs, which recognize and assemble on exon–intron boundaries to catalyse intron processing of the pre-mRNA. The spliceosome recognizes the pre-mRNA based on certain conserved signals: the exon–intron junctions (potential splice sites (SS)) at the 5' and 3' ends of introns called as 5' SS and 3' SS respectively, the branch site sequence located upstream of the 3' SS and the polypyrimidine tract (PPT) located between the 3' SS and the branch site. Basic mechanism of splicing is usually initiated by binding of U1 and U2 snRNPs to the splice sites(Chen and Manley 2009).

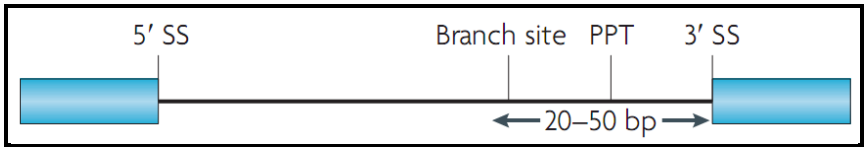


Figure 32. Schematic picture of conserved signals that enable pre-mRNA recognition by the spliceosome. Blue blocks represent exons and solid line represents intron between the exons. [Hadas Keren et al. *Nature (2010) 11, 345-355*].

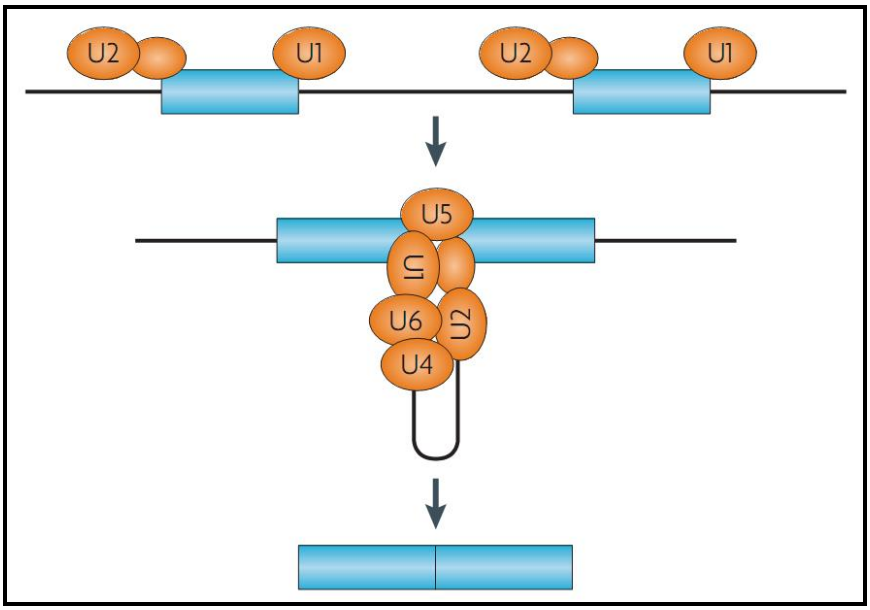


Figure 33. Schematic that explains spliceosome mediated splicing mechanism. Blue blocks represent exons, solid lines represent introns, and orange ovals represent the components of spliceosome. [Hadas Keren et al. *Nature (2010) 11, 345-355*].

Splice site recognition is in turn mediated by certain specific proteins like serine/arginine (SR) proteins, heterogenous nuclear ribonucleoproteins (hnRNPs), polypyrimidine tract-binding (PTB) proteins, the TIA1 RNA-binding protein, Fox proteins, Nova proteins, and such others(Licatalosi and Darnell ; Hui 2009).

5.2.1 Types of alternate splicing.

Alternative splicing (AS) can be broadly classified into four main subgroups viz, Exon skipping, alternative 3' SS (splice site), alternative 5' SS and intron retention.

a) Exon skipping is a splicing mechanism where an exon along with the flanking introns is removed during the splicing process. Exon skipping is known to be more prevalent in higher eukaryotes and accounts for nearly 40% of AS (Sugnet, Kent et al. 2004; Alekseyenko, Kim et al. 2007).

b and c) Alternative 3' splice site (3'SS) & Alternative 5' splice site (5'SS) are splicing events where two or more splice sites are recognized at one end of an exon during the splicing process. These account for 18.4% and 7.9% of AS events in higher eukaryotes respectively.

d) Intron retention is the situation where an intron is not spliced out during the AS event and therefore retains its structure in the mature mRNA transcript. The prevalence of intron retention is more in plants, fungi, protozoa and other lower organisms than the higher eukaryotes (<5%)(Galante, Sakabe et al. 2004; Hong, Scofield et al. 2006; Sakabe and Souza 2007; Kim, Goren et al. 2008).

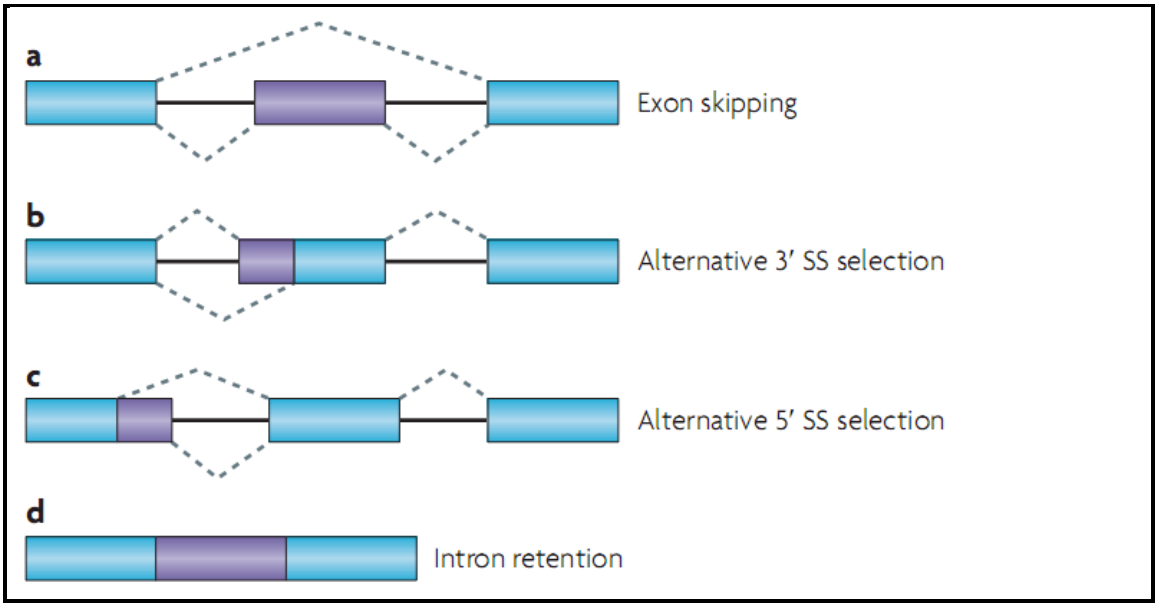


Figure 34. Schematic showing the major subgroups of Alternative splicing. Constitutive exons (exons that are not skipped out) are shown in blue, alternatively spliced regions in purple, solid lines represent introns and dashed lines represent the splicing options. [Hadas Keren et al. *Nature (2010) 11, 345-355*].

Other forms of AS include mutually exclusive exons, alternative promoter usage and alternative polyadenylation which are depicted in the figure 34.

e) Mutually exclusive exons mean that two exons involved in the splicing mechanism become mutually exclusive for their existence in the final transcript.

f) Alternative promoter usage indicates the processing of one exon by another upstream promoter than its own.

g) Alternative polyadenylation means the usage of alterantive poly (A) signal during the AS event.

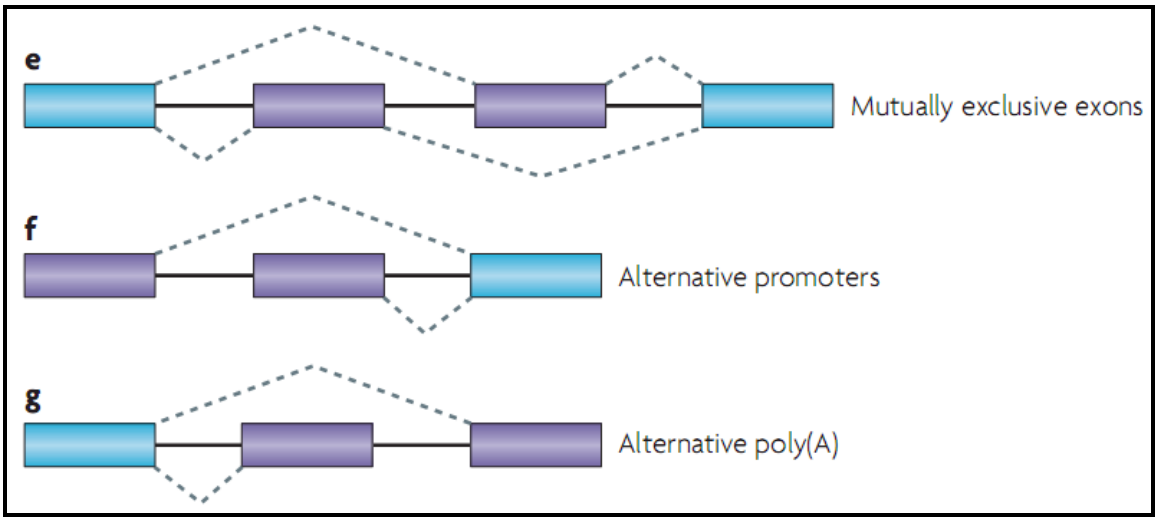


Figure 35. Schematic showing other forms of AS events. Constitutive exons (exons that are not skipped out) are shown in blue, alternatively spliced regions in purple, solid lines represent introns and dashed lines represent the splicing options.
[Hadas Keren et al. *Nature (2010) 11, 345-355*].

Based on reported laboratory studies and evolutionary computational sequence analysis, it is clear that exon skipping is the most prevalent and frequently occurring alternate splicing event in higher eukaryotes, especially in humans. This provided the basis for us to focus on exon skipping for the rest of this work.

5.3 A Database of Kinases generated from Exon skipping (KINEXON).

Significant evidence reveals that exon skipping is the predominant alternate splicing event and plays a major role in generating nearly 40% of human splice variants. With this rationale, we intended to manually generate a library of TK isoforms generated due to skipped exons. The concepts of exon symmetry and exonic end phase are important to understand the consequences of exon skipping on open reading frames.

5.3.1 Exon symmetry.

Exon symmetry is based on whether the exon's length is a multiple of 3, and whether it is flanked by introns of the same phase. If an exon's length is a multiple of 3 and it is flanked by introns of the same phase, it is referred to as a symmetric exon. Exon length influences the reading frame within the exon, because of the triplet codon rule. When we manually perform

exon skipping on a given gene, the reading frame of the exon-skipped transcript would depend on exon symmetry. Skipping of asymmetric exons causes a higher probability of frame shift and hence the production of novel protein sequences. The preceding and succeeding exons of a skipped exon are also important to consider for analyzing the generation of abnormal protein products. Figure 35 shows the different kinds of exon skipping causing frameshifts and the potential for occurrence of novel protein products.

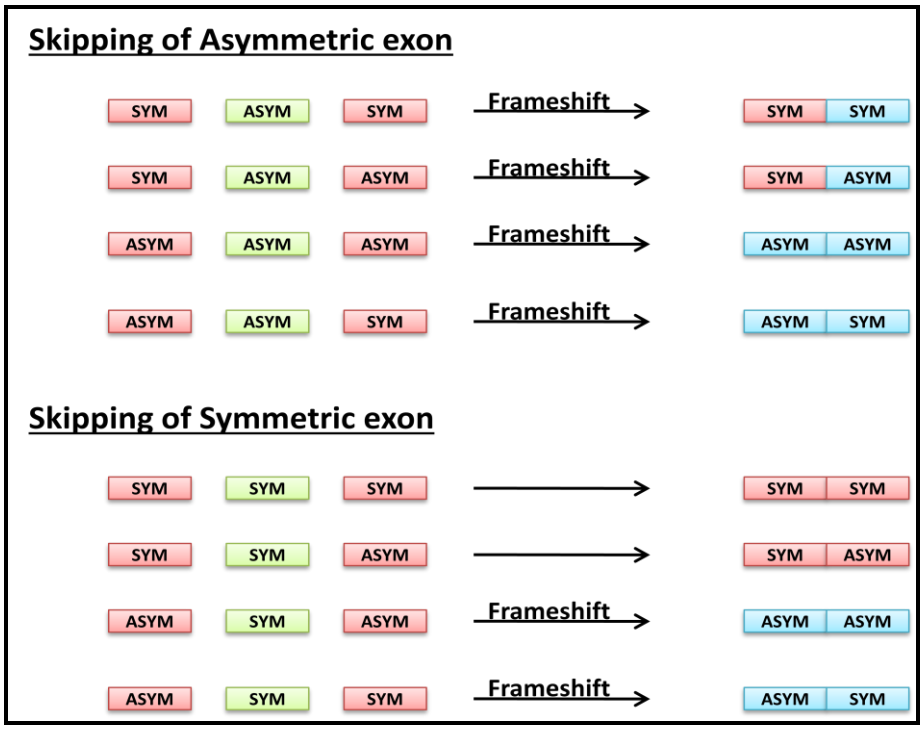


Figure 36. Schematic showing the possibilities of frameshift and production of novel protein sequences. SYM = symmetric; ASYM = asymmetric. Boxes in red represent preceding and succeeding exons (left) and unchanged protein sequences (right); Boxes in green indicate skipped exons. Boxes in blue indicate novel protein sequences produced as a result of frame shift.

It is important to note that unusual combinations of symmetric exons can also lead to novel protein sequences. This is determined by the concept of exon (or exonic) end phase.

5.3.2 Exonic end phase

In protein-coding exons, the end phase is the place where the intron lands inside the codon. End phase 0 means the intron lands between codons, End phase 1 indicates that the intron lands between the 1st and second base, 2 indicates intron landing between the second and 3rd base. Exons therefore have a start phase and an end phase, but introns have just one phase. So exons can be classified into 9 groups (symmetrical: 0-0, 1-1, 2-2 and asymmetrical: 0-1, 0-2, 1-0, 1-2, 2-0, 2-1) depending on the phases of their flanking introns (Kolkman and Stemmer 2001).

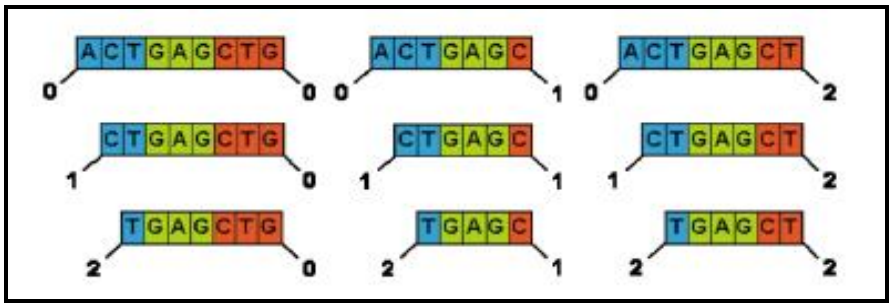


Figure 37. Schematic of exons showing 9 groups of exon phase combinations that include start phase and end phase. [Kolkman and Stemmer. Nature (2001) 19, 423-428]

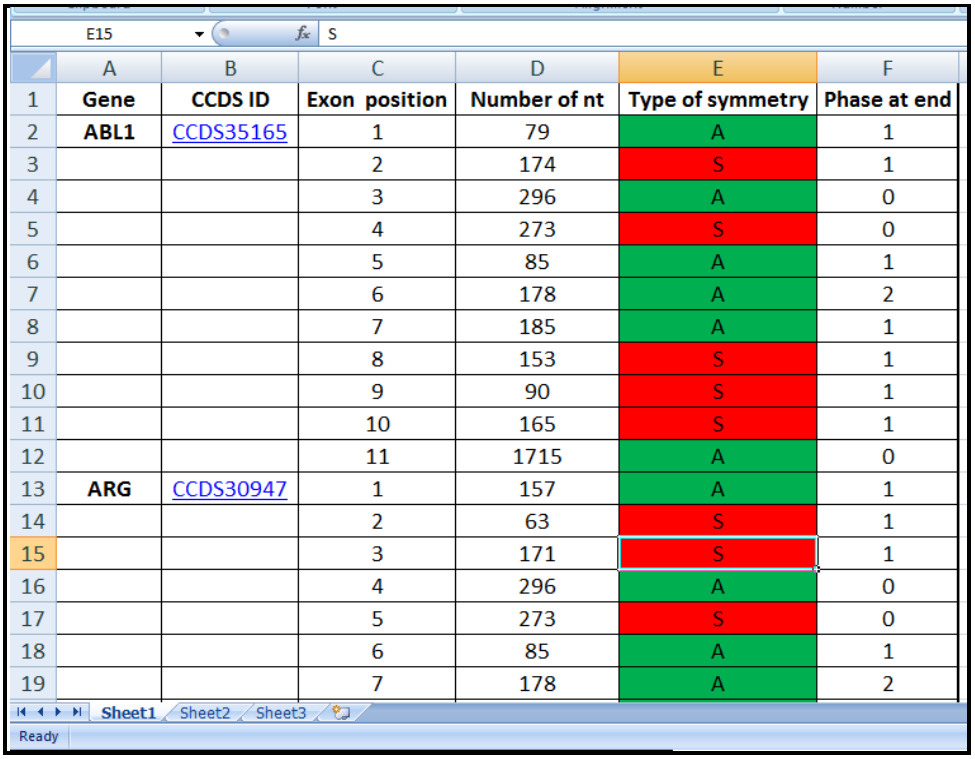
Among the 9 groups, except (0-0) all other combinations can potentially cause a frameshift and end up in novel protein sequences.

5.3.3 Protocol for computational Exon skipping.

We carried out exon skipping of all tyrosine kinase genes with their natural exon phases taken into consideration and the information generated was hosted on a publicly accessible database KINEXON (Kinases generated from Exon skipping) at the website www.kinexon.in. The protocol involved the following steps.

- The tyrosine kinase gene sequences were retrieved from the Consensus Coding Sequence Project (CCDS) database that describes a standard set of gene annotations and ensures consistent representation by NCBI (National Centre for Biotechnology Information), Ensemble and UCSC (University of California, Santa Cruz) Genome Browser.

- Gene sequences retrieved from CCDS contain exon demarcations which was helpful for our study.
- For a selected TK gene, each exon length was measured to determine if it was a Symmetric or an Asymmetric exon based on whether it was a multiple of 3 (Triplet codon concept).
- An excel sheet was maintained which contained the details of gene name, CCDS ID, Exon number/position, Number of nucleotides and type of symmetry.
- With the above details, the gene sequence was manually curated by removing each exon at a time and translating the rest of the sequence using the Expasy translate tool to obtain the protein product indicated by ORF1 (Open Reading Frame 1).
- All the protein sequences were documented gene-wise in FASTA format with code names.
- Example 1 Code name >ABL1_CCDS represents the gene name ABL1 kinase and CCDS represents the consensus coding sequence without any skipped exon.
- Example 2 Code name >ABL1_SKP_ASYME1_ORF1 represents the gene name as ABL1 kinase; SKP represents skipping; ASYME1 represents Aymmetric Exon 1 and ORF 1 represents the protein product of Open Reading Frame 1. Overall, the code name indicates that the Protein product (indicative of ORF1) is of ABL1 kinase gene after skipping exon 1 which is asymmetric in nature. Likewise, if the skipped exon is symmetric in nature, the code would be 'SYME'.
- In parallel, all the predicted protein products were processed using SMART (Simple Modular Architecture Research Tool) online tool in order to obtain the domain structure details.



E15 | S

	A	B	C	D	E	F
1	**Gene**	**CCDS ID**	**Exon position**	**Number of nt**	**Type of symmetry**	**Phase at end**
2	**ABL1**	CCDS35165	1	79	A	1
3			2	174	S	1
4			3	296	A	0
5			4	273	S	0
6			5	85	A	1
7			6	178	A	2
8			7	185	A	1
9			8	153	S	1
10			9	90	S	1
11			10	165	S	1
12			11	1715	A	0
13	**ARG**	CCDS30947	1	157	A	1
14			2	63	S	1
15			3	171	S	1
16			4	296	A	0
17			5	273	S	0
18			6	85	A	1
19			7	178	A	2

Sheet1 Sheet2 Sheet3

Ready

Figure 38. Screenshot of excel sheet showing the Exon symmetry details of ABL1 kinase. Asymmetric and symmetric exon types are indicated in green and red colour respectively.

ATGTTGGAGATCTGCCTGAAGCTGGTGGGCTGCAAATCCAAGAAGGGGCTGTCCTCGTCCTCCAGCTGTT
ATCTGGAAGAAGCCCTTCAGCGGCCAGTAGCATCTGACTTTGAGCCTCAGGGTCTGAGTGAAGCCGCTCG
TTGGAACTCCAAGGAAAACCTTCTCGCTGGACCCAGTGAAAATGACCCCAACCTTTTCGTTGCACTGTAT
GATTTTGTGGCCAGTGGAGATAACACTCTAAGCATAACTAAAGGTGAAAAGCTCCGGGTCTTAGGCTATA
ATCACAATGGGGAATGGTGTGAAGCCCAAACCAAAAATGGCCAAGGCTGGGTCCCAAGCAACTACATCAC
GCCAGTCAACAGTCTGGAGAAACACTCCTGGTACCATGGGCCTGTGTCCCGCAATGCCGCTGAGTATCTG
CTGAGCAGCGGGATCAATGGCAGCTTCTTGGTGCGTGAGAGTGAGAGCAGTCCTGGCCAGAGGTCCATCT
CGCTGAGATACGAAGGGAGGGTGTACCATTACAGGATCAACACTGCTTCTGATGGCAAGCTCTACGTCTC
CTCCGAGAGCCGCTTCAACACCCTGGCCGAGTTGGTTCATCATCATTCAACGGTGGCCGACGGGCTCATC
ACCACGCTCCATTATCCAGCCCCAAAGCGCAACAAGCCCACTGTCTATGGTGTGTCCCCCAACTACGACA
AGTGGGAGATGGAACGCACGGACATCACCATGAAGCACAAGCTGGGCGGGGGCCAGTACGGGGAGGTGTA
CGAGGGCGTGTGGAAGAAATACAGCCTGACGGTGGCCGTGAAGACCTTGAAGGAGGACACCATGGAGGTG
GAAGAGTTCTTGAAAGAAGCTGCAGTCATGAAAGAGATCAAACACCCTAACCTGGTGCAGCTCCTTGGGG
TCTGCACCCGGGAGCCCCCGTTCTATATCATCACTGAGTTCATGACCTACGGGAACCTCCTGGACTACCT
GAGGGAGTGCAACCGGCAGGAGGTGAACGCCGTGGTGCTGCTGTACATGGCCACTCAGATCTCGTCAGCC
ATGGAGTACCTGGAGAAGAAAAACTTCATCCACAGAGATCTTGCTGCCCGAAACTGCCTGGTAGGGGAGA
ACCACTTGGTGAAGGTAGCTGATTTTGGCCTGAGCAGGTTGATGACAGGGGACACCTACACAGCCCATGC
TGGAGCCAAGTTCCCCATCAAATGGACTGCACCCGAGAGCCTGGCCTACAACAAGTTCTCCATCAAGTCC
GACGTCTGGGCATTTGGAGTATTGCTTTGGGAAATTGCTACCTATGGCATGTCCCCTTACCCGGGAATTG
ACCTGTCCCAGGTGTATGAGCTGCTAGAGAAGGACTACCGCATGGAGCGCCCAGAAGGCTGCCCAGAGAA
GGTCTATGAACTCATGCGAGCATGTTGGCAGTGGAATCCCTCTGACCGGCCCTCCTTTGCTGAAATCCAC
CAAGCCTTTGAAACAATGTTCCAGGAATCCAGTATCTCAGACGAAGTGGAAAAGGAGCTGGGGAAACAAG
GCGTCCGTGGGGCTGTGAGTACCTTGCTGCAGGCCCCAGAGCTGCCCACCAAGACGAGGACCTCCAGGAG

Figure 39. Screenshot of word document showing a part of ABL1 kinase gene sequence retrieved from CCDS. Sequences highlighted with alternative blue colour represent alternative exons.

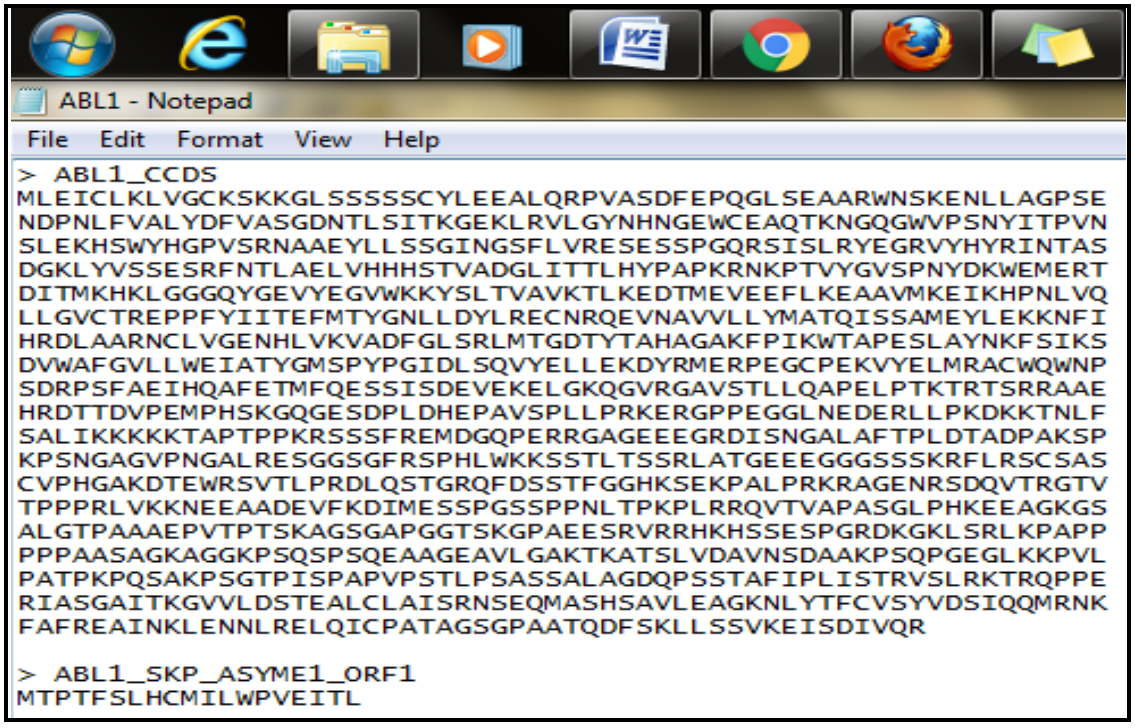


Figure 40. Screenshot showing the protein products of original coding sequence and skipped exon 1 of ABL1 kinase represented in FASTA format along with code names.

As per the above mentioned exon skipping protocol, a library of >1000 predicted alternative isoforms was generated from the 90 human TK genes. Complete sequence translation without any stop codon was observed when symmetric exons preceded by asymmetric were removed unlike the removal of asymmetric exon preceded by asymmetric exon. This apparently proves the concept that skipping of asymmetric exons imparts more probability for frameshifts during alternative splicing. In general, on an average all the TK genes contained more asymmetric exons than symmetric ones. However, it was specifically observed that cellular receptor coding genes like EGFR, VEGFR, PDGFR and FGFR that have oncogenic implications contained more asymmetric exons (>60%) than symmetric exons. Interestingly, we found that most of the TK genes conserved their kinase domain upon all exon skipping events, except for the skipping of the exon coding for kinase domain itself.

5.3.4 The KINEXON Database.

Based on the library, we developed an algorithm that catalogues the information of predicted TK isoforms and their protein properties (Work done in collaboration with the group of Dr. Niyaz Ahmed, Head, Dept. of Biotechnology and Bioinformatics, School of Life Sciences, University of Hyderabad). The physical and chemical properties of the predicted proteins were theoretically calculated using the "Peptides" package version 1.1.1 of R programming language. The package was a well-validated user friendly version that calculates physicochemical properties and indices from amino-acid sequences of peptides and proteins.

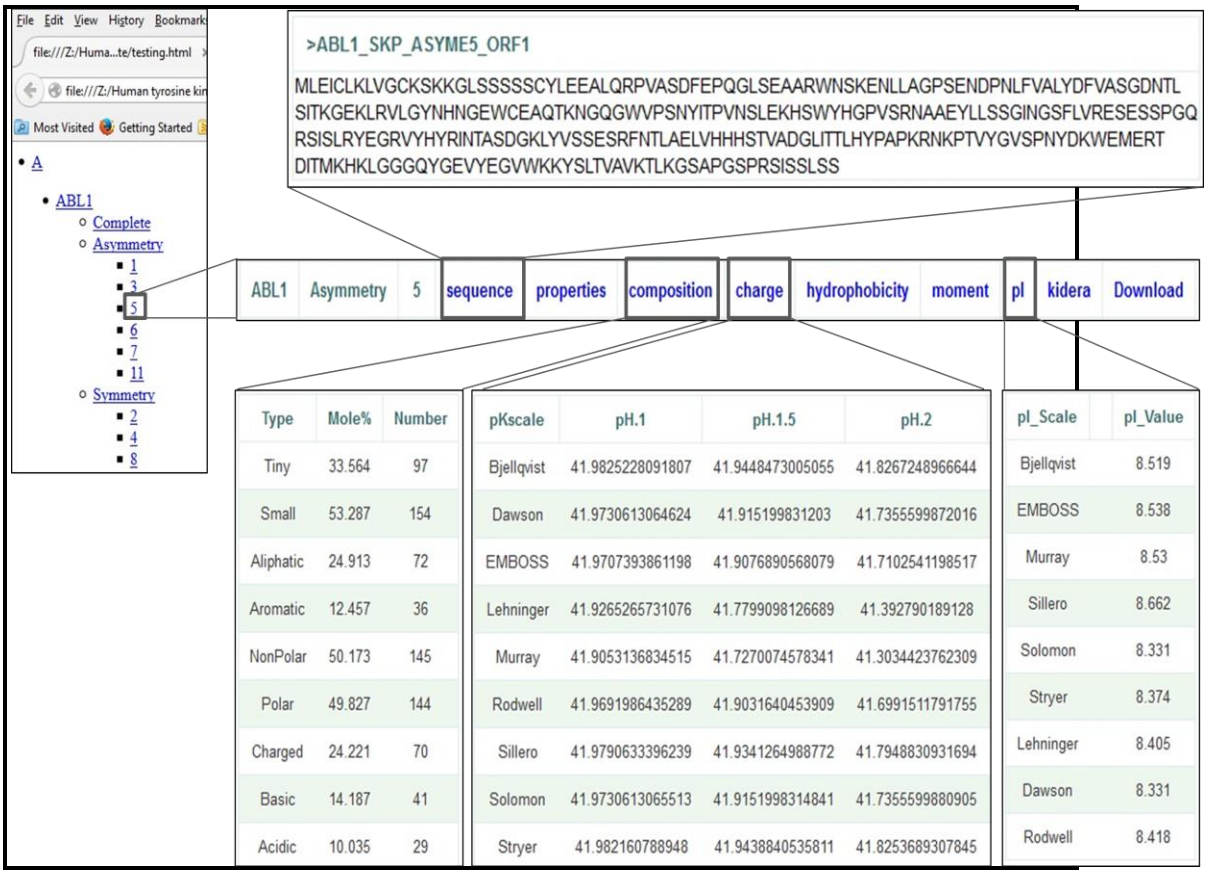


Figure 41. A snapshot of the database of Kinases generated from Exon skipping (KINEXON). The snapshot shows ABL1 kinase as an example.

5.3.5 Potential applications of KINEXON.

- The KINEXON database currently documents the potential proteome diversity of the human tyrosine kinome due to the predominant alternative splicing event, exon skipping. A comparable database does not exist for any gene family as of today.
- Comprehensive human proteome maps published by two research groups in 2014 report the presence of many protein isoforms with several of them unannotated or unknown (the coding gene unidentified). This is attributed to their novel protein sequence features. Such unique protein isoforms if formed due to exon skipping events could be efficiently annotated or mapped back to their parent gene(s) using the database we generated.
- We observed that some TK genes retained the kinase domain upon all exon skipping events, except when the kinase domain-coding exon was skipped. Exon symmetry and/or exonic end phase are likely to have a potential role in kinase domain retention over the course of evolutionary genomic alterations. Our database is a potential source of information regarding evolutionary adaptations in specific TK genes or families to retain kinase activity despite genomic alterations.
- The KINEXON database, in its present form, provides information on predicted physicochemical properties of the protein isoforms. Properties like molecular weight, charge, hydrophobicity, isoelectric point (pI), measure of polarity and other structural information would be useful in biochemical, biophysical and proteomics studies.
- The clinical picture of abnormal TK isoforms obtained from genome sequencing studies could be further validated by our database.
- Comparison of protein isoform sequences from our database with that of validated (reported) ones, if any would reveal specific sequences susceptible/resistant to exon skipping events.

5.4 Identification of all splicing events in the human tyrosine kinome.

Though we explored all the possibilities of tyrosine kinase isoform generation due to exon skipping, it is equally important to identify all types of splicing events occurring in the tyrosine kinome. For achieving this, we collected human RNAseq data and Riboseq data from

accessible databases such as like NCBI and Riboseq genome browser. Riboseq data is derived from the sequencing of mRNA following translation arrest, RNA digestion and isolation of mRNA regions protected by bound ribosomes. Such data serves as additional evidence for mRNA translation into proteins, beyond RNAseq data. We considered both resources for our study. The work was carried out on high configuration computer workstations running Ubuntu Linux (Work done in collaboration with the group of Dr. Niyaz Ahmed, Head, Dept. of Biotechnology and Bioinformatics, School of Life Sciences, University of Hyderabad). The brief protocol is depicted in the figure below.

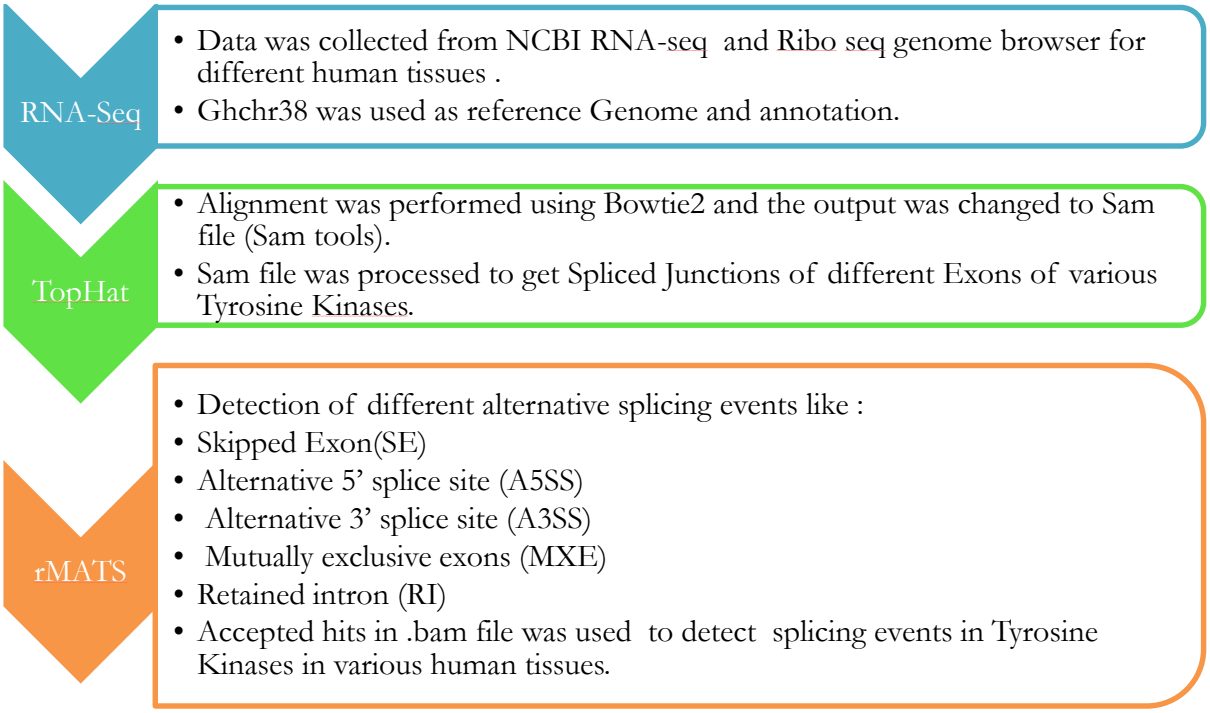


Figure 42. Scheme of the protocol for identifying all splicing events in human tyrosine kinases.

The rMATS tool version 3.0.9 was used for identification of splicing events. MATS - Multivariate Analysis of Transcript Splicing is a computational tool used to detect differential alternative splicing events from RNAseq data. It can automatically detect and analyze all major alternate splicing events (Park, Tokheim et al. ; Shen, Park et al.).

Our initial attempts did not yield results because the number of hits of tyrosine kinase genes in the RNAseq data and the Riboseq data were very low, not meeting the criteria of the analyzing software program. However, this research direction remains a future goal.

5.5 Summary.

Our findings from the computational studies described in this chapter are unique and have never reported earlier. To our knowledge, this is the first such study globally. We found that the 90 Tyrosine kinase (TK) genes of the human genome produce 835 splice variants among which 438 (52.46%) can produce proteins due to the presence of ORFs (Open Reading Frames). Interestingly, 47.76% of Receptor TK splice variants and 47.64% of non-receptor TK splice variants were observed to be non-catalytic in nature, like the BRK splice variant λm5. Unique sequences were identified manually by comparing with the main transcript encoded protein sequences. When all the unique sequences were aligned, we observed that unique sequences are conserved only within the respective genes but not globally. We also found some unique sequences which contained proline-rich sequences which may have functionalities similar to λm5. Overall, we generated information that highlights the coding potential of splice variants and the unique sequences that can be produced adding to human proteome diversity. The recent reports on human proteome mapping stand supportive of our work on the BRK splice variant λm5 and its potential relevance in disease. We developed a library of predicted tyrosine kinase protein isoforms generated from skipped exons, underpinned by the concepts of exon symmetry and exon end phase. The library, hosted on an online database www.kinexon.in, foresees potential use in validating or mapping unidentified/unannotated TK isoforms. Our initial attempts in identifying all splicing events from human RNAseq data and Riboseq data using rMATS did not yield fruitful results, probably due to insufficient data quality. We intend to continue this study by considering other sources of sequencing data.

Chapter 6
Discussion and Future Perspectives.

BRK/PTK6, a non-receptor tyrosine kinase was discovered in the year 1993. During its research history of nearly two and half decades, it attained clinical importance attributed to its implications in various cancers and attracted the attention of researchers to design BRK inhibitors. The BRK/PTK6 story began with its identification in normal skin and breast cancers (which explains its dual nomenclature). Its expression was then found to extend to several anatomical sites of human malignancies such as head and neck, gastro-enteric cancers, etc. BRK's involvement in specific oncogenic and non-oncogenic signaling pathways was identified. However, the hidden story of its splice variant λm5 which lacks the kinase domain was completely unknown. The work described in this thesis began in 2010. While the work was well under way, a report in 2011 demonstrated that λm5 negatively regulates growth of prostate cancer cells and enhances PTK6 mediated inhibition of β-catenin transcriptional activity. This literature report partially validated our hypothesis and provided clues for our study. We focused on three aspects of λm5 function: *In vitro* biochemical and cellular effects, its cellular interacting partners and its clinical relevance in cancer. This led us to the final part of the work where we addressed the broader relevance of our findings: Exploring all possible human tyrosine kinase isoforms using computational strategies.

In chapter 2, we successfully demonstrated that λm5 regulates the kinase activity of BRK in a dose-dependent biphasic manner showing maximum and minimum activities of 2 fold and 0.5 fold respectively. A similar phenomenon was observed under cellular conditions, where a constant amount of active Brk mutant DNA was transfected with increasing doses of λm5 DNA into non-cancerous human embryonic kidney fibroblast cells (HEK-293T cells). It was shown that 48 hours post-transfection, cell proliferation and whole cell tyrosine phosphorylation of HEK-293T cells followed a biphasic response effect consistent with the *in vitro* kinase assay data. The specific role of λm5 on active BRK was confirmed using an inactive BRK mutant under similar experimental conditions where no significant effect was

noticed. This confirmed that the biphasic effect of λm5 was mediated via the kinase activity of BRK, indicating a physical interaction between them. Though we were not successful in directly demonstrating an interaction between BRK and λm5 using techniques such as immunoprecipitation and co-localization, we obtained indirect evidence for their interaction in the *in vitro* kinase assay. A possible explanation for our observation (and that of the earlier reported study) is that the interaction is of low affinity or is transient with a high 'off' rate.

An explanation for the biphasic response effect of λm5 can be given based on the established literature on intra and intermolecular interactions of tyrosine kinases and their activity regulation. It is well known in the literature that the SH3 domain interacts with proline-rich sequences and in fact one of the regulatory points in tyrosine kinases is the interaction between the SH3 domain and the poly-proline (PP) motif present in the linker region between SH2 domain and the kinase domain. Because λm5 has an SH3 domain similar to BRK but a unique proline rich (PR) region at the C-terminal end, our premise was that the unique proline rich region of λm5 could interact with SH3 domain of BRK restricting it from binding to its own PP motif.

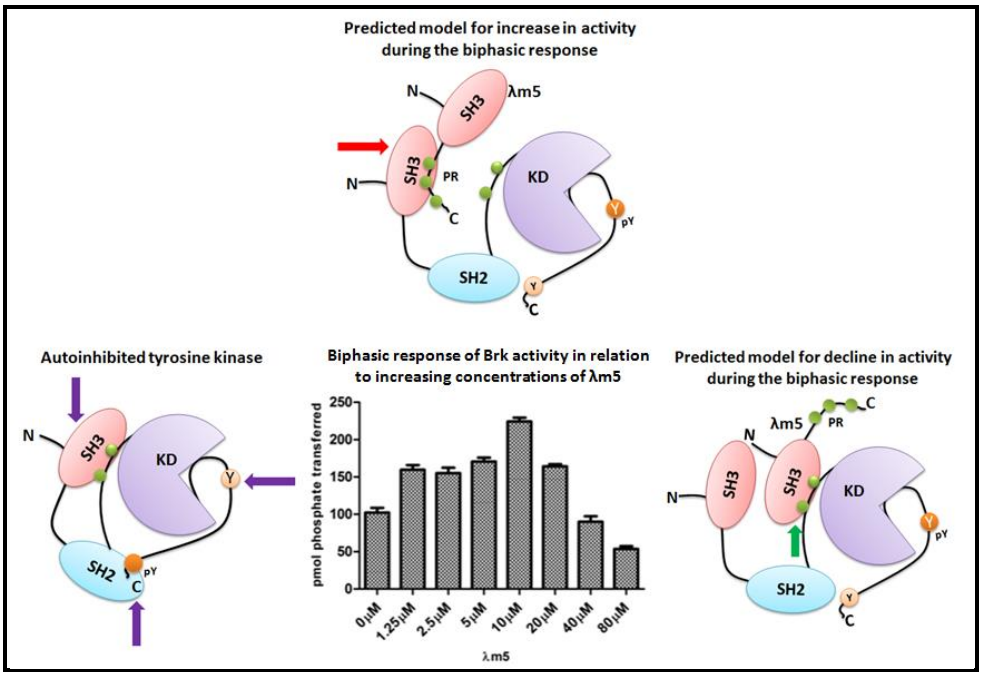


Figure 43. Schematic model explaining the biphasic response.

Purple arrows (left) indicate regulatory check points of a typical tyrosine kinase applicable to BRK as well. The red arrow (middle) indicates the predicted interaction of λm5 proline-rich (PR) region with BRK SH3 domain. The green arrow (right) indicates the predicted interaction of λm5 SH3 domain with the poly-proline motif of BRK.

This broken interaction activates the kinase which could be the potential cause for increase in the activity during the initial phase of the biphasic response. The decline in the activity after reaching a maximum can be explained by observations from the literature that mutation of prolines in the linker between SH2 and kinase domains decreases the kinase activity for unknown reasons. In the present context, when the kinase is in a maximally active state, the increasing concentration of λm5 could be allowing the accumulation of SH3 domains of λm5 onto the prolines of BRK, and bringing about a decline in activity. To address this specific question, we made an attempt to conduct a Surface Plasmon Resonance (SPR) study that can detect subtle interactions but were not successful due to excessive non-specific interactions. Overall, the novelty of our findings in this part of the work lie in identifying an unusual biphasic regulatory role of a non-catalytic splice variant on its parent kinase which has not been reported earlier.

In chapter 3, we successfully developed PCR-based methods to differentiate *brk* and *λm5* transcripts using selective primers, despite their sequence constraints. Such selective differentiation of the two transcripts has not been reported earlier. We used our method to profile their expression levels among a cohort of Indian cancer patients for the first time. The expression analysis was initially carried out in a set of cancer cell lines. These included 6 cell lines which represent cancers of different anatomical origins. It was clear from the study that the splice variant λm5 was found to be higher in expression levels than Brk in all except A549 (lung cancer cell line) with fold difference range of 1-14 times. This led us to extend the study with a collection of 14 matched pairs (cancer vs. normal) patient biopsy samples which represented cancers of various anatomical origins and a cancer tissue cDNA array. The patient samples were acquired from a local hospital Basavatarakam Indo-American Cancer Hospital & Research Institute (BIACHRI) following all required guidelines of human ethics. We observed that the expression of λm5 was higher than Brk in most cases with a trend towards metastatic samples. The fold difference in expression (λm5/Brk) was in the range of 2-75 times depending on the cancer type. A mixed pattern of relative λm5 and Brk expression was noticed across samples in the cancer tissue cDNA array which represented 4 different cancers viz., breast, kidney, lung and ovary.

Our study of cancer patient samples of different anatomical origins has provided clues to the potential relevance and utility of λm5 expression. The expression data clearly show that λm5 expression and its relation to Brk is likely to have discriminatory power, because the expression is not constant. The main limitations of the present study were the lack of biochemical, hematological, disease progression and mortality data, as well as the relatively limited sample size (14 and 24). Our data has clear indications about the higher expression of λm5 in advanced stage cancers, including metastatic cancers. This highlights its potential clinical importance. Thus, this part of our study adds to the growing need for considering the clinical relevance of splice variants in cancer research and drug discovery programs, and provides a rational basis for large-scale case studies of λm5/Brk among different cancer types.

In chapter 4, we addressed the question of other potential functions of the splice variant λm5 in the cell, apart from the role towards its parent kinase. We performed an immunoprecipitation (IP) assay by over-expressing λm5 in non-cancerous HEK-293T cells. A similar IP was performed with BRK as well. We found that Nucleolin (NCL) is a potential strong interacting partner of λm5 and BRK. The interaction is probably mediated via the SH3 domain which is common to both their protein domain structures. Moreover, λm5 abrogated the proliferative function of overexpressed NCL in HEK cells. NCL is a well established onco-target for many notable cancers. It was known that over-expression of NCL in chronic lymphocytic leukemia cells induces stabilization of bcl-2 mRNA. Recently, it was demonstrated that NCL suppresses Fas-mediated apoptosis in vivo via a surface Nucleolin-Fas complex. Further, a tumour homing peptide F3 was shown to target Nucleolin which is gaining clinical importance. Interestingly, a proline-rich region of F3 peptide shares 56.2% of similarity with the C-terminal unique proline rich region of λm5. This fact raises the potential possibility that the proline-rich region of λm5 can be used as a NCL-targeting peptide.

Functionally, interaction of BRK and NCL could effect proliferative functions under malignant conditions and it would be interesting to study if λm5 abrogates the hypothesized synergism.

Future directions from this part of the study would include the evaluation of λm5 and NCL expression levels in cancer tumor samples; understanding the relevant protein-protein interactions and deciphering possible mechanistic insights in the context of NCL for cancer therapy and diagnosis.

In chapter 5, we expanded our study using computational strategies, to investigate all possible TK splice variants, based on the interesting results from our work on λm5. Curation of human RNAseq data from Ensemble revealed that the 90 Tyrosine kinase (TK) genes of the human genome produce 835 splice variants among which 438(52.46%) can produce proteins due to the presence of ORFs (Open Reading Frames). Interestingly, 47.76% of Receptor TK splice variants and 47.64% of Non-receptor TK splice variants were observed to be non-catalytic in

nature, similar to the BRK splice variant λm5. Unique sequences were identified by alignment and comparison with the main transcript-encoded protein sequences. When all the unique sequences were aligned, it was observed that unique sequences are conserved only within the respective genes but not across gene families. The purpose of exploring the unique sequences was to find any consensus signature sequences among the variants that might have been conserved during evolution. We could not find such conserved signatures, but we noticed that some unique sequences contained proline-rich motifs which may have potential functionalities like λm5. Overall, we generated information that highlights the coding potential of TK splice variants, unique sequences present in them and the complement of non-catalytic splice variants that can be produced leading to human proteome diversity.

Recent reports of the human proteome map stand supportive of our findings regarding the expression of the BRK splice variant λm5 in normal and disease proteomes. In fact, the human proteome maps confirmed the existence of λm5 and support the notion of exploring other splice variants.

Based on the importance of exon skipping as the predominant form of alternative splicing, we developed a library of predicted protein isoforms generated from skipped exons. The basis of novel protein sequence generation due to exon skipping is explained by the concepts of exon symmetry and exon end phase, which influence the occurrence of a frameshift. Notably, we could observe that some TK genes conserved the kinase domain upon all exon skipping events, except when the kinase domain-encoding exon is skipped. This observation may be explained by evolutionary adaptations in specific TKs or TK families leading to kinase domain retention despite genomic alterations. The library of TK isoforms has been hosted on a publicly accessible database KINEXON (Kinases generated from Exon skipping), at www.kinexon.in. The library foresees several applications, including validating or mapping unidentified/unannotated TK isoforms, and being a general resource for protein biochemistry and biophysics researchers who work on protein variants. Attempts to identify all possible splicing events from human RNAseq data and Riboseq data using rMATS did not yield results due to a very low number of hits of TK isoforms from the datasets we could access. This

study is of future interest and will be extended by considering other sources of RNAseq data. A study focused on validating the predicted TK splice variants from the KINEXON is also under consideration.

In summary, the work performed in this Ph.D thesis successfully demonstrated novel functions, clinical relevance and other potential roles of a non-catalytic BRK splice variant λm5. We identified an important oncoprotein Nucleolin as a potential functional partner of λm5. Our study provides a rational basis for new translational research directions in cancer diagnosis and therapy involving λm5, BRK and Nucleolin. Our computational efforts yielded new information on the hidden complement of all potential non-catalytic splice variants encoded by human tyrosine kinome. We developed a publicly accessible database of Kinases generated from Exon skipping (KINEXON) that envisions being of practical utility as a tool in splice variant protein research and human proteome mapping studies.

Chapter 7
Materials and Methods.

7.1 Cloning and expression.

7.1.1 Cloning of Brk and λm5.

Bacterial expression vector pET28a (+) from Novagen (cat#69864-3) was used to clone PTK6 and λm5. cDNA clone (Acc No: AK301364; Clone ID: SYNOV2007726) of λm5 was procured from NITE Biological Resource Center (NBRC), Japan. Brk cDNA was prepared by RT-PCR from HEK-293T cells. Both Brk and λm5 were cloned between the restriction sites NdeI (NEB cat# R0111S) and XhoI (NEB cat#R0146S) of pET28a(+) respectively. Both Brk and λm5 were also cloned into mammalian expression vectors pCMV-HA (Clontech, cat#635690) and pcDNA3.1/myc-His(-)A (Invitrogen, cat#V855-20) respectively. PTK6-K219M kinase inactive mutant clone was generated from pCMV-HA-PTK6 clone using QuikChange site-directed mutagenesis kit (Agilent technologies, cat#200519). pEGFP-C1-PTK6-Y447F (Lukong and Richard 2008) constitutively active kinase mutant plasmid was a generous gift from Dr. Kiven Erique Lukong, University of Saskatchewan, Canada.

7.1.2 Protein expression.

BRK purification.

His-tagged wild-type Brk cDNA cloned into a pet28a(+) vector was expressed in E.coli BL21 (DE3) cells. For protein production, 1mM IPTG was added to a log-phase culture and induced overnight at 18°C. Cells were harvested and washed with phosphate-buffered saline two times. Cells were then lysed using a sonicator in 50 ml of buffer A (20 mM Tris, pH 8.5, 10% glycerol, 5 mM β-mercaptoethanol) containing protease inhibitors (5 mg/liter aprotinin, 5 mg/liter leupeptin, 0.1 mM phenylmethylsulfonyl fluoride), 0.1 mM vanadate, and 1 mM EDTA. Cell lysate was diluted to 200 ml, centrifuged, and filtered. The lysate was then mixed with 40 ml of Macro-Prep High Q resin (Bio-Rad) that had previously been equilibrated in buffer A. After rocking at 4 °C for 1 h, the resin was washed with 100 ml of buffer A four times. BRK was then eluted from the High Q resin with 40 ml of buffer B (buffer A plus 1 M NaCl). The protein was then loaded onto a 3 ml column of Ni-NTA resin (Qiagen) that had

been pre-equilibrated in buffer A. The column was washed with three column volumes of washing buffer (buffer B plus 0.01 M imidazole). BRK was eluted with buffer A containing 0.2 M imidazole, dialysed and confirmed on SDS-PAGE.

λm5 purification.

His tagged λm5 was expressed under similar conditions as of BRK but in 100ml culture and processed with 1ml bed volume of Ni-NTA affinity column. λm5 was eluted with buffer containing 0.25M imidazole, dialysed and run through anion exchange chromatography using the working of Fast Pressure Liquid chromatography (FPLC from GE biosystems – ACTA explorer). Software used for computer-instrument interface was UNICORN 5.11 (Build 407). Anion exchanger column used was Mono Q 5/50_GL from GE biosystems. FPLC conditions included pH 8.2 with 0% NaCl in buffer A and 100% NaCl in buffer B. Purified fractions were achieved at a gradient of 37% to 57% of salt concentration. All the fractions were dialysed, quantified, confirmed through SDS-PAGE and preserved at -80^{0}C with 50% glycerol for future use.

7.2 In vitro kinase assay.

1 μM purified BRK protein was incubated with increasing doses of λm5 (1.25 μM - 80 μM) along with [γ- 32P] ATP (2 μCi/reaction) in a total reaction volume of 50 μl that contained 150 mM Tris/HCl pH 7.5, 100 mM MgCl2, 800 μM ATP, 20 μM poly(Glu, Tyr), at 30^o C, for 30 min. Reactions were terminated by spotting 35 μl of the mixture onto 1 x 1 inch square pieces of phophocellulose filter paper, which were washed three times in 600 mL of hot (65o C) 5% trichloroacetic acid, dried, and analyzed in a scintillation counter. Control reaction involved only BRK without λm5.

Autokinase assays were performed as above, but without the substrate peptide poly(Glu, Tyr). Reactions were terminated by adding Laemmli buffer and analyzed by SDS-PAGE and autoradiography.

7.3 Cell proliferation assay – Sulforhodamine B (SRB) assay.

4hours post—transfection, HEK-293T cells were trypsinised, counted and 5000cells were seeded in two 96 well plates in triplicates per sample type. One of the plates serves as a 0 hour plate which is processed after 24h incubation. The optimum cell number to be seeded was determined by a growth curve analysis for HEK-293T cell line. After 48h incubation relative to the 0h plate, cell monolayers were fixed with 10% (wt/vol) trichloroacetic acid (TCA) for 1h at 4^0C. Then the plate was washed and stained with 0.4% SRB dye (Sigma, cat#S1402) prepared in 1% (vol/vol) acetic acid solution for 30 min at room temperature, after which the excess dye is removed by washing repeatedly with 1% (vol/vol) acetic acid and finally air dried. The protein-bound dye is dissolved in 10 mM Tris base solution for OD determination at 510 nm using a microplate reader. Percentage growth was determined by the formula [(Ti-Tz/C-Tz)] X 100, where Ti=absorbance after 48h of test plate, Tz=Absorbance at time 0, C=Absorbance of the untransfected sample after 48h.

7.4 Western blotting.

Tranfected cells were lysed in RIPA buffer along with protease inhibitor cocktail mix at 4^0C for 5min, centrifuged to remove cell debris and collected the cell lysate. The lysate was quantified using Pierce BCA protein assay kit (cat#23225) and 50µg of total cell lysate was loaded on to 12% polyacrylamide gel electrophoresis set up. Seperated protein bands were transferred on to PVDF membrane and blocked with 5% non-fat milk for 1h. Blocking was followed by primary antibody incubation overnight at 4^0C. Before developing, the blot was exposed to horse radish peroxidase conjugated secondary antibody for 45min. fluorescence was detected by using Amersham ECL prime western blotting detection reagent (cat#RPN2232). Densitometric scanning analysis of the X-ray films was done using ImageJ software.

7.5 Semi-quantitative PCR.

Semi-quantitative PCR was performed using Taq DNA polymerase (SIGMA cat # D4545-1.5KU) with standard PCR composition detailed below. 'X' represent amounts that depend on

final volume of PCR mix. The PCR amplifications were optimized to 60^0C annealing temperature.

PCR component	Final composition
Template DNA	X ng
10X PCR buffer with MgCl2	1X
10mM dNTP mix	0.2mM
Forward primer	0.2mM
Reverse primer	0.2mM
Taq DNA polymerase	0.5U
Nanopure water	X µl

7.6 Quantitative PCR.

2µg of total RNA was converted into cDNA using superscript first-strand synthesis system for RT-PCR, invitrogen (lot#1041283). Quantitative PCR was done in triplicates using the primers 5'-TGGTTCTTTGGCTGCATCT-3' (forward) and 5'-CAGGACGTAGTCGGCACTC-3' (reverse) for BRK/PTK6 exon 2 and 5'-AAGGAGGAGCAGTGGTG-3' (forward) and 5'-TCCCGCAGGTTCCGA-3' (reverse) for λm5. The primer efficiency for specific amplification of λm5 was optimized with sequence lengths which span the junction of exon 1 and 3 since λm5 lacks exon 2 unlike the full length BRK/PTK6. The quantitative PCR reactions were set with KAPA SYBR FAST qPCR kit master mix (2X) universal (cat#KK4601). Fold expressions of BRK/PTK6 and λm5 were generated by calculating $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen 2001).

7.7 Cell culture and transfections.

Mammalian cells HEK-293T (ATCC cat#CRL-3216), A549 (ATCC cat#CCL-185), CAL27 (ATCC cat#CRL-2095), HepG2 (ATCC cat#HB-8065), K562 (ATCC cat#CCL-243) and TZM-BL (ATCC cat#PTA-5659) were maintained in complete Dulbecco's modified eagle's medium (DMEM) and T47D (ATCC cat#HTB-133) cells were maintained in complete RPMI with 1X concentration of Penstrep antibiotic solution (Himedia cat#A018) and 10% fetal

bovine serum (Gibco cat#16000-044). The cells were grown in 5% CO_2 incubator maintained at 37^0C. Transfections were performed with Polyethylenimine (PEI) protocol optimized for HEK-293T cells in the lab.

7.8 Handling and processing of cancer patient biopsy samples.

Cancer tissue biopsy samples were collected and processed for total RNA isolation at Basavatarakam Indo-American Cancer Hospital and Research Institute (BIACHRI) with prior approval from institutional ethics committee (IEC), following good clinical practice (GCP) guidelines. Total RNA isolation was done with TRIzol reagent from life technologies (lot#47114) and stored at -80^0C for subsequent usage.

7.9 Immunoprecipitation & Mass spectrometry.

1. Cells under investigation were trypsinized from a 100mm dish, pelleted and washed with 1X PBS. PBS was replaced with 500µl of RIPA along with protease inhibitors. Resuspended by pipetting and kept on ice for 20min.
2. The lysed mix was centrifuged at maximum speed for 20 min at 4^0C.
3. The supernatant was collected and quantified for protein concentration.
4. 500 – 1000µg protein was taken and 1-2ug of Antibody was added and the volume was made to 500µl with lysis buffer.
5. Kept for gentle rocking overnight at 4^0C.
6. Next day morning Protein G resin was washed thrice with lysis buffer and the overnight protein lysate antibody mix was added to 20µl resin bed volume and kept for gentle rocking at 4^0C for 2 hours.
7. The protein G resin was pelleted and washed 5 times with the lysis buffer with gentle spins.
8. Finally, the resin beads were harvested by boiling in SDS sample buffer for 10min and then run on 10% Tricine gel (SDS-PAGE).

The IP experiment was performed three times separately and consistent bands were excised and analyzed by Trypsin digestion and LC-MS/MS analysis at the Centre for Cellular and Molecular platforms facility, National Centre for Biological Sciences, Bengaluru, India.

References

Alekseyenko, A. V., N. Kim, et al. (2007). "Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes." RNA 13(5): 661-70.

Alexandropoulos, K., G. Cheng, et al. (1995). "Proline-rich sequences that bind to Src homology 3 domains with individual specificities." Proc Natl Acad Sci U S A 92(8): 3110-4.

Aubele, M., G. Auer, et al. (2007). "PTK (protein tyrosine kinase)-6 and HER2 and 4, but not HER1 and 3 predict long-term survival in breast carcinomas." Br J Cancer 96(5): 801-7.

Aubele, M., M. Spears, et al. (2010). "In situ quantification of HER2-protein tyrosine kinase 6 (PTK6) protein-protein complexes in paraffin sections from breast cancer tissues." Br J Cancer 103(5): 663-7.

Babon, J. J., I. S. Lucet, et al. "The molecular regulation of Janus kinase (JAK) activation." Biochem J 462(1): 1-13.

Barker, K. T., L. E. Jackson, et al. (1997). "BRK tyrosine kinase expression in a high proportion of human breast carcinomas." Oncogene 15(7): 799-805.

Blume-Jensen, P. and T. Hunter (2001). "Oncogenic kinase signalling." Nature 411(6835): 355-65.

Bouyain, S., P. A. Longo, et al. (2005). "The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand." Proc Natl Acad Sci U S A 102(42): 15024-9.

Brauer, P. M., Y. Zheng, et al. (2011). "The alternative splice variant of protein tyrosine kinase 6 negatively regulates growth and enhances PTK6-mediated inhibition of beta-catenin." PLoS One 6(3): e14789.

Brignatz, C., M. P. Paronetto, et al. (2009). "Alternative splicing modulates autoinhibition and SH3 accessibility in the Src kinase Fyn." Mol Cell Biol 29(24): 6438-48.

Burgess, A. W., H. S. Cho, et al. (2003). "An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors." Mol Cell 12(3): 541-52.

Chang, Y. M., H. J. Kung, et al. (2007). "Nonreceptor tyrosine kinases in prostate cancer." Neoplasia 9(2): 90-100.

Chen, H. Y., C. H. Shen, et al. (2004). "Brk activates rac1 and promotes cell migration and invasion by phosphorylating paxillin." Mol Cell Biol 24(24): 10558-72.

Chen, M. and J. L. Manley (2009). "Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches." Nat Rev Mol Cell Biol 10(11): 741-54.

Cho, H. S. and D. J. Leahy (2002). "Structure of the extracellular region of HER3 reveals an interdomain tether." Science 297(5585): 1330-3.

Cooper, J. A. and T. Hunter (1985). "Major substrate for growth factor-activated protein-tyrosine kinases is a low-abundance protein." Mol Cell Biol 5(11): 3304-9.

Coyle, J. H., B. W. Guzik, et al. (2003). "Sam68 enhances the cytoplasmic utilization of intron-containing RNA and is functionally regulated by the nuclear kinase Sik/BRK." Mol Cell Biol 23(1): 92-103.

David J. Easty1, Philip J. Mitchell2, Ketan Patel3, Vivi Ann Florenes4, Richard A. Spritz5 andDorothy C. Bennett1 (1997). "Loss of expression of receptor tyrosine kinase family genes PTK7 and SEK in metastatic melanoma." International Journal of Cancer Volume 71, Issue 6, pages 1061–1065.

Derry, J. J., G. S. Prins, et al. (2003). "Altered localization and activity of the intracellular tyrosine kinase BRK/Sik in prostate tumor cells." Oncogene 22(27): 4212-20.

Derry, J. J., S. Richard, et al. (2000). "Sik (BRK) phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding ability." Mol Cell Biol 20(16): 6114-26.

Druillennec, S., C. Dorard, et al. "Alternative splicing in oncogenic kinases: from physiological functions to cancer." J Nucleic Acids 2012: 639062.

Eckhart, W., M. A. Hutchinson, et al. (1979). "An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates." Cell 18(4): 925-33.

Ferguson, K. M., M. B. Berger, et al. (2003). "EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization." Mol Cell 11(2): 507-17.

Foudah AI , J. S., Busnena BA , El Sayed KA . (2013). "Optimization of marine triterpene sipholenols as inhibitors of breast cancer migration and invasion.." Chem Med Chem. 8(3):497-510. .

Fox-Walsh, K. L., Y. Dou, et al. (2005). "The architecture of pre-mRNAs affects mechanisms of splice-site pairing." Proc Natl Acad Sci U S A 102(45): 16176-81.

Galante, P. A., N. J. Sakabe, et al. (2004). "Detection and evaluation of intron retention events in the human transcriptome." RNA 10.

Garrett, T. P., N. M. McKern, et al. (2002). "Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha." Cell 110(6): 763-73.

Ghigna, C., C. Valacca, et al. (2008). "Alternative splicing and tumor progression." Curr Genomics 9(8): 556-70.

Ginisty, H., H. Sicard, et al. (1999). "Structure and functions of nucleolin." J Cell Sci 112 (Pt 6): 761-72.

Haegebarth, A., D. Heap, et al. (2004). "The nuclear tyrosine kinase BRK/Sik phosphorylates and inhibits the RNA-binding activities of the Sam68-like mammalian proteins SLM-1 and SLM-2." J Biol Chem 279(52): 54398-404.

Hanks, S. K. and R. A. Lindberg (1991). "Use of degenerate oligonucleotide probes to identify clones that encode protein kinases." Methods Enzymol 200: 525-32.

Hanks, S. K. and A. M. Quinn (1991). "Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members." Methods Enzymol 200: 38-62.

Hantschel, O. and G. Superti-Furga (2004). "Regulation of the c-Abl and Bcr-Abl tyrosine kinases." Nat Rev Mol Cell Biol 5(1): 33-44.

Heinzelmann-Schwarz, V. A., M. Gardiner-Garden, et al. (2006). "A distinct molecular profile associated with mucinous epithelial ovarian cancer." Br J Cancer 94(6): 904-13.

Hmitou, I., S. Druillennec, et al. (2007). "Differential regulation of B-raf isoforms by phosphorylation and autoinhibitory mechanisms." Mol Cell Biol 27(1): 31-43.

Hong, E., J. Shin, et al. (2001). "Complete sequence-specific 1H, 13C and 15N resonance assignments of the human PTK6 SH2 domain." J Biomol NMR 19(3): 291-2.

Hong, X., D. G. Scofield, et al. (2006). "Intron Size, Abundance, and Distribution within Untranslated Regions of Genes." Mol Biol Evol 23.

Hui, J. (2009). "Regulation of mammalian pre-mRNA splicing." Sci China C Life Sci 52(3): 253-60.

Hunter, T. and J. A. Cooper (1985). "Protein-tyrosine kinases." Annu Rev Biochem 54: 897-930.

Hunter, T. and G. D. Plowman (1997). "The protein kinases of budding yeast: six score and more." Trends Biochem Sci 22(1): 18-22.

Hunter, T. and B. M. Sefton (1980). "Transforming gene product of Rous sarcoma virus phosphorylates tyrosine." Proc Natl Acad Sci U S A 77(3): 1311-5.

Ikeda, O., Y. Miyasaka, et al. (2009). "STAP-2 is phosphorylated at tyrosine-250 by Brk and modulates Brk-mediated STAT3 activation." Biochem Biophys Res Commun 384(1): 71-5.

Ishiguro, Y., T. Iwashita, et al. (1999). "The role of amino acids surrounding tyrosine 1062 in ret in specific binding of the shc phosphotyrosine-binding domain." Endocrinology 140(9): 3992-8.

Kamalati, T., H. E. Jolin, et al. (1996). "Brk, a breast tumor-derived non-receptor protein-tyrosine kinase, sensitizes mammary epithelial cells to epidermal growth factor." J Biol Chem 271(48): 30956-63.

Kasprzycka, M., M. Majewski, et al. (2006). "Expression and oncogenic role of Brk (PTK6/Sik) protein tyrosine kinase in lymphocytes." Am J Pathol 168(5): 1631-41.

Kazem Ahmed Mahmoud, T. W., Inna Slynko, Frank Totzke, Christoph Schächtele, Markus Oelze, Wolfgang Sippl, Christoph Ritter and Andreas Hilgeroth (2014). "Novel inhibitors of breast cancer relevant kinases Brk and HER2

" *Med. Chem. Commun.*, 5, 659-664.

Kim, E., A. Goren, et al. (2008). "Alternative splicing: current perspectives." Bioessays 30(1): 38-47.

Kolkman, J. A. and W. P. Stemmer (2001). "Directed evolution of proteins by exon shuffling." Nat Biotechnol 19(5): 423-8.

Lee, S. T., K. M. Strunk, et al. (1993). "A survey of protein tyrosine kinase mRNAs expressed in normal human melanocytes." Oncogene 8(12): 3403-10.

Lemmon, M. A. and J. Schlessinger "Cell signaling by receptor tyrosine kinases." Cell 141(7): 1117-34.

Licatalosi, D. D. and R. B. Darnell "RNA processing and its regulation: global insights into biological networks." Nat Rev Genet 11(1): 75-87.

Lin, H. S., G. J. Berry, et al. (2004). "Identification of tyrosine kinases overexpressed in head and neck cancer." Arch Otolaryngol Head Neck Surg 130(3): 311-6.

Liu, L., Y. Gao, et al. (2006). "Identification of STAT3 as a specific substrate of breast tumor kinase." Oncogene 25(35): 4904-12.

Liu, L. N., P. Y. Huang, et al. (2013). "Protein tyrosine kinase 6 is associated with nasopharyngeal carcinoma poor prognosis and metastasis." J Transl Med 11: 140.

Liu, X. K., X. R. Zhang, et al. (2013). "Low expression of PTK6/Brk predicts poor prognosis in patients with laryngeal squamous cell carcinoma." J Transl Med 11: 59.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods 25(4): 402-8.

Llor, X., M. S. Serfas, et al. (1999). "BRK/Sik expression in the gastrointestinal tract and in colon tumors." Clin Cancer Res 5(7): 1767-77.

Lorenzo, M. J., G. D. Gish, et al. (1997). "RET alternate splicing influences the interaction of activated RET with the SH2 and PTB domains of Shc, and the SH2 domain of Grb2." Oncogene 14(7): 763-71.

Lucini, C., L. Castaldo, et al. (1999). "Neuronal and non-neuronal Trk neurotrophin receptor-like proteins in Eisenia foetida (Annelida Oligochaeta)." Neurosci Lett 261(3): 163-6.

Lukong, K. E., M. E. Huot, et al. (2009). "BRK phosphorylates PSF promoting its cytoplasmic localization and cell cycle arrest." Cell Signal 21(9): 1415-22.

Lukong, K. E., D. Larocque, et al. (2005). "Tyrosine phosphorylation of sam68 by breast tumor kinase regulates intranuclear localization and cell cycle progression." J Biol Chem 280(46): 38639-47.

Lukong, K. E. and S. Richard (2008). "Breast tumor kinase BRK requires kinesin-2 subunit KAP3A in modulation of cell migration." Cell Signal 20(2): 432-42.

Mahmoud, K. A., M. K. , et al. (2014). "Discovery of 4-anilino alpha-carbolines as novel Brk inhibitors." Bioorganic & Medicinal Chemistry Letters 24: 948–1951.

Manning, G., D. B. Whyte, et al. (2002). "The protein kinase complement of the human genome." Science 298(5600): 1912-34.

Mi, Y., S. D. Thomas, et al. (2003). "Apoptosis in leukemia cells is accompanied by alterations in the levels and localization of nucleolin." J Biol Chem 278(10): 8572-9.

Miller, M. A., I. A. Malik, et al. (2000). "The Src/Csk regulatory circuit arose early in metazoan evolution." Oncogene 19(34): 3925-30.

Mitchell, P. J., K. T. Barker, et al. (1994). "Cloning and characterisation of cDNAs encoding a novel non-receptor tyrosine kinase, brk, expressed in human breast tumours." Oncogene 9(8): 2383-90.

Mitchell, P. J., K. T. Barker, et al. (1997). "Characterisation and chromosome mapping of the human non receptor tyrosine kinase gene, brk." Oncogene 15(12): 1497-502.

Mohamed R. Akl, A. I. F., Hassan Y. Ebrahim, Sharon A. Meyer and Khalid A. El Sayed (2014). "The Marine-Derived Sipholenol A-4-O-3',4'-Dichlorobenzoate Inhibits Breast Cancer Growth and Motility in Vitro and in Vivo through the Suppression of Brk and FAK Signaling" *Marine drugs* 12, 2282-2304.

Morrison, D. K., M. S. Murakami, et al. (2000). "Protein kinases and phosphatases in the Drosophila genome." J Cell Biol 150(2): F57-62.

Muller, W. E., M. Kruse, et al. (1999). "Gene structure and function of tyrosine kinases in the marine sponge Geodia cydonium: autapomorphic characters in Metazoa." Gene 238(1): 179-93.

Ogiso, H., R. Ishitani, et al. (2002). "Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains." Cell 110(6): 775-87.

Ono, H., M. D. Basson, et al. (2014). "PTK6 promotes cancer migration and invasion in pancreatic cancer cells dependent on ERK signaling." PLoS One 9(5): e96060.

Ostrander, J. H., A. R. Daniel, et al. (2007). "Breast tumor kinase (protein tyrosine kinase 6) regulates heregulin-induced activation of ERK5 and p38 MAP kinases in breast cancer cells." Cancer Res 67(9): 4199-209.

Ota, T., Y. Suzuki, et al. (2004). "Complete sequencing and characterization of 21,243 full-length human cDNAs." Nat Genet 36(1): 40-5.

Otake, Y., S. Soundararajan, et al. (2007). "Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA." Blood 109(7): 3069-75.

Papin, C., A. Denouel-Galy, et al. (1998). "Modulation of kinase activity and oncogenic properties by alternative splicing reveals a novel regulatory mechanism for B-Raf." J Biol Chem 273(38): 24939-47.

Park, J. W., C. Tokheim, et al. Identifying Differential Alternative Splicing Events from RNA Sequencing Data Using RNASeq-MATS. Deep Sequencing Data Analysis. N. Shomron. Totowa, NJ, Humana Press: 171-179.

Petro, B. J., R. C. Tan, et al. (2004). "Differential expression of the non-receptor tyrosine kinase BRK in oral squamous cell carcinoma and normal oral epithelium." Oral Oncol 40(10): 1040-7.

Plotnikov, A. N., S. R. Hubbard, et al. (2000). "Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity." Cell 101(4): 413-24.

Plowman, G. D., S. Sudarsanam, et al. (1999). "The protein kinases of Caenorhabditis elegans: a model for signal transduction in multicellular organisms." Proc Natl Acad Sci U S A 96(24): 13603-10.

Popovici, C., R. Roubin, et al. (1999). "The family of Caenorhabditis elegans tyrosine kinase receptors: similarities and differences with mammalian receptors." Genome Res 9(11): 1026-39.

Qiu, H. and W. T. Miller (2002). "Regulation of the nonreceptor tyrosine kinase Brk by autophosphorylation and by autoinhibition." J Biol Chem 277(37): 34634-41.

Qiu, H. and W. T. Miller (2004). "Role of the Brk SH3 domain in substrate recognition." Oncogene 23(12): 2216-23.

Qiu, H., F. Zappacosta, et al. (2005). "Interaction between Brk kinase and insulin receptor substrate-4." Oncogene 24(36): 5656-64.

Rikke, B. A., S. Murakami, et al. (2000). "Paralogy and orthology of tyrosine kinases that can extend the life span of Caenorhabditis elegans." Mol Biol Evol 17(5): 671-83.

Robinson, D. R., Y. M. Wu, et al. (2000). "The protein tyrosine kinase family of the human genome." Oncogene 19(49): 5548-57.

Rossel, M., A. Pasini, et al. (1997). "Distinct biological properties of two RET isoforms activated by MEN 2A and MEN 2B mutations." Oncogene 14(3): 265-75.

Sakabe, N. and S. J. Souza (2007). "Sequence features responsible for intron retention in human." BMC Genomics 8(1): 1-14.

Sakuma, M., H. Onodera, et al. (1997). "The protein tyrosine kinases of the sea urchin Anthocidaris crassispina." Zoolog Sci 14(6): 941-6.

Schmandt, R. E., M. Bennett, et al. (2006). "The BRK tyrosine kinase is expressed in high-grade serous carcinoma of the ovary." Cancer Biol Ther 5(9): 1136-41.

Scott, R. P., S. Eketjall, et al. (2005). "Distinct turnover of alternatively spliced isoforms of the RET kinase receptor mediated by differential recruitment of the Cbl ubiquitin ligase." J Biol Chem 280(14): 13442-9.

Serfas, M. S. and A. L. Tyner (2003). "Brk, Srm, Frk, and Src42A form a distinct family of intracellular Src-like tyrosine kinases." Oncol Res 13(6-10): 409-19.

Shen, C.-H. (2008). "Breast Tumor Kinase Phosphorylates p190RhoGAP to Regulate Rho and Ras and Promote Breast Carcinoma Growth, Migration, and Invasion." American Association for Cancer Research 10.1158/0008-5472.CAN-08-0997.

Shen, S., J. W. Park, et al. "MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data." Nucleic Acids Research 40(8): e61.

Sicheri, F. and J. Kuriyan (1997). "Structures of Src-family tyrosine kinases." Curr Opin Struct Biol 7(6): 777-85.

Srivastava, M. and H. B. Pollard (1999). "Molecular dissection of nucleolin's role in growth and cell proliferation: new insights." FASEB J 13(14): 1911-22.

Stauber, D. J., A. D. DiGabriele, et al. (2000). "Structural interactions of fibroblast growth factor receptor with its ligands." Proc Natl Acad Sci U S A 97(1): 49-54.

Suga, H., D. Hoshiyama, et al. (1999). "Protein tyrosine kinase cDNAs from amphioxus, hagfish, and lamprey: isoform duplications around the divergence of cyclostomes and gnathostomes." J Mol Evol 49(5): 601-8.

Sugnet, C. W., W. J. Kent, et al. (2004). "Transcriptome and genome conservation of alternative splicing events in humans and mice." Pac Symp Biocomput: 66-77.

Ullrich, A. and J. Schlessinger (1990). "Signal transduction by receptors with tyrosine kinase activity." Cell 61(2): 203-12.

Weaver, A. M. and C. M. Silva (2007). "Signal transducer and activator of transcription 5b: a new target of breast tumor kinase/protein tyrosine kinase 6." Breast Cancer Res 9(6): R79.

Wehrman, T., X. He, et al. (2007). "Structural and mechanistic insights into nerve growth factor interactions with the TrkA and p75 receptors." Neuron 53(1): 25-38.

Wilhelm, M., J. Schlegl, et al. "Mass-spectrometry-based draft of the human proteome." Nature 509(7502): 582-587.

Wise, J. F., Z. Berkova, et al. (2013). "Nucleolin inhibits Fas ligand binding and suppresses Fas-mediated apoptosis in vivo via a surface nucleolin-Fas complex." Blood 121(23): 4729-39.

Witte, O. N., A. Dasgupta, et al. (1980). "Abelson murine leukaemia virus protein is phosphorylated in vitro to form phosphotyrosine." Nature 283(5750): 826-31.

Yuzawa, S., Y. Opatowsky, et al. (2007). "Structural basis for activation of the receptor tyrosine kinase KIT by stem cell factor." Cell 130(2): 323-34.

Zeng, H., D. B. B. , et al. (2011). "Discovery of novel imidazo[1,2-a]pyrazin-8-amines as Brk/PTK6 inhibitors." Bioorganic & Medicinal Chemistry Letters 21: 5870–5875.

Zhang, B., H. Wang, et al. "Nucleolin/C23 is a negative regulator of hydrogen peroxide-induced apoptosis in HUVECs." Cell Stress Chaperones 15(3): 249-57.

Zhang, P., J. H. Ostrander, et al. (2005). "Regulated association of protein kinase B/Akt with breast tumor kinase." J Biol Chem 280(3): 1982-91.

Zhao, C., Y. Chen, et al. (2013). "Expression of protein tyrosine kinase 6 (PTK6) in nonsmall cell lung cancer and their clinical and prognostic significance." Onco Targets Ther 6: 183-8.

Zheng, Y. and A. L. Tyner (2013). "Context-specific protein tyrosine kinase 6 (PTK6) signalling in prostate cancer." Eur J Clin Invest 43(4): 397-404.

Zheng, Y., Z. Wang, et al. (2013). "PTK6 activation at the membrane regulates epithelial-mesenchymal transition in prostate cancer." Cancer Res 73(17): 5426-37.

Appendix

Vamshi Krishna Irlapati

Education

2010-till date: **Ph.D. Biotechnology.** Title: Functional characterization of a Breast Tumour Kinase splice variant, λm5: The significance of tyrosine kinome diversity due to alternate splicing, Research Supervisor: Dr. Kiranam Chatti, Dr. Reddy's Institute of Life Sciences, University of Hyderabad, India.

2006-2008: **M Sc,** Biotechnology, University of Mysore, Karnataka, India (Distinction).

2002-2005: **B Sc,** Biotechnology, Botany, Chemistry, Osmania University, Hyderabad, India.

Research Experience

2010-till date: **Ph.D (Thesis to be submitted),** Dr. Reddy's Institute of Life Sciences, University of Hyderabad, Hyderabad, India in **Dr. Kiranam Chatti's lab**.

Area of Research: Cancer Biology. The project is focused on exploring the function of the Breast Tumour Kinase (BRK) splice variant λm5 in cancer, and the relevance of novel splice variant production in the tyrosine kinome.

2009-2010: **Research Trainee,** Dr. Reddy's Institute of Life Sciences, University of Hyderabad, Hyderabad, India.

Area of Research: Molecular Biology: Cloning and expression of the Human glucose transporter GLUT4.

2008-2009: **Industrial Project Trainee,** BCIL fellowship**,** sponsored by Department of Biotechnology, Jupiter Biosciences Pvt. Ltd, India.

Area of Research: Biotechnology: Optimization studies in the production of optically pure Amino acids.

2008: **Master's project dissertation,** Basavatarakam Indo-American Cancer Hospital and Research Institute, Hyderabad**,** India.

Area of Research: Biotechnology: Cytotoxicity and genotoxicity evaluation of anti-neoplastic agents on human peripheral blood mononuclear cells (PBMCs), the results of which were published in ***Cancer Biomarkers*.**

Awards/Achievements

2016: My view for a question asked by science magazine has been selected as one of the top online essays in **Science Magazine** – Next Gen VOICES Global survey. Question: ***How do political priorities (or political sensitivities to particular groups) affect your ability to do or communicate science?***
http://science.sciencemag.org/content/351/6268/20/suppl/DC1

2015: My idea for an invention has been selected as one of the top online essays in **Science Magazine**- Next Gen VOICES Global survey. Title: ***NextGen Theranostic Tool for Fighting Global Pandemics.***
http://www.sciencemag.org/content/348/6230/32/suppl/DC1

2014: Qualified National Eligibility Test **(**NET**)** in Plant Biotechnology conducted by Agricultural
Scientists Recruitment Board (ASRB), Indian Council of Agricultural Research (ICAR), Govt.
of India.

2014: Qualified Telangana and Andhra Pradesh State Eligibility Test (SET-TS & AP) – 2014.

2009-(till Feb 2012): Junior Research Fellowship Award (JRF), Indian Council of Medical Research (ICMR), Govt. of India

2009-(till Feb 2015): Senior Research Fellowship Award (SRF), Indian Council of Medical Research (ICMR), Govt. of India

2009: National Eligibility Test for Lectureship (NET) - Council of Scientific and Industrial Research, (CSIR), Govt. of India.

2009: Cleared All-India Ph.D. entrance test conducted by NII – National Institute of Immunology, New Delhi, Govt. of India.

2008: Graduate Aptitude Test in Engineering (GATE) - 74.37%.

2008: Cleared All India entrance test (M.Tech. Bioinformatics) conducted by J.N.U, New Delhi, Govt. of India.

2006: Cleared All India combined entrance (Biotechnology) conducted by J.N.U., New Delhi, Govt. of India and achieved 29th Rank.

2006: Cleared Hyderabad Central University (HCU) M.Sc. Entrance (Biochemistry) and achieved All India 4th Rank

Publications

1. **Vamshi Krishna I**, Vanaja G.R, Srihari Kirmani Kumar.N, Suman G. Cytotoxic Studies of anti-neoplastic drugs on Human lymphocytes - in vitro studies, ***Cancer Biomarkers.*** **2009**; 5(6):261-272.

2. Alvala Ravi, Mallika Alvala, Venkatesh Sama, Arunasree M Kalle, **Vamshi K Irlapati,** B Madhava Reddy. Anticancer activity of Pupalia lappacea on chronic myeloid leukemia K562 cells. ***DARU Journal of Pharmaceutical Sciences,*** **2012,** 20:86. http://www.darujps.com/content/pdf/2008-2231-20-86.pdf

Manuscripts under review:

1. Soujanya D. Yelamanchi, Anand Srinivasan, Jayshree Advani, Satwant Kumar, Sandip Chavan, M. Rajyalakshmi, Nandini A. Sahasrabuddhe, Abhishek Chaturvedi, Amit Kumar, Apabrita Ayan Das, Dhiman Ghosh, Gajendra Jogdand, Gururao S. Desai, Haritha H. Nair, Keshav Saini, Manoj Panchal, Mansi Sarvaiya, Mohan Raj, Nabonita Sengupta, Priti Saxena, Pradeep Annamalai Subramani, Pradeep Kumar, Rakhil Akkali, Reshma S. V., Ramachandran Sarojini Santhosh, Sangita Rastogi, Sudarshan Kumar, Susanta Kumar Ghosh, **Vamshi Krishna Irlapati**, Parthasarathy Satishchandra, Akhilesh Pandey, Harsha Gowda, Kanchan K. Mukherjee, S. K. Shankar, Anita Mahadevan, T. S. Keshava Prasad. Extended proteome map of human pituitary glands. (***Pituitary-Springer publications***).

Manuscripts under preparation:

1. **Vamshi Krishna Irlapati and Kiranam Chatti**, Regulation of Breast Tumour Kinase activity by its catalytically inactive splice variant, λ m5.
2. **Vamshi Krishna Irlapati and Kiranam Chatti,** Splice variant production in the tyrosine kinome and its relevance.

Workshops/Symposium/Seminars

1. Secured first place in poster competition. International Conference on Translational Cancer Research (ICTCR) December 14-16, 2015. Avinashilingam University in coordination with Ohio State University.
2. Participated in DRILS Science Café-2015 held at Dr. Reddy's Institute of Life Sciences on August 10th, 2015.
3. International Conference on Genome Architecture and Cell Fate Regulation, held from 1st – 4th December, 2014, at University of Hyderabad, India.

4. Workshop on Flow Cytometry Basics and Immunophenotyping held from 10th - 12th March, 2014 organized by DBT-OU-ISLARE at Osmania University, Hyderabad, India.
5. 82nd Annual Meeting of the Society of Biological Chemists (India) and International Conference on "Genomes: Mechanism and Function" held from 2nd - 5th December, 2013 at School of LifeSciences, University of Hyderabad, Hyderabad, India.
6. Workshop on "Clinical Proteomics" held from 29th July - 01st August, 2013 at Institute of Bioinformatics, Bangalore, India.
7. Conference on "Animal Genetics and Genomics" held from 16 th - 19th December, 2012 by National Institute of Animal Biotechnology, Hyderabad, India.
8. XXXII All India Cell Biology Conference and International Workshop on "Cell Cycle Regulation", 2009 at University of Hyderabad, Hyderabad, India.

CLINICAL AND IN VITRO STUDIES ON A BREAST TUMOUR KINASE

ORIGINALITY REPORT

16%	14%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet Source	3%
2	www.jbc.org Internet Source	1%
3	www.tau.ac.il Internet Source	1%
4	www.biolchem.ucla.edu Internet Source	<1%
5	Sabine Druillennec. "Alternative Splicing in Oncogenic Kinases: From Physiological Functions to Cancer", Journal of Nucleic Acids, 2012 Publication	<1%
6	Submitted to University of Hyderabad, Hyderabad Student Paper	<1%
7	edoc.ub.uni-muenchen.de Internet Source	<1%

8	Submitted to Bilkent University Student Paper	<1%
9	www.neoplasia.com Internet Source	<1%
10	www.ensembl.org Internet Source	<1%
11	epub.uni-regensburg.de Internet Source	<1%
12	ecommons.usask.ca Internet Source	<1%
13	www.mcponline.org Internet Source	<1%
14	emboj.embopress.org Internet Source	<1%
15	www.bionewsonline.com Internet Source	<1%
16	opus.bibliothek.uni-wuerzburg.de Internet Source	<1%
17	Submitted to University of Adelaide Student Paper	<1%
18	mozigo.risko.org Internet Source	<1%
19	innovareacademics.in Internet Source	

		<1%
20	www.dailyexcelsior.com Internet Source	<1%
21	breast-cancer-research.com Internet Source	<1%
22	www.researchgate.net Internet Source	<1%
23	lib.bioinfo.pl Internet Source	<1%
24	Submitted to Universidade do Porto Student Paper	<1%
25	Ko, S.. "Structural basis of the auto-inhibition mechanism of nonreceptor tyrosine kinase PTK6", Biochemical and Biophysical Research Communications, 20090626 Publication	<1%
26	www.omicsonline.org Internet Source	<1%
27	www.dsi.univ-paris5.fr Internet Source	<1%
28	etheses.bham.ac.uk Internet Source	<1%
29	lra.le.ac.uk	

	Internet Source	<1%
30	f1000.com Internet Source	<1%
31	Submitted to University of Birmingham Student Paper	<1%
32	www.cancer-genetics.org Internet Source	<1%
33	Hadas Keren. "Alternative splicing and evolution: diversification, exon definition and function", Nature Reviews Genetics, 04/08/2010 Publication	<1%
34	publications.polymtl.ca Internet Source	<1%
35	Submitted to La Trobe University Student Paper	<1%
36	www.serology.sexmamuski.kaszuby.pl Internet Source	<1%
37	Submitted to Lehigh University Student Paper	<1%
38	www.alhadapedia.com Internet Source	<1%
39	Submitted to University of Queensland	

	Student Paper	<1%
40	www.nature.com Internet Source	<1%
41	jcs.biologists.org Internet Source	<1%
42	cdn.intechopen.com Internet Source	<1%
43	Miah, Sayem, Raghuveera Kumar Goel, Chenlu Dai, Natasha Kalra, Erika Beaton-Brown, Edward T. Bagu, Keith Bonham, and Kiven E. Lukong. "BRK Targets Dok1 for Ubiquitin-Mediated Proteasomal Degradation to Promote Cell Proliferation and Migration", PLoS ONE, 2014. Publication	<1%
44	www.enexpress.net Internet Source	<1%
45	europepmc.org Internet Source	<1%
46	genomebiology.com Internet Source	<1%
47	d-scholarship.pitt.edu Internet Source	<1%

48	nslij-genetics.org Internet Source	<1%
49	Receptor Tyrosine Kinases Family and Subfamilies, 2015. Publication	<1%
50	www.stamms-lab.net Internet Source	<1%
51	Submitted to University College London Student Paper	<1%
52	ec.asm.org Internet Source	<1%
53	cloneresources.com Internet Source	<1%
54	www.hg17.com Internet Source	<1%
55	Ostrander, J.H.. "Brk/PTK6 signaling in normal and cancer cell models", Current Opinion in Pharmacology, 201012 Publication	<1%
56	www.nssanu.org Internet Source	<1%
57	www.molecular-cancer.com Internet Source	<1%
58	www.sdbonline.org	

	Internet Source	<1%
59	Submitted to McMaster University Student Paper	<1%
60	gbpuat-cbsh.ac.in Internet Source	<1%
61	Submitted to Academia Sinica Life Science Library Student Paper	<1%
62	J M S Bartlett. "Prognostic value of protein tyrosine kinase 6 (PTK6) for long-term survival of breast cancer patients", British Journal of Cancer, 10/07/2008 Publication	<1%
63	www.rgcb.res.in Internet Source	<1%
64	www.pierce-antibodies.com Internet Source	<1%
65	"A draft map of the human proteome.(Report)", Nature, May 29 2014 Issue Publication	<1%
66	Submitted to University of New South Wales Student Paper	<1%
67	B. Xiang. "Brk is coamplified with ErbB2 to promote proliferation in breast cancer",	<1%

Proceedings of the National Academy of Sciences, 08/26/2008
Publication

68	Submitted to University of Texas Health Science Center Student Paper	<1%
69	www.intechopen.com Internet Source	<1%
70	Burmi, RS and Harvey, A. "Future therapeutic strategies: Implications for Brk targeting", InTech, 2012. Publication	<1%
71	Submitted to University of KwaZulu-Natal Student Paper	<1%
72	Submitted to Macquarie University Student Paper	<1%
73	real-d.mtak.hu Internet Source	<1%
74	www.science.gov Internet Source	<1%
75	www.plob.org Internet Source	<1%
76	www.bacterio.cict.fr Internet Source	<1%

www.s113616465.onlinehome.us

77	Internet Source	<1%
78	Receptor Tyrosine Kinases Structure Functions and Role in Human Disease, 2015. Publication	<1%
79	Submitted to Indian Institute of Science Education and Research Student Paper	<1%
80	jb.asm.org Internet Source	<1%
81	aac.asm.org Internet Source	<1%
82	www.amedeo.com Internet Source	<1%
83	Submitted to Brunel University Student Paper	<1%
84	www.mdpi.com Internet Source	<1%
85	documents.irevues.inist.fr Internet Source	<1%
86	www.uniprot.org Internet Source	<1%
87	www.molbiolcell.org Internet Source	<1%

88	www.thesis.bilkent.edu.tr Internet Source	<1%
89	165.112.7.20 Internet Source	<1%
90	www.citeulike.org Internet Source	<1%
91	smart.embl-heidelberg.de Internet Source	<1%
92	www.ijps.ir Internet Source	<1%
93	www.tufts-mori.org Internet Source	<1%
94	www.bioscience.org Internet Source	<1%
95	paperity.org Internet Source	<1%
96	settsap.org Internet Source	<1%
97	www.n-genetics.com Internet Source	<1%
98	mcb.asm.org Internet Source	<1%
99	www.freepatentsonline.com Internet Source	

		<1%
100	www.scialert.net Internet Source	<1%
101	tampub.uta.fi Internet Source	<1%
102	intl.molbiolcell.org Internet Source	<1%
103	www.tumorbiologie-freiburg.de Internet Source	<1%
104	www.hcnet.usp.br Internet Source	<1%
105	Chastkofsky, Michael, Wenjun Bie, Susan M Ball-Kell, Yu-Ying He, and Angela L Tyner. "Protein Tyrosine Kinase 6 Regulates UVB Induced Signaling and Tumorigenesis in Mouse Skin", Journal of Investigative Dermatology, 2015. Publication	<1%
106	Krauss, . "Signal Transmission via Transmembrane Receptors with Tyrosine-Specific Protein Kinase Activity", Biochemistry of Signal Transduction and Regulation, 2014. Publication	<1%
107	journal.frontiersin.org	

	Internet Source	<1%
108	lindinglab.org Internet Source	<1%
109	Stéphane Richard. "Reaching for the STARs", Advances in Experimental Medicine and Biology, 2010 Publication	<1%
110	Rosemarie E. Schmandt. "The BRK tyrosine kinase is expressed in high-grade serous carcinoma of the ovary", Cancer Biology & Therapy, 09/01/2006 Publication	<1%
111	ediss.uni-goettingen.de Internet Source	<1%
112	doc.rero.ch Internet Source	<1%
113	archiv.ub.uni-marburg.de Internet Source	<1%
114	doras.dcu.ie Internet Source	<1%
115	medlibrary.org Internet Source	<1%
116	www.plantcell.org Internet Source	<1%

117	en.wikipedia.org Internet Source	<1%
118	compbio.soe.ucsc.edu Internet Source	<1%
119	e-sciencecentral.org Internet Source	<1%
120	www.sciencemag.org Internet Source	<1%
121	crick-jacobs.salk.edu Internet Source	<1%
122	weight-loss.fitness.com Internet Source	<1%
123	www.pnas.org Internet Source	<1%
124	www.biovision.com Internet Source	<1%
125	repository.ias.ac.in Internet Source	<1%
126	Wang, Eric T. Sandberg, Rickard Luo, Shu. "Alternative isoform regulation in human tissue transcriptomes.", Nature, Nov 27 2008 Issue Publication	<1%
127	jcb.rupress.org Internet Source	<1%

128	digitool.library.mcgill.ca Internet Source	<1%
129	X Li. "Brk/PTK6 sustains activated EGFR signaling through inhibiting EGFR degradation and transactivating EGFR", Oncogene, 01/09/2012 Publication	<1%
130	Handbook of Anticancer Drugs from Marine Origin, 2015. Publication	<1%
131	"Findings from China Pharmaceutical University in Breast Cancer Reported (Inhibition of metastatic tu", Biotech Week, August 5 2015 Issue Publication	<1%
132	Park, Sun Hee, Koichi Ito, William Olcott, Igor Katsyv, Gwyneth Halstead-Nussloch, and Hanna Y. Irie. "PTK6 inhibition promotes apoptosis of Lapatinib-resistant Her2+ breast cancer cells by inducing Bim", Breast Cancer Research, 2015. Publication	<1%
133	Ikeda, Osamu, Yuichi Sekine, and Tadashi Matsu. "Interactions of STAP-2 with BRK and STAT3/5 in Breast Cancer Cells", Breast Cancer - Carcinogenesis Cell Growth and Signalling Pathways, 2011.	<1%

Publication

134	Goel, Raghuveera Kumar, and Kiven Erique Lukong. "Tracing the footprints of the breast cancer oncogene BRK — Past till present", Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 2015. Publication	<1%

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	< 5 WORDS
EXCLUDE BIBLIOGRAPHY	ON		