Genome-wide analysis of Sam68 as a prognostic biomarker in cancer and fabrication of a biosensor for clinical diagnosis

THESIS

Submitted in partial fulfillment of the requirements
for the award of the degree of

Doctor of Philosophy

IN
BIOTECHNOLOGY

BY

SUMITHRA B
(Roll No. 701435)

Dr. ASIM BIKAS DAS

RESEARCH SUPERVISOR

DEPARTMENT OF BIOTECHNOLOGY
NATIONAL INSTITUTE OF TECHNOLOGY
WARANGAL-506004, (T. S.), INDIA
JULY – 2020

NATIONAL INSTITUTE OF TECHNOLOGY

WARANGAL

DEPARTMENT OF BIOTECHNOLOGY

CERTIFICATE

This is to certify that the thesis entitled **"Genome-wide analysis of Sam68 as a prognostic biomarker in cancer and fabrication of a biosensor for clinical diagnosis"** that is being submitted by **Ms. SUMITHRA B** (Roll No.701435) in partial fulfilment for the award of Doctor of Philosophy (**Ph.D.**) in the Department of Biotechnology, National Institute of Technology, Warangal, is a record of bonafide work carried out by her under my guidance and supervision. The results embodied in this thesis have not been submitted to any other Universities or Institutes for the award of any degree or diploma.

Dr. Asim Bikas Das
(Supervisor)
Assistant Professor
Department of Biotechnology
NIT-Warangal

NATIONAL INSTITUTE OF TECHNOLOGY
WARANGAL

DECLARATION

This is to certify that the work presented in the thesis entitled **"Genome-wide analysis of Sam68 as a prognostic biomarker in cancer and fabrication of a biosensor for clinical diagnosis"**, is a bonafide work done by me under the supervision of Dr. Asim Bikas Das was not submitted elsewhere for the award of any degree.

I declare that this written submission represents my idea in my own words and where other's ideas or words have been included, I have adequately cited and referenced the sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misinterpreted or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not taken when needed.

Date:

Place: Warangal

(Ms. Sumithra B)
Research Scholar
(Roll No. 701435)

ACKNOWLEDGMENT

*First of all, I would like to express my deepest gratitude to my supervisor **Dr. Asim Bikas Das**, Assistant Professor, Department of Biotechnology, NIT Warangal for the continuous support & inspiration throughout my Ph.D. study and research. It is because of his immense knowledge, motivation and scientific freedom has always encouraged me to continue work with new ideas and writing of this thesis. Under his able guidance, I have learn a lot, overcame many difficulties and have been inspired for which I would always be indebted. I could not imagine having a better advisor and mentor for my Ph.D. study.*

*I would like to thank my Doctoral scrutiny committee: **Prof. K. Laxma Reddy,** Department of Chemistry**, Dr. R. Satish Babu,** Department of Biotechnology and **Dr. P. Srinivasa Rao**, Department of Biotechnology, National Institute of Technology, Warangal for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.*

*I sincerely thank, **Prof. N.V. RAMANA RAO**, Director, National Institute of Technology, Warangal and other authorities who gave me this opportunity to carry out my research work.*

*I wish to extend my heartfelt thanks to **Dr. Onkara Perumal**, previous HOD, Department of Biotechnology, for providing computational facility and also understanding and guiding me properly in this journey. I also thank **Dr. Amitava Bhadhu,** Dept. of Biotechnology, for all those valuable suggestions and infrastructure facilities provided for my study.*

*I will always be indebted to **Dr. V Kohila,** Dept. of Biotechnology, for being my source of support and guidance. She has always inspired me and have being there during all those tough situations not only in this PhD journey but also in life. Thank you for helping me till this day.*

*I also owe a special thanks to **Dr. Urmila Saxena,** Dept. of Biotechnology, for helping me every possible way to complete this work and also for guiding me throughout my journey.*

*I would like to extend my thanks to **all the faculty members** in the **Department of Biotechnology** who has directly or indirectly helped me during my work. I also acknowledge all the **supporting and technical staff** of the **Department of Biotechnology,** for providing research environment and necessary facilities when required.*

*I am blessed to meet **Kaushik & Nirupama Medavaram** and all their family members, who have taken a lot of care and efforts during my stay here at NIT Warangal. I cannot count on the help and affection but express my heartfelt thanks.*

*I thank and appreciate all my friends, scholars and genetic engineering labmates for all those fun, support and advice. I also thank my batchmates **Bhaskar, Manasa, Aditya,** for the valuable discussions and the fun we have had in the past. Special thanks to all my wonderful hostel-mates who have all made this experience enjoyable.*

*Most of all, my heartfelt gratitude and indebtedness to my parents **Shri. Jayadasu B** & **Smt. Anjamma,** my Husband **Mr. Ram Sagar B and** my Son **Master Sai Tanay.** Whatever I'm today is all because of their immeasurable sacrifices, endless love and support. Even in the most difficult situations, unflinching courage & encouragement of my father & husband have inspired me to continue my research work, without which this thesis would not have been possible. Words will not suffice to express my reverence and thankfulness, but I dedicate this thesis to you all. I also owe special thanks to my sister **Revathi** and my mother in-law, **Smt. Sasirekha** for their sincere prayers & blessings.*

*Finally, I thank, "**ALMIGHTY**" whose divine light and warmth showered upon me the perseverance and enough strength to keep the momentum of work high even at tough moments of research work.*

(Sumithra B)

SUMMARY

Sam68 is an oncogenic splicing factor involved in cell signaling pathways and pre-mRNA splicing. Altered expression of Sam68 mis-regulates mRNA splicing events and generate cancer-specific transcripts that contribute to Oncogenesis. However, underlying molecular mechanisms of its association with cancer phenotype and the significance of its expression is unclear.

In this study, we used high throughput RNA sequencing Datasets of four different types of Cancer including kidney renal papillary cell carcinoma (KIRP), lung adenocarcinoma (LUAD), acute myeloid leukemia (LAML), and ovarian cancer (OV) obtained from The Cancer Genome Atlas (TCGA). We initially analyzed the expression of Sam68 and its functional consequence by assessing the prognostic impact on the overall survival of cancer patients. Then performed Genome-wide co-expression analysis both at gene and transcript level to identify correlating proteins that form a functional cancer-specific network of Sam68. For further understanding of Sam68 protein interactors and target transcripts, we explored the molecular mechanisms behind differential biological functions and prognostic value in these cancer types. This computational analysis revealed Sam68 as a prognostic biomarker and a potential therapeutic target in KIRP and LUAD due to its cancer-specific interaction partners and functional correlation networks. Additionally, we analyzed lung cancer stage-specific patient data which reveals that differential expression of Sam68 mRNA is associated with the clinical tumor stage. Based on these observations, an electrochemical immunosensor was developed for the quantification of Sam68 protein in LUAD. The target protein was captured by the Anti-Sam68 antibody that was immobilized on the modified Glassy carbon electrode. This fabricated immunosensor displayed good analytical performance in comparison to the commercial ELISA kit with sensitivity and lower detection limits (LOD).

Herein, we report the first study using Genome-wide analysis that shows Sam68 as a cancer-specific & patient-specific Biomarker in KIRP and LUAD. Further, Sam68 is a stage-specific diagnostic and prognostic biomarker in lung cancer. Then finally, the development of a sensitive antibody-based sensor for detection of Sam68 protein, a novel patient-specific early biomarker in lung cancer.

Contents

List of Symbols & Abbreviations

µl	Microliter
µg/mL	Microgram per milliliter
2XTY	Yeast Extract tryptone medium
Abs	Antibodies
AS	Alternative Splicing
Bp	Base pairs
BSA	Bovine serum albumin
CV	Cyclic voltammetry
CYFRA21-1	Cytokeratin 19 fragment 21-1
DEGs	Differentially expressed genes
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
E.coli	Escherichia coli
EGF	Epidermal growth factor
HGF	Hepatocyte growth factor
EIS	Electrochemical impedance spectroscopy
ELISA	Enzyme-linked immunosorbent assay
ERK	extracellular-signal-regulated kinase
FISH	Fluorescence in-situ hybridization
GAF	Genomic Annotation File
GCE	Glassy carbon electrode
GDAC	Broad's Genome Data Analysis Center
GDC	Genomic Data Commons
GEO	Gene expression omnibus
GO	Gene ontology
GSG	GRP33, Sam68, GLD-1
GSH	Glutathione
GST	Glutathione S-transferase
hnRNP	Heterogeneous ribonucleic proteins
HRP	Horseradish peroxidase
HRP-Ab2	Horseradish peroxidase-labeled secondary antibody
Hrs.	Hours
Hz	Hertz
IHC	Immunohistochemistry
IPTG	Isopropyl β- d-1-thiogalactopyranoside
IS	Immunosensor

Kbp	Kilo base pairs
KDa	Kilo Dalton
KH	K Homology
Sam68	KH RNA Binding Domain Containing, Signal Transduction Associated 1
KIRP	Kidney renal papillary cell carcinoma
LAML	Acute myeloid leukemia
LB	Luria-Bertani
LH	luteinizing hormone
LOD	Limit of detection
LOQ	Limit of quantification
LUAD	Lung adenocarcinoma
MCODE	Molecular complex detection
Mg	Milligram
Min	Minute
miRNA	MicroRNA
Ml	Milliliter
mM	Milli molar
MRI	Magnetic resonance imaging
mRNA	Messenger Ribonucleic Acid
mV	Millivolts
mV/sec or mV s-1	Millivolts per second
MWCNT-NH2	Amino-functionalized carbon nanotubes
NCBI	National center for biotechnology information
NCI	National Cancer Institute
NCI-H23	Human lung adenocarcinoma cell line
ng/mL	Nano gram/Milliliter
NGS	Next-generation sequencing
NLS	Nuclear localization signal
Nm	Nanometer
NSCLC	non-small cell lung cancer
NSE	Neuron-specific enolase
OD	Optical density
OS	Overall survival
OV	ovarian cancer
P-value	Probability value
PAGE	Polyacrylamide gel electrophoresis
PANI	Polyaniline
PBS	Phosphate buffered saline
PB-SiO2	Prussian blue doped silica dioxide
PBST	Phosphate buffered saline with tween 20
PC	prostate cancer

PCR	Polymerase chain reaction
pg/mL	Pico gram per milliliter
pH	Potential of Hydrogen
PMSF	Phenylmethylsulfonyl fluoride
POCD	Point of care diagnosis
PPI	Protein-Protein Interactions
PSA	Prostate-specific Antigen
PVDF	polyvinylidene difluoride
RasGAP	Ras GTPase-activating protein
RBM4	RNA-binding motif protein 4
RBP	RNA-binding protein
Rct	charge transfer resistance
RG	Arginine- glycine motif
RIPA	Radio immunoprecipitation assay
RNA	Ribonucleic acid
RPKM	Reads Per Kilobase Million
Rs	Spearman correlation coefficient
RSD	Root-mean Square Deviation
RSEM	RNA-Sequencing by Expectation-Maximization
RT-PCR	real-time polymerase chains
Sam68	Src-Associated substrate in Mitosis of 68 kDa
SD	Standard Deviation
SDS	Sodium dodecyl sulphate
Src	Src proto-oncogene, non-receptor tyrosine kinase
PAGE	Polymerase Agarose Gel Electrophoresis
SELEX	Systematic evolution of ligands by exponential enrichment
SFK	Src-family kinases
SF	Splicing factor
SH	Src homology
SNBs	Sam68 nuclear bodies
SPCE	screen-printed carbon electrode immunosensor
SR	Serine/arginine-rich
SRSF1	Serine And Arginine Rich Splicing Factor 1
STAR	Signal Transduction and Activation of RNA
TAQ	Thermus aquaticus (thermostable DNA polymerase I)
TCGA	The Cancer Genome Atlas
UCSC	The University of California, Santa Cruz (genome browser)
UTR	Untranslated Region
UV	Ultraviolet
V	Voltage

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Chapter 1: Introduction

1.1. Alternative Splicing

Eukaryotic genes are copied into a primary transcript known as pre-mRNA that harbors both noncoding introns and coding exons. A regulatory mechanism of gene expression, Splicing is a post-transcriptional, ubiquitous & essential processing step of the preliminary transcript (pre-mRNA) of genes that involves excision of introns and joining of exons to produce shortened functional mRNA. In the human genome, the number of protein encoding genes was estimated to be 25,000 to 30,000. However, the cellular processes involve a farther number of different proteins (~150,000) for proper structural and functional regulations (Oltean S. *et al.,* 2013, Black DL. 2003). By choosing alternative mechanisms of RNA splicing known as Alternative Splicing (AS) improves the coding capacity of the gene. AS involves choosing differential 3'/5' splice sites, cassette exon, exon shuffling, intron retention, alternative exons in the primary transcript, thus, generating multiple isoforms of a single protein that differ in structure and functions and sometimes have opposing functions. By this strategy, different isoforms of a protein from a single gene can be produced, thus amplifying the coding potential of the genome and complexity of transcriptome and proteome (Liu S. 2013, Wang *et al.,* 2015). Plasticity *of splicing gives the advantage of "One gene, multiple mRNA's encoding different polypeptides" having distinct structural and, different functions.* Thus, AS is important in varying the protein complexity and cellular functions depending on developmental stages and cellular/ tissue phenotypes. Significantly, misregulation in Alternative Splicing profiles can produce aberrant isoforms that contribute to developing new phenotypes of the cell which has implications in developing diseases including cancer (David CJ. *et al.,* 2010, Chen J *et al.,* 2014).

1.2. Signaling Pathways

On the other hand, cells use a large number of intracellular signaling pathways to act in sequence or parallel to regulate their activity and adapt to their surroundings to maintain homeostasis. Signaling systems responds to internal metabolic messengers or external stimuli as hormones, morphogens that are secreted from other cell or tissue types and activates transduction components and transcription factors, thereby regulating target gene expressions important for cellular processes including cell proliferation, apoptosis, and differentiation. So, the diverse

functions of signaling pathways are mainly by altering potency of transcriptional regulators and further regulating transcriptional activation of genes in a context-dependent manner (Weigelt B. *et al.*, 2010, Parikh *et al.*, 2010).

1.3. Splicing and Signaling Connection in Cancer

Cancer cells adapt and evolve to the surrounding pressures, which monitor the cell division and ultimately lead to the death of the old cells through apoptotic mechanisms, by employing several strategies and thus achieve replicative immortality. The development of cancer is often connected with dysregulations in the signaling pathways to modulate various gene expression profiles, either sustained expression or repression of the target genes (Oltean S. *et al.*, 2014). In addition, cancer cells also took the advantage of alternative splicing, another major regulatory mechanism to express the isoforms that stimulate the continued proliferation of the cells by counteracting the growth regulatory processes (Liu S. *et al.*, 2013, Bonomi S. *et al.*, 2013). For instance: aberrant splicing of proto-oncogenes can produce constitutively active or even gain-of-function variant, in contrast, a tumor suppressor with aberrant splicing could result in loss of function that may confer new survival or proliferative abilities. (Chen J. *et al.*, 2015). Also, accumulating evidences have shown alternative splicing regulation is linked with signal transduction pathways and contextual isoforms are produced as a cellular response to internal and external stimuli (Black DL. 2003, Venables JP. *et al.*, 2004, David CJ. *et al*., 2010, Oltean S. *et al.*, 2014, Liu S. *et al.*, 2013, Bonomi S. *et al.*, 2013, Chen J. *et al.*, 2015). For example, β-catenin induced changes in expression of the splicing regulator SRSF3 (SRP20), affects the alternative splicing of the oncogenic cell adhesion molecule: CD44 mRNA. Altered expression of CD44 isoforms is known to play a critical role in anomalous cellular proliferation, tumor initiation and metastasis in colon cancer (Sumithra *et al.*, 2016). Further, Changes in splicing patterns occur in a context-dependent manner for genes that are needed in every step of the transformation process in cancer development, progression and metastasis. However, these tumor-associated splicing changes reflect alterations in splicing machinery, particularly splicing factors which are RNA binding proteins (RBPs) involved in splicing regulations (El Marabti Ettaib, Younis Ihab. 2018, Anczuków *et al.*, 2016). Only in recent times, these splice factors are being investigated, but they play a prominent role in cancer. So, inhibiting splice factors could pave a new path for oncogenic control and can be a powerful technique to combat cancer progression. (Koedoot E. *et al.*, 2019).

A better understanding of alternative splicing regulations in a given condition involves a systematic approach or Genome-wide analysis in which the expression and multiple targets of RBPs are evaluated for a better understanding of cancer etiology. (El Marabti Ettaib and Younis Ihab. 2018).

1.4. Sam68

In this respect, the present research focuses on "Sam68": Src substrate associated in mitosis of 68 KDa, is a member of Signal Transduction and Activation of RNA (STAR) family of nuclear RNA binding proteins. Sam68 protein is encoded by KHDRBS1 gene: KH domain-containing, RNA-binding, signal transduction-associated protein 1. This protein is a synonym of KHDRBS1. Sam68 is implicated in a wide variety of cellular processes including signal transduction, mitosis, cell cycle progression, as well as RNA processing as pre-mRNA splicing, mRNA export, mRNA stability, and protein translation. (Bielli P. *et al.*, 2011 and Lukong KE. 2003). Sam68 is mainly involved in linking pre-mRNA splicing and signal transduction pathways (Lukong KE. 2003). The function of Sam68, in turn, is highly regulated by the cell signaling pathway, thus provides the link between signaling and mRNA splicing. The dual function of Sam68 is due to the presence of highly conserved KH (K-homology) & SH (Src homology) domain, which are involved in RNA binding and signal transduction pathway respectively (Najib S. *et al.,* 2005). Therefore, external cues could influence the splicing pattern of the Sam68 target gene. Matter *et al.,* have shown that phosphorylation of Sam68 via the ERK pathway modulates the alternative splicing of the CD44 gene (Matter N. *et al.,* 2002). Evidently in a cancer cell, RNA splicing machinery receives aberrant signaling response via Sam68 and results in the generation of oncogenic splicing variants (Frisone P. *et al., 2015*). In addition to functional domains, the post-transcriptional modifications as phosphorylation, methylation, acetylation and sumoylation within regulatory regions of Sam68 also finely control the subcellular localizations, target RNA binding affinity and functions and the interplay with other signaling components (Bielli P. *et al.,* 2011). Crucially, the multimodular structure of Sam68 contributes to cross-talk between cell signaling, transcription, and RNA processing in a context-dependent manner.

Involvement of Sam68 in cancer-specific cellular processes including apoptosis, cell proliferation, chemo-resistance, and metastasis are well established (Paronetto MP. *et al.*, 2011, Sánchez-Jiménez F. *et al.,* 2013). Particularly, Sam68 mediated cancer-specific events are

regulated by misregulation of alternative splicing in cancer which results in the generation of the oncogenic splicing isoforms (Stockley J. *et al.,* 2015, Paronetto MP. *et al.,* 2007). Several studies illustrate Sam68 as an oncogenic protein with good prognostic value in various cancers (Matter N. *et al.,* 2005, Stockley J. et al, 2015, Paronetto MP. *et al.,* 2007. However, none of these studies used a holistic approach to understand the molecular mechanisms of Sam68 involved in the pathogenesis of different cancer types. Moreover, despite its prognostic value, there is no single study that reports a detection system which utilizes Sam68 as a biomarker.

In this thesis, we study the prognostic value of Sam68 as a cancer biomarker and focused to develop a sensitive detection system as electrochemical biosensor based on PANI to accurately quantify Sam68. PANI based immunosensor show good advantages as in neutral pH, PANI enables good conductivity, increased charge-transfer and matrix stability for immobilization (Cho IH. *et al.,* 2018).

- **Organization of the Thesis**

The thesis presents the work in five chapters and the following section gives the outline.

Chapter 1: Presents a General introduction to work: Alternative splicing, Signaling, Signaling and Splicing in Cancer, splicing factor in cancer, Introduction to Sam68, and Organization of the thesis.

Chapter 2: Presents literature review on splicing mis-regulations, the role of splicing factor in cancer, splicing factors as cancer biomarkers, Sam68: Domain structure, biological functions, expression and prognostic value in cancer. Detection methods for biomarkers in cancer. Also, the role of RNA-Seq and microarray data and genome-wide analyses for new biomarker identification are discussed. Finally, the different types of electrochemical immunosensors reported earlier for the detection of cancer biomarkers are broadly covered in this section and finally Aim of the Work.

Chapter 3: Presents a detailed description of the materials and methods used in this work in three parts. The first part discusses the data retrieval, analysis tools and methodology used for the *In-silico* approach (for both RNA-Seq and Microarray analysis). The second part is about protocols of Sam68 plasmid preparation, Restriction digestion, ligation, polymerase chain reaction (PCR),

transformation, Sam68 protein expression, purification and confirmation by western blot. This part also covers Culturing conditions of lung adenocarcinoma cell line (NCI-H23), whole-cell protein lysate of NCI-H23 preparation and then, confirmation of expression of Sam68 using western blot. Finally, the last part discusses methods of Indirect ELISA, description of the electrochemical station used, Immunosensor fabrication, characterization, spike and recovery are broadly discussed.

Chapter 4: Presents the results and discussion related to the *in-silico* work, identification of Sam68 as a potential biomarker in multiple cancer types. Our results show higher expression reduced survival of the patient in KIRP and LUAD but not in LAML and OV. Genome-wide correlation analysis both at gene and transcript level was performed in four different cancer to screen direct interactors that have a significant correlation. Next, we identified recurrent network modules involved in cancer driven biological processes are KIRP and LUAD not in LAML and OV. This presents the Sam68 prognostic value in KIRP and LUAD. Thus, we choose to develop an immunosensor for Sam68 in LUAD. The second part of this section consists of results related to the expression and purification part of Sam68 recombinant protein. The polyclonal anti-Sam68 antibody was purchased and tested to check their interaction with purified recombinant protein by western blot and indirect ELISA. All the results showed that recombinant Sam68 protein was interacting with anti-Sam68 antibody and no non-specific band was present. In the final part, immunosensor fabrication and characterization are discussed. Immunosensor fabrication was carried out using glassy carbon electrode, poly-aniline as immobilization matrix, and glutaraldehyde as a crosslinker, polyclonal anti-Sam68 antibody as a recognition molecule and BSA as blocking agent. Characterization of this immunosensor was carried out using cyclic voltammetry (CV) and electrochemical impedance spectroscopy (EIS). This immunosensor was calibrated with different concentrations of recombinant Sam68. Lower detection limits and limits of quantification of immunosensor were achieved. Then immunosensor was also used to detect Sam68 present in Lung cancer cell line NCI-H23 cell lysate. All the results of this immunosensor are explained in this section.

Chapter 5: Presents the conclusions drawn from the analysis of the prognostic value of Sam68 and biosensor fabrication. Potential future work and the scope of this work are also summarized here.

CHAPTER 2

REVIEW OF LITERATURE

Chapter 2: Review of Literature

Studies on Genome-wide deep sequencing and exome analysis reveals the presence of cancer-specific splicing patterns, which contributes to re-annotate the proteome, that accounts for altered cellular programs and provides opportunity for cancer cells to escape from their surrounding environment. Recent studies have revealed splicing as a hallmark of cancer (Liu S. and Cheng C. 2013, Martínez-Montiel, N. *et al.,* 2017), since the tumor-specific splicing variants contribute to various characteristics of tumor biology including proliferation, angiogenesis, metastasis and invasion, aberrant metabolism, resistance to apoptosis, immune response and chemotherapy (Chen J. and Weiss WA. 2015, Lukong KE. *et al.,* 2003). The Spatio-temporal expression and biological functions of mRNA splice variants are influenced by different developmental stages and extracellular cues. Moreover, pieces of evidence have shown that alternative splicing is linked with signal transduction pathways and contextual isoforms are produced as a result of cellular response (David CJ. *et al.,* 2010, Liu S. *et al.*, 2013).

2.1. Splicing Mis-regulations in Cancer:

Alternative splicing is a tightly regulated multi-stage process that produces different isoforms with a function unique to the cell type or disease. Certain splicing isoforms are expressed when required e.g.: stress, metastasis (Younis I. *et al.,* 2013). Splicing is altered in several proto-oncogenes, tumor suppressor genes, thus producing tumor-specific isoforms that vary from normal cells. (Oltean S. and Bates D. 2014). Splicing regulation is catalyzed by a complex made of five small nuclear ribonucleoproteins (U1, U2, U4, U5 and U6), spliceosome and more than 250 additional regulatory proteins that are splicing factors (Hegele MN. *et al.,* 2012). Studies from the past two decades suggest splicing regulations are significantly altered in cancer (David CJ. *et al.,* 2010, Chabot B. and Shkreta L. 2016, Scotti MM. *et al.,* 2016). From our understanding, we provide a comprehensive list of mis-regulations, which alters the combination of splice site selection and leads to abnormal splicing events. Thus, tumor-specific splicing changes can result from:

1. Alterations in core spliceosomal components can lead to global splicing deregulation and result in a large number of aberrant isoform.

2. Alterations, e.g.: Mutations in splice sites within the pre-mRNA, make the splicing factors to produce different splicing transcripts of that particular pre-mRNA.
3. Alterations in regulatory proteins as splicing factors can deregulate the splicing for a set of transcripts where that specific factor is needed for accurate splicing.

2.2. Role of Splicing Factors in Cancer

AS (Alternative splicing) mechanism is largely regulated by splicing factors. Splicing factors (SF) are RNA-binding proteins whose function is to recognize and bind to splice site donors and acceptors with varying specificity. Thereby, recruits the splicing machinery to decide the fate of surrounding sequences of the pre-mRNA. Splice factors are significantly different considering their structure and function, which can either block or guide the association between target pre-mRNAs and Spliceosome. Thus, SF's can activate, inhibit or modulate the splicing. (Long LC. *et al.,* 2009). Moreover, the expression of cancer-specific isoforms relates to that of certain splice factors. This suggests that the characteristic features of cancer, including sustaining proliferation, inhibiting cell death, angiogenesis, mis-regulating cellular energetics and metastasis, could also be regulated by splice factors, by affecting major signaling pathways such as epithelial-to-mesenchymal transition, proliferation due to hormone receptor, DNA damage response and the Warburg effect (Koedoot E. *et al.,* 2019).

Historically splicing factors belong two different families: a) serine/arginine-rich (SR) proteins which act as coactivator's and b) heterogeneous ribonuclear proteins (hnRNPs) that could either activate Exon splicing enhancer (ESEs) and inhibit splicing by binding introns (Long LC. *et al.,* 2009, Dvinge *et al.,* 2016, Yang Q. *et al.,* 2019). SF's regulate splicing through different routes which get dysregulated in cancer. Several aspects related to the splice factors dictate the distinct outcome of splicing events in cancer, like changes in expression levels, structural variations due to posttranscriptional modifications, localization, functional activity, mutations and finally the modulations of signaling pathways that transduce the change in splice factors. (Wang Y. *et al.,* 2014, Yang Q. *et al.,* 2019, Baimonti G. *et al.,* 2014, El Marabti Ettaib, Younis Ihab. 2018). In-addition, strength and context of binding to the target and stotiometric relation that is either competitive or cooperative with other RBPs of the target are crucial in deciding the end splicing pattern (El Marabti Ettaib, Younis Ihab. 2018). Such combinatorial regulations are interdependent, thus, it is very difficult to comprehend the difference is splicing

outcomes of single splicing factors in diseased states compared to normal (Koedoot E. *et al.*, 2019, Grosso AR. *et al.,* 2008). Significantly, complex roles of SF's demand a systemic approach to understand their functions in any cancer.

In the last decade, much focus was on oncogenic splicing isoforms as cancer drivers, despite the prominent role of splicing factors. Only the recent reports begin to unravel the regulatory mechanism underlying the generation of cancer-specific isoforms including splicing factor dysregulations. Interestingly to date, compared to recurrent somatic mutations in SF's, which are known to affect only a few transcripts, the de-regulated expression of splicing factors have altered the splicing of many cancer-related genes. Overexpression of these RBP's are well documented and thus, elicit splicing deregulations in a concentration-dependent manner. For instance: A prototypical splicing factor protein, SRSF1, involves both constitutive and alternative splicing (Das S. *et al.,* 2014). Overexpression of SRSF1 is frequently observed in many solid tumors including lung (25%), colon (25%), breast (13%), as well as thyroid, small intestine, kidney, and ovarian tumors. Moreover, this upregulated expression is enough for cancer transformation and progression (Bejar R. 2016, Anczuków O. *et al.,* 2016). In addition, the co-regulated expression of splicing factors and other RBP's, which act together in driving the modulation of cancer-promoting splicing events, crucial in developing cancer. Mostly, SF's are assembled from different macromolecular complexes (e.g. nuclear bodies) that are dynamic in composition, time and space (Koedoot E. *et al.,* 2019, Fackenthal JD. and Godley LA. 2008). For example, SR proteins and hnRNPs are known to antagonize each other function in a concentration-dependent manner (Fackenthal JD. and Godley LA. 2008).

2.3. Splicing Factors as Cancer Biomarkers

In recent years, several studies emphasized that differential alternative splicing changes and splicing regulators which specifically associates with cancer are a new class of prognostic biomarkers (Anczuków O. *et al.,* 2016, Zong Z. *et al.,* 2018, Sveen A. *et al.,* 2016, Lopez-Bigas N. *et al.,* 2005). For instance, oncogenic splicing factor SRSF1 is a master splicing regulator and is upregulated in several tumor types including breast cancer (Karni et al., 2007). Splicing alterations mediated by SRSF1 results in epithelial mammary cell transformation, cell proliferation, metastasis, and drug resistance in different cancers. In prostate cancer (PCa), this

SF induces the production of oncogenic cyclin D1b variant of cyclin that is enough to promote cell transformation and associated with progression and poor prognosis (Sette C. 2013, Olshavsky NA. *et al.,* 2010). It also drives the production of anti-apoptotic variants of pro-apoptotic genes as BIM & BIN1 (Olga Anczuków *et al.,* 2012) in breast cancer and caspase 9 in non-small cell lung cancer (Shultz JC. *et al.,* 2011). Other targets of SRSF1 are E-cadherin in head and neck cancer cells (Sharma S. *et al., 2011),* Ron, CD44, Rac1b in different cancer types all promote metastatic events as migration and invasion (Brown RL. *et al.,* 2011, Radisky DC. *et al., 2005,* Todaro M. *et al.,* 2014). Similarly, several other SF's are also identified as crucial in the development of cancer and serves as potential targets, such as SRSF7 in colorectal cancer (Wan L. *et al.,* 2017), USP39, PTBP1 and HNRNPs, SNRPB and CELF2 are prognostic splicing factors in glioma progression (Li Y. *et al.,* 2019, Ding K. *et al.,* 2019). In addition to aberrant expression mutation in SF's are also correlated with mainly in hematological malignancies. E.g.: Somatic mutation in the genes SRSF2 (serine/arginine-rich splicing factor 2), SF3B1 (splicing factor 3b subunit 1) and ZRSR2 (zinc finger RNA binding motif and serine/arginine-rich 2) occur commonly in chronic lymphocytic leukemia (CLL) (Anczuków O. *et al.,* 2016, Yoshida *et al*., 2011). It is important to note that all these studies have used genome-wide analysis to gain comprehensive insights on different splicing factors as prognostic predictors in cancer. The advent of high throughput genomic technology, such as RNA sequencing, has made genomic profiling possible, which has greatly improved our understanding of mis-regulations in AS and SF's. In addition, it is important to perceive the effect of dysregulations in splicing regulators as it dictate not one but several of its target's splicing and thereby affect different biological processes which result in disease pathophysiology. (Wang Q. *et al.,* 2019).

2.4. Genomic Approaches to Identify Biomarkers

The advent of high-throughput molecular technologies such as microarrays and next-generation sequencing (NGS) is important to understand the nature of cancer and introduced a new method for the detection of biomarkers. With the advantages of these technologies, several candidate biomarkers are being discovered for tumor screening, diagnosis, prognosis and therapy assessment (Sienel W. *et al.,* 2006, Metzker ML. 2010)

Before the emergence of NGS, for more than a decade DNA Microarray technology had a huge impact on cancer research. In this chip-based method, detection of relative abundance of

nucleic acid is through hybridization of target to probe (short DNA or cDNA) and quantified by detection of chemiluminescence or fluorophore signals. Several studies with DNA microarrays facilitated the acquisition of genome-wide gene expression data. Thus, crucial advantages of microarray are as follows: a) Identification of DEG's between cancer and normal cells by simultaneously comparing thousands of genes and identification of biomarkers (Adomas *et al.,* 2008). b) Tumor classification is crucial for cancer therapy (Golub TR. *et al.,* 1999). c) To understand genetic changes in cancer e.g.: identification of SNPs through SNP arrays (So AY. *et al.,* 2014). d) To study Epigenetic changes e.g.: global patterns of methylation in cancer were studied using DNA methylation arrays (Bibikova M. *et al.,* 2011, Shi H. *et al.,* 2003). Low throughput, high noise, heterogeneity of sample processing has restrained the usage of a microarray.

NGS technologies involve deep sequencing of all types of RNA, thus providing isoform level information of the transcriptome (Trapnell *et al.,* 2012). This technology has revolutionized our understanding of the cancer genome and offered great advantages in cancer prevention, diagnostic, prognostics and treatment (Metzker ML. 2010, Mardis ER. 2011). RNA-seq is highly accurate for quantifying expression levels and the amount of human transcriptome data has grown tremendously over the past decade. Unlike microarray, NGS data has a very low background signal (Zhao *et al.,* 2014). Like microarrays, NGS can also be used for RNA profiling, genome-wide genetic changes that occur upon cell transformation, identifying genomic elements that are bound by transcription factors and deciphering the epigenetic makeups of cancer cells. NGS has paved the path for discovering microRNAs, non-coding RNAs including, long non-coding RNAs, and circular RNAs. It is now appreciated some of these non-coding RNAs play crucial roles in tumorigenesis and tumor suppression (So AY. *et al.,* 2014, So AY. *et al.,* 2013). NGS technology has been critical in discovering new somatic mutations and signal pathways and abnormal mRNA splicing that are associated with cancer pathology.

Following are few studies based on these genomic technologies:

In prostate cancer, GABPB1-AS1, DDC and HEATR5B were identified as potential biomarkers and eight genes NREP, PTGFR, DOCK9, SCARNA22, IK2F3, CLASP1 and FLVCR2USP13 were identified as prognostic indicators to predict progression of prostate cancer, from early-

stage II to subsequent metastatic stages. These genes are significant in directing the treatment strategies (Alkhateeb A. *et al.,* 2019). In another prostate cancer study, Pflueger *et al.,* identified 7 novel gene fusions of prostate cancer from the analysis of 25 human prostate cancer RNA-Seq data. Gene fusion TMPRSS2-ERG was identified as the early molecular event is associated with disease invasion and is present in 50% ~ 90% of prostate cancers (Tomlins. *et al.,* 2008).

In gastric cancer, nine prognostic gene signatures were identified: TOP2A, TPX2, COL3A1, NDC80, COL1A1, CEP55, COL1A2, CDKN3 and TIMP1 as crucial genes linked with the prognosis and pathogenesis of gastric cancer. In this study, DEG's - Differentially expressed genes were identified using both RNA Seq and microarray data of human normal and GC samples. Also, using Survival analysis, network protein-protein interaction (PPI) and functional enrichment analysis key genes which affect GC patients pathogenesis and prognosis were identified (Liu. X. *et al.*, 2018).

In pancreatic cancer, authors have analyzed RNA-seq data from TCGA to investigate the prognostic value of Alternative splicing (AS). The study includes analysis of survival, gene ontology and correlation network. Results indicate alternate splicing events of splicing factors RBM4, DAZAP1, ESRP1, SF1 and QKI, transcription factors SP1, GANPA and KLF7 and 13 cancer driver genes including TP53 and CDC27 were largely associated with overall survival and serve as prognostic predictors (Yu M. *et al.*, 2019).

In lung cancer, Yanaihara et al, performed Cox proportional hazard regression analysis to assess the genome-wide expression profile of miRNAs within 104 pairs of primary lung cancer patients and corresponding noncancerous tissues. This led to the identification of five differentially expressed miRNAs; hsa-mir-145, hsa-mir-17-3p, hsa-mir-2, hsa-let-7a-2 and hsa-mir-155 that are linked to adenocarcinoma patient survival. Significantly, hsa-mir-155 was reported as prognostic marker in lung adenocarcinoma (Yanaihara N. *et al.,* 2006). In addition, in chronic lymphocytic leukemia, a unique 13 miRNA expression pattern which includes hsa-mir-155 was also identified as a prognostic factor. (Calin *et al.,* 2005).

2.5. Research Focus: Sam68

Src associated in mitosis of 68 KDa (Sam68) was initially referred to as p62, which is a 62 KDa phosphorylated protein associated with p120-RASGTPase-activating protein. Sam68 was

initially found associated with v-Src tyrosine kinase during mitosis. Also, Sam68 is a substrate, which is tyrosine phosphorylated by Src tyrosine kinases" (Fumagalli et al., 1994, Taylor and Shalloway 1994). It was identified as the first mitotic substrate of the v-Src tyrosine kinase in fibroblasts and is present in cells transformed by oncogenic tyrosine kinases including v-Src (Fumagalli *et al.,* 1994, Taylor and Shalloway 1994). Sam68 is a proto-typical member of signal transduction and activation of RNA (STAR) family of RNA binding proteins. Other than Sam68, this family also includes mammalian proteins as QkI (quaking), SF1 (splicing factor 1), orthologs of Sam68: slm2/T-star (also known as KHDRBS3), Drosophila HOW, slm1/KHDRBS2 and C. elegans GLD-1, KEP1, Sam50, gld1 and Artemia Salina GRP33. STAR proteins owe their name to the existence of a conserved domain of 200 amino acids that is referred to as the GSG (Sam68, GRP33, and GLD-1) / STAR domain that harbors the binding activity of RNA. In addition, they also contain motifs recognizable by several signaling proteins. Crucially, these ribonucleoproteins, including Sam68, functions in developmental processes and links extracellular signals to changes in transcriptional and post-transcriptional regulations and processing of RNA. Thus, the STAR proteins including Sam68 crucially link signal transduction and splicing regulation. This family of proteins is involved in a multitude of the developmental process as cell proliferation and differentiation (Sette et. al. 2010). STAR proteins have a role in many processes such as splicing (Arning S *et al.,* 1996, Berglund JA *et al.,* 1998), tumorigenesis (Jones AR *et al.,* 1995, Liu K *et al.,* 2000), apoptosis (Di Fruscio M *et al.,* 1998, Chen T *et al.,* 1998, Pilotte J *et al.,* 2001), cell cycle progression (Barlat I *et al.,* 1997), translation (Jan E *et al.,* 1999, Clifford R *et al.,* 2000), and development (Jones AR. *et al.,* 1995, Zorn AM. *et al.,* 1997 Baehrecke EH. 1997, Zaffran S. *et al.,* 1997). The implication of Sam68 in biological process and disease development is better understood by describing its putative functional domains.

2.5.1. Domain Structure and Post-transcriptional Modifications

Sam68 protein contains a GSG domain for RNA binding which is flanked by regulatory elements. Moreover, the structure also includes proline-rich motifs, arginine-glycine rich regions and tyrosine-rich motifs in c-terminal tails that are important for protein-protein interactions in signal transduction pathways. Figure 2.1 represents the Domain structure of Sam68

RNA binding domain (GSG domain):

The structure of Sam68 contains a single KH domain of 70-100 amino acids, the second most prevalent protein motif, and thus Sam68 also belongs to the K homology protein family. This domain is in homology with domain found in heteronuclear ribonucleoprotein particle (horn) K protein, hence the name KH. The presence of a conserved GXXG loop is the classical feature of the KH domain, responsible for direct Protein-RNA binding (Lukong KE. 2003). The Flanking regulatory elements, ~25 amino acids C-terminal sequences (NK, CK of KH domain) and ~75 amino acids N-terminal are also required for RNA binding. Collectively, this tripartite region responsible for sequence-specific RNA binding is referred to as the GSG/ STAR domain. GSG domain attributes several properties to the protein, including homodimerization/self-association, RNA binding, heterodimerization and protein localization (Chen T *et al.,* 1997, Di Fruscio M *et al.,* 1998, Chen T *et al.,* 1998, Zorn AM. *et al.,* 1997, Chen T *et al.,* 1999, Wu J *et al.,* 1999). Sam68 binds to nonspecifically to ribonucleoprotein homopolymers at poly (U) and poly (A) (Taylor and Shalloway 1994, Chen T *et al.,* 1997). The consensus sequence for Sam68 is a four-nucleotide A/U-rich motif. 3'-UTR (3'-untranslated region) have an AU-rich sequence, which is a probable candidate as a Sam68 target. Recombinant Sam68 was found to bind RNA with high affinity to UAAA or UUUA motifs using SELEX (Lin *et al.,* 1997, Garneau, A. *et al.,* 2009). Indeed, Sam68 seems to favor a UAAA motif preferentially surrounded by A, for instance, sequences as AAAUAA and AAUAAA are optimal (Jaelle and foot). Further many potential binding sites in Pre-mRNA or splicing targets were identified by several studies suggesting Sam68 involvement in Post-transcriptional regulation of these genes (Chawla G. *et al.,* 2009, Itoh *et al.,* 2002, Sánchez-Jiménez F. *et al.,* 2013). Itoh *et al.* identified 29 pre-mRNA binding targets that prominently include hnRNP A2/B1 and Beta-actin (Itoh *et al.,* 2002)

Protein Binding Domains:

SH3 Domain

Sam68 includes arginine-glycine regions, proline-rich motifs and tyrosine motifs in c-terminal tails that regulate the protein-protein interactions. Six short proline-rich sequences (P0-P5) are

potential binding sites of WW and SH3 domain-containing proteins. Amongst the proline-rich motifs, P0, P1, P3, P4 and P5 are identified to interact with tyrosine kinases of the Src family. These short proline-rich regions consist of a core consensus PXXP sequence that lies outside of the GSG domain. Several groups identified that Sam68 in cytoplasm binds to proteins containing the SH3 domain such as tyrosine kinases. The Shalloway group noticed P62 migrated at 68Kda and is tyrosine phosphorylated by Src during mitosis, hence renamed as Sam68. Further, Sam68 was identified to interact with tyrosine-phosphorylated by several tyrosine kinases containing SH3 domain including Src kinases, Sik/BRK (Derry JJ. *et al.*, 2000), PRMT2 (Espejo A. *et al.*, 2002), Grb-2 (Trub T. *et al.*, 1997), p85 PI3K (Taylor SJ. *et al.*, 1995), Grap (Trub T *et al.*, 1997), Itk/Tec/BTK (Andreott AH. *et al.*, 1997, Bunnell SC. *et al.*, 1996), Nck (Lawe DC. *et al.*, 1997) and PLCγ-1 (Richard S. *et al.*, 1995, Maa MC. *et al.*, 1994 and Weng A. *et al.*, 1994). Interaction with Src kinases is important for Sam68 phosphorylation which in turn either decreases the RNA binding or increases the specificity of binding to RNA targets.

WW Domain

The WW domain is a single or tandem repeat series of around ~40 conserved amino acids with two characteristic tryptophan residues positioned 22 or 23 residues apart (Bedford MT. *et al.*, 1998, Macias MJ. *et al.*, 2002). Like the SH3 domain, proline motifs specifically P3 and P4 have been identified interacting with proteins having the WW domain. Thus, both these domains may contend for the same ligands in vivo (Bedford MT. *et al.*, 1997, Espejo A. *et al.*, 2002). WW domains of the cytoskeleton FBP21 and FBP30 (formin binding proteins) are known to interact with Sam68 (Bedford MT. *et al.*, 2000). FBP21, spliceosome proteins are involved in transcription and splicing in nuclear speckles (Klippel S. *et al.*, 2011). Nuclear functions of Sam68 as RNA binding and association with core spliceosome proteins & splicing factors probably occur through binding WW domains of nuclear proteins.

SH2 Domain

Tyrosine residues present in the C-terminal region of Sam68 are potential sites of phosphorylation. As mentioned earlier, numerous soluble tyrosine kinases are found to phosphorylate Sam68 including p60src, p59fyn, p56lck, ZAP-70 (Lang V. *et al.*, 1997) and Sik/BRK. Tyrosine-phosphorylated Sam68 associates with numerous SH2 domain-containing

proteins including Src family kinases, Grb2 (Najib S. *et al.,* 2002), Sik/BRK, Nck (Lawe D. C. *et al.,* 1997), Grap (Trub T. *et al.,* 1996) RasGAP (Guitard E. *et al.,* 1998), PLCg-1, PI3K p85a and Itk/Tec family kinases (Fumagalli S. *et al.,* 1994, Weng A. *et al.,* 1994, Richard S. *et al.,* 1995, Vogel LB. *et al.,* 1995, Andreott AH. *et al.,* 1997, Bunnell SC. *et al.,* 1996). These observations support the Sam68 role as an Adaptor protein.

RG-rich binding Domain

Sam68 has sequences rich in arginine–glycine that often surrounds proline-rich sequences. Proteins containing "RG-rich regions and RGG boxes" are involved in RNA metabolism which includes the STAR family of proteins (Burd CG. *et al.,* 1994). Arginine methylation is an important post-transcriptional modification that possibly modulates several biological processes including gene transcription, intracellular localization and protein-protein interactions (Gary GD. *et al.,* 1998, McBride A. *et al.,* 2002, Stallcup MR. *et al.,* 2001).

Nuclear Localization signal

The last 24 amino acids (RPSLKAPPARPVKGAYREHPYGRY) from 420 to 443 in the C-terminal end of the Sam68 form a nonconventional nuclear localization signal (NLS) that dictates the nuclear localization of the protein (Ishidate T. *et al.,* 1998). Prominently, two nuclear targeting motifs: PPXXR (Ishidate T. *et al.,* 1998) and RXHPYQ/GR are present (Wu J. *et al.,* 1999). It has been shown that mutating the arginines at both ends of the RXHPYQ/GR motif to alanines abolishes nuclear targeting, thus they are important for nuclear localization of Sam68 polypeptide (Wu J. *et al.,* 1999).

Thus, the multi-modular structure of Sam68 is crucial in crosslinking signaling transduction pathways and splicing regulations. The post-transcriptional modifications such as phosphorylation, methylation, acetylation and sumoylation within the regulatory regions of Sam68 finely control its interplay with several signaling components, subcellular localizations, target RNA binding affinity and functions (Bielli P. *et al.,* 2013). For instance, during mitosis, the association of Sam68 with Ras-GAP is enhanced due to Tyrosine phosphorylation by Src kinase, however, this prevents its binding with RNA. Furthermore, acetylation on lysine residues by histone acetyltransferases enhances RNA binding (Babic I. *et al.,* 2004, Meyer NH. *et al.,*

2010). In addition, arginine methylation of Sam68 that negatively controls SH3 but not WW domain interactions, this can be a process by which the cytoplasmic activities of the protein are transferred to nuclear functions by preventing the association with the cytoplasmic protein-containing SH3 domain (Cote J. et al., 2003, Bedford MT. et al., 2000).

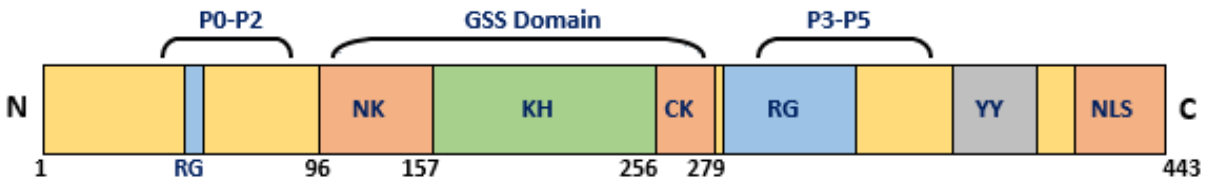


Figure 2.1*: Domain structure of Sam68.*
GSG: GRP33/Sam68/GLD-1 is required for RNA binding is composed of NK =N-terminal of KH domain, KH =hnRNP K homology domain and CK =C-terminal of KH domain. P0-P5 are six consensus proline-rich regions responsible for binding SH3 and WW domain-containing proteins as signaling proteins, tyrosine kinases. , RG =arginine/glycine-rich region potential sites for arginine methylation, YY =C terminal tyrosine rich region potential for tyrosine phosphorylation, NLS =nuclear localization sequence vital for nuclear localization. The size of each motif is indicated as the number of amino acids.

2.5.2. Biological Functions of Sam68 in Cancer

Sam68 is a transcriptional and Post-transcriptional regulator of gene expression. It is a versatile protein with multiple functions in cancer that are context/cell type-dependent. Activities that determine cell fate. Notably, all the functional activities of Sam68 are linked to carcinogenesis, so it is important to note that most of the functions described here are also related to cancer.

Sam68 in the signaling pathway

Sam68 is a key protein in signal transduction pathways where it acts as a scaffold protein in response to activation of different membrane-bound receptors including T-cell receptor, insulin receptor, leptin receptor, TNF-alpha, EGF or HGF/Met signaling pathway activation. Sam68, a specific target of the Src tyrosine kinase in mitosis. Tyr phosphorylation by Src-family kinases (SFKs) caused the accumulation of Sam68 in nuclear granules, named Sam68 nuclear bodies

(SNBs). A recent study revealed that Sam68 modulates nuclear transcription factor-kappa B (NF-dB) activation, thus inducing inflammation.

Sam68 supports interaction between PLCγ1 and Src-related kinase Fyn (Paronetto *et al.*, 2003), which results in phosphorylation and phospholipase activation (Sette *et al.,* 2002, Paronetto *et al.,* 2003). Sam68/PLCγ1/Fyn assembly was triggered by the expression of the truncated form of the c-KIT tyrosine kinase receptor. Noticeably, in a subgroup of patients with prostate cancer (PCa), this receptor is abnormally expressed, and its expression is linked with increased Src activation and tyrosine phosphorylation of Sam68 (Paronetto *et al.,* 2004). Sam68 is a substrate of FYN, a soluble nRTKs, Moreover, FYN-dependent Tyr-phosphorylation negatively affected the interaction hnRNP A1 and also reduced the affinity of Sam68 binding to target RNAs, BCL-X and CCND1 genes, thereby altering the outcome of AS events.

In breast cancer cells, tyrosine phosphorylation of Sam68 may also play a role. Breast tumor kinase BRK is an excessively-expressed, non-receptor tyrosine kinase (nRTK) in human breast cancer cells (Barker *et al.,* 1997) which promotes proliferation and anchorage-independent growth (Ostrander *et al.,* 2010). One of the first substrates of BRK has been identified as Sam68, which overlaps the nuclear localization signal of Sam68 by phosphorylating tyrosine residues. BRK-related phosphorylation induces transient sub-cellular re-localization of Sam68 following mitogenic stimulation of breast cancer cells with an epidermal growth factor (EGF) (Lukong *et al.,* 2005). Further phosphorylation reprograms the functional activity of Sam68 as it decreases the RNA-binding activity and increases signaling protein interactions (Lukong & Richard 2003). Notably, the expression of both Sam68 and BRK is upregulated in breast cancer and supports cell proliferation and invasiveness (Barker *et al.,* 1997).

In contrast to Tyr phosphorylation, the Ser/Thr phosphorylation of Sam68 has reportedly increased binding and splicing to its RNA targets. Increased expression of Sam68 promotes the inclusion of the variable exon v5 in the CD44 mature mRNA upon T-cell receptor activation followed by the RAS-RAF-MEK-ERK signaling cascade. ERKs-mediated phosphorylation of Sam68, a target of this pathway, has increased the ability to promote exon v5 inclusion. Further, the interaction of Sam68 with splicing factor U2AF65 enhances the recognition of the 3-splice-site. Also, in prostate cancer cells upregulated Sam68 has shown to promote cyclin D1b, a variant of the CCND1 gene. This activity was enhanced by again through the activation of the RAS/ERK

pathway but counteracted by SFKs. It was clear that ERK-dependent (Ser/Thr) phosphorylation increased binding to CCND1 intron 4, whereas SFK-dependent (TYR) phosphorylation abolished the same. Notably, posttranslational modifications have an opposite impact on the splicing activity of Sam68.

Transcription Regulation by Sam68

/The transcriptional role of Sam68 directly affects cancer cell biology. Sam68 was identified as a coactivator of the p53 tumor suppressor in DNA damage-induced cell cycle arrest and apoptosis. Sam68 physically interacts with p53 in an RNA dependent manner, then the complex was recruited at promoter regions to transcriptionally coactivate the p53 target genes involved in negative feedback (*MDM2*), cell cycle arrest (*P21*) and apoptosis (*BAX, PUMA*) (Naomi Li. *et al.*,2016). Sam68 modulates transcription by associating with coactivator CBP (Hong W. *et al.*, 2002), the androgen receptor (Rajan P. *et al*., 2008) and NF-κB (Fu K. *et al*., 2013). In another instance, this ribonucleoprotein behaves as a competitive inhibitor of positive regulators of transcription. Binding of Sam68 to transcriptional activator horn K inhibits the activation of a reporter driven by the CT promoter element of the proto-oncogene c-myc. In breast cancer cells, Sam68 competes with transcriptional co-regulators to bind the CBP cofactor. This co-localization represses CBP-dependent expression of cell cycle regulators (cyclin D1 and cyclin E Transcripts), independent of its RNA-binding activity (Hong *et al.*, 2002, Taylor *et al.*, 2004). In mammary development and tumorigenesis, Sam68 acts as a co-activator of ER-dependent transcription. Further, Sam68 has been showed to directly bind the androgen receptor and androgen-responsive elements (AREs) within the promoter region of the prostate-specific antigen (PSA) gene. Independent of its splicing regulatory properties, Sam68 affects the AR-regulated transcriptional activity in prostate cancer cells. Moreover, Sam68 is essential for proper expression of the gonadotropin receptor transcripts in pre-ovulatory follicles from the adult ovary, where Sam68 possibly upregulates both the FSH and LH receptor transcripts.

Sam68 in Alternative Splicing

Another pivotal role of Sam68 is the regulation of alternative splicing of multiple genes. Sam68, a decisive splicing regulator exploits splicing decisions in response to extracellular

stimuli to pathological development and progression of tumors (Figure 2.2) (Cristina V. *et al.*, 2010). Indeed, the molecular mechanisms through which Sam68 regulates AS decisions in response to signaling cascades are poorly understood. The expression of Sam68 positively correlates with the levels of these oncogenic splice variants in human cancers. Upon activation of ERK1/2 pathway, phosphorylated Sam68 transduces the signal response to stimulate the production of an oncogenic and constitutively active variant of CD44 with exon V5 inclusion (CD44 V5) in prostate cancer (Matter *et al.*, 2002, Cappellari M. *et al.*, 2014), Proliferation-associated survivin (BIRC5) has 6 isoforms with antagonists' function. Sam68 regulates the exclusion of exon3 of this gene and leads to the production of survivin DEx3, anti-apoptotic isoform which is highly expressed in advanced breast & cervical cancers. Androgen receptor with exon 3b inclusion in Castration-resistant (CR) prostate cancer (PC) (Stockley J. et al., 2015) both promote proliferation. Other than exon inclusion, cryptic retention of introns as intron 4 in cyclin D1 (cyclin D1b) variant is also seen in prostate cancer but this Sam68 mediated Cyclin D1b splicing is abolished by coexpression Fyn kinases (Paronetto *et al.*, 2010). In addition, intron retention in 3' UTR of SRSF1, has increased the mRNA stability of this proto-oncogenic splicing factor by preventing AS-NMD degradation. This further contributes to the production of a constitutively active splice variant of RON (Cristina V. *et al.*, 2010) (Figure. 2). On the other hand, tyrosine phosphorylation of Sam68 by Src kinase Fyn, promotes the expression of BCL-Xs, a pro-apoptotic isoform of BCL-X (Paronetto MP. *et al.*, 2007). Splicing factor HnRNPA2B1 in pancreatic cancer (Chen ZY. *et al.*, 2011) and other transcriptional factors FBI-1 (Bielli P. *et al.*, 2014) are all shown to regulate Sam68 tumor suppressor BCL-X splicing. Further, over-expression of Sam68 in murine fibroblast induces cell cycle arrest and apoptosis. Liu *et al.*, showed expression of Sam68 is 25% less in RHKO NIH3T3 in comparison to wild type, and therefore these cells exhibited anchorage-independent growth and metastatic tumor formation in nude mice (Liu *et al.*, 2000). But to date, no supporting information about tumor suppressor activity of Sam68 in Vivo is published. However, Sam68-Knockout mice do not exhibit tumor formation in vivo (Lukong KE. *et al.*, 2007) and further Sam68 haploinsufficiency delays mammary tumorigenesis and metastasis in MMTV-PYMT transformed cell lines (Richard S. *et al.*, 2008). This implicates the role of Sam68 majorly as pro-oncogenic and also tumor-suppressor. This can be explained as Sam68 also modulates the transcription, mRNA translation and also as part of protein complexes regulates cell transformation in either splicing dependent

or independent manner. Thus, multiple functions and regulations of Sam68 make it difficult to distinguish the prominent role of Sam68 in cancer.

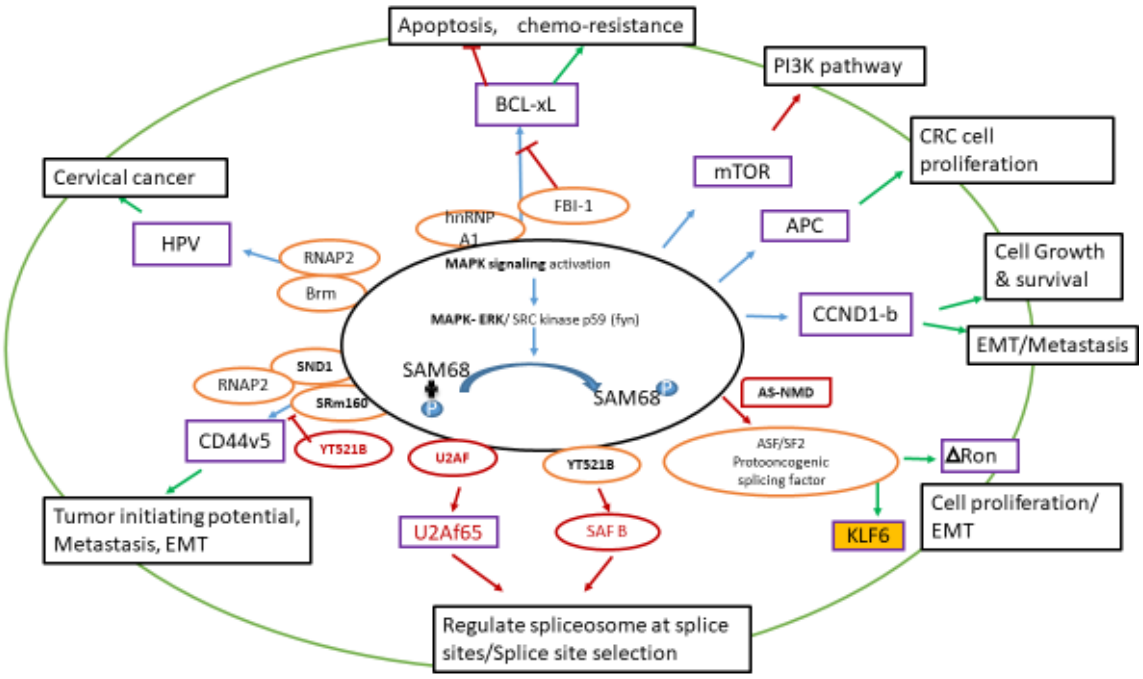


***Figure 2.2**: Molecular functions of Sam68. This picture Sam68 regulation of splicing of target genes which contributes to cancer initiation and progression.*

Table 2.1: List of Sam68 splicing targets and their biological functions in the development of other diseases.

Splicing targets of Sam68	**Splicing event**	**Biological Function**	**Reference**
mTOR	Diminish the retention of intron 5 containing a premature termination codon.	Defects in adipogenesis	Huot MÉ. *et al.*, 2012
Tenascin C	Promotes larger isoform.	Promotes the proliferation of Neural stem cells.	Moritz S. *et al.*, 2008
SMN2	Exon-7 skipping	Development of spinal muscular atrophy	Pedrotti S. *et al.*, 2010,

Neurexin (Nrxn)	Skipping of cassette exon20 at AS4.	Associated with synaptogenesis and neurodevelopment disorder	Iijima T. *et al.*, 2011
SGCE (sarcoglycan epsilon)	Exclusion of exon 8	Atypical role in complex of the muscle cell and movement disorder myoclonus dystonia	Chawla G. *et al.*, 2009
β-tropomyosin	Regulates splicing of exon 7 by selection long-range branch point.	Contributes to mutually spliced (exon 6/exon7) in Non-muscle cells.	Grossman JS. *et al.*, 1998
HIV-1	Increases unspliced RNA and also stimulates 3' end processing for REV dependent transport.	Sustains HIV-1 protein expression in the host cell, important for HIV-1 replication.	McLaren M. *et al.*, 2004

2.5.3. Sam68 as a Biomarker

A series of recently mounting reports stated that Sam68 expression was up-regulated in a variety of human cancers, including prostate carcinoma (Busa *et al.,* 2007), renal cell carcinoma (Zhang *et al.,* 2009), non-small cell lung carcinoma (NSCLC) (Zhang *et al.,* 2014), colorectal cancinoma (Liao *et al.,* 2013, Kai Fu *et al.,* 2016), cervical cancinoma (Li Z *et al.,* 2012), Ovarian cancinoma (Dong L *et al.,* 2016, Wang Y *et al.,* 2016), bladder cancinoma (Zhiling Zhang *et al.,* 2015), liver cancinoma/hepatocellular carcinoma (Tingting Zhang *et al.*, 2015), breast cancinoma (Song *et al.,* 2010, S Richard *et al.,* 2008), endometrial carcinoma (Qingying Wang *et al.,* 2015), esophageal squamous cell carcinoma (Yayun Wang *et al.*, 2015), neuroblastoma (Xiaohong Zhao *et al.,* 2013) and T-acute lymphoblastic leukemia (Qi Wang *et al.,* 2016) together indicating Sam68 as an oncogene which promotes tumor progression.

2.6. Detection of Cancer Biomarkers

According to National Cancer Institute (NCI), a biomarker is "a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process or a condition or disease. A biomarker may be used to see how well the body responds to a treatment for a disease or condition" (https://www.cancer.gov/).Iin simple terms, Biomarker is an indicator that could be measured to detect and assess a biological condition. Cancer biomarkers are

biological indicators specific to cancer that are widely considered and used. Broadly, cancer biomarkers are classified into diagnostic, prognostic, predictive and therapeutic biomarkers based on their application. Diagnostic biomarkers detect the type of cancer whereas prognostic biomarkers predict the cancer stage and reoccurrence of cancer. Predictive biomarkers predict the therapy/ drug response and therapeutic biomarkers are possible therapeutic targets of the disease (Carlomagno *et al.,* 2017). A biomarker can be a form of nucleic acids (DNA, RNA), protein (enzymes, hormone, antigens, antibody, tumor suppressors & oncogenes), biochemical molecules (e.g.: glucose) or specific cells (tumor cells). But, widely used are protein biomarkers that are detected as the change in the expression of a certain protein. These biomarkers are measured in tissues, body fluids such as blood, serum, urine, sputum and blood and that are present within or on the surface of the tumor cells.

Detection of Cancer biomarkers is most valuable for early cancer detection, diagnosis, cancer staging (grading), prognosis, selection & response to treatment, and disease recurrence (Basil CF. *et al.,* 2006, Clinical practice guidelines. 1996). However, in the early stages of cancer, biomarkers are present in trace levels with other biological molecules, hence diagnostic tests must be extremely reliable and efficient. Currently, clinical detection of a cancer diagnosis is mainly based on imaging techniques such as X-ray, mammography, computational tomography, visualization through magnetic resonance, endoscopy and ultrasound. Although imaging techniques have advantages most of the techniques need a biopsy, which is invasive (Altintas Z. *et al.,* 2015). In addition, these techniques present low sensitivity and their ability to differentiate between benign and malignant lesions are limited (Boice JD Jr. *et al.,* 1991). This led to emergence of the Genomic and Proteomic approaches for tumor identification by detecting biomarkers. Generally used techniques includes polymerase chain reactions (PCR), southern blotting, real-time polymerase chains (RT-PCR), fluorescence in-situ hybridization (FISH) for genetic modification or immunohistochemistry (IHC) or high-concentration (HCS) analysis for protein expression and subcellular localization. Moreover, techniques such as enzyme-related immunosorbent assays (ELISA) are extensively used in hospitals. These techniques are highly sensitive and selective, but they can take time and cost. Furthermore, ELISA studies are not sensitive enough to detect low biomarker concentrations at the early stage of cancer, which leads to false positives. Recently, biomarker exploration has been investigated through fluid

chromatography-mass spectrometry tests, but these procedures are cost-intensive and advanced technology to be used in clinical and Point of Care diagnostics (Khanmohammadi A. *et al.,* 2020).

Due to various technical difficulties in current technologies used for biomarker detection, the potential of biomarkers are not explored completely (Sawyers C.L. 2008, Nimse SB. *et al.,* 2015). Moreover, it is important to note that cancer detected in early-stage is typically cured with the greatest probability of success. Therefore early, sensitive and specific detection of biomarkers is crucial in cancer treatment (Filella X *et al.,* 2015, Kulasingam V. *et al.,* 2008). Further, all the recent attempts have focused to identify the biomarkers without a biopsy. In this respect, Biosensors are very useful for easy, accurate and non-invasive detections of cancer biomarkers. A biosensor is made up of three components: a recognition element to interface with the analyte, a signal transducer that produces a measurable signal from the interaction of the analyte-biomolecular substrate, and a signal processor that relays and displays measurable output signal. The molecular recognition component detects a 'signal' from the environment in the form of an analyte, and the transducer then converts the biological signal to an electrical output (Chaplin M. 2010). Based on Transducers used, biosensors are classified as Electro-chemical (amperometric, potentiometric, conductometric, impedance), optical (colorimetric, fluorescence, luminescence, interferometry), calorimetric (thermistor) and mass sensitive (piezoelectric, acoustic wave). A biosensor is a rapid and easy to use tool that offers critical advantages as highly sensitive, specific, minimum detection limits and real-time measurement (Zhang *et al.,* 2017). These devices are cost-effective and emerged to detect multiple cancer biomarkers at low concentrations in biological fluids. Moreover, the current biosensors technology is expanding with promising tools as nanomaterials, artificial receptors, surface imprinting techniques to provide a point of care diagnosis (POCD), disease care with therapeutics, and personalized medicine (Altintas Z. *et al.,* 2015).

2.6.1. Electrochemical Immunosensor in Lung cancer

Immunosensors (IS) are miniaturized measuring devices, which selectively detect their targets utilizing antibodies (Abs) and provide concentration-dependent signals. Electrochemical immunosensor are biosensors in which the antibody acts as the capture agent that is coupled to an electrochemical transducer (Figure 2.3). This immunosensor employs an immobilized antibody interaction with a specific target analyte then the concentration of a targeted analyte is

quantified as electrical change by applying a potential. Generally, immunosensors involve a sandwich type of immunoassay. Among other types, this electrochemical detection is preferred due to less complexity in signal generation and output, portability, etc. (Cho IH. *et al.,* 2018). Based on the measured signals, electrochemical immunosensors are categorized as amperometry (current), potentiometry (voltage), impedimetry (impedance/resistance) (Ronkainen, NJ. *et al.*, 2010, Grieshaber D. *et al.,* 2008). Direct detection of the analyte without any antibody labeling is well performed by cyclic voltammetry and impedimetry. These methods detect the change in capacitance and/or resistance, induced by the binding of the target protein. Electrochemical impedance spectroscopy (EIS) is the most often used impedance method. Using a redox couple, typically a mixture of ferricyanide and ferrocyanide, the change in the charge transfer resistance (Rct) is obtained. Cyclic Voltammetry belongs to a category of electro-analytical methods, through which information about an analyte is obtained by varying potential and then measuring the resulting current (Feiyun C. *et al.,* 2019).

Electrochemical biosensors based on conducting polymers are widely used and has offered new possibilities, fast, label free and sensitive detection. Conducting polymers are themselves sensing elements and transducers that convert the biological event such as antibody-antigen interaction or enzyme-substrate reaction into electrical signal. Surface functionalization of the electrode surface using conducting polymers increases the performance of the biosensor both in terms of sensitivity and specificity (Cho, I et al., 2020, Aydemir, N et al., 2016). Amongst several conducting polymers, polyaniline (PANI) is intensively investigated and widely used to detect different biological compounds. PANI has excellent chemical and electrical behavior. Unique features of PANI makes it not only a good mediator for electron transfer in redox reactions but also appropriate immobilization matrix for biomolecules (Shoaie, N., et al., 2019). Immobilization matrix is crucial to maintain the biological activity and orientation of the recognition element, as inappropriate immobilization could result in less specificity, loss of activity (Cho, I et al., 2020). Other properties of PANI including longterm stability in different solutions, biocompatibility, pH sensitivity, high conductivity, redox reversibility and easy modifiable, processible and printable on diverse electrode surface has made PANI continuously used in designing of biosensors (Dhand C et al., 2015, Shoaie, N., et al., 2019).

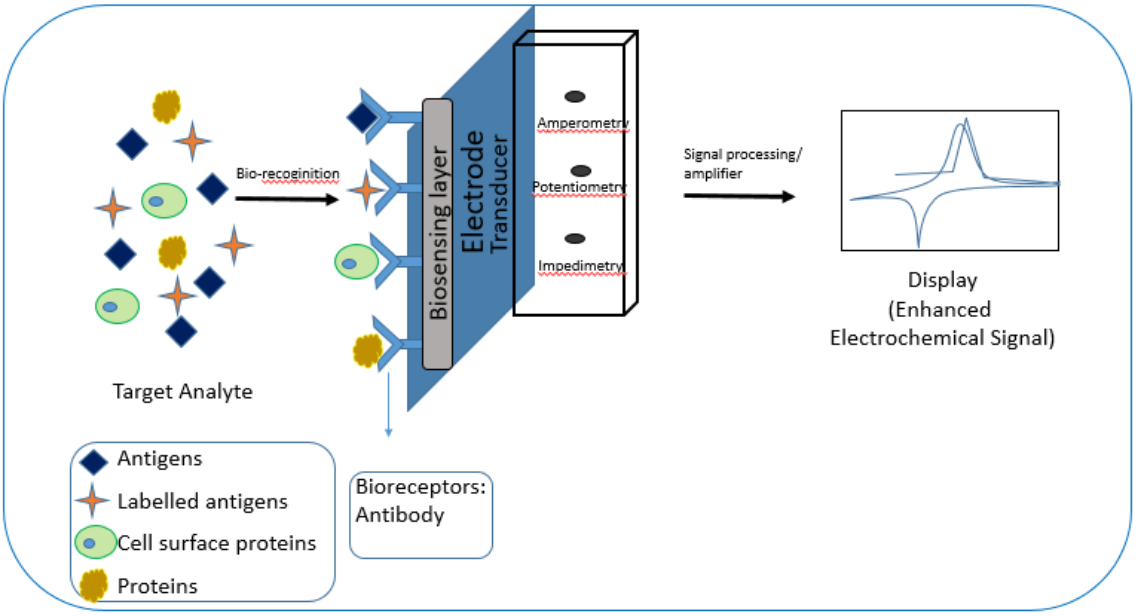


***Figure 2.3**: Schematic representation of electrochemical biosensor. The bioreceptor recognizes the target analyte and produces a signal that is transduced to produce readable output.*

Different electrochemical immunosensors were reported for the detection of lung cancer biomarkers. Zhong *et al.* reported label-free electrochemical immunosensor for measurement of Lung cancer-specific biomarker, Neuron-specific enolase (NSE). The proposed immunosensor employs anti-NSE antibody adsorbed onto chitosan stabled gold nanoparticle and attached to Prussian blue doped silica dioxide (PB-SiO2) through 3-Aminopropyltriethoxy saline (APTES), an amino-functionalized interface. This GCE modified immunosensor exhibited limit of detection 0.08ng/mL (Zhong Z. *et al*., 2010).

Another lung cancer-specific biomarker, Cytokeratin 19 fragment 21-1 (CYFRA21-1) is highly correlated with non-small cell lung cancer (NSCLC). Zheng *et al.* reported a sandwich-type electrochemical immunosensor for the sensitive detection of CYFRA21-1. The proposed GCE immunosensor is modified with three-dimensional graphene (3D-G), chitosan (CS) to immobilize the primary antibody (Ab1). To enhance conductivity and electrochemical signal, the horseradish peroxidase-labeled secondary antibody (HRP-Ab2) was captured onto gold nanoparticles (AUNPs) coated through amino-functionalized carbon nanotubes (MWCNT-NH_2). The

developed immunosensor reported excellent analytical performance with LOD 43pg/mL at a linear range from 0.1 to 150ng/mL (Zeng Y. *et al*., 2018).

Zhen *et al.* reported an Enzyme-free electrochemical immunosensor using an anti-p53 antibody tagged with gold nanoparticles was reported for sensitive detection of p53 protein. The Nanogold particles were doped with Prussian blue and then labeled with the antibody. The modified screen-printed carbon electrode immunosensor (SPCE) showed a detection limit of 0.1U/mL in the range from 0.5 to 80 U/mL (Liu Z. *et al.,* 2014).

In this line, to detect carcinoembryonic antigen (CEA), a tumor marker for familiar cancers including Lung cancer, an antibody-based immunosensor was developed. The electrochemical CEA sensor was fabricated with HRP labeled anti-CEA antibody adsorbed successively onto gold nanoparticle-decorated graphene composites (AU-GN) reported detection limit of 0.04ng/mL at a concentration from 0.10 to 80ng/mL (Zhu L. *et al.,* 2014)

Sudeshna *et al.* reported the fabrication of a sandwich immunosensor for detection IgG molecule. A Glassy carbon electrode (GCE) modified with synthesized redox-active ferrocenyl dendrimer as a functional moiety to immobilize the antibody, anti-IgG was efficient in identifying IgG. Immunosensor was sensitive in detecting IgG concentration as low as 2ng/mL **(**Chandra S. *et al., 2016***).**

- **Aim of the Study**

Sam68, oncogenic splicing factor has been studied lately and gaining importance for its multiple functions in support of tumor development. Significantly it has shown implications in all stages of cancer as transformation, proliferation and invasion. Sam68 and its other associated factors are involved in perturbations of splicing and signaling regulations thereby causing transcriptome alterations. These alterations are partly caused by switching the splicing alterations to produce cancer-specific splicing isoforms which is either driven by pathway regulated splicing or splicing alone. The up-regulated expression of this splicing factor is positively associated with the expression of oncogenic splice variants and also the progression of tumors. Depending on the expression levels, Sam68 and other associated splicing factors form an interaction network to dictate the resulting isoform. Thus, cancer-specific interacting partners play a significant role in Sam68 divergent biological roles and prognostic value.

Recently, studies illustrate the increased expression of Sam68, an oncogenic protein with good prognostic value in various cancers. However, mechanisms underlying frequent Sam68 upregulation in these cancers are largely unknown and it is important to understand the clinical significance of such increased expression. Furthermore, it remains a challenge to understand the genome-wide associations that regulate Sam68 prognostic value in cancer. Moreover, despite its prognostic value, there is no single study that reports a detection system for splicing factors including Sam68 as a biomarker.

Based on present lacunas, we have focused on the role of Sam68, as a potential biomarker of cancer. In the present study, we initially aimed to investigate the prognostic value of Sam68 in four different cancer type's kidney renal papillary cell carcinoma (KIRP), lung adenocarcinoma (LUAD), acute myeloid leukemia (LAML), and ovarian cancer (OV), Using Genome-wide analysis to understand the molecular mechanism of Sam68 in cancer-development. Thereby, to develop a sensitive and selective detection system to quantify Sam68 in cancer.

- **Objectives**

The objective of the work is to understand the molecular mechanism of Sam68 (Sam68 encoding gene) to be a prognostic marker in four different cancers. Then, the development of an antibody-based immunosensor for Sam68 and its application in real-time sample analysis. This thesis focuses on the following objectives.

1. Genome-wide co-expression analysis, construction of correlation network and functional evaluation of Sam68 in different cancers with gene-level data.
2. Genome-wide co-expression analysis of Sam68 target transcripts and analysis of functional correlation with transcript level data in different cancers
3. In-Vitro expression, purification and characterization of recombinant Sam68 in E.coli. In-Vitro expression and purification of Sam68.
4. Fabrication of Biosensor to detect Sam68 for clinical diagnosis of lung cancer

CHAPTER 3

MATERIALS & METHODS

Chapter 3: Materials and methods

Part 1 - Data, tools and methodology used for Genomic analysis (RNA-Seq and Microarray analysis).

3.1.1 Retrieval of TCGA Datasets

The Cancer Genome Atlas (TCGA) RNA sequencing data of Lung Adenocarcinoma (LUAD), kidney Renal papillary cell carcinoma (KIRP), Acute Myeloid Leukemia (LAML), Ovarian Serous Cystadenocarcinoma (OV) with clinical annotations, were obtained from Broad GDAC Firehose Stddata (https:// gdac.broadinstitute.org/). To avoid reanalysis of raw datasets, level 3 gene and transcript expression data from 'illuminahiseq_rnaseqv2-RSEM_normalized' were retrieved. Normalized "scaled_estimates", RSEM counts of genes & isoforms were used for analysis. The mapping of the raw data to the reference assembly was done by MapSplice v12_07 (Wang K. *et al.,* 2010). The reference genome assembly used was hg19 (GAF2.1) and the reference transcriptome annotation is of UCSC hg19 GAF2.1 for known genes standard tracks. Read quantification was performed by Expectation-Maximization, RSEM package (RSEM v1.1.13) (Li B. and Dewey CN. 2011) and then normalized using upper quantile normalization. Sample sequencing methods and detailed description of processing can be found from the previous publication (Li B. and Dewey CN. 2011, Weinstein JN. *et al.,* 2013) and downloaded from Broad firehose (http://gdac.broadinstitute.org/) or GDC (Weinstein JN. *et al.,* 2013). Each patient sample has sequencing reads for 73,599 transcripts of over 20,531 genes (GAF 2.1). Therefore, both gene & transcript expression data in a large cohort of samples over chosen cancers (LUAD, KIRP, LAML, OV) was collected.

Table 3.1: Details the tumor, normal and Paired samples in all four cancers.

Cancer Type	**Samples**	**Cancerous**	**Normal**	**Normal-tumor pairs**
LUAD	576	518	59	59
KIRP	322	290	32	32
LAML	179	179	0	0
OV	304	304	0	0

3.1.2 Classification of Data

Patient samples within cancer datasets were categorized into Sam68 high and low group based on the Standard deviation in expression of Sam68. Z-scores in samples of Sam68 high & low have at least one standard deviation above & below its mean expression respectively. Moreover, in Sam68 high only cancer samples with Z = +1 and above (higher expression of Sam68) whereas Sam68 low cancer samples with Z = -1 and below (lower expression of Sam68) were analyzed. For example, a sample is said to have a high expression of a gene if its expression is at least one standard deviation above its mean expression in the subtype.

3.1.3 Measurement of Co-expression

The co-expression enrichment between any two genes or isoforms was investigated by computing the Spearman's rank correlation coefficient. It is a measure of nonparametric association. It assesses the nonlinear monotonic relationship between the two variables by the linear relationship between the ranks of the values of the two variables. The correlation was calculated using the following formula

$$r_s = \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

Where, n = the number of pairs of values, di = the difference between the ranks of the i^{th} observations of the two variables. Under the null hypothesis of statistical independence of the variables, for a sufficiently large sample, the quantity follows a student's t-distribution with n-2 degree of freedom (Kumari S. *et al.*, 2012).

$$t = \frac{r_s}{\sqrt{(1-r_s^2)/(n-2)}}$$

Spearman correlation and significance level (P-value) was measured using R Package: Hmisc (Harrell Miscellaneous),

3.1.4 Survival Analysis

For performing the survival analysis, the clinical data from Broad GDAC Firehose Stddata was collected. The patients were categorized into two groups based on the mRNA expression level of Sam68 as Z = 1 and above (Sam68 high) and Z = −1 and below (Sam68 low). To understand the clinical relevance of Sam68 difference in expression, compared the effect of high and low expression of Sam68 on patient survival using the Kaplan and Meier method

(Kaplan EL. and Meier P. 1958) and significance (P-value) was calculated using a Log-rank test (Mantel-Cox). Survival curves were generated using GraphPad Prism 7 software.

3.1.5 Enrichment Analysis and Transcript Annotation

Gene ontology analysis both pathway and process enrichment were determined using the Meta-analysis tool, Metascape (Tripathi S. *et al.,* 2015). Both pathways and protein networks within a selected group of genes were built from the metascape ontology sources: GO Biological Processes, KEGG Pathway and Gene Sets. The transcript annotation was done using hg19 as a reference genome, which is available in the UCSC genome browser database (http://genome.ucsc.edu/cgibin/hgGateway).

3.1.6 Measurement of Functional Semantic Similarity

To evaluate the functional similarity between genes, the semantic similarity was measured between sets of Gene ontology terms with which they were annotated. The method described by Wang *et al.,* (Wang JZ. *et al.,* 2007) was used to calculate the functional similarity. A semantic similarity score was determined using the R package "GOSemSim" (Yu G. *et al.,* 2008). Considering any two genes G1 and G2 enriched by GO term sets

GO1 = [go11, go12….go1m] and GO2 = [go21, go22……go2n] respectively

Their semantic similarity score of Wang's methods is defined as:

$$\text{Sim (G1, G2)} = \sum_{1\leq i\leq m} Sim\left(go_{1i,}GO_2\right) + \left(\sum_{\leq j\leq n} Sim\,(go_{2j,}\,GO_1)/m + n\right.$$

3.1.7 Prediction of Sam68 Target Transcripts

The genomic coordinates of the genome-wide binding specificity of Sam68 were obtained from previously published RNAcompete pull-down assay (Ray D. *et al.,* 2013). Only the experimentally determined binding sites were considered. The binding coordinates were then annotated to corresponding hg19 transcripts using the UCSC Genome browser (http://genome.ucsc.edu/cgi bin/hgGateway). All the transcripts that are present within the binding coordinates of Sam68 were identified and further filtered only the coding transcripts, which are Sam68 binding targets. Thus, we have screened all possible UCSC transcripts which have a Sam68 binding site.

3.1.8 Retrieval of Microarray Datasets

Microarray datasets of OV: GSE18520 (Mok SC. *et al.,* 2009), LAML: GSE9476 (Stirewalt DL. *et al.,* 2008) and Non-small-cell lung carcinoma (NSCLC): GSE30219 (Rousseaux *et al.,* 2013) and GSE31210 (*Lin Q. et al.1997)* with clinical information were obtained from GEO database ((*https://www.ncbi.nlm.nih.gov/geo/*). For survival analysis, the patient samples were categorized into two groups: Sam68 high group and Sam68 low group based on the expression median. Both overall survival and stage-wise survival (Stage I-II and Stage III-IV) were compared between these two groups.

3.1.9 Statistical Analysis

The expression value of Sam68 mRNA in different TCGA tumors was shown as Mean + SD as a result of three independent analyses. Differences in Sam68 expression between low and high groups within the TCGA tumors were evaluated by the Mann-Whitney test, with $P < 0.05$ considered to be statistically significant. All the statistical analysis was at least carried out thrice to avoid any insignificant conclusions. Sigma plot was used for all statistical analysis and graphs and survival curves were generated in Graph pad prism 7 software. Other packages, coding was done in R platform -3.2.2 version (R Development Core Team. *http://www.R-project.org)*.

Part 2 – In-vitro Expression, Purification of Sam68 and Extraction of whole-cell protein lysate of NCI-H23.

Materials

All general chemicals are of highest-grade purity and preparation of common reagents is described in the Appendix.

Common chemicals: Sodium chloride (NaCl), Potassium chloride (KCl), sodium hydroxide (NaOH), Disodium hydrogen phosphate anhydrous (Na2HPO4), potassium dihydrogen phosphate (KH2PO4), Glycerol, Absolute ethyl alcohol (99%), methanol, hydrochloric acid (HCl), were purchased from Merck or Himedia (India)

Molecular grade chemicals: Triton X 100, sodium fluoride, sodium orthovanadate, Bradford Reagent (Coomassie Blue G250), ethidium bromide (EtBr), Ampicillin were all procured from Himedia. Isopropyl-β-D-thiogalactopyranoside (IPTG), bovine serum albumin (BSA) phenylmethylsulfonyl fluoride (PMSF) and ELISA- 3, 3', 5, 5'--tetramethylbenzidine TMB substrate, Ethylenediaminetetraacetic acid (EDTA), Trypsin were procured from Sigma (India). Glutaraldehyde from Merck. Tris-HCL and other SDS-PAGE (SDS, acrylamide, bis-acrylamide, Aps, Temed) reagents were obtained from Amersco or Sigma (India) and Western Blot reagents substrate (Luminol enhancer and peroxidase solution) from thermo scientific. PVDF membrane (0.2 µm) from Millipore, USA, 3mM filter Paper from Whatman, USA

Bacterial culture medium: Luria-Bertani (LB) media, Bato agar, Yeast extract, Tryptone, glucose was purchased from Merck or Himedia.

Mammalian cell culture medium: DMEM (himedia) and 1X- an antimycotic antibiotic from Thermo scientific and fetal calf serum purchased from GIBCO.

DNA and Protein markers: 100bp and 1kb DNA ladder and protein molecular weight marker from Banglore Genei OR Thermo-scientific India.

Enzymes: BamH1, EcoR1 from Thermo-scientific, Taq DNA polymerase from Himedia

Plastic and Glasswares: Micropipette, Micropipette tips, microcentrifuge tubes, Petri dishes and other plastic wares- from Tarson Products Pvt. Ltd., India. Glassware – from Borosil International, India.

Recombinant Sam68 Plasmid

Plasmid pGEX-2T containing the insert, full-length Sam68 cDNA of mouse origin in E.*coli* DH5α was kindly gifted by David Shalloway. The construct was described in Lin *et al.*, 1997.

Nucleotide Primers

Name	Sequence	Tm value
Sam68 for	5'-GAAGTCTTTGGACCCCCGTG-3'	60 °C
Sam68 Rev	5'- TTCCACGAACCAAAGCTCCTC-3'	60 °C

Antibodies

Anti-Sam68 rabbit polyclonal, Sigma Aldrich, India (primary antibody)

Anti-IgG rabbit monoclonal, Sigma Aldrich, India (secondary antibody)

Bacterial Cell Cultures

Bacterial strains and colonies of E.*coli* DH5α and BL21 (DE3) were used in this work. Glycerol stocks of these clones were stored at -80°c and cultured in LB or 2XTY medium under suitable conditions and antibiotics. Composition of culture medium are given in appendix table

Culture of NCI-H23 Cell line

Human Lung adenocarcinoma cell line, NCI-H23 was procured from National Centre for Cell Sciences, (NCCS), India and cultured in DMEM (GIBCO) supplemented with 10% fetal calf serum under standard conditions at 37 °C and 5% CO2 supply. When confluence was reached, the culture passaged in 1:20 dilution by trypsin digestion and the further whole-cell extract was used in experiments.

3.2.1 Plasmid DNA Isolation

Plasmid mini preparation was done using a QIAGEN Miniprep kit, following manufacturer's protocol and maxipreparation was done by alkaline lysis method. Individual colonies carrying recombinant plasmid were picked from LB Agar plate (appendix) and inoculated in 3 mL LB medium containing Ampicillin (appendix). After 12-16hrs growth at 37 °C with 180 rpm shaking speed, the culture was pelleted by centrifugation at 10,000rpm for 1min. The harvested pellet was resuspended in 250µl of solution-I, lysed by addition of 250 µl solution-II and mixed by slowly inverting Eppendorf tubes for 4-6 times until the solution becomes Viscous and clear. Then, 350µl of Solution -3 was added and again gently inverted tubes for 4-8 times until this solution turns cloudy. The white pellet containing protein and Genomic DNA complexes were cleared by centrifuging for 10 mins at 13,000rpm. The supernatant was added to a pre-equilibrated Qiagen spin column and centrifuged for 30-60seconds which allows the binding of plasmid DNA to the resin. The spin column was washed by adding 750µl of wash buffer PE and centrifuged for 30-60seconds. Flow-through was discarded and centrifuged for an additional one minute to remove any residual particulates. The spin column was placed in a clean eppendorf tube and 100 µl of ddH_2O was applied to the center of the column. After the column was allowed to stand for 1min, DNA was eluted by centrifuging 5mins at 4000rpm speed.

For maxipreparation of Plasmid DNA, 200mL of LB medium containing ampicillin was inoculated with 200ul of culture medium left from mini preparation and incubated overnight (12-16 hrs.) at 37 °C with 180 rpm shaking speed. The culture was harvested by centrifuging at 8000 rpm for 5mins. Resuspended Bacterial pellet in 20 mL of solution 1 (resuspension solution) and cells was lysed by the addition of 40 mL freshly prepared solution 2 (lysis solution). The mixture was gently mixed by inversion for less than 5 mins, followed by the addition of 30mL ice-cold solution 3(neutralization solution). The tube contents were thoroughly mixed, incubated on ice for 15mins and centrifuged at 15000 x g for 30mins at 4°C. To the obtained supernatant, an equal volume of phenol: chloroform: isoamyl alcohol (1:24:1) was added and mixed by inversion. The contents were spin at 12000 x g for 10mins to separate aqueous and organic phases. The aqueous phase was transferred into a fresh tube, to which 2 volumes of ethanol were added to precipitate out plasmid DNA. After 10mins of incubation at room temperature, plasmid DNA was isolated

by centrifuging at 12000xg for 10mins at 4°C. The DNA was washed with 70% ice-cold ethanol, air-dried and resuspended in nuclease-free water.

3.2.2 Agarose Gel Electrophoresis

For visualizing the DNA, samples were resolved in an agarose gel matrix (plasmid DNA on 0.8% and PCR products on 1% gel, stained with 0.5 µg/mL of ethidium bromide). This gel was run in 1X TAE under 80Volts until the dye line reaches 70-80% of the gel. The resolved DNA fragments were visually compared with standards under UV-Trans-Illuminator and imaged using a gel documentation system.

3.2.3 Quantification of DNA

DNA Fragments in samples were quantified using UV-spectrophotometry. The concentration of double-stranded DNA was calculated with formula 1 OD at 260nm is 50µg/mL.

3.2.4 Restriction Digestion of DNA Fragment

A plasmid containing the desired gene of interest was cut using Restriction enzymes. An aliquot of a plasmid containing Sam68 (~80-90ng/ µl) was incubated with restriction endonucleases (EcoR1, BamH1) and the suitable buffer was kept at 37℃ incubator for 3 hrs. This sample was resolved in agarose gel, alongside uncut plasmid and molecular marker. Details of the double digestion reaction mix are given below.

Table 3.1: The composition of Double Digestion reaction mix

Components	**Volume**
Plasmid	8µl
2X Tango buffer	3µl
EcoR1	0.3µl
BamH1	0.3µl
H2O	3.4µl
Total	**15 µl**

3.2.5 Polymerase Chain Reaction (PCR)

PCR was carried out using TAQ DNA polymerase and primers prescribed earlier. For amplification, 30-50ng (~80ng concentrated) of DNA was used in 20 µl reaction, reaction mix is

given in Table 3.2. PCR was performed in a thermal cycler and typical thermal cycling conditions include three main steps: initial denaturation, Extension and final extension. Details of conditions carried in thermal cycling are given in Table 3.3.

Table 3.2: The composition of PCR reaction mix

Components	**Volume**
dH2O	*10.75*µ*l*
10X Assay Buffer	*2*µ*l*
MgCl2	*2*µ*l*
Dntp mix	*2*µ*l*
Taq Pol	*0.25*µ*l*
Forward primer	*1*µ*l*
Reverse primer	*1*µ*l*
Template plasmid	*1*µ*l*
Total	***20***µ***l***

Table 3.3: Thermal cycling conditions

Initial DNA denaturation	95˚c for 2mins	1 cycle
Denaturation	95˚c for 30secs	30 cycles
Primer annealing	56˚c for 30secs	
Extension	72˚c for 30secs	
Final extension	72˚c for 10mins	1 cycle

3.2.6 Colony PCR

Individual transformants obtained after transformation were suspended in 5ul of nuclease-free water and lysed with heating at 95°c for 5mins. This short heating step releases the inserted plasmid which acts as a target for PCR amplification. Except for initial heating cycling conditions, colony PCR is like normal PCR and conditions followed are as shown in Table 3.4.

Table 3.4: The composition of Colony PCR reaction mix

Components	**Volume**
dH2O	*6.75ul*
Denatured colony	*5ul*
10X Assay Buffer	*2ul*
MgCl2	*2ul*
Dntp mix	*2ul*
Taq Pol	*0.25ul*
Forward primer	*1ul*
Reverse primer	*1ul*
Total	*20 ul*

3.2.7 Preparation of *E.coli* Competent Bacteria Cells

Competent E.*coli* DH5α and BL21 (DE3) were prepared using $CaCl_2$ method as follows. The single bacterial colony was inoculated into 3mL of LB medium and cultured overnight at 37°c with shaking (180rpm). 500µl of this primary culture was then subcultured in 25mL of LB medium and incubated at standard conditions until the OD at 600nm reaches 0.5-0.6. The culture flask was chilled on ice for 5 mins before harvesting the culture by centrifuging at 3000g for 10mins at 4°C. The medium was decanted and the cell pellet was by gentle swirling in 12.5mL of sterile ice-cold 50mM $CaCl_2$. The sample was chilled on ice for 45 mins and centrifuged at 3000g for 10mins at 4°c. The pellet was resuspended in 1-2mL of ice-cold sterile 50mM $CaCl_2$. To this suspension, cold glycerol of 20% final volume was added and aliquoted 200 µl each in sterile eppendroff vials before storing at -80°c until further use.

3.2.8 Transformation in Competent Bacterial cells

One aliquot of Competent *E.coli* BL21 bacterial cells (200 µl) was thawed on ice. 10 µl of Ligated recombinant plasmids were gently mixed to competent bacterial aliquots by pipetting up and down. This mixture was allowed on the ice for 45 minutes. Then this suspension was subjected to heat shock at 42°C for 90 seconds and immediately placed on ice for 2-3 mins. To these cells, 800 µl of prewarmed LB medium was added and incubated for an hour at 37°C with 180rpm shaking. the grown culture was centrifuged at 3000g for 10 minutes at 4°C and resuspended in minimal (50-100 µl) LB medium. These transformants were then plated on agar plates containing specific antibiotic for selection of plasmid containing cells. The plates were

incubated for 12-16 hrs. at 37°C and subsequently positive colonies were further screened for plasmid presences by colony PCR and restriction digestion.

3.2.9 Protein Expression

A single colony of *E.coli* Bl21 (DE3) carrying pGEX-2T - GST-Sam68 was inoculated into 2mL of 2X-TY medium with ampicillin (1mg/mL) and 1% glucose and incubated with shaking at 30'c. 1mL of this overnight pre-culture was subcultured into 100mL of 2X-TY medium (ampicillin(1mg/mL) followed by incubation at 30°C with 180rpm shaking until the OD at 600nm reached 0.5-0.6. Expression of GST tagged Sam68 protein (GST-Sam68) was induced by 0.5 mM IPTG and incubated for 3hrs at 30°C with 180 rpm shaking. Subsequently, the cells were pelleted at 5000rpm (centrifuge model) at 4°C and resuspended in ice-cold PBST lysis buffer (1X Phosphate buffered saline pH 7.4 containing 0.5% Triton X-100, 5mM DTT, 50ul PMSF). Then the cells were sonified at 35% amplitude for 6 times with 10 secs pulse and 10secs break in-between sonication steps. The cell lysate was centrifuged at 10,000 rpm for 20 mins to separate supernatant and pellet and proceeded for purification.

3.2.10 GST Affinity-tag Purification

Sam68 is a GST-tagged fusion protein. Affinity purification of GST-Sam68 was done using Glutathione–Sepharose 4B (Sigma) beads as per the manufacturer's protocol (Sigma). To activate the Lyophilized Glutathione agarose beads powder, initially, the lyophilized powder was soaked overnight in autoclaved deionized water at 4°C to obtain 100% beads slurry. After swelling, the slurry was equilibrated by washing thoroughly with 5 volumes of autoclaved deionized water and then by 5-10 volumes of equilibration buffer (Phosphate buffered saline (PBS), 10 mM phosphate buffer, pH 7.4, 150 mM NaCl, PMSF). 250ul of this glutathione beads slurry was added to 1mL of clarified supernatant of induced *E.coli* BL21 sample. This sample was incubated with slow rotation for 30mins in cold conditions or 4°C to allow efficient binding. Followed by binding, the beads were washed twice with 5 volumes of PBS-T (PBS containing 1% Triton X-100). Finally, one volume of elution buffer (50mM Tris-HCl buffer containing 20mM reduced glutathione (GSH), pH-8) was added to the beads and after 10mins of incubation, the protein was eluted. This elution step was repeated thrice. To estimate recombinant protein concentration Bradford assay was performed and an aliquot of this protein sample was analyzed using SDS PAGE.

3.2.11 Estimation of Protein Concentration

Protein concentration was estimated by Bradford assay (*Bradford M. M. 1976)* using BSA as a standard. As per manufacturers protocol, to different concentrations of BSA: 1ug, 2ug, 5ug, 10ug, 20ug, 50ug and different amounts of recombinant protein: 10ul, 40ul, 1 mL of Bradford reagent (Sigma) was added. The volumes of the samples were made up using PBS buffer or deionized water and then incubated for 5 mins at room temperature. Followed by incubation, the optical density of all the samples was measured at 595nm.

3.2.12 Sodium Dodecyl Sulfate - Polyacrylamide gel electrophoresis (SDS PAGE).

Proteins were separated electrophoretically based on their molecular weight, to assess the expression of recombinant Sam68 protein in a bacterial system and to study the native Sam68 expression in NCI-H23 (human Lung adenocarcinoma cell line). A sieving matrix with 12% resolving and 5% stacking gel was cast for electrophoresis according to Sambrook (Sambrook *et. al.,*). The required amount of purified Protein or whole-cell extracts were mixed with 6x SDS PAGE loading buffer (appendix) containing reducing agents and kept in boiling water bath for 5mins. The denatured samples along with Protein markers were resolved using the SDS PAGE method of Lamelli (Laemmli UK. 1970). Samples are run in Tris-Glycine Tank buffer (appendix) at constant 100V and 120V for stacking and separating in a Vertical electrophoresis system. Following the completion of electrophoresis, the Gels were stained with Coomassie Brilliant Blue method (appendix).

3.2.13 Western Blotting

Following Sam68 separation by SDS-Page, the proteins were confirmed using Western Blotting. Initially, proteins were electrically blotted on the PVDF membrane (pre-soaked in methanol) for 4hrs at 60V in transfer buffer (recommended by the manufacturer) using semi-dry transfer apparatus. The completion of protein transfer was confirmed by Ponca S staining. The membrane was then washed twice with PBS containing - 0.1% tween (PBST) to remove stain and then incubated in a Blocking solution, PBST containing 5% BSA for 2hrs at room temperature. For the detection of the protein of interest, the membrane was exposed to appropriate (manufacturer's recommended dilution (1:1000) primary antibody in 5% BSA in PBST for 2hrs at room temperature with shaking or overnight incubation at 4°C. After three washes with PBST, the membrane was incubated with HRP conjugated secondary antibody for 2hrs at room

temperature. This was again followed by three washes of PBST for a minimum of 5 minutes each at room temperature with shaking. Finally, the blot was developed by enhanced chemoluminescence using Pierce™ ECL Western blotting substrate (Thermo Fischer Scientific) and then imaged by the Gel documentation system to visualize the protein band.

3.2.14. Cell Viability Assessment using Trypan Blue method

For measuring the viability of the cells, the cells were stained with trypan blue, followed by cell counting using a hemocytometer. Trypan blue exclusively stains the dead cells as it cannot enter the intact cell membrane of the live cells (Mosmann T. 1983). Adherent Cells suspension was trypsinized to detach from the culture plate and resuspended in fresh medium. To this equal volume of 0.4% trypan blue was added, gently pipetted and incubated for 2mins at room temperature. The cell suspension was visualized under an inverted phase-contrast microscope, and viable cells were counted. The percentage of viable cells was determined using the below formula.

Percentage viability = (No. of. viable cells/Total No. of cells) *100

3.2.15. Isolation of Protein from the Cell line

Cell lysis and extraction of whole-cell proteins of human lung adenocarcinoma cell line (NCI-H23) were done using RIPA (radioimmunoprecipitation assay) buffer (appendix). At 80-90% confluency of NCI-H23, the media was discarded, and the cells were washed with PBS buffer. The cell suspension was centrifuged at 600Xg for 5 mins at 4°C and the pellet was resuspended into pre-chilled RIPA buffer containing 1mM PMSF, 50mM sodium fluoride and 1mM sodium orthovanadate. After 10mins of incubation on ice, cells were sonicated for 10secs and total protein suspension was obtained by centrifuging cell lysate at 10,000rpm for 10 mins at 4°C. The total protein sample was stored at 4°C for further applications.

Part 3 - Indirect ELISA, Immunosensor Fabrication, Characterization and Application.

3.3.1 ELISA

Indirect Elisa was performed to quantify Sam68 protein in NCI-H23 using Purified recombinant GST-Sam68 as the standard known concentration of BSA, various predetermined concentrations of GST-Sam68 (10^{-3}, 10^{-2}, 10^{-1}, 1, 5, 10 µg/mL) and different dilutions of NCI-H23 protein extract as 1:100, 1:10 and 1:1 were used for detection. 250ul of all the above said concentration was initially coated in 96 well plates. Then the plate was washed with PBS buffer containing 0.1% tween 20 (PBS-T) and blocking solution (0.5% BSA (1mg/mL) in PBS-T was added to the plate. After 1 hr. incubation, primary antibody prepared in blocking solution was added to the plates and left for untouched for immunoreaction to take place for 1hr under slow rotation at 4°C. The next step is washing, and the addition of secondary antibody conjugated with HRP enzyme was diluted (according to manufacturer's protocol) with blocking buffer and left for 1hr of incubation. Finally, the 50ul of 3, 3′, 5, 5′-Tetramethylbenzidine (TMB) liquid substrate system (Sigma) was added as a substrate. Enzyme-substrate reaction was stopped by the addition of 0.2M sulphuric acid and optical density was measured at 370 nm using UV-Vis spectrophotometry in Multiskan GO microplate reader (ThermoFisher Scientific, USA).

3.3.2 Electrochemical Measurements

All the Electrochemical measurements (Cyclic Voltammetry (CV) and Electrochemical Impedance Spectrometry (EIS)) were performed on Auto lab Potentiostat/Galvanostat (Met Rohm, India) comprising of conventional three-electrode system. Glassy carbon electrode (GCE) as the working electrode, Platinum wire as Counter electrode and saturated Ag/Agcl (3 M KCl) filled with Kcl as a reference electrode. Analysis of Electrochemical results was carried out with NOVA software (version 1.11.0).

EIS and CV were performed in 0.1 M PBS (pH 7.4) with 5.0 mM redox couple $K_3Fe(CN)^6$ / $K_4Fe(CN)^6$ and 0.1 M KCl. All CV experiments were performed in the potential range from -1 to +1 V at a scan rate of 0.1 V/s. For EIS measurements an alternating wave of 10 mV amplitude at 0.18 V step potential was applied. Impedance spectra was recorded between 10,000-0.05 Hz

frequency range. The impedance values were fitted to the Randle's equivalent circuit using NOVA software. The charge transfer resistance (*Rct*) at the electrode surface was calculated from the diameter of the semicircle portion at higher frequencies in the impedance spectra.

3.3.3 Fabrication of Electrochemical Immunosensor

The glassy carbon electrode (GCE) electrode surface was polished mechanically using alumina slurry until a mirror finish was obtained and cleaned thoroughly by sonication using Milli-Q water for 10 mins. To modify Bare GCE, polymer polyaniline (PANI) films were electro-deposited by cyclic voltammetry in 0.1 M aniline and 0.5 M sulphuric acid solution, which was performed from 0 to 1V potential range with a scan rate of 100mV/sec for 10 cycles onto the surface of the electrode to form polyaniline (PANI) modified GCE (GCE/PANI). Subsequently, the PANI/GCE electrode was rinsed with ultrapure water to remove any un-polymerized aniline and then 5µl of 4% glutaraldehyde was immobilized onto the PANI/GCE surface and incubated for 2hrs. Glutaraldehyde is a crosslinked and helps in good binding of the biomolecule. After washing with sterile deionized water, the surface of the Glutaraldehyde/PANI/GCE electrode was modified by drop-casting, 5 µl of 3µg/mL anti-Sam68 antibody in PBS and left for 2 hrs. at room temperature to get immobilized. Finally, the electrode was thoroughly rinsed with sterile deionized water, PBS (0.1M) and incubated with 5µl of 0.1% BSA solution for 1hrs. to avoid any non-specific binding. The modified electrode (GCE/PANI/Glue/Sam68-Ab/BSA) was stored at 4 ºC.

3.3.4 Electrochemical Determination of Target Proteins

Increasing concentrations of purified Sam68 protein (10^{-3} ,10^{-2} ,10^{-1} ,1,5 µg/mL) and different dilutions of NCI-H23 whole cell lysate (1:100, 1:10, crude extract) were incubated with the modified working electrode with for 1 h at room temperature. After the immunoreaction has occurred, the unbound proteins on the electrode surface were washed off with PBS and EIS, CV measurements were acquired.

3.3.5 Spike and Recovery

To perform Spike and recovery, a known concentration of reference Sam68 protein was spiked with a proper dilution of the NCI-H23 whole-cell protein sample (1:10). This sample was incubated for 1hr at room temperature with the modified electrode and electrochemical signals were obtained.

CHAPTER 4

RESULTS & DISCUSSION

Chapter 4: Results and discussion

Part 1 – Identification of Sam68 as a potential biomarker for Cancer

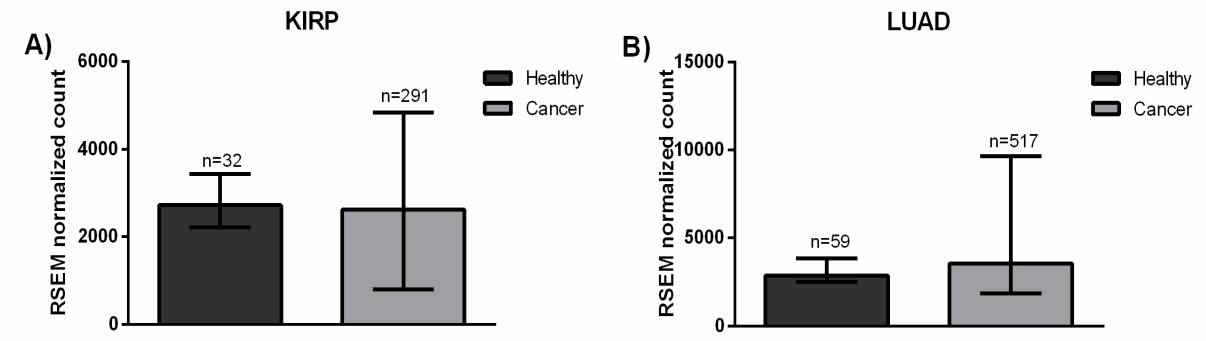
4.1.1 Differential Expression of Sam68 in Different Cancer Types

To comprehensively analyze the expression pattern of the Sam68 gene (Sam68) in different cancer types, four previously unexplored cancer types as KIRP, LUAD, OV, and LAML were chosen. High throughput genomic sequencing data of normal tissues, primary tumors and adjacent normal tissues of TCGA were extracted from BROAD Institute (http://gdac.broadinstitute.org/). Normalized RNA-Sequencing by Expectation-Maximization (RSEM) counts of Sam68 were compared for expression differences between different samples. Although other alternative–forms of normalization methods as RPKM, FPKM are available, RSEM expression values were preferred as exon reads are proportionally assigned based upon the ratio of reads mapping to isoform unique regions and its gained popularity among TCGA data analysis (Zhao *et al.*, 2014).

First, the Sam68 expression between healthy and cancerous tissues of KIRP and LUAD patients were compared. The basal RSEM counts of Sam68 in LUAD cancer patients are minimalistically more but No significant expression difference was observed in cancer compared to normal samples (Figure 4.1A and B). To reconfirm the observation, the difference in the expression of Sam68 between cancer and adjacent healthy tissues was assessed. Similar to the previous observation, no valid difference in Sam68 expression was identified. Substantially, Sam68 exhibited highly scattered expression within the cancer samples of KIRP and LUAD (Figure 4.1C and D). Due to the unavailability of normal tissue samples for LAML and OV in TCGA, microarray datasets of OV (GSE18520) and LAML (GSE9476) were obtained from GEO (Gene expression omnibus) to explore the Sam68expression level and no further analysis was carried out using these microarray datasets. Like, KIRP and LUAD, in LAML and OV also Sam68 has no difference in expression (Figure 4.2A and 4.2B). From the observation, in all four cancers, there is no difference in Sam68 expression, but it is highly scattered within the cancer samples.

To further explore, the cancer patient samples were subdivided based on Sam68 mRNA expression. Samples were dichotomized is based on the Z-score value as $Z \geq +1$ and $Z \leq -1$ as Sam68 overexpression and low expression group respectively. The z-score value indicates several standard deviations away from the mean expression. Z-score values of Sam68 for KIRP, LUAD, OV, LAML samples were obtained from TCGA. In all the four cancers, the Z-score is broadly distributed from negative to positive values with exceptionally many patients within cancer showing significantly high or low Z-score values (Figure 4.3A). Moreover, Z-scores of Sam68 is not widely distributed in normal adjacent tissues tissue of KIRP and LUAD in comparison to Cancerous tissues (Figure 4.3C). This observation indicates, that Sam68 mRNA expression doesn't follow a pattern (high or low) with any particular cancer type. Further, Sam68 expression is patient-specific within particular cancer and not cancer-specific. Therefore, samples of each cancer were categorized based on Z-score as Z = 1 and above as high expression group and Z =-1 and below as low expression group (Figure 4.3B). Following patient selection in two groups, differential expression analysis was carried out and the significance of the expression difference was assessed by performing a non-parametric Mann-Whitney test. Sam68 exhibited, statistically significant ($P < 0.0001$) expression difference between $Z \geq +1$ and $Z \leq -1$groups in KIRP, LUAD, OV and LAML (Figure 4.4). However, this stratification of patients in higher and lower expression based on the Z-score of Sam68 expression is limited to specific cancer patients within a cancer type.

Pooled Healthy Vs Pooled Cancer Samples



Adjacent Healthy Vs Cancer tissue

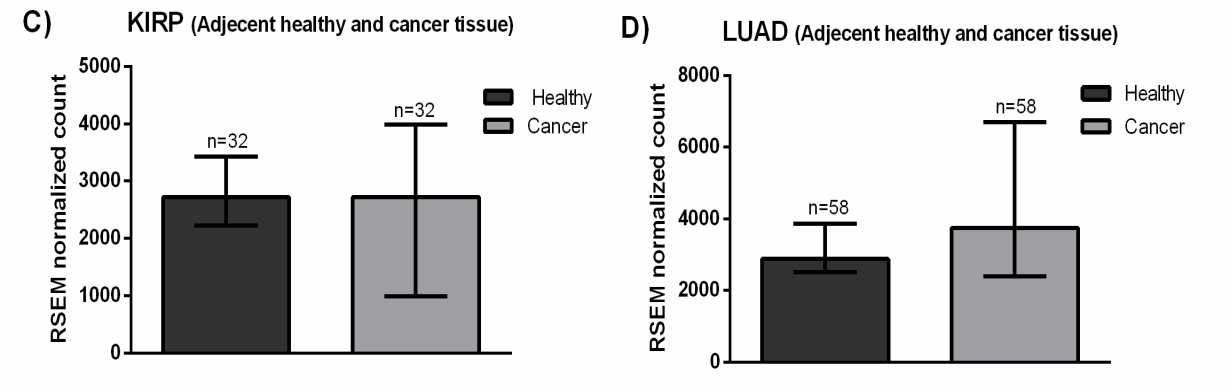


Figure 4.1. *Expression of Sam68 mRNA in* ***KIRP and LUAD****: (A, B) mRNA expression in the healthy and cancerous tissue of KIRP & LUAD patients. (Error bar in each diagram represent the maximum and minimum value of RSEM normalized count. KIRP: kidney renal papillary cell carcinoma, LUAD: lung adenocarcinoma). (C, D) mRNA expression in adjacent healthy and cancer tissue from the same patient in KIRP & LUAD respectively (Error bar in each diagram represents the maximum and minimum value of RSEM normalized count).*

Pooled Healthy Vs Pooled Cancer Samples

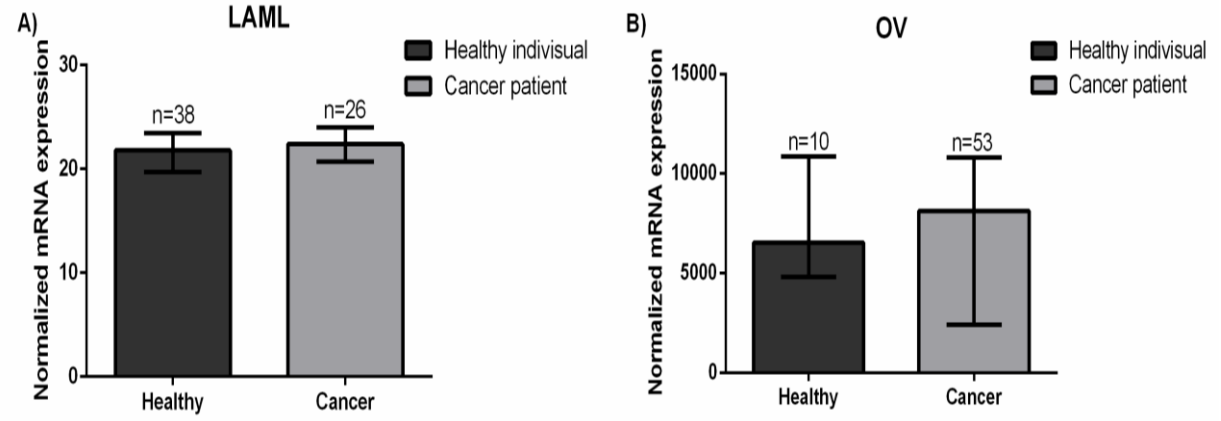


Figure 4.2*: Expression of Sam68 mRNA in* ***LAML and OV****: A) & B) mRNA expression in the healthy and cancerous tissue of LAML and OV patients respectively. The normalized mRNA expression data of the microarray experiment were collected from the Gene Expression Omnibus (Error bar in each diagram represent the maximum and minimum value of normalized mRNA expression).*

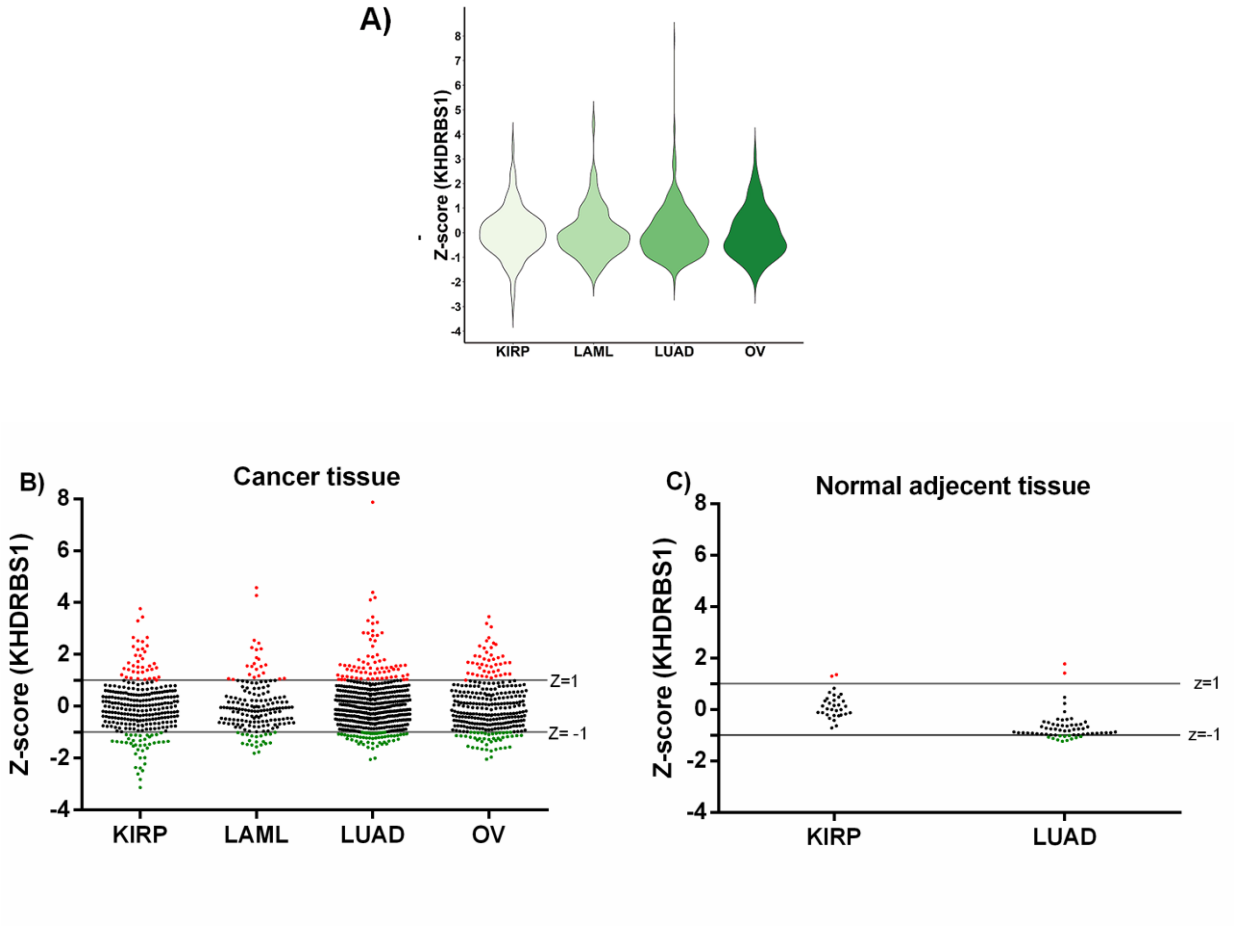


Figure 4.3*: Z-score distribution of Sam68 expression in cancer and normal tissue*: *(**A**) Volcano plot summarizing the Z-score distribution of Sam68 expression in different cancer B) Dot plot summarizing the Z-score distribution of Sam68 expression in four different cancers. The horizontal line indicates the chosen threshold value of the Z-score (Z = 1 and Z = -1). The red color dots indicate the patient with Z-score>1 and green color indicates the patient with Z-score < -1. C) The dot plot is summarizing the Z-score distribution of Sam68 expression in the normal adjacent tissue of KIRP and LUAD (a similar dot plot for LAML and OV are not shown due to non-availability of normal adjacent tissue data). It is observed that the Z-score distribution of KHDR BS1 expression in the normal adjacent tissue of KIRP and LUAD is not widely distributed compared to cancer tissue.*

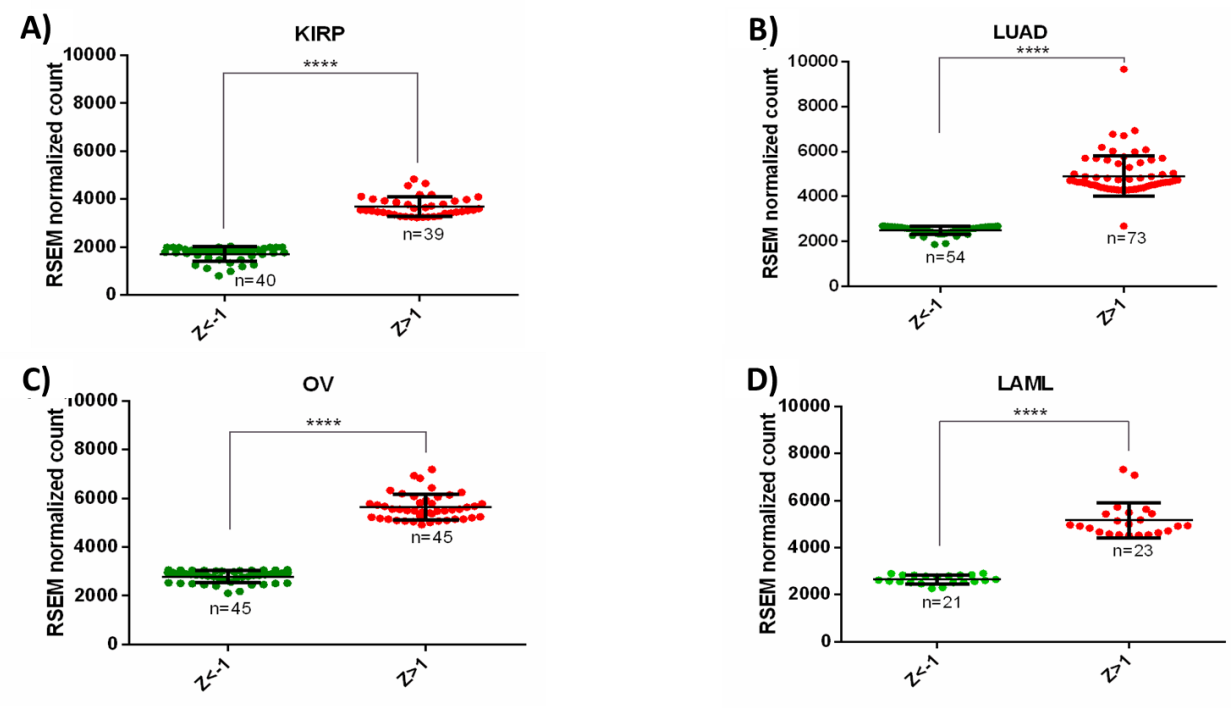


Figure 4.4: *Dot Plot shows the difference in Sam68 mRNA expression level in Z > 1 and Z < −1 sample in KIRP, LUAD, OV and LAML respectively (****P < 0.0001).*

4.1.2 Sam68 Higher Expression is associated with the Survival of KIRP and LUAD Patients.

To understand the clinical relevance of the higher expression of Sam68, Kaplan-Meier survival analysis and log-rank test (Mantel-Cox) were performed. Clinical data of cancer patients selected into $Z \geq +1$ and $Z \leq -1$ were collected from TCGA and compared the overall survival between the groups. This analysis indicates, increased Sam68 expression has substantially ($P < 0.05$) decreased the overall survival in $Z \geq +1$ patients of KIRP, and LUAD (Figure 4.5A, 4.5B). Even though, in LAML and OV, $Z \geq +1$ patients have higher expression of Sam68 ($P < 0.0001$) compared to $Z \leq -1$, which doesn't affect the overall survival of the patients (Figure 4.5C, 4.5D). This significant correlation indicates the Prognostic value of Sam68 in specific patients of KIRP and LUAD but not in LAML and OV. Thus, Sam68 could be a molecular target in KIRP and LUAD. Moreover, this gives us fascinating evidence that the over-expression of Sam68 may not always be accountable for cancer progression and patient survival.

The gene or protein function in any particular environmental or diseased context depends on its specific interacting partners. In this scenario, Sam68 interacting partners in KIRP and LUAD could be possibly different from that of LAML and OV, which results in a different outcome. Moreover, each cancer has a unique phenotypic property which is evolved due to distinct gene expression and molecular interactions inside a cell. Thus, knowledge of interacting partners of Sam68 is a pre-requisite, to the light-up exact mechanism of Sam68 function in different cancer types.

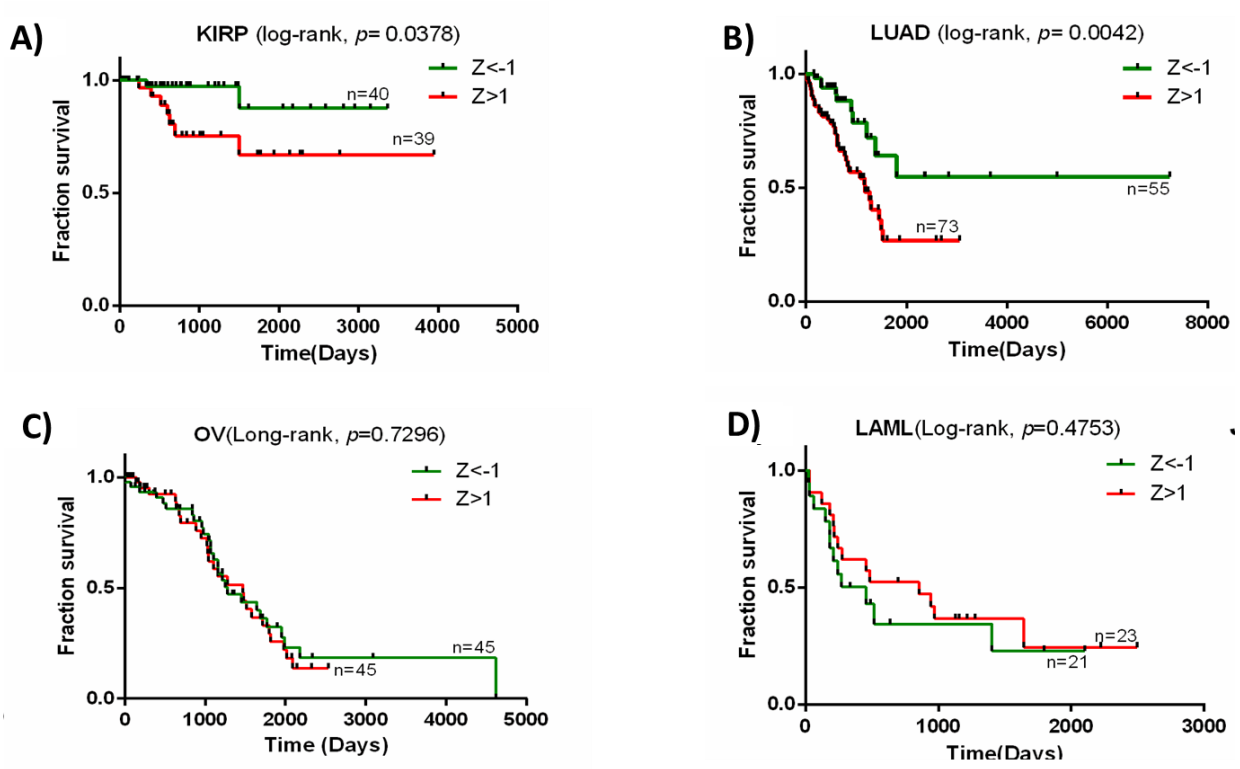


***Figure 4.5**: Kaplan-Meier curves showing the comparison of fraction survival in higher expression (Z > 1) and lower expression (Z < −1) group in all four cancers. In KIRP and LUAD, the higher expression of Sam68 affects the patient survival (P < 0.05), whereas in OV and LAML there is no difference in patient survival (P > 0.05) in higher and lower expression group.*

4.1.3 Genome-wide Coexpression Analysis and Functional Clustering of Sam68 Coexpressed genes

Given the importance of Interacting partners, to address the patient and cancer-specific role of Sam68, genome-wide correlation analysis was performed. Correlation networks are an essential framework for identifying sets of genes coexpressed along Sam68, that respond in a coordinated way to diseased conditions such as cancer. Such networks also elucidate the regulatory and functional relationships and determine the potential of Sam68 as a molecular target in Cancer. Sam68 Correlation Networks associates genome-wide gene sets with a significant positive correlation in cancer. To find the gene sets, the spearman correlation of Sam68 to all other expressed genes (20,531 genes) in a specific cancer was calculated. In all four cancer types, $Z \geq +1$ patients (Sam68 high expression) were only selected for correlation analysis. Genes with correlation coefficient (*rs*) > 0.3 with statistical significance, $P < 0.05$ were filtered for network analysis. Distribution plot of correlation coefficient (*rs* > 0.3 and $P < 0.05$) (Figure 4.6) shows that the no of genes correlating at median coefficient values for KIRP, LUAD and OV are almost equal, but it is high in case of LAML. However, the higher number of correlated genes in LAML does not play any significant role in the overall function, because in the subsequent experiment, the Functional similarity index, (Figure 4.8) we have observed that the functional similarity between these genes is less.

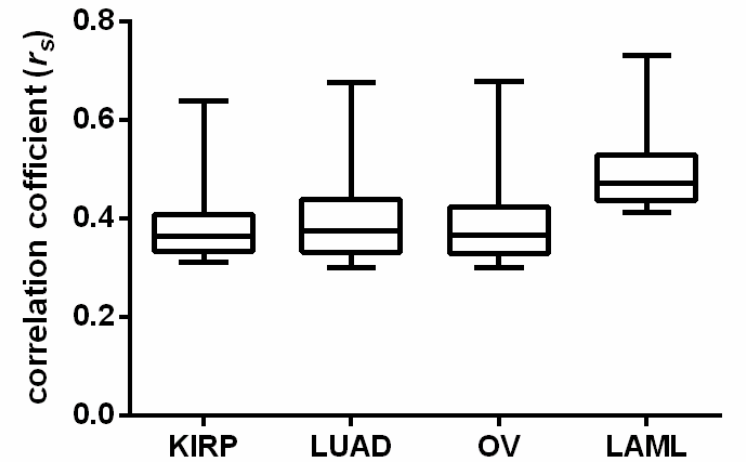


Figure 4.6*: Boxplot summarizing the distribution of the correlation coefficient of Sam68 to all other genes (rs > 0.3, P < 0.05). In boxplot, the median is indicated by the horizontal line dividing the interquartile range (Q25, Q75). Upper and lower ticks represent the maximum and minimum value (KIRP: kidney renal papillary cell carcinoma, LUAD: lung adenocarcinoma, LAML: acute myeloid leukemia, and OV: ovarian cancer).*

A common observation in gene expression is that many genes which show similar expression patterns frequently clustered according to their biological functions (Eisen M. B. et al.1998, Reynier F. *et al.,* 2011) Therefore analysis of functional clustering of all genes which are co-expressed with Sam68 can provide a clear view of predominant functions associated with the group of genes expressed in a specific cellular context. Next, protein-protein interaction enrichment analysis for all co-expressed genes ($rs > 0.3$, $P < 0.05$) in each cancer was performed using Metascape tools, which fetch the interaction data from BioGrid (Stark C. *et al.,* 2006), In Web_IM (Li T. *et al.,* 2017), and OmniPath (Turei, D., *et al.,* 2016). The resulting network was again used to identify densely connected network components using molecular complex detection (MCODE) algorithm (Bader G.D. *et al.,* 2003). Pathway and process enrichment analysis find the function of each densely connected component (Figure 4.7). Interestingly, the co-expressed genes in KIRP and LUAD are mostly involved in the cell cycle, and cell division related processes such as chromatin assembly and organization, cell cycle checkpoint control. As many of these densely connected genes are co-expressed with Sam68, it can be presumed that probably Sam68 is also involved in a similar function in KIRP and LUAD (Figure 4.7). However, in OV and LAML, the network components are less densely connected and several gene clusters that are present in KIRP and LUAD and involved in cell proliferation are absent in OV and LAML (Figure 4.7). It is now comprehensible that Sam68 driven molecular processes are similar in the case of KIRP and LUAD but different in OV and LAML for a specific group of patients. To gain insights into whether the genes which are coexpressed with Sam68 are involved in similar biological functions or not, Gene Ontology (GO) semantic similarity was performed. This analysis was used to quantify the functional association of coexpressed genes. The co-expressed genes in KIRP and LUAD tend to have significantly high ($P < 0.001$) functional relationships compared to OV, LAML and random set (Figure 4.8). It explains coexpressed genes in KIRP and LUAD are involved in the functionally similar biological processes and pathways, which support our previous observation of functional clustering of coexpressed genes (Figure 4.7) as most of the enriched processes in KIRP and LUAD are linked to cell proliferation.

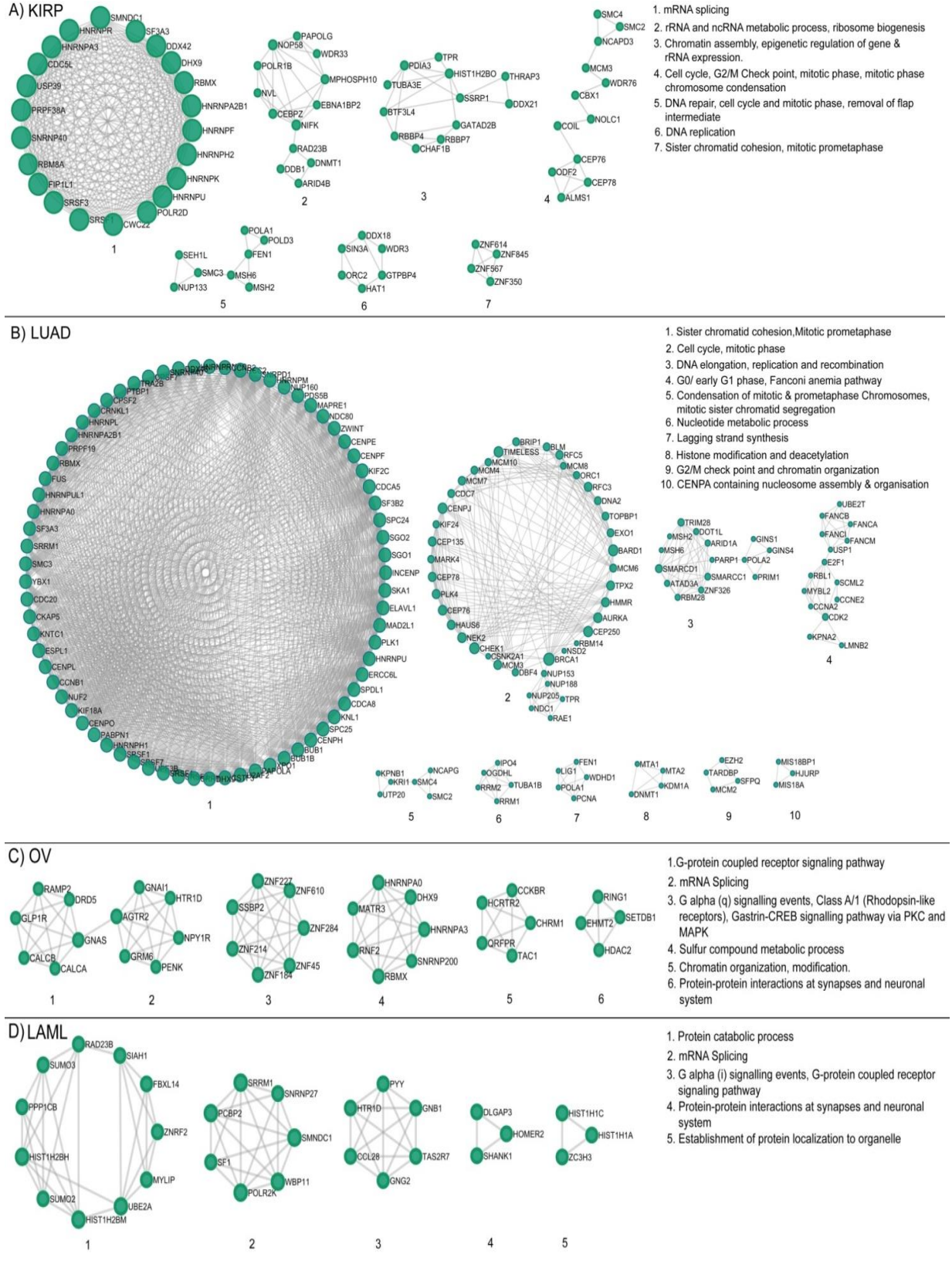

A) KIRP
1. mRNA splicing
2. rRNA and ncRNA metabolic process, ribosome biogenesis
3. Chromatin assembly, epigenetic regulation of gene & rRNA expression.
4. Cell cycle, G2/M Check point, mitotic phase, mitotic phase chromosome condensation
5. DNA repair, cell cycle and mitotic phase, removal of flap intermediate
6. DNA replication
7. Sister chromatid cohesion, mitotic prometaphase
B) LUAD
1. Sister chromatid cohesion,Mitotic prometaphase
2. Cell cycle, mitotic phase
3. DNA elongation, replication and recombination
4. G0/ early G1 phase, Fanconi anemia pathway
5. Condensation of mitotic & prometaphase Chromosomes, mitotic sister chromatid segregation
6. Nucleotide metabolic process
7. Lagging strand synthesis
8. Histone modification and deacetylation
9. G2/M check point and chromatin organization
10. CENPA containing nucleosome assembly & organisation
C) OV
1.G-protein coupled receptor signaling pathway
2. mRNA Splicing
3. G alpha (q) signalling events, Class A/1 (Rhodopsin-like receptors), Gastrin-CREB signalling pathway via PKC and MAPK
4. Sulfur compound metabolic process
5. Chromatin organization, modification.
6. Protein-protein interactions at synapses and neuronal system
D) LAML
1. Protein catabolic process
2. mRNA Splicing
3. G alpha (i) signalling events, G-protein coupled receptor signaling pathway
4. Protein-protein interactions at synapses and neuronal system
5. Establishment of protein localization to organelle

Figure 4.7*: Functional clustering of coexpressed genes in different cancer tissue*: *A), B), C) and D) show densely connected components in the coexpressed gene (rs>0.3, P<0.05) of Sam68 in KIRP, LUAD, OV, LAML and their functions (gene ontology) in each of cancer tissue respectively.*

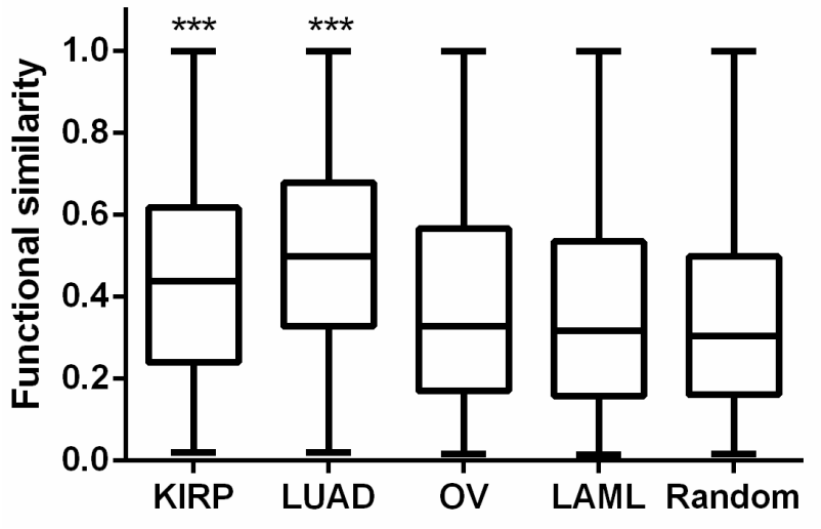


Figure 4.8: *Distribution of functional similarities between the coexpressed genes in different cancer. The functional similarities between coexpressed genes (rs > 0.3, p < 0.05) with Sam68 is calculated based on GO semantic similarity. The random set of genes (Random) is used as a negative control. The functional similarity is high in the case of KIRP and LUAD compared to the OV, LAML and random set (n = 500) of genes (box boundaries represent the first and third quartile (Q.25, Q.75). The median is indicated by the horizontal line dividing the interquartile range. Upper and lower ticks represent the maximum and minimum value). Mann-Whitney test was performed separately in between KIRP vs. OV, LAML, Random and LUAD vs. OV, LAML, Random (***P < 0.001).*

Next, the protein interaction map of Sam68 was constructed using a different database Resources (Bader GD. *et al.,* 2003, Zanzoni A. *et al.,* 2002, Stark C. *et al.,* 2006, Xenarios I. *et al.,* 2000, Peri S. *et al.,* 2004, Razick S. *et al.,* 2008). Only the direct physical interactions between Sam68 and other human protein and experimentally determined binary interactions, which are generated using yeast two-hybrid or high-throughput experiments (ABI2, ACTB, AGO1, AHI1, AMPH, APBB1, APC, ARHGEF4, ARHGEF9, AZIN1, BAIAP2L1, BMI1, BTK, BZRAP1, CAND1, CASP8, CBL, CBX6, CCDC8, CD2AP, CD2BP2, CD81, CDK1, CIRBP, CREB3L3, CREBBP, CRK, CRKL, CSK, CUL3, CUL7, DDX5, DHX9, DLG1, DLG2, DLG3, DLG4, DNMBP, DOCK2, DOCK3, DROSHA, EFEMP1, EMG1, ESR1, EZH2, FADD, FGR, FNBP4, FRK, FYN, GAS7, GPHN, GRAP, GRAP2, GRB2, HCK, HNRNPA1, HNRNPC,

HNRNPK, HNRNPR, INSR, IRS1, ITK, ITSN1, ITSN2, JAK3, LCK, LGR4, LYN, MAPK1, MIA2, MPP6, MYO1C, MYO7A, NCF1, NCK1, NCK2, NCKIPSD, NPHP1, OBSL1, OSTF1, PACSIN1, PACSIN2, PACSIN3, PIK3R1, PIK3R3, PLCΓ1, POT1, PPP1R13B, PRMT1, PSTPIP1, PTBP2, PTK6, PTPN6, RALY, RAPSN, RASA1, RBFOX2, RBM7, RBMX, RIPK1, RNF2, RPA1, RPA2, RPA3, RUSC2, SASH1, SCG5, SH3GL1, SH3KBP1, SH3PXD2A, SH3YL1, SKAP2, SMAD2, SMARCA2, SMARCAD1, SNRPN, SNX30, SNX9, SORBS1, SPATA13, SRC, SRRM1, SSFA2, STAT3, STUB1, SUMO1, SUZ12, TBL1X, TJP1, TNFRSF1A, TNFSF11, TNS3, TOB1, TUBB3, TUBB4A, U2AF2, UBA52, UBASH3B, UBC, VAV1, VCL, WBP4, YES1, YTHDC1, ZBTB7A, ZDHHC6) were considered for the network. Following, a weighted gene list was generated from screening genes with significant correlation ((*rs*) > 0.3, P < 0.05), which have physical interaction with Sam68 for each cancer. Both criteria were chosen to increase the stringency of a selection of Sam68 interacting partners in a specific cancer cell. Venn diagrams (Figure 4.9, 4.10) show that each cancer type has overlapping genes that are co-expressed and physically interact with Sam68. Interestingly, most of these coexpressed and interacting genes of Sam68 are different across the four cancers. Moreover, we observed that the numbers of these overlapping genes are less in OV and LAML compared to KIRP and LUAD. Further, this curated gene list was used for gene set enrichment analysis aimed to find the cancer-specific biological function of these genes. For this, Gene ontology: process and pathway category enrichment were performed. In case of KIRP and LUAD, the cancer-specific processes such as regulation of signaling by cbl (Liyasova MS. *et al.*, 2015) SUMOylation of RNA binding protein (Kota V. *et al.*, 2018, Yang Y. *et al.*, 2017, Seeler JS. and Dejean A. 2017), ras protein signal transduction pathway (Downward J. 2003), microRNAs in cancer (Peng Y. and Croce CM. 2016) are predominant (Figure 4.9). However, in the case of OV, a pathway of RNA splicing is the only predominant event and no process or pathway enrichment is found in the case of LAML (Figure 4.10). It is interesting to notice that overexpression of Sam68 leads to enrichment of cancer-specific events in KIRP, LUAD but not in OV and LAML. The result indicates a positive correlation between Sam68 expression status and cancer phenotype in KIRP and LUAD. The results also show a similar expression pattern of a gene differentially affects the disease state, probably due to cancer and patient-specific genetic profile. Therefore, genes which are coexpressed and interact with Sam68 are mostly different in KIRP and LUAD,

although they are involved in cancer-specific biological processes that are accountable for patient mortality.

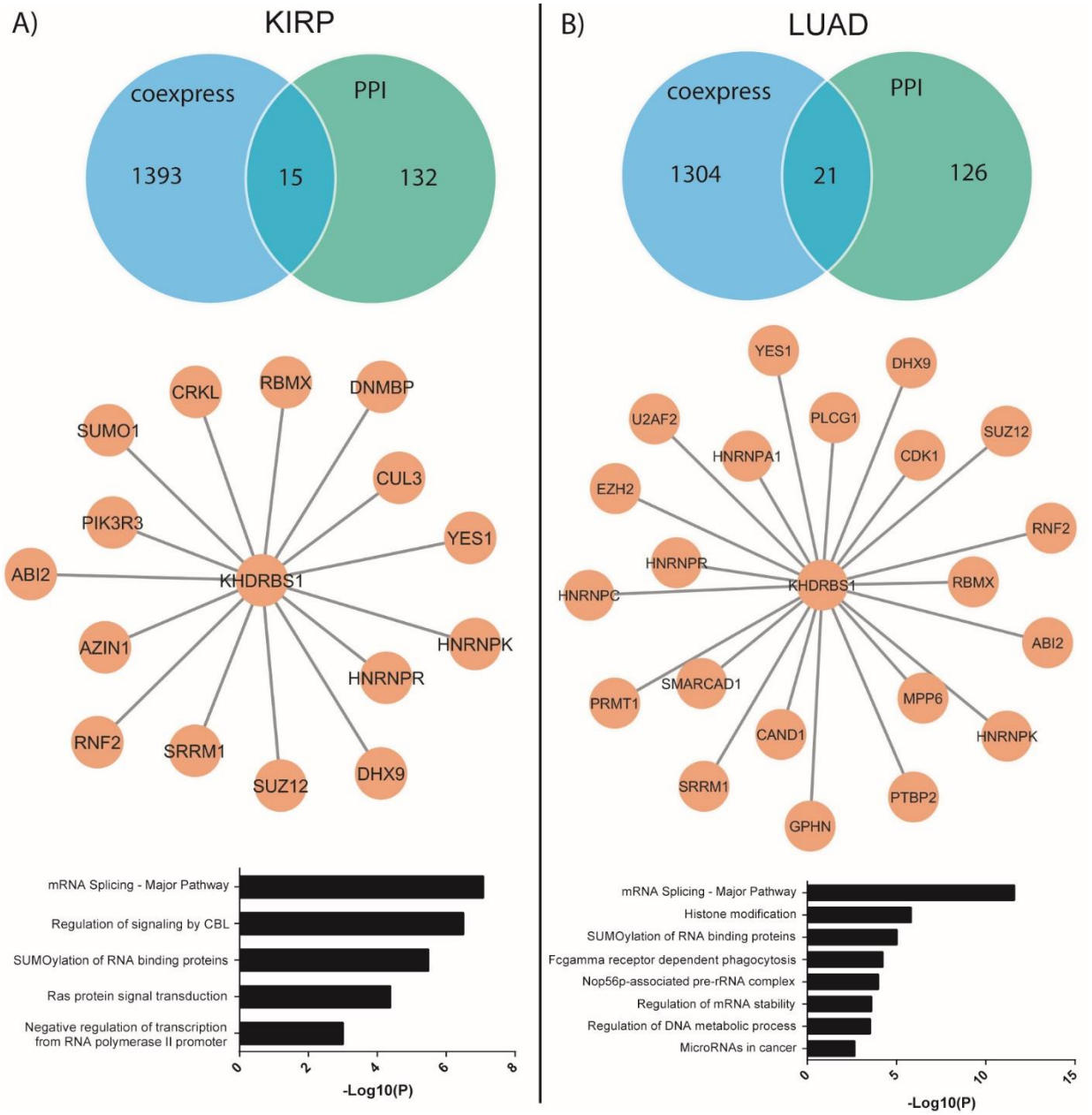


Figure 4.9: *Overlap of protein-protein interactions (PPI) dataset and coexpressed gene of Sam68 and processes and pathway enrichment analysis in KIRP and LUAD: (**A**-B) Venn diagram and network figure shows the overlapping genes which coexpress and interact with Sam68 in KIRP and LUAD respectively. The bar diagram indicates the process and pathway enrichment analysis of the overlapping gene in the respective cancer. Logarithmic corrected p-values for significant overrepresentation are shown.*

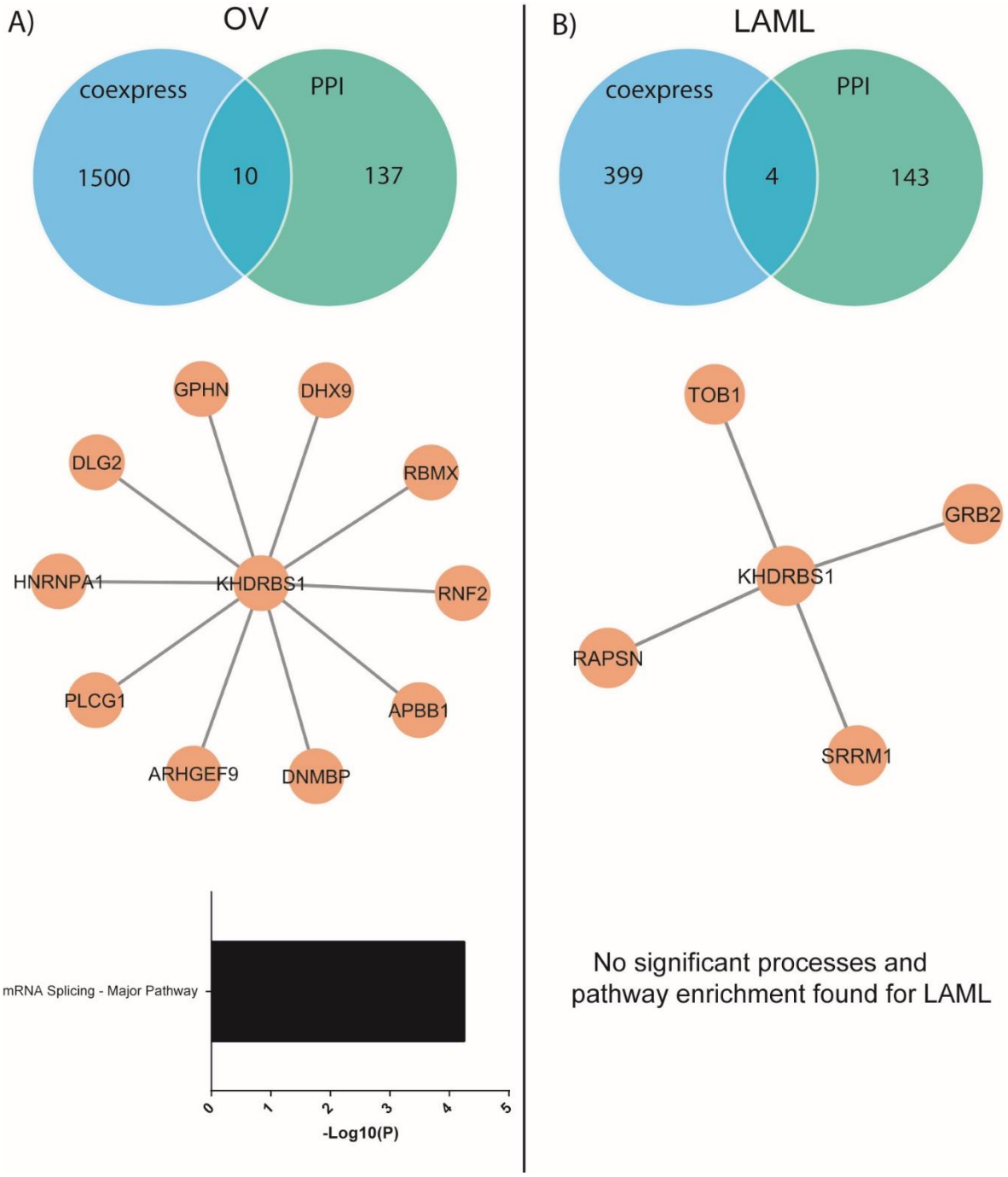


***Figure 4.10**: Overlap of protein-protein interactions (PPI) dataset and coexpressed gene of Sam68 and processes and pathway enrichment analysis in OV and LAML: (**A**-B) Venn diagram and network figure shows the overlapping genes which coexpress and interact with Sam68 in OV and LAML respectively. The bar diagram indicates the process and pathway enrichment analysis of the overlapping gene in the respective cancer. Logarithmic corrected p-values for significant overrepresentation are shown*

4.1.4 Genome-wide Transcript Correlation Analysis confirms that Sam68 is a Prognostic Marker in KIRP and LUAD

Through, gene-level expression data analysis allowed the identification of the genome-wide coexpressed genes of Sam68 and their prevailing cellular function in different cancer. However, Sam68 is an RNA binding protein and involved in RNA splicing. Indeed, Sam68 driven oncogenic splicing isoform is reported in the development of many Cancer phenotypes (Matter N. *et al.,* 2002, Paronetto MP. *et al.,* 2007). Hence, it is necessary to address other possible reasons at the transcript level, to understand the underlying mechanism for differential behavior of Sam68 in cancers: a) whether expression variation amongst different transcripts of Sam68 gene in different cancer phenotypes b) what is the pattern of correlation among Sam68 gene isoform and other relative coexpressed isoforms. C) Moreover, investigating the downstream target transcripts of Sam68 could provide the clues of differential behavior in different cancer cells. Initially, isoform variants of Sam68 was obtained from UCSC. Sam68 can be spliced in three different splice isoforms namely uc001bua, uc001bub, and uc001buc (Figure 4.11). To examine the relative expression of these isoforms in different cancer datasets the transcript level expression data of TCGA were obtained from Broad institute and analyzed for differential expression in different cancer types. The results indicate, out of three isoforms, uc001bub has a higher mean expression level in all cancer. Moreover, the other two transcripts as uc001bua and uc001buc exhibit minimal basal expression in al cancer types. Thus, the expression of uc001bub alone contributes to the overall expression of Sam68 mRNA at all conditions. Additionally, uc001bub expression is significantly high in $Z > 1$ compared to the $Z < -1$ sample in all cancer (Figure 4.12). This suggests that higher expression of Sam68 is mainly contributed by uc001bub isoform.

***Figure 4.11**: Transcript structure of KHDRBS1 (Sam68 Gene)*

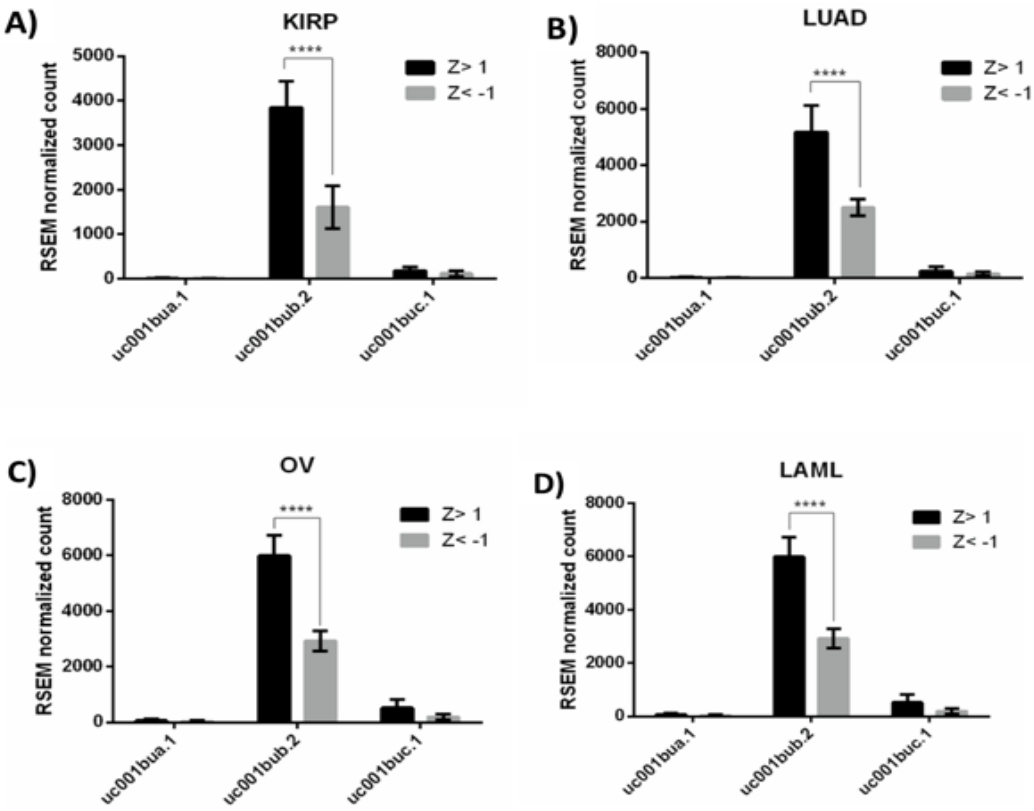


***Figure 4.12**: Bar diagram showing the relative expression of uc001bua, uc001bub, and uc001buc transcripts in KIRP, LUAD, OV, and LAML respectively. The error bar represents the standard deviation.*

Next, the Spearman correlation coefficient (*rs*) between uc001bub and all transcripts (73,599 transcripts) was calculated. The pattern of association of uc001bub transcript to all other correlated transcripts in all four cancers has no observable trend (Figure 4.13). Further, all highly correlated transcripts ($rs > 0.6$, $P < 0.05$) were analyzed for process and pathway enrichment using Metascape tools. In KIRP and LUAD, the coexpressed transcripts are mostly involved in cell division, and proliferation, which are highly interconnected (Figure 4.14 A, B). However, in LAML (Figure 4.14C), the prevailing pathway and processes are not directly linked to the cancer-specific events, and in OV process enrichment was found. The results of both gene and transcript level correlation analysis show that even though the Sam68 expression pattern is the same in KIRP, LUAD, OV, and LAML for a specific group of patients, its higher expression has different clinical outcomes due to the change in interaction partners and correlation network.

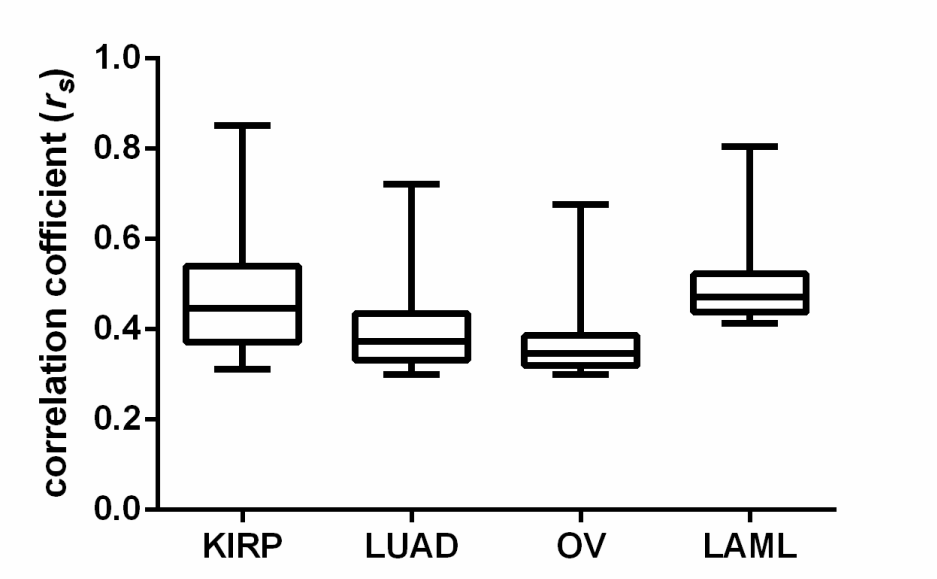


Figure 4.13*: Boxplot is summarizing the distribution of the correlation coefficient of uc001bub with all other transcripts (rs > 0.3, P < 0.05) in all four cancers.*

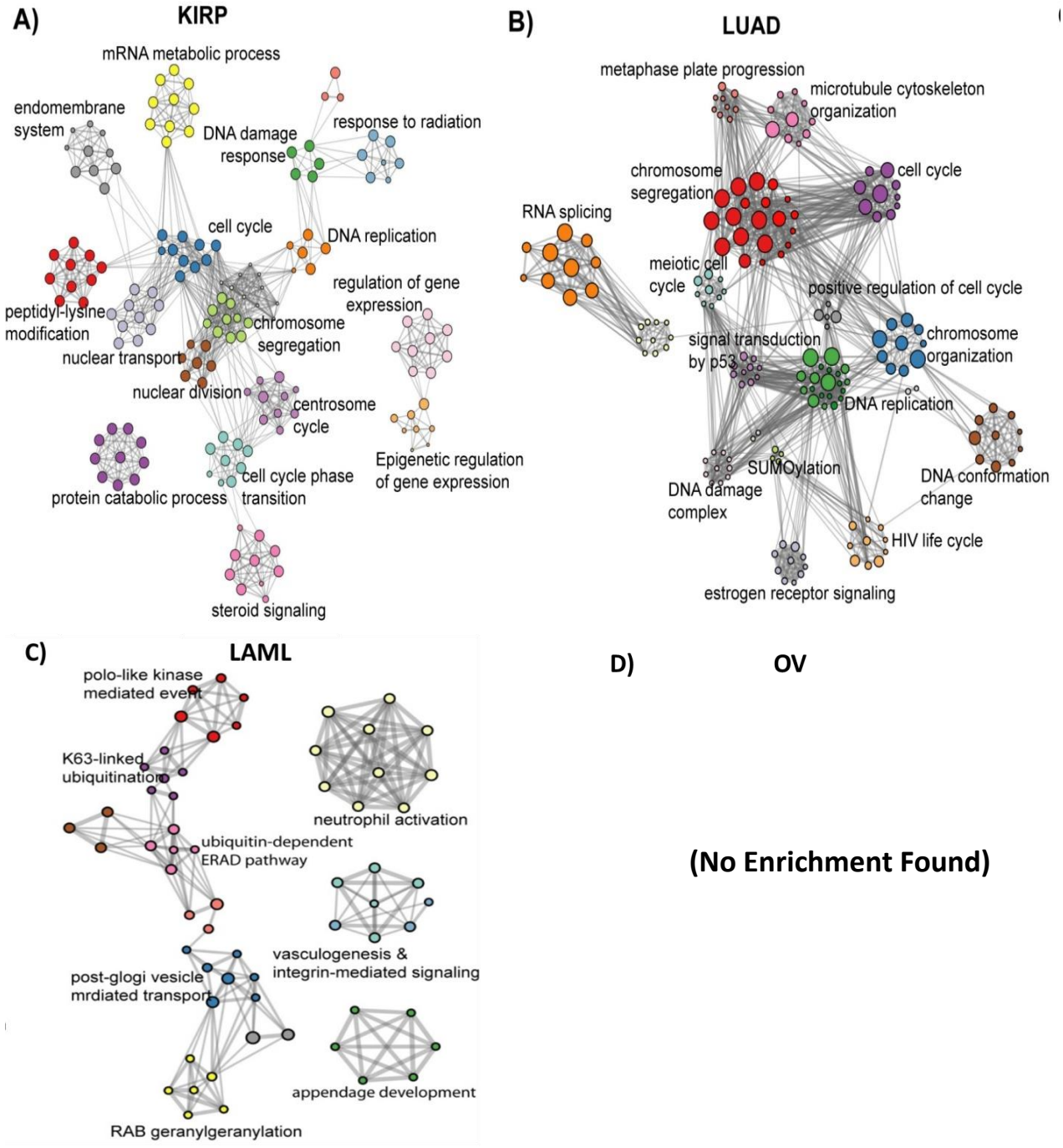


Figure 4.14*: Process and pathway enrichment analysis of the highly correlated transcript of Sam68. A) & B) shows over-represented biological processes, which regulate cell proliferation such as, cell cycle, chromosome segregation in KIRP and LUAD. C) Over-represented biological processes in LAML are not linked with the cancer-specific event. Due to the insufficient number of highly correlated transcripts (rs>0.6), no process and pathway enrichment are found in OV. The P-value < 0.01, minimum count 3 and enrichment factor > 1.5 are considered for enrichment analysis.*

Further, top 2000 transcripts with correlation coefficient (*rs*) > 0.3 and P < 0.05 were screened for each cancer type. However, many of these UCSC transcripts do not code for protein. Therefore, to identify the protein-coding transcript, the UCSC transcripts were matched to the RefSeq accession number of NCBI, and subsequently, coding transcripts were chosen for analysis. To find the target transcripts of Sam68, which are also co-expressed with uc001bub, the genome-wide binding region of Sam68 was obtained from RNA complete experiment by Ray *et al.*, 2013. The study shows that Sam68 can bind to a total of 268 sites in the human genome (human genome version hg19). From the co-ordinate of the binding region and using hg19 as the reference genome, a total of 1036 different transcripts were predicted that could be produced by Sam68 (Figure 4.15). Further, the coding and non-coding transcripts were screened out of 1036 transcripts, among which 562 are coding transcripts. Target transcripts (coding), which are present in the top 2000 correlated transcript data were screened and subjected to process and pathway enrichment analysis (Figure 4.16). Interestingly, like gene-level data analysis, coexpressed target transcripts of Sam68 are also involved in cancer-specific processes such as cell cycle, protein N-terminal acetylation, cell cycle phase transition, E2F6 transcription regulation in KIRP and LUAD (Figure 4.16 A, B) (Kalvik TV. & Arnesen 2013, Giangrande, P. H. *et al.*, 2004, Sherr CJ. 1996). However, in OV and LAML, cancer-linked biological processes are absent (Figure 4.16 C, D)

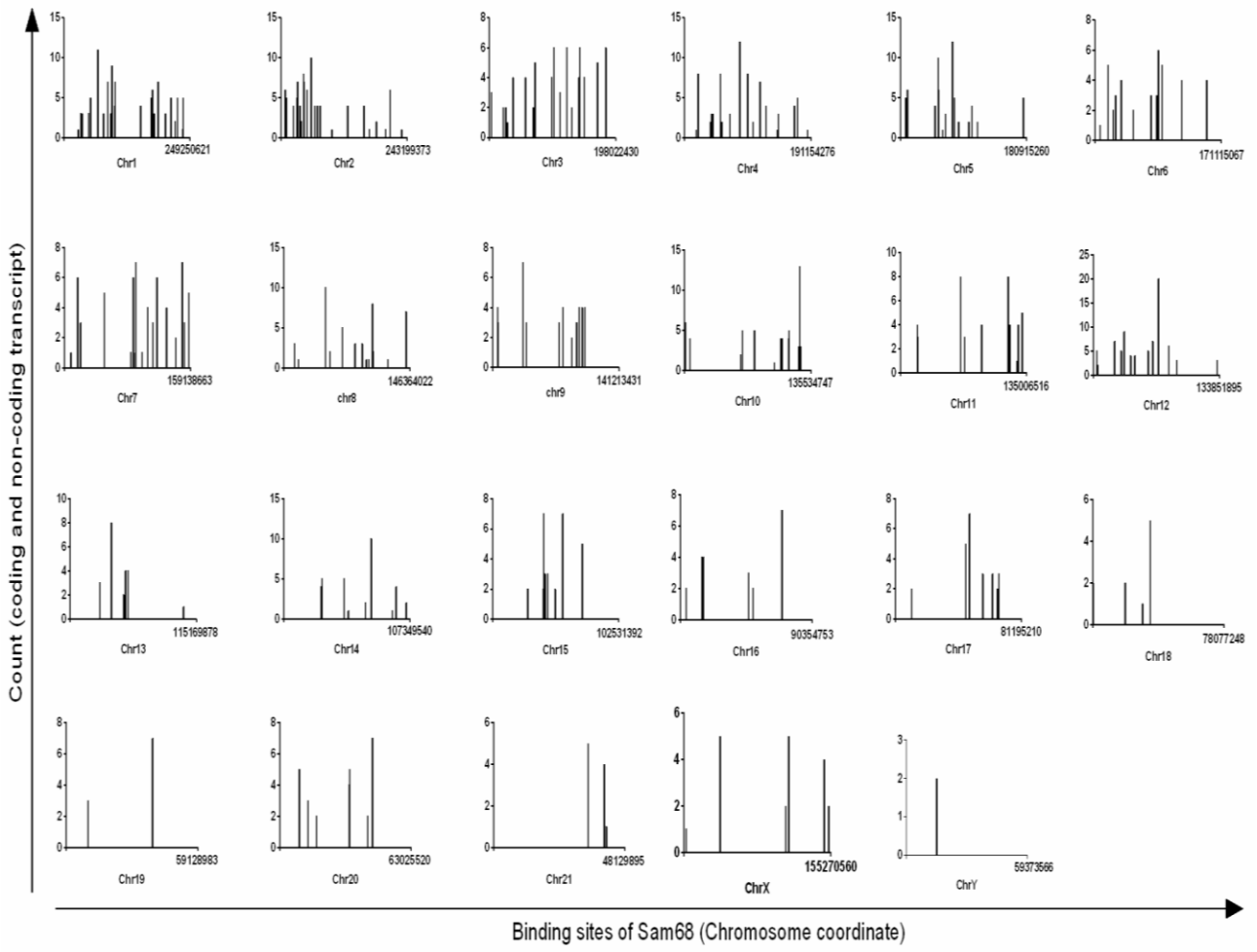


Figure 4.15*: Genome-wide binding region and count of predicted target transcripts of Sam68. Each histogram represents a chromosome and the x-axis represents chromosome coordinate. The binding sites of Sam68 are indicated by a bar on x-axis and length of the bar corresponds to the number of predicted transcripts*

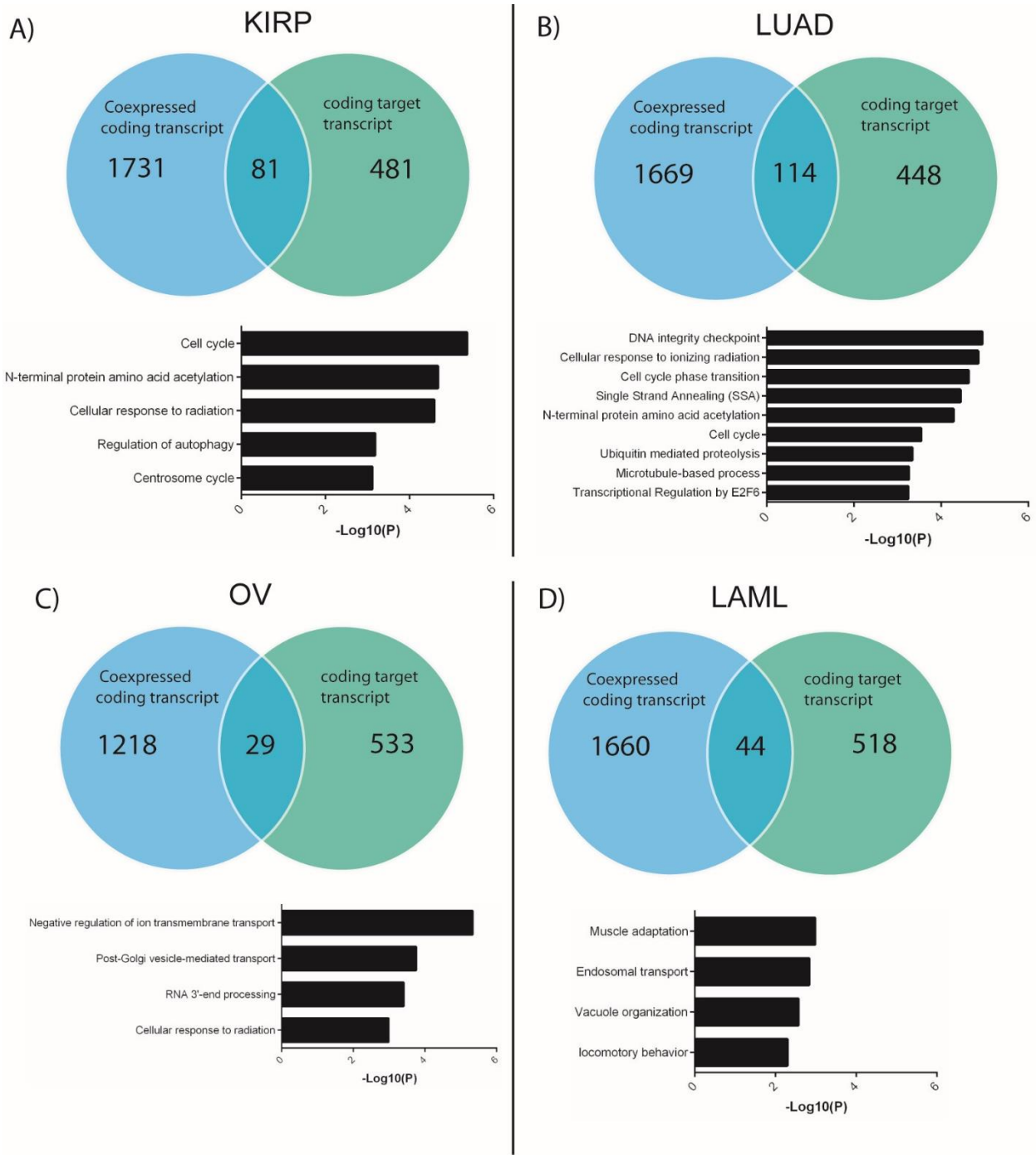


Figure 4.16*: Venn diagram representing overlapping coexpressed and target transcript of Sam68 in KIRP, LUAD, OV and LAML respectively. The bar diagram indicates the process and pathway enrichment analysis of overlapping genes in specific cancer. (Logarithmic corrected P-values for significant overrepresentation are shown).*

We explored the molecular mechanism of Sam68 to be a prognostic marker in four different cancers. Within specific cancer, including KIRP, LUAD, LAML, and OV, Sam68 expression is heterogeneous and patient-specific. In KIRP and LUAD, higher expression of Sam68 affects patient survival, but not in LAML and OV. Genome-wide coexpression analysis reveals genes and transcripts which are coexpressed with Sam68 in KIRP and LUAD, form the functional modules which are majorly involved in cancer-specific events. However, in the case of LAML and OV, such modules are absent. In-depth analysis of both gene and transcript level correlation analysis shows that even though the Sam68 expression pattern is same in KIRP, LUAD, OV, and LAML but in a specific group of patients of KIRP and LUAD, its higher expression has different clinical outcomes due to the change in interaction partners and correlation network. Our study shows a molecular network of Sam68 is patient-specific and varies across the cancer tissue. The essentiality of a gene in disease progression is determined by its interaction partners (Ashworth A. *et al.*, 2011). Similarly, our study shows that higher expression & clinical outcomes are not always a proportionally linked event, rather it depends on network architecture in a cell. Therefore, irrespective of the higher expression of Sam68, the significant divergence of its biological roles and prognostic value is due to its cancer-specific interaction partners and correlation networks. The result indicates a positive correlation between Sam68 expression status and cancer phenotype in KIRP and LUAD. Altogether, our study demonstrates the potential prognostic value of Sam68 in both LUAD and KIRP as it is involved in crucial molecular processes, which are specific to the cancer progression.

4.1.5 Sam68 is a Prognostic Marker in Lung Cancer:

Based on previous results, we validated Sam68 as a prognostic biomarker in lung cancer. We performed an expression analysis of Sam68 and its clinic-pathological significance using microarray datasets. Moreover, it is reasonable to assess the prognostic value of Sam68 using a dataset of different platforms (RNA-seq & Microarray), to evaluate the variability in our results. Two independent microarray datasets of non-small-cell lung carcinoma (NSCLC): GSE30219 and GSE31210 were chosen for further analysis. Initially, the expression profile of Sam68 mRNA was evaluated in a cohort of 519 Lung tumoral and 34 normal patient samples of GSE30219 and GSE31210 microarray datasets. In line with previous results, we observed that the expression of Sam68 in healthy cells is not sufficiently lower compared to the cancer cells (Figure 4.17). In

general, the genes which are expressed in early-stage and continue the higher expression in advance stages of cancer could be the best target for diagnosis and treatment. So, to answer the question of how Sam68 is linked with different stages of Lung Cancer patients. The cancer stage/grade-specific expression of Sam68 was evaluated. The cancer patients of GSE30219 were sub-grouped according to their stages (stage1 to stage4) and then compared the expression. Interestingly, the mean expression of Sam68 is monotonically and significantly increased as cancer progressed to advance stages (Figure 4.18a). In addition, NSCLC patients of GSE31210 were sub-grouped into two pathological stages according to TNM staging as early cancer (stage I-II) and advanced cancer (Stage III-IV). The results suggested poor prognosis of patients of all stages is linked with upregulated Sam68 (Figure 4.18b). Also, the increased levels of Sam68 substantially correlate with advanced and lymph node metastatic tumors (Stage III-IV). This significant relationship between Sam68 and NSCLC pathological stages highlights the prognostic importance of Sam68.

Next, we asked whether this aberrant expression pattern is correlated with the clinical outcomes of the patient. Like previous results, Kaplan Meier survival analysis and Log-rank test confirm elevated expression of Sam68 is associated with poor overall survival of patients (Figure 4.19 A). In addition, stage-specific survival analysis reconfirms that the overall survival of the patient significantly reduced as cancer progressed to a higher stage (Figure 4.19 B). Therefore, these results indicate higher expression of Sam68 is associated with pathogenic stages of lung and increases the lung cancer patient mortality. Our observation from these results supports Sam68 as a good prognostic and diagnostic marker in lung cancer.

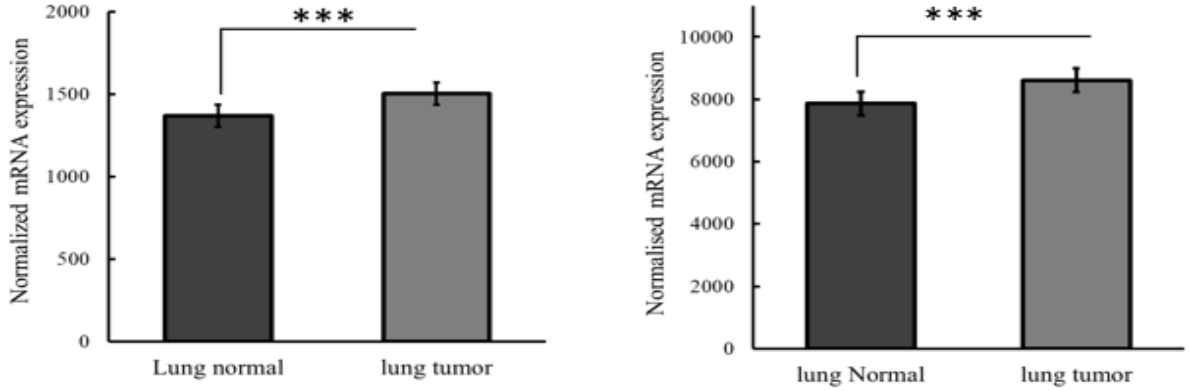


Figure 4.17: *Analysis of Sam68 mRNA expression. Elevated expression of Sam68 mRNA is visualized as the box plot comparing the expression between healthy and cancer patients of a) GSE30219 B) GSE31210.*

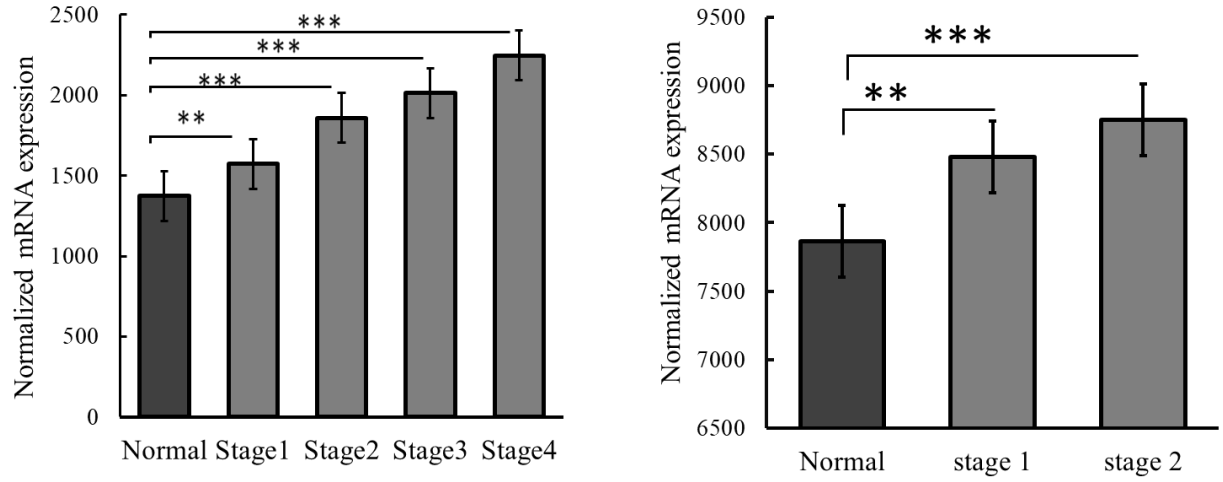


Figure 4.18: *Comparison of Sam68 expression between TNM stages & normal samples. Box plot showing expression difference between a) four tumor stages (stage1, 2, 3 &4) of lung cancer & normal samples of GSE30219 b) Early-stage (stage 1-2), Advanced stage (stage 3-4) &normal samples from GSE31210. The expression of Sam68 mRNA linearly increased as the disease progressed to advanced stages. Error bar represents SEM. Statistical significance is shown as * P<0.05, ** P< 0.01, *** P<0.001 by Rank sum test*

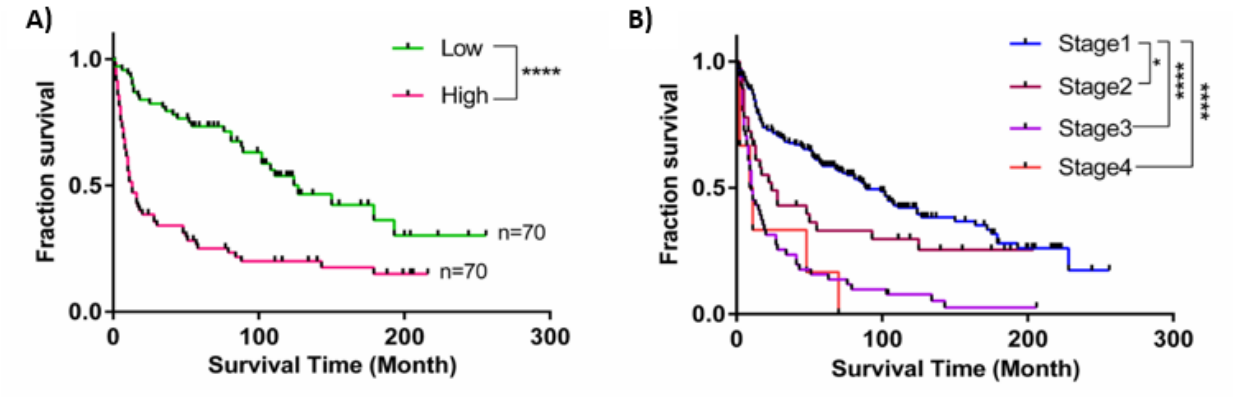


Figure 4.19: *Survival analysis a) Kaplan-Meier curve of survival plot shows the comparison of fraction survival in higher and lower expression group of Sam68. B) Kaplan-Meier curve shows the comparison of fraction survival in different stages of lung cancer. The significance level in each plot is represented as * P< 0.0001.*

Indeed, this analysis with a microarray dataset further signifies our finding, "Sam68 as a potential early prognostic biomarker in Lung cancer". However, early diagnosis and efficient treatment of lung cancer remains a challenge, as the disease remains latent in the early stage and is typically diagnosed at an advanced stage.

To meet this challenge

1. Identification of early diagnostic biomarkers is of urgent need and importance, which is a critical determinant for the prognosis of diseases and initiation of appropriate treatment.
2. A simple and efficient biomarker detection system is required, as low levels of the biomarker are expressed in early pathological stages.

Given these challenges and our approach which demonstrated Sam68 protein as an early tumor stage-specific diagnostic and prognostic biomarker in lung cancer, it is essential to develop an efficient detection system to quantify Sam68 protein. Thus, we focused to fabricate a sensitive antibody-based immunosensor to accurately measure the Sam68 protein level and which can be clinically applied for Lung Cancer patients.

Part 2 – Expression, Purification and Characterization of Sam68

4.2.1. Confirmation of the Sam68 Clone

Full-length mouse origin Sam68 cDNA cloned in the pGEX-2T plasmid was a gift from David Shalloway, Cornell University, NY (Add gene plasmid # 17687). To understand the similarity between human & mouse origin Sam68, the sequence of human (NP_035447.3) & mouse Sam68 (NP_006550.1) was compared using the NCBI protein blast. The results indicate 94 % identity (418/443) over 100% query coverage with 96% positives and 0% gaps (36) (Figure 4.20). This assures Sam68 of mouse origin replicates the Sam68 of human origin. With this initial confirmation, the clone was used for further analysis.

Full-length Sam68 contains 443 amino acids and the calculated molecular weight of this gene is 49KDa with 2713bp (NM_006559.3). However, this recombinant Sam68 is a GST (glutathione S-transfers) tagged protein, which accounts for an additional 25 KDa molecular weight. The vector construct containing Sam68 is shown in Figure 4.21. Agar stabs of E.coli Dh5alpha containing the plasmid were revived with suitable ampicillin resistance. From the overnight cultures, the plasmid was eluted and visualized on 0.8% of Agarose gel (Figure 4.22A). The presence of insert, Sam68, in the vector was confirmed by restriction digestion (Figure 4.22B) and PCR (Figure 4.22C). Single digestion with EcoR1 resulted in 7.5 kb band and double digestion with EcoR1 and BamH1 enzymes generated ~5kb, ~2.7kb bands, which indicate the sizes of the “vector with insert”, “vector without insert” and “insert” respectively. Further, to confirm the alignment, a fragment of the Sam68 gene was PCR amplified using primers with flanking sites of KH Domain. The amplified product in Figure 24c. Presents the 200kb region of KH Domain in the right frame.

Score	Expect	Method	Identities	Positives	Gaps
721 bits(1860)	0.0	Compositional matrix adjust.	418/443(94%)	428/443(96%)	0/443(0%)

```
Query  1    MQRRDDPASRLTRSSGRSCSKDPSGAHPSVRLTPSRPSPLPHRPRGGGGGPRGGARASPA  60
            MQRRDDPA+R++RSSGRS S DPSGAHPSVR TPSR  PLPHR RGGGGG RGGARASPA
Sbjct  1    MQRRDDPAARMSRSSGRSGSMDPSGAHPSVRQTPSRQPPLPHRSRGGGGGSRGGARASPA  60

Query  61   TQPPPLLPPSTPGPDATVVGSAPTPLLPPSATAAVKMEPENKYLPELMAEKDSLDPSFTH  120
            TQPPPLLPPS  GPDATV G APTPLLPPSATA+VKMEPENKYLPELMAEKDSLDPSFTH
Sbjct  61   TQPPPLLPPSATGPDATVGGPAPTPLLPPSATASVKMEPENKYLPELMAEKDSLDPSFTH  120

Query  121  AMQLLSVEIEKIQKGESKKDDEENYLDLFSHKNMKLKERVLIPVKQYPKFNFVGKILGPQ  180
            AMQLL+ EIEKIQKG+SKKDDEENYLDLFSHKNMKLKERVLIPVKQYPKFNFVGKILGPQ
Sbjct  121  AMQLLTAEIEKIQKGDSKKDDEENYLDLFSHKNMKLKERVLIPVKQYPKFNFVGKILGPQ  180

Query  181  GNTIKRLQEETGAKISVLGKGSMRDKAKEEELRKGGDPKYAHLNMDLHVFIEVFGPPCEA  240
            GNTIKRLQEETGAKISVLGKGSMRDKAKEEELRKGGDPKYAHLNMDLHVFIEVFGPPCEA
Sbjct  181  GNTIKRLQEETGAKISVLGKGSMRDKAKEEELRKGGDPKYAHLNMDLHVFIEVFGPPCEA  240

Query  241  YALMAHAMEEVKKFLVPDMMDDICQEQFLELSYLNGVPEPSRGRGVSVRGRGAAPPPPPV  300
            YALMAHAMEEVKKFLVPDMMDDICQEQFLELSYLNGVPEPSRGRGV VRGRGAAPPPPPV
Sbjct  241  YALMAHAMEEVKKFLVPDMMDDICQEQFLELSYLNGVPEPSRGRGVPVRGRGAAPPPPPV  300

Query  301  PRGRGVGPPRGALVRGTPVRGSITRGATVTRGVPPPPTVRGAPTPRARTAGIQRIPLPPT  360
            PRGRGVGPPRGALVRGTPVRG+ITRGATVTRGVPPPPTVRGAP PRARTAGIQRIPLPP
Sbjct  301  PRGRGVGPPRGALVRGTPVRGAITRGATVTRGVPPPPTVRGAPAPRARTAGIQRIPLPPP  360

Query  361  PAPETYEDYGYDDTYAEQSYEGYEGYYSQSQGESEYYDYGHGELQDSYEAYGQDDWNGTR  420
            PAPETYE+YGYDDTYAEQSYEGYEGYYSQSQG+SEYYDYGHGE+QDSYEAYGQDDWNGTR
Sbjct  361  PAPETYEEYGYDDTYAEQSYEGYEGYYSQSQGDSEYYDYGHGEVQDSYEAYGQDDWNGTR  420

Query  421  PSLKAPPARPVKGAYREHPYGRY  443
            PSLKAPPARPVKGAYREHPYGRY
Sbjct  421  PSLKAPPARPVKGAYREHPYGRY  443
```

***Figure 4.20**: Blast alignment of Human (NP_035447.3) & mouse Sam68 (NP_006550.1) sequence.*

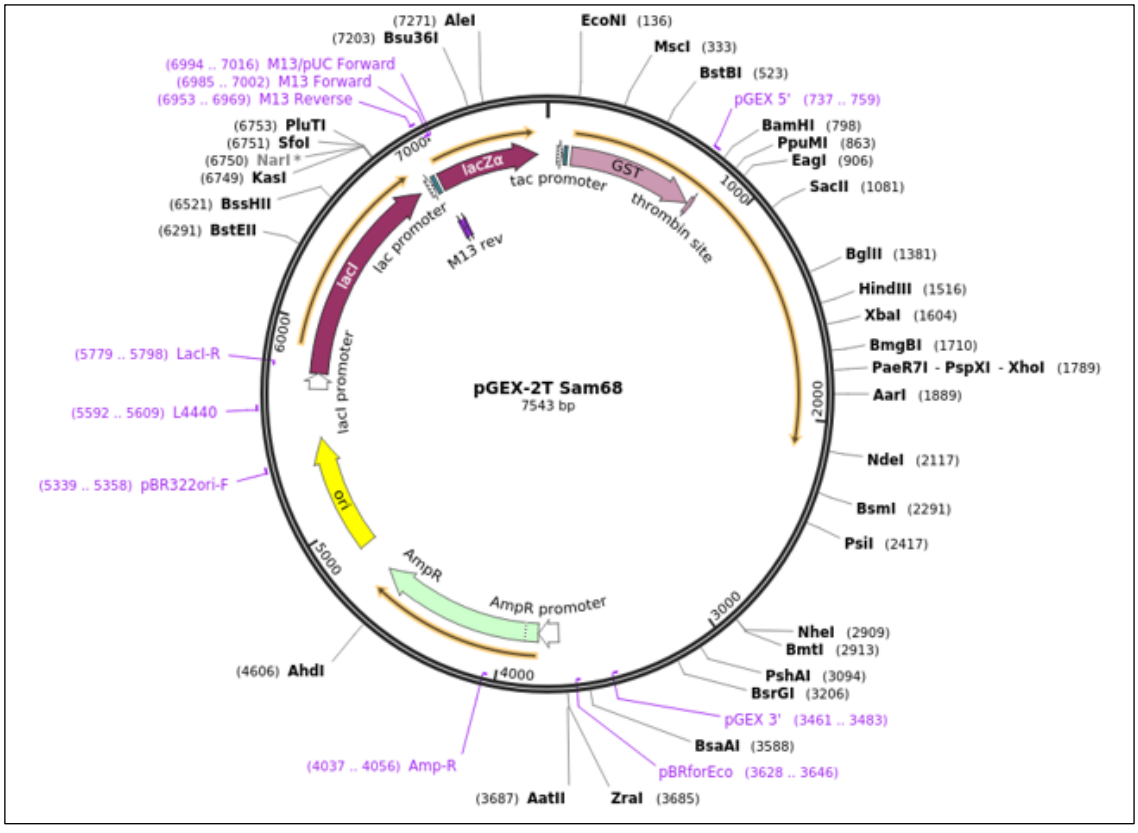


***Figure 4.21**: Full sequence map for pGEX-2T-Sam68 adapted from Adgene (plasmid no: 17687)*

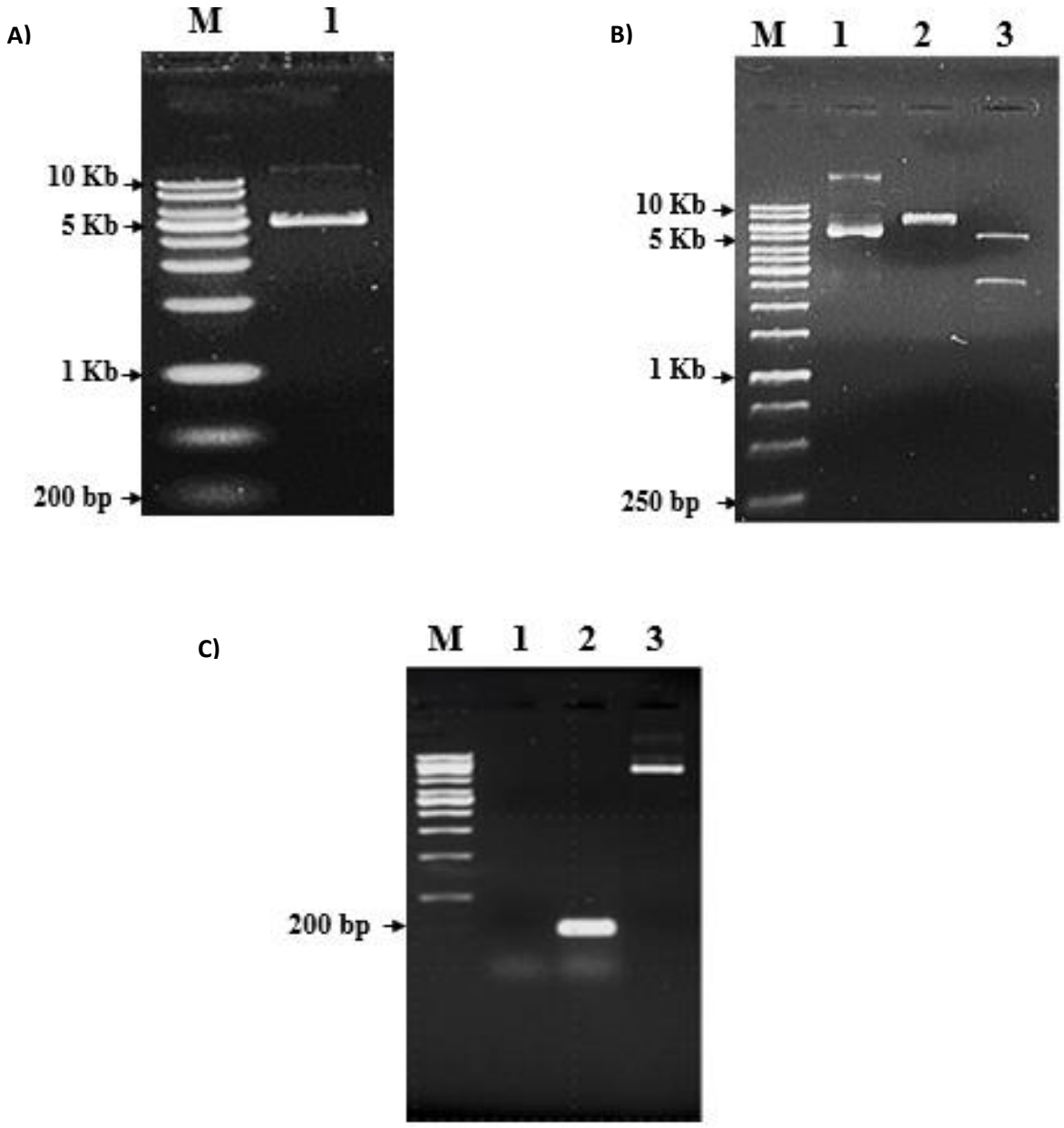


Figure 4.22*: Purification and Characterization. A) Isolation of pGEX-2T-Sam68 plasmid. Lane M: 1 Kbp DNA ladder & lane 2: Isolated plasmid B) confirmation of the clone by restriction digestion analysis. Lane M: 1Kbp DNA ladder, Lane 2: undigested plasmid, Lane 3: pGEX-2T-Sam68 digested with EcoR1 (7.5 Kbp band), Lane 4: pGEX-2T-Sam68 digested with EcoR1 and BamH1 (5kb, 2.7kb) C) pGEX-2T-Sam68 was amplified by PCR. Lane M: 1Kbp DNA ladder, Lane 1: negative, Lane 2: 200bp amplified KH domain of Sam68, Lane 3: pGEX-2T-Sam68 plasmid as a control. All the samples were resolved on 0.8% or 1% agarose gel.*

4.2.2 Cloning, Expression, Purification and Characterization of Recombinant Sam68

E.coli Bl21 DE3 was used to express the Sam68 protein. BL21 Competent cells were generated using the Cacl2 method. The recombinant plasmid, pGEX-2T Sam68 was cloned into competent cells of Bl21 by heat shock. The cloned cells were then selected by addition ampicillin antibiotic in the culture media. Transformation efficiency was high, and the cloned colonies were screened by colony PCR (Figure 4.23). Further, the selected clone (C3, Figure 4.23) was used for the expression of the recombinant protein. Overnight culture of a single colony of the selected clone was subcultured in 2XTY media (supplemented with 1% glucose and ampicillin) and incubated at 37°C with 180rpm shaking. Once the growth reached an absorbance of 0.6- 0.7, the sample was induced with 0.5 mm IPTG for 3 h at 30°C. Subsequently, the cells were pelleted, resuspended in lysis buffer (Phosphate buffered saline pH 7.4 containing 0.1% Triton X-100, PMSF) and then sonicated Aliquot of both induced and unindicted samples were characterized by SDS Page (Figure 4.24). In the induced sample, a ~ 70 KDa band was visualized, which corresponds to the Sam68 protein. Next, this Sam68 fused with GST at N-terminal (GST-Sam68) was successfully purified from clarified cell lysate using Sepharose 4B -Glutathione beads (Sigma) as per the manufacturer's protocol. The recombinant protein was eluted in freshly prepared 20 mM reduced GSH in 50 mM Tris-HCl, pH 8.0. The purified protein was characterized using SDS Page and further confirmed by western blot using specific anti-Sam68 antibody followed by anti-rabbit HRP conjugated antibody and BSA protein used as a marker (Figure 4.25). SDS Page analysis shows Sam68 protein was successfully purified whereas, in western blot, a ~ 70KDa band above the BSA marker confirms the purified product as Sam68 protein. To determine the purified protein concentration, Bradford assay was performed using BSA as a standard. The estimated concentration of purified Sam68 was 672 μg/mL.

Although initially for Induction, we used LB media and maintained growth at 37’c, this resulted in poor induction, and also found protein cleavage. To resolve the problem, we shifted to rich media as 2xty, reduced both incubation temperature to 30’c and OD at 600 to 0.6. This increased the induction and made protein more stable after purification. Further, the addition of glucose to media also helped to attain higher cell densities. Moreover, both the strains of E.coli, Bl21-DE3 and DH5alpha, were used for induction, amongst which the earlier showed better expression of the desired protein. The induction time was standardized by incubating the culture with 0.5 mm

IPTG concentration at 30'c from 3hr to overnight. The expression was checked at every 3hrs interval, as a result, the expression didn't improve with time.

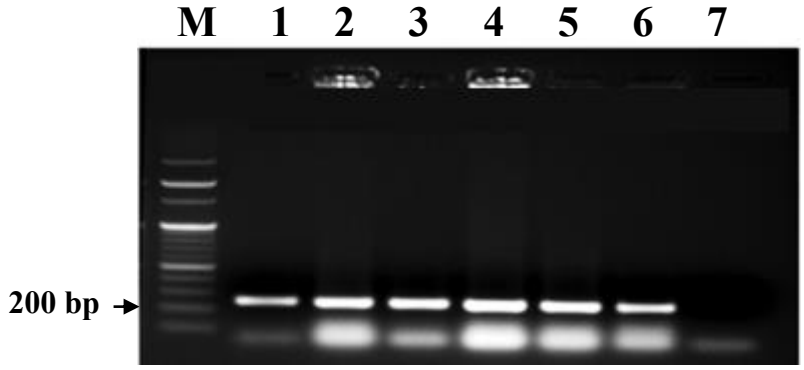


***Figure 4.23**: Colony PCR was performed to screen the transformed clones. Amplification was done using KH domain-specific primers. Lane M: 100bp DNA ladder, Lane 1: positive control, Lane 2-6: 5 different clones of BL21 containing pGEX-2T Sam68 (C1, C2, C3, C4, C5), Lane 7: Negative control.*

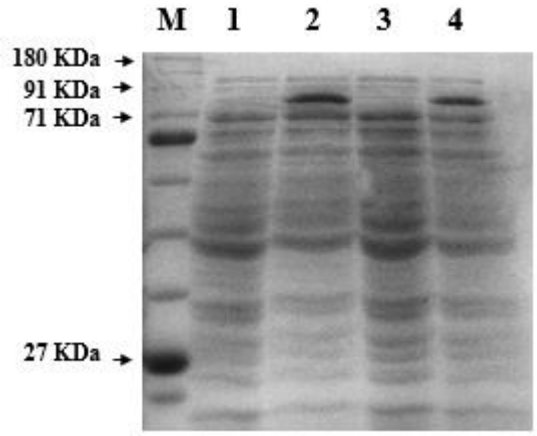


***Figure 4.24**: SDS-Page to detect the expression of Sam68 in E.coli BL21 (DE3). The C3 & C4 clones were induced with 0.5 mm IPTG for 3 h at 30°C. The clarified protein lysate of both induced and unindicted was resolved in 12% denaturing gel. Sam68 protein was expressed in induced samples of both C3 & C4 clones. Lane M: Protein Marker, Lane 1:C3 unindicted, Lane2: C3 induced, Lane 3:C4 unindicted, Lane 4: C4 induced samples.*

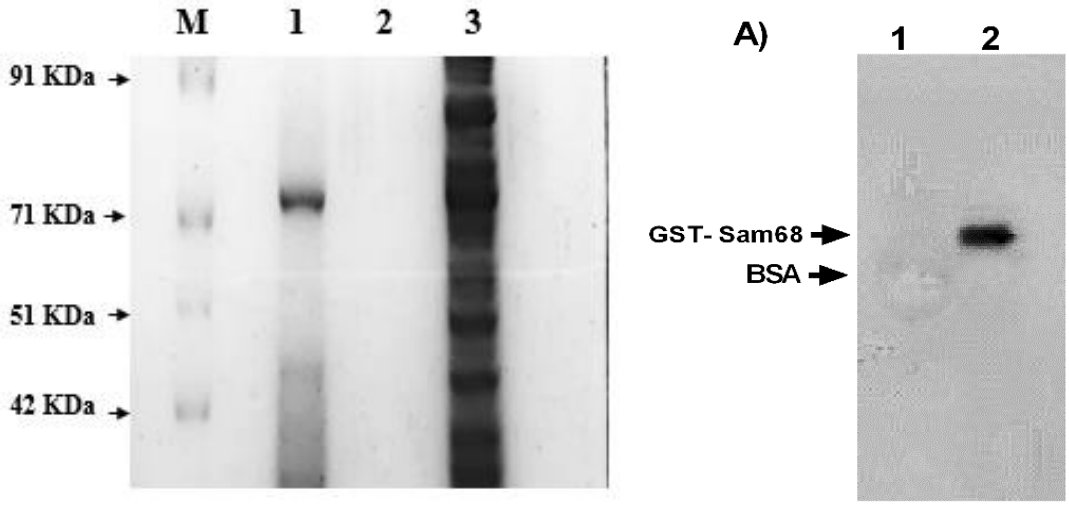


Figure 4.25*: Characterization of Purified Sam68 protein.* a*) SDS Page shows a purified protein product above 71 KDa band. Lane M: 180KDa Protein marker, Lane 1: Eluted product (20mM Glutathione), Lane 2: Wash, Lane 3: Flow-through. Proteins were resolved on 12% denaturing polyacrylamide gel were used and staining as done by Coomasie brilliant blue. b) Western blot confirms the purified protein as Sam68. The protein was detected using a polyclonal rabbit anti-Sam68 antibody followed by anti-rabbit HRP conjugated antibody.*

4.2.3 Sam68 Expression in Lung Cancer Cell Line

NCI-H23 was used as a lung cancer model for the real-time application of fabricated immunosensor. NCI-H23 cells were maintained in DMEM with 10% fetal calf serum and 1× antimycotic-antibiotic subculturing was done using the trypsinization method and the number of cells was counted by a hemocytometer. Once 80-90% confluence was reached, cell lysis and total protein were extracted using chilled RIPA buffer, which contains nonionic and ionic detergents for efficient cell lysis and protein solubilization. To avoid protein degradation, Protease inhibitor, 1 mm PMSF was added to the lysis buffer. Further, the cell suspension was sonicated and centrifuged for the clarified supernatant which contains the total cell protein. Followed by separation of the total proteins by SDS-PAGE and then transferred to a nitrocellulose membrane for Western blotting. The analysis specifically detected the endogenously expressed Sam68 in NCI-H23 cell lysate (Figure 4.26). However, in contrast to purified protein, the mammalian Sam68 migrated below the BSA standard, this due to Post-translational modifications.

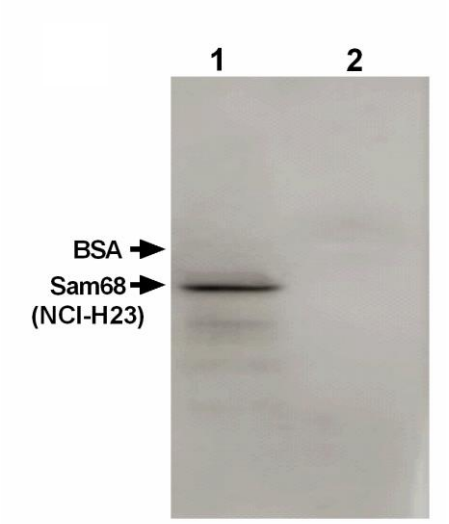


Figure 4.26*: Detection of Sam68 protein in the NCI-H23 cell line. Western blotting analysis confirms the endogenous expression of Sam68 in NCI-H23. The protein was detected using an anti-Sam68 antibody followed by anti-Rabbit IgG HRP conjugate.*

Part 3 - Fabrication of Biosensor to detect Sam68

4.3.1 ELISA

As Sam68 is having a high prognostic value in lung cancer, it is essential to have an efficient detection system for Sam68 which can be clinically applied. We first explored conventional Indirect ELISA technique to detect Sam68 protein using an anti-Sam68 antibody (1:1000 dilutions) as primary antibody and anti-rabbit HRP conjugated antibody as a secondary antibody. All the measurements were performed in triplicates with Sam68 in the concentration range of 0.001-10 μg/mL. The results as shown in Figure 4.27A demonstrate that the ELISA was not sensitive at lower concentrations (0.001 -0.1μg/mL) as the signal was a week and failed to differentiate concentrations at these levels. But then the signal increased linearly from 1- 10 μg/mL of Sam68 protein. This validates a good binding of purified Sam68 with the anti-Sam68 antibody. A calibration curve was plotted from the detected values as presented in Figure 4.27B The limit of detection (LOD) and limit of quantification (LOQ) were calculated from the expression LOD = (3×SD)/m and LOQ = (10×SD)/m (where SD is the estimated standard deviation from the points used to construct the calibration curve and m, its slope) (Lu *et al.*, 2016, Srivastava and Gupta, 2011). LOD and LOQ of ELISA were determined as 1.381119 μg/mL and 4.603035 μg/mL respectively. Next, Indirect ELISA was also carried out with different dilutions (1: 100 dilution, 1: 10 dilution and crude extract) of NCI-H23 whole cell lysate containing total cell proteins. The concentration of NCI-H23 Sam68 was estimated using the calibration curve plotted earlier. As the concentration of Sam68 in different dilutions of NCI-H23 cell extract is very low, Again ELISA method was unable to detect and differentiate between the lower amounts of protein in different dilutions as noted in Figure 4.29c Thus confirms ELISA was not sensitive enough to quantify the Sam68 concentration specifically at lower levels. Further induced us to switch to develop a more reliable detection system as electrochemical immunosensor that is sufficiently sensitive to address the clinical needs.

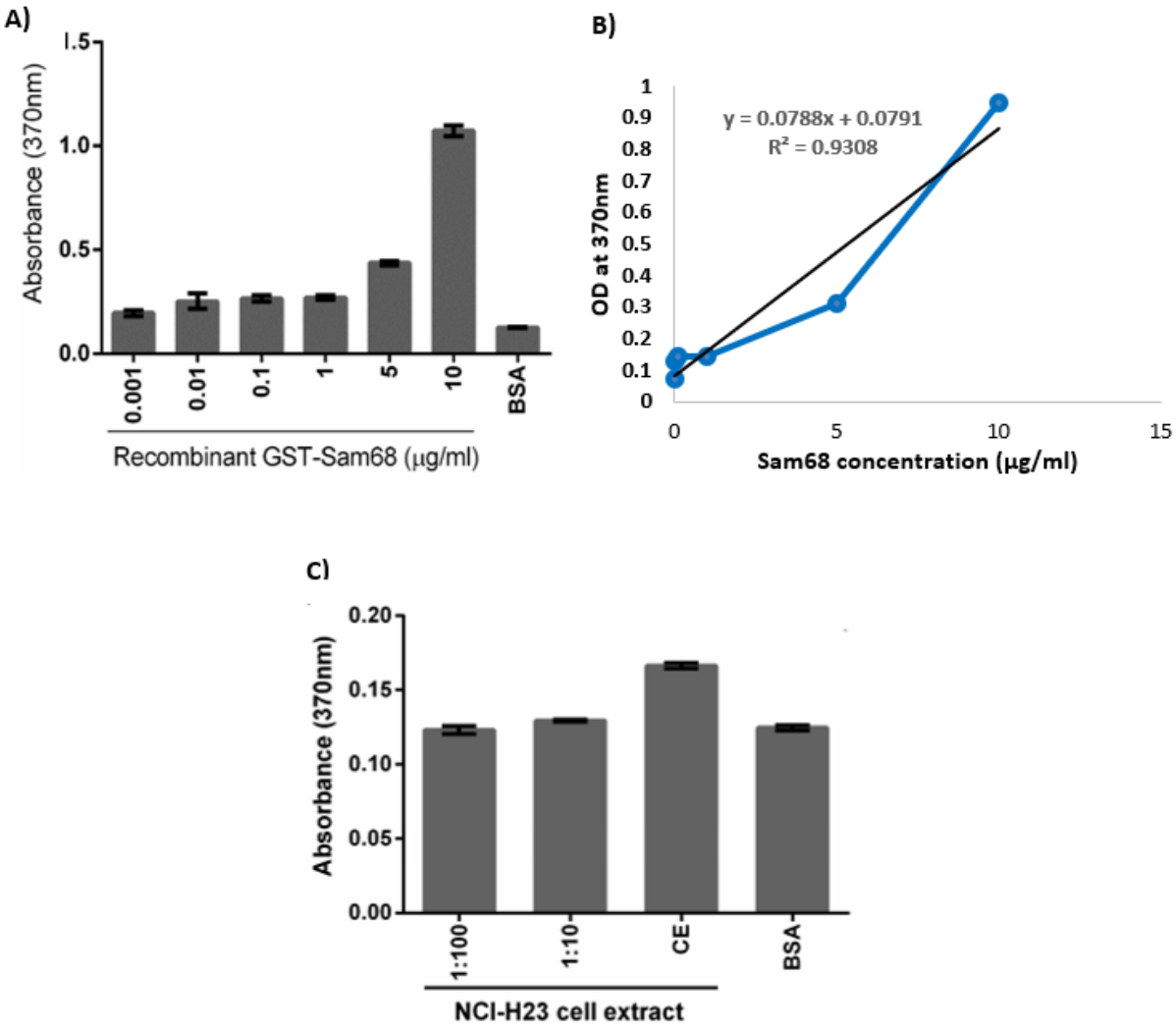


***Figure 4.27**: Indirect ELISA a) Detection of Purified Sam68 at different concentrations (0.001, 0.01,0.1,0,5,10 µg/mL) with mean absorption value at 370 nm and BSA as a control Absorbance increased Linearly from 1 to 10 µg/mL concentration of protein. But, at lower concentrations (0.001 to 0.1 µg/mL), Elisa failed to quantify the low levels of concentration accurately. b) Standard curve of Sam68 C) Detection of NCI-H23 Sam68 at different dilutions (1:100, 1:10, crude extract (CE)), with BSA as a control. All experiments were done in triplicates and error bars represent the standard deviations.*

4.3.2 Characterization of Immunosensor

An overview of the fabrication process of the proposed immunosensor is shown in Figure 4.28. To characterize the step by step fabrication of the immunosensor both CV and EIS were performed to study the change in the electrical interface. As shown in Figure 2.29A, the CV measurements were performed in 0.1 M PBS (pH 7.4) with 5.0 mm *[$K_3Fe(CN)^6$ / $K_4Fe(CN)^6$]*

and 0.1 M KCl at the potential range from 0.2 to 0.6 V with Scan rate: 100 mV s-1. The bare GCE showed a typical quasi-reversible electrochemical response, indicating good electron transfer at the surface (curve A). We have used PANI as immobilization support because of its excellent conductivity, stability and biocompatibility (Dhand *et al.,* 2011, Fang *et al.,* 2017). The modification of the GCE with highly conductive PANI increased electron transfer kinetics at the electrode interface evident from the drastic increase in the oxidation peak current of the redox probe (curve b: GCE/PANI). The steepness of the peak then decreased when PANI was cross-linked with glutaraldehyde (curve c: GCE/PANI/GLU). Due to the nonconductive nature of the protein, a decrease in the peak current was observed after the immobilization of the anti-Sam68 antibody on the electrode surface (Curve d: GCE/PANI/GLU/Sam68-Ab). For the similar property of BSA, the peak current was furthermore decreased when the electrode was blocked with BSA (curve e: GCE/PANI/GLU/Sam68-Ab/BSA). Moreover, the change in the dielectric properties at the electrode/electrolyte interface at different stages of immunosensor fabrication was further examined with Electrical impedance spectroscopy (EIS). The EIS measurements were done in 0.1 M PBS (pH 7.4) with 5.0 mm [$K_3Fe(CN)^6$ / $K_4Fe(CN)^6$] at a frequency range between 10,000 and 0.05 Hz with an alternating wave of 10 mV amplitude. The Nyquist plots in Figure 4.29B represent the quantitatively measured electron transfer resistance (Rct) which was similar to the semicircle diameter in the impedance spectra. The insert in Figure 2.29B represents the enlarged spectra. The randle's equivalent circuit obtained using NOVA software is given in appendix. The bare GCE showed a diffusional limiting electrochemical process represented by a small semicircle with a straight line in the impedance spectra. As the highly conducting PANI facilitates the electron transfer process, expectedly, a very low Rct was obtained on PANI modified GCE. The crosslinking of PANI with glutaraldehyde slightly increased the Rct value. When the anti-Sam68 antibody was immobilized on the electrode, there was a significant increase in the Rct value and the Rct value further increased when the antibody modified electrode surface was blocked with BSA. The increase in the Rct value is due to the insulating effect of protein molecules on the electron transfer process.

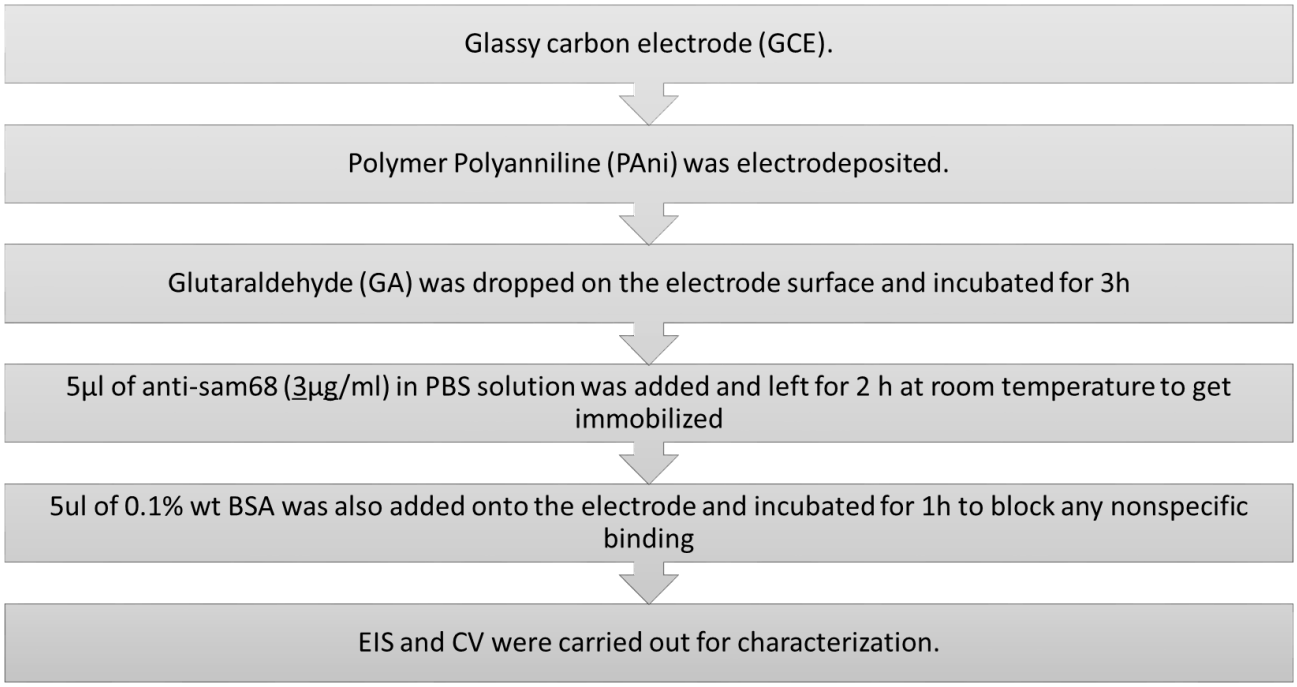


***Figure 4.28**: Schematic illustration of the Fabrication process of electrochemical immunosensor.*

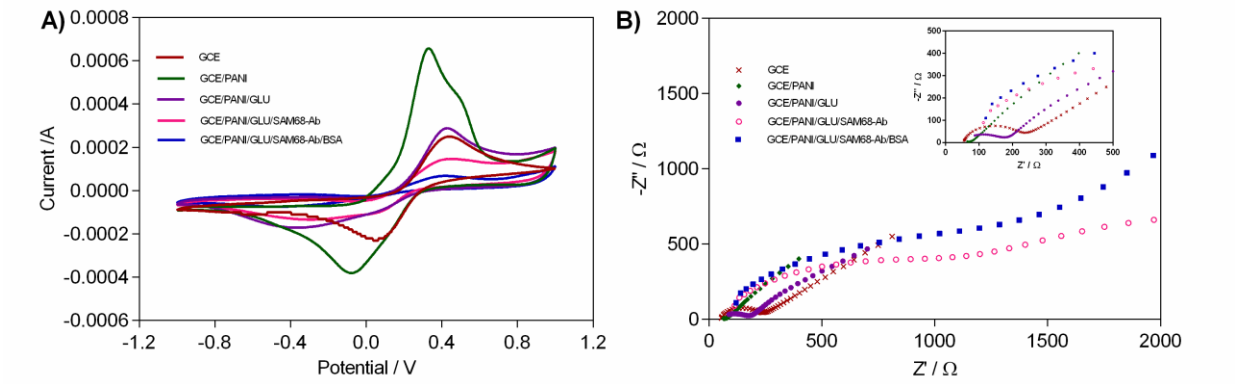


***Figure 4.29**: Stepwise characterization of the immunosensor assembly process. A) Cyclic voltammograms and B) Electrochemical impedance spectra (Insert: enlarged spectra) of a) GCE b)GCE/PANI c)GCE/PANI/GLU d)GCE/PANI/GLU/Sam68-Ab e)GCE/PANI/GLU/Sam68-Ab/BSA. Measuring solution: 0.1 M PBS (pH 7.4) with 5.0 mm [$K_3Fe(CN)^6$ / $K_4Fe(CN)^6$] and 0.1 M KCl. CV performed at Scan rate: 100 mV s-1. EIS was done with an alternating wave of 10 mV amplitude in the frequency range between 10,000 and 0.05 Hz*

4.3.3 Analytical Detection of Recombinant Sam68 Protein

Electrochemical detection of Sam68 was carried out by incubating the fabricated immunosensor (GCE/PANI/GLU/Sam68-Ab/BSA) with different concentrations of purified Sam68 protein ranging from 0.001 - 5 µg/mL for 1 hr. Generally, the affinity of antigen- antibody interaction is affected by pH, temperature and the ionic strength of the solvent. So throughout the study, we used manufacturer's (sigma) recommended conditions for anti-Sam68 antibody interaction that is pH 7.4 and room temperature (~ 25 - 30 °C). Electrochemical measurements, EIS and CV were performed in triplicates (n=3) at each concentration. The resulting cyclic voltammograms and impedance spectra are shown in Figure 2.30. Between each analysis, the immunosensor surface was regenerated by incubation in a solution containing 6 mm NaOH and 0.6% ethanol for 5 min and washing with PBS three times.

RCT refers to the opposition offered by biological samples to the flow of electrical current in a particular frequency spectrum. Analysis of EIS data demonstrated that with an increase in the relative concentration of Sam68, the impedance/opposition to the flow of current (*R*ct values) gradually increased (Table 4.1). Due to the recognition and formation of immunocomplexes (anti-Sam68 antibody - Sam68 protein) by target protein on the surface of the electrode, the amount of non-electroactive proteins at the interface increased. Thus, the resistance to charge transfer (*R*ct) increased between the solution containing active-redox species ($[Fe(CN)_6]^{3-/4-}$) and the electrode surface (Figure 3.30B). However, at higher concentrations, the antibody molecules were saturated and showed a less steep peak. A standard Sam68 response curve was generated with log-transformed concentration values (x-axis) and obtained Rct values (y-axis) (Figure 3.30C). The plot exhibited good linearity between Sam68 concentration and impedance with linear equation y=660.69x + 966.62 and measure of correlation, r^2 is 0.9656. The sensor response was also characterized using another technique, CV, which measures the change in current intensity at different concentrations of Sam68. The redox current peaks gradually decreased as the Sam68 protein concentration increased (Figure 3.30A). This indicates CV results are in agreement with the impedance results where impedance increases and correlated current decreases in Sam68 concentration-dependent manner. Next, to understand the performance of the immunosensor, As used in ELISA, both the LOD and the LOQ were calculated from the same expression (LOD = (3 × SD)/*m* and LOQ = (10 × SD)/*m* (where SD is the estimated standard deviation from the points used to

construct the calibration curve and *m*, its slope. (Lu *et al.*, 2016, Srivastava and Gupta 2011). The immunosensor gave a LOD of 10.54549 pg/mL and a LOQ of 35.15165 pg/mL which are much lower compared to ELISA. Table 4.2 compares the LOD and LOQ of these two detection methods for Sam68. The immunosensor has overcome the limitations of ELISA by being sensitive and efficient in determining the concentrations of Sam68 in the Pico molar range that is required for early diagnosis of cancer. Moreover, the sensitivity of the GCE/PANI/GLU/Sam68-Ab/BSA immunosensor was compared with other electrochemical GCE modified biosensors used for the detection of Lung cancer biomarkers (Table 4.3). Therefore, this immunosensor quantifies the target protein with excellent selectivity and specificity at a lower and broad range of concentrations.

Table 4.1: Showing the mean Rct values obtained at different concentrations of Sam68 protein. RSD ranges between 0.8- 2.2% which validates the values obtained.

The concentration of Recombinant Sam68	**Average Rct (ohms) (resistance to charge transfer)**	**RSD %**
0.001 µg/mL	1029.666667	2.024021
0.01 µg/mL	1524.666667	1.279108
0.1 µg/mL	2336.333333	2.29554
1 µg/mL	2839.666667	1.430601
5 µg/mL	3599	0.891304

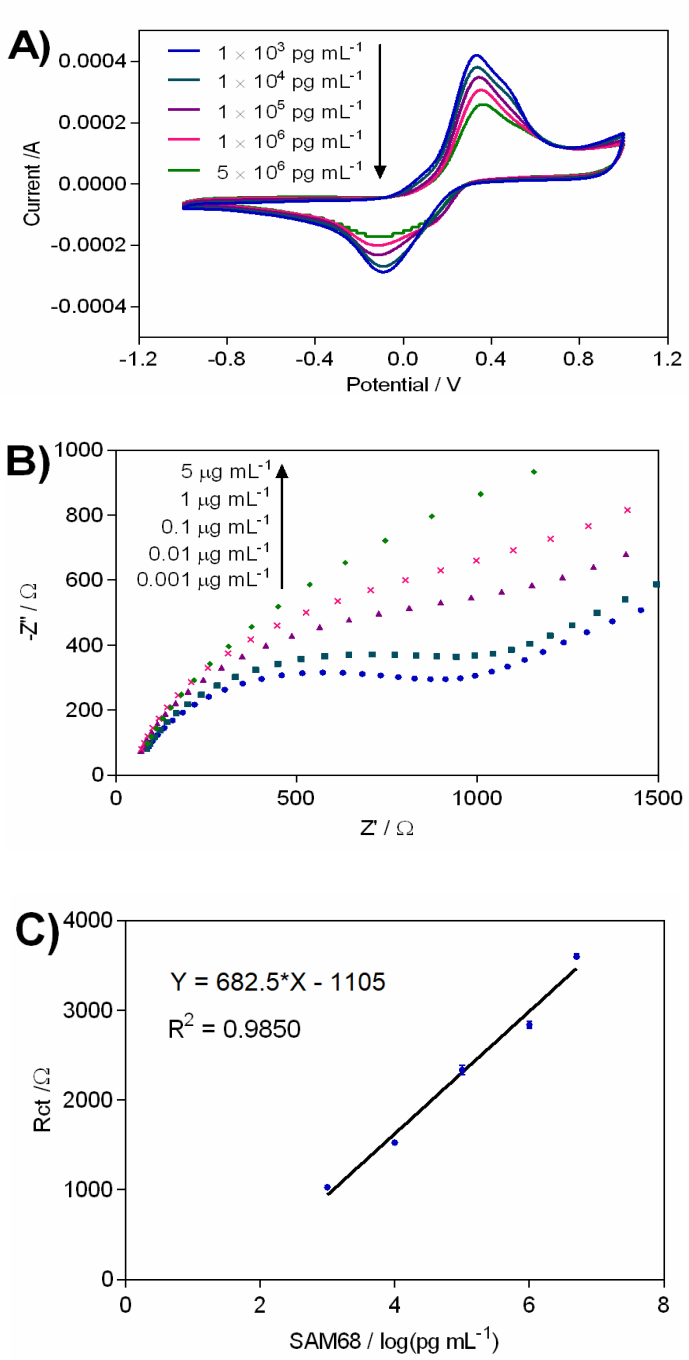


***Figure 4.30**: CV and EIS Response of the fabricated immunosensor (GCE/PANI/GLU/Sam68-Ab/BSA) on incubation with increasing concentrations of purified Sam68 protein as (10^{-3}, 10^{-2}, 10^{-1}, 1, 5 µg/ml). A) Cyclic voltammograms and B) Electrochemical impedance spectra. Measuring solution: 0.1 M PBS (pH 7.4) with 5.0 mm [$K_3Fe(CN)^6$ / $K_4Fe(CN)^6$] and 0.1 M KCl. C) Calibration plot of the immunosensor with Sam68. Each datum point represents the average of triplicate values (n=3). Error bars correspond to the standard deviation (SD) of three measurements (n=3) at each concentration.*

Table 4.2**:** compares the LOD and LOQ of these two detection methods for Sam68

	ELISA	**GCE modified Sensor**
Limit of detection	1.381119 µg/mL	**10.54549 pg/mL**
Limit of Quantification	4.603035 µg/mL	**35.15165 pg/mL**

Table 4.3: Comparison of the present GCE/PANI/GLU/Sam68-Ab/BSA immunosensor with other GCE modified biosensors used for the detection of Lung cancer biomarkers.

Biomarker	**Biosensor components**	**Linear range**	**Detection limit**
Sam68	Polyaniline/ Glutaraldehyde	1 ng/mL to 5 µg/mL	10.545 pg/mL
Cytokeratin 19 fragment 21-1 (CYFRA21-1)	Three-dimensional graphene (3D-G)/chitosan (CS) /glutaraldehyde (GA) composite	0.1 to 150 ng/mL	43 pg/mL
carcinoembryonic antigen (CEA)	Gold nanoparticle–graphene composites (Au–GN)	0.10 to 80 ng/mL	40 pg/mL
carcinoembryonic antigen (CEA)	Au nanoparticles (AuNPs)/multi-walled carbon nanotubes (MWCNTs)/chitosans composite	0.3 to 20 ng/mL	0.01 ng/mL
carcinoembryonic antigen (CEA)	AU-TiO2 nanoparticles	0.02 to 120 ng/mL	12 pg/mL
cytokeratin-19 (CK-19)	chitosan stabilized AuNPs/Aminosalicylic acid-based resin (AAR) microspheres/thionine(thi)	0.05 to 80 ng/mL	40 pg/mL
neuron-specific enolase (NSE)	Au nanoparticle/mesoporous silica nanoparticles (Au-MSNs)	0.1 to 2000 ng/mL	50 pg/mL

4.3.4 Determination of Sam68 Protein in NCI-H23 Whole Cell lysate.

Next, to evaluate the diagnostic application in real sample analysis, the developed immunosensor was employed to detect Sam68 in lung cancer whole-cell extract. Initially, the expression of the Sam68 protein in NCI-H23 was confirmed by Western blot (Figure 4.26). The results show the endogenous expression of Sam68 protein in NCI-H23 cell extract. To detect the target protein using fabricated immunosensor, the working electrode was incubated in three different dilutions of NCI-H23 whole cell lysate (1:100 dilution, 1:10 dilution and crude extract) at room temperature for one hour and then the electrochemical response was recorded. The impedance spectroscopy was performed and the concentration of Sam68 in these dilutions was identified from the Standard curve plotted earlier (Figure 4.30C). The concentration of Sam68 protein detected in different dilutions of cell line extract is shown in table 4.4. The determination of these dilutions was also carried out with ELISA (Figure 4.27C). As the concentration of Sam68 in different dilutions of NCI-H23 cell extract is very low, the ELISA method was unable to differentiate between the amounts of protein. Thus, the proposed immunosensor can be applied to quantify the target protein in complicated clinical serum samples.

Table 4.4: Determination of Sam68 with the immunosensor in different dilutions of NCI-H23 cell extract

The concentration of cell line extract (serial dilutions)	**Average RCT (ohms) (resistance to charge transfer)**	**RSD %**	**Concentration (ug/ml)**
1:100	947	3.932676758	0.001017081
1:10	1517	1.618725004	0.006959364
Crude extract	2792	1.092847156	0.513819864

4.3.5 Spiking and Recovery

To access the viability and accuracy of this bio-sensing system for Sam68, a spiking and recovery experiment was performed. Different known dilutions of purified Sam68 protein (0.01, 0.1 and 1 µg/mL) were spiked in 1:100 dilution of NCI-H23 cell extract and the immunosensor response

was measured. The two sets of responses observed concentration calculated from the Standard curve and expected concentration were compared as in Table 4.5. Good recovery of Sam68 was obtained in the range of 92.37-109.37 and RSD in the range of 0.8 – 2.3%. In general, an acceptable recovery range is 80-120%. The optimum recovery obtained assures the reliability of immunosensor performance with a negligible effect of the Sample matrix. Further, GCE based immunosensor (GCE/PANI/Glu/Sam68-Ab/BSA) was storied in 0.1M PBS (pH 7.4) at 4 °C. During this period, the immunosensor retained ~ 90% of the initial activity till three weeks.

Table 4.5: Determination of Sam68 spiked in NCI-H23 cell extract with the immunosensor

S.no	**The concentration of Sam68 (µg/mL)**	**The concentration of Sam68 spiked (µg/mL)**	**Expected Conc. of Sam68 (µg/mL)**	**Observed Sam68 using fabricated biosensor (µg/mL)**	**Recovery %**	**RSD %**
1	0.01	0.00101708	0.011017	0.01205	109.37	2.33946
2	0.1	0.00101708	0.101017081	0.102329	101.29	2.333013
3	1	0.00101708	1.001017081	0.924698	92.37	0.852126

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

Chapter 5

Conclusion and future scope

In this study, we present genome-scale evidence for Sam68 to be a prognostic or non-prognostic marker in four different human cancers. Our result represents that higher expression of a gene is not always a cause of the pathogenesis of cancer. A gene can be labeled as a prognostic maker if it is involved in crucial molecular processes, which are specific to the disease progression. In the present work, we first evaluated the expression level of Sam68 in four different types of cancer: KIRP, LUAD, LAML and OV cancer. For the first time, we have shown that the expression of Sam68 in all four cancers is heterogeneous and patient-specific. However, our results show that higher expression of Sam68 causes reduced survival of the patient in KIRP and LUAD but not in LAML and OV. This indicates, in KIRP and LUAD, higher expression of Sam68 possibly plays a critical role in the cancer-specific event. To understand the cancer-specific behavior of Sam68, we performed the genome-wide correlation analysis in all four cancers for the patients with higher expression of Sam68 and screened the genes which have significant correlation and direct interaction with Sam68. It is noticed that the common genes, which are coexpressed and interact with Sam68 are involved in the cancer-specific processes in KIRP and LUAD, but not in LAML and OV. This provides us the lead to do a further experiment to find the cancer-specific module in all co-expressed genes of Sam68. We identified that several recurrent network modules are involved in cell cycle and division linked processes in KIRP and LUAD. These network modules contain a core set of genes, which, when highly expressed are sufficient for cell proliferation and metastasis. Additionally, the functional similarity shows that more significant numbers of coexpressed genes are involved in similar molecular functions in KIRP and LUAD compared to OV and LAML. For an additional layer of understanding, we have calculated the genome-wide correlation of isoform level data as Sam68 is involved in RNA splicing. These results also confirm that cancer driven biological processes are enriched in KIRP and LUAD not in LAML and OV, although Sam68's predominant isoform uc001bub is highly expressed in all four cancers. The change of the cellular environment drives the rewiring of the molecular network of a particular gene which can result in alteration of gene function (Billmann M. 2018). We observed a similar result in the case of Sam68 in the different cancer cells. It should

be noted that the observation is restricted to a specific group of patients, either in LUAD or KIRP. This is not generalized observation for specific cancer type rather it is patient-specific. Therefore, the present work supports the need for personalized medicine and diagnosis in cancer treatment. In general, a gene is identified as a prognostic cancer biomarker when its mRNA expression level is significantly correlated with overall patient survival (Yang Y. *et al.,* 2014). However, our observations suggest that besides higher expression, a prognostic biomarker should directly or indirectly be associated with the cancer-specific network and event. Therefore, to understand the prognostic value of a target molecule a detailed landscape of possible molecular events should be studied, which will lead to improved cancer diagnosis and therapy.

To further evidence the significance of Sam68 as a cancer biomarker in lung cancer pathogenesis, we examined microarray expression data, to understand the Stage-specific (Stage1-4) expression of Sam68 and its clinical relevance to the survival of Lung cancer patients. In accordance with the above results, the upregulated expression of Sam68 is specific to the cancer stage (grade), which monotonically increased from the early cancer stage to the late metastatic stage. Moreover, this differential expression of Sam68 reduced the overall survival of lung cancer patients. This reanalysis with microarray datasets is equally important, as analysis with two differently processed and well-characterized datasets increased the reliability of our results. Together results demonstrated, Sam68 as a stage-specific diagnostic and prognostic biomarker in lung cancer.

On the other hand, Lung cancer remains latent in the early stage and is typically diagnosed at an advanced stage. This makes both early diagnosis and treatment challenging. So, identification of early diagnostic biomarkers is of urgent need, which is a critical determinant for the initiation of appropriate treatment. Moreover, only trace levels of the biomarker are expressed in early pathological stages**.** This has motivated us to design a simple & efficient detection system to quantify Sam68 protein in cancer, which can be applied to clinical diagnosis. In this line, recombinant Sam68 was produced in E.coli to fabricate biosensor. A GCE modified immunosensor was successfully developed, which employs glutaraldehyde cross-linked with PANI as immobilization support for adsorption of anti-Sam68 antibody. In comparison with ELISA, the analytical performance of the immunosensor was found to be superior in terms of a wide detection range (1 ng/mL to 5 μg/mL), very low LOD (10.54549 pg/mL) and LOQ

(35.15165 pg/mL). We validated sensor performance with a lung cancer cell line (NCI-H23). The immobilization strategy applied here shows the promising result to determine the Sam68 protein concentration in a complex biological sample, which shows a promise for better lung cancer patient care.

In the present study, we have analyzed the whole-genome expression data. However, analysis of other genomics data such as methylation and mutation can provide in-depth knowledge of oncogenesis by Sam68. Moreover, the similar Genome-wide analysis can be applied in other cancer types, to understand the prognostic role of Sam68 in other cancers. In addition, the same study can be used to explore the prognostic value of other oncogenic splicing factors, e.g., SRSF1. In the study, we identified & presented genome-wide targets transcripts that are co-expressed with Sam68 in each cancer type (Figure 5). From this result, new splicing targets of Sam68 can be identified and their role in cancer could be possibly studied. Furthermore, the proposed immunosensor can be developed as a miniaturized point of care diagnostic tool for monitoring the disease progression after proper validation of the novel biomarker, Sam68.

REFERENCES

Adomas A, Heller G, Olson A et al, Comparative analysis of transcript abundance in Pinus sylvestris after challenge with a saprotrophic, pathogenic or mutualistic fungus. Tree Physiol. 2008; 28:885-97.

Alkhateeb A, Rezaeian I, Singireddy S et al, Transcriptomics Signature from Next-Generation Sequencing Data Reveals New Transcriptomic Biomarkers Related to Prostate Cancer. Cancer informatics. 2019; 18.https://doi.org/10.1177/1176935119835522

Altintas Z, Tothill I. Molecular biosensors: promising new tools for early detection of cancer. Nanobiosensors in Disease Diagnosis. 2015; 4:1-10 https://doi.org/10.2147/NDD.S56772

Anczuków O, Krainer AR. Splicing-factor alterations in cancers. RNA. 2016; 22(9):1285–1301. DOI:10.1261/rna.057919.116

Anczuków O, Rosenberg AZ, Akerman M, et al. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. Nat Struct Mol Biol. 2012; 19(2):220–228. Published 2012 Jan 15. DOI:10.1038/nsmb.2207

Andreotti AH, Bunnell SC, Feng S, Berg LJ, Schreiber SL. Regulatory intramolecular association in a tyrosine kinase of the Tec family, Nature. 1997; 385:93–97.

Arning S, Gruter P, Bilbe G, Kramer A. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. RNA (N. Y.). 1996; 2:794–810.

Ashworth A, Lord CJ and Reis-Filho JS. Genetic interactions in cancer progression and treatment. Cell. 2011; 145:30–38.

Aydemir, N., Malmström, J., & Travas-Sejdic, J. Conducting polymer based electrochemical biosensors. Physical Chemistry Chemical Physics 2016; 18(12):8264–8277. doi:10.1039/c5cp06830d

Babic I, Cherry E, Fujita DJ. SUMO modification of Sam68 enhances its ability to repress cyclin D1 expression and inhibits its ability to induce apoptosis. Oncogene 2006; 25:4955–4964.

Babic I, Jakymiw A, Fujita DJ. The RNA binding protein Sam68 is acetylated in tumor cell lines, and its acetylation correlates with enhanced RNA binding activity. Oncogene 2004; 23:3781–3789.

Bader GD and Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics.2003; 42. https://doi.org/10.1186/1471-2105-4-2.

Bader GD, Betel D and Hogue CW. BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 2003; 31:248–250.

Baehrecke EH, who encodes a KH RNA binding protein that functions in muscle development. Development 1997; 124:1323–1332.

Barlat I, Maurier F, Duchesne M, Guitard E, et al, A Role for Sam68 in Cell Cycle Progression Antagonized by a Spliced Variant within the KH Domain. J. Biol. Chem. 1997; 272:3129–3132. DOI: 10.1074/jbc.272.6.3129

Basil CF, Zhao Y, Zavaglia K, et al, Common cancer biomarkers. Cancer Res. 2006; 66(6):2953–2961.

Bedford MT, Chan DC, Leder P. FBP WW domains and the Abl SH3 domain bind a specific class of proline-rich ligands, EMBO J. 1997; 16:2376 – 2383.

Bedford MT, Chan DC, Leder P. WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of protein-rich motif: the proline glycine and methionine-rich motif, Proc. Natl. Acad. Sci. U. S. A. 1998; 95:10602 – 10607.

Bedford MT, Frankel A, Yaffe MB, Clarke S, Leder P, Richard S. Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. J. Biol. Chem. 2000; 275:16030–16036.

Bejar R. Splicing Factor Mutations in Cancer. Adv Exp Med Biol. 2016; 907:215–228. DOI:10.1007/978-3-319-29073-7_9

Biamonti G, Catillo M, Pignataro D, Montecucco A, Ghigna C. The alternative splicing side of cancer. Semin Cell Dev Biol. 2014; 32:30–36. doi:10.1016/j.semcdb.2014.03.016

Bibikova M, Barnes B, Tsan C, et al, High density DNA methylation array with single CpG site resolution. Genomics 2011; 98:288-95.

Bielli P, Busà R, Di Stasi SM, et al. The transcription factor FBI-1 inhibits Sam68-mediated BCL-X alternative splicing and apoptosis. EMBO Rep. 2014; 15(4):419–427. DOI:10.1002/embr.201338241

Bielli P, Busa R, Paronetto MP, Sette C. The RNA-binding protein Sam68 is a multifunctional player in human cancer. Endocr. Relat. Cancer 2011; 18:R91-R102. https://doi.org/10.1530/ERC-11-0041.

Billmann M, Chaudhary V, ElMaghraby MF, Fischer B, Boutros M. Widespread Rewiring of Genetic Networks upon Cancer Signaling Pathway Activation. Cell Syst. 2018; 6:52–64.

Black DL. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem. 2003; 72:291–336. DOI:10.1146/annurev.biochem.72.121801.161720

Bo Li and Colin N Dewey “RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome”. BMC Bioinformatics. 2011; 12:323

Boice JD Jr, Preston D, Davis FG, Monson RR. Frequent chest X-ray fluoroscopy and breast cancer incidence among tuberculosis patients in Massachusetts. Radiat Res. 1991; 125:214–222.

Bonomi S, Gallo S, Catillo M, Pignataro D, Biamonti G, Ghigna C. Oncogenic alternative splicing switches: role in cancer progression and prospects for therapy. Int J Cell Biol. 2013; 962038. doi:10.1155/2013/962038

Boon-Unge K, Yu Q, Zou T, Zhou A, Govitrapong P, Zhou J. Emetine regulates the alternative splicing of Bcl-x through a protein phosphatase 1-dependent mechanism. Chemistry & biology. 2007; 14(12):1386–1392.

Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976; 72:248-254.

Brown RL, Reinke LM, Damerow MS, et al. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. J Clin Invest. 2011; 121(3):1064-74. DOI: 10.1172/JCI44540. PubMed PMID: 21393860; PubMed Central PMCID: PMCPMC3049398.

Bunnell SC, Henry PA, Kolluri, Kirchhausen T, Rickles RJ, Berg LJ. Identification of Itk/Tsk Src homology 3 domain ligands, J. Biol. Chem. 1996;271:25646 – 25656.

Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins, Science. 1994; 265:615 – 621.

Burd CJ, Petre CE, Morey LM, Wang Y, Revelo MP, Haiman CA, Lu S, Fenoglio-Preiser CM, Li J, Knudsen ES, Wong J, Knudsen KE. Cyclin D1b variant influences prostate cancer growth through aberrant androgen receptor regulation. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(7):2190–2195. https://doi.org/10.1073/pnas.0506281103

Cappellari M, Bielli P, Paronetto MP, et al. The transcriptional co-activator SND1 is a novel regulator of alternative splicing in prostate cancer cells. Oncogene. 2014; 33(29):3794–3802. DOI:10.1038/onc.2013.360

Chabot B, Shkreta L. Defective control of pre-messenger RNA splicing in human disease. J Cell Biol. 2016; 212(1):13–27. DOI:10.1083/jcb.201510032

Chandra S, Gäbler C, Schliebe C; Lang H; Bahadur, D. Fabrication of a label-free electrochemical immunosensor using a redox-active ferrocenyl dendrimer. New J. Chem. 2016;40:9046–9053.

Chaplin M. What are biosensors? Available from: http://www.lsbu.ac.uk/biology/enztech/biosensors.html. 2004.

Chawla G, Lin CH, Han A, Shiue L, Ares M Jr, Black DL. Sam68 regulates a set of alternatively spliced exons during neurogenesis. Molecular and cellular biology. 2009;29(1):201–213. https://doi.org/10.1128/MCB.01349-08

Chen J, Weiss WA. Alternative splicing in cancer: implications for biology and therapy. Oncogene. 2015,34(1):1–14. DOI:10.1038/onc.2013.570

Chen T, Boisvert FM, Bazett-Jones DP, Richard S. A Role for the GSG Domain in Localizing Sam68 to Novel Nuclear Structures in Cancer Cell Lines. Mol. Biol. Cell 1999;10:3015–3033. https://doi.org/10.1091/mbc.10.9.3015

Chen T, Damaj BB, Herrera C, Lasko P, Richard S. Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain. Mol. Cell. Biol. 1997; 17:5707–5718 DOI: 10.1128/MCB.17.10.5707.

Chen T, Richard S. Structure-Function Analysis of Qk1: a Lethal Point Mutation in Mouse quaking Prevents Homodimerization. Mol. Cell. Biol. 1998; 18:4863–4871. DOI: 10.1128/MCB.18.8.4863

Chen ZY, Cai L, Zhu J, et al. Fyn requires HnRNPA2B1 and Sam68 to synergistically regulate apoptosis in pancreatic cancer. Carcinogenesis. 2011; 32(10):1419–1426. DOI:10.1093/carcin/bgr088

Chen J, Weiss WA. Alternative splicing in cancer: implications for biology and therapy. Oncogene. 2015; 34(1):1–14. doi:10.1038/onc.2013.570

Cho IH, Lee J, Kim J et al, Current Technologies of Electrochemical Immunosensors: Perspective on Signal Amplification. Sensors (Basel). 2018; 18(1):207. DOI:10.3390/s18010207

Clifford R, Lee MH, Nayak S, Ohmachi M, Giorgini F, Schedl T. FOG-2, a novel F-box containing protein, associates with the GLD-1 RNA binding protein and directs male sex determination in the C. elegans hermaphrodite germline. Development. 2000; 127:5265–5276.

Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. Adopted on 1996 May 17 by the American Society of Clinical Oncology. J Clin Oncol. 1996; 14(10):2843–2877.

Côté J, Boisvert FM, Boulanger MC, Bedford MT, Richard S. Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1. Molecular biology of the cell. 2003; 14(1):274–287. https://doi.org/10.1091/mbc.e02-08-0484

Das S, Krainer AR. Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. Mol Cancer Res. 2014; 12(9):1195–1204. DOI:10.1158/1541-7786.MCR-14-0131

David CJ and Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev. 2010; 24:2343–64.

Derry JJ, Richard S, Carvajal HV, Ye X, Vasioukhin V, Cochrane AW, Chen T, Tyner AL. Sik (BRK) phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding activity. Mol. Cell. Biol. 2000; 20:6114 – 6126.

Dhand C, Das M, Datta M, Malhotra BD. Recent advances in polyaniline based biosensors. Biosens. Bioelectron. 2011; 26: 2811-2821. https://doi.org/10.1016/j.bios.2010.10.017.

Di Fruscio M, Chen T, Bonyadi S, Lasko P, Richard S. The Identification of Two Drosophila K Homology Domain Proteins KEP1 AND SAM ARE MEMBERS OF THE Sam68 FAMILY OF GSG DOMAIN PROTEINS. J. Biol. Chem. 1998; 273:30122–30130.

Di Fruscio M, Chen T, Richard S. Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. Proc Natl Acad Sci U S A. 1999; 96(6):2710–2715. doi:10.1073/pnas.96.6.2710

Ding K, Ji J, Zhang X. et al. RNA splicing factor USP39 promotes glioma progression by inducing TAZ mRNA maturation. Oncogene. 2019; 38: 6414–6428. https://doi.org/10.1038/s41388-019-0888-1

Dole MG, Clarke MF, Holman P, Benedict M, Lu J, Jasty R, Eipers P, Thompson CB, Rode C, Bloch C, Nuñez, Castle V. P. Bcl-xS enhances adenoviral vector-induced apoptosis in neuroblastoma cells. Cancer research. 1996; 56(24):5734–5740.

Downward J. Targeting RAS signaling pathways in cancer therapy. Nat. Rev. Cancer 2003; 3:11–22.

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA. 1998; 95:14863–14868.

El Marabti Ettaib, Younis Ihab. The Cancer Spliceome: Reprogramming of Alternative Splicing in Cancer. Frontiers in Molecular Biosciences. 2018; 5:80. DOI=10.3389/fmolb.2018.00080

Espejo A, Coˆte´ J, Bednarek A, Richard S, Bedford MT. A protein-domain microarray identifies novel protein-protein interactions. Biochem. J. 2002; 367:697 – 702.

Fackenthal JD, Godley LA. Aberrant RNA splicing and its functional consequences in cancer cells. Dis Model Mech. 2008; 1(1):37–42. DOI:10.1242/dmm.000331

Fang L, Liang B, Yang G, Hu Y, Zhu Q, Ye X. A needle-type glucose biosensor based on PANI nanofibers and PU/E-PU membrane for long-term invasive continuous monitoring. Biosens. Bioelectron. 2017; 97:196-202.https://doi.org/10.1016/j.bios.2017.04.043.

Feiyun Cui and Zhiru Zhou and H. Susan Zhou. Review—Measurement and Analysis of Cancer Biomarkers Based on Electrochemical Biosensors.The Electrochemical Society. 2019; 167

Filella X, Foj L. Emerging biomarkers in the detection and prognosis of prostate cancer. Clin Chem Lab Med. 2015; 53(7):963–973. DOI:10.1515/cclm-2014-0988

Frisone P. et al, Sam68: Signal Transduction and RNA Metabolism in Human Cancer. Biomed. Res. Int. 2015; 528954, https://doi.org/10.1155/2015/528954 (2015).

Fu K, Sun X, Zheng W, Wier E.M, Hodgson A, Tran DQ, Richard S, Wan F. Sam68 modulates the promoter specificity of NF-κB and mediates expression of CD25 in activated T cells. Nat. Commun. 2013;4:1909.

Fumagalli S, Totty NF, Hsuan JJ, Courtneidge SA. A target for Src in mitosis. Nature. 1994;368(6474):871–874. DOI: 10.1038/368871a0.

Galarneau A, Richard S. The STAR RNA binding proteins GLD-1, QKI, Sam68 and SLM-2 bind bipartite RNA motifs. BMC Mol Biol. 2009;10:47.

Gary JD, Clarke S, RNA and protein interactions modulated by protein arginine methylation, Prog. Nucleic Acid Res. Mol. Biol. 1998; 61:65 – 131.

Gaughan L, Dalgliesh C, El-Sherif A, Robson CN, Leung HY, Elliott DJ. The RNA-binding and adaptor protein Sam68 modulates signal-dependent splicing and transcriptional activity of the androgen receptor. J. Pathol. 2008; 215:67–77.

Giangrande PH. et al, A role for E2F6 in distinguishing G1/S- and G2/M-specific transcription. Genes Dev. 2004; 18:2941–2951.

Golub TR, Slonim DK, Tamayo P, et al, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999; 286:531-7.

Grieshaber D, MacKenzie R, Vörös J, Reimhult E. Electrochemical Biosensors—Sensor Principles and Architectures. Sensors. 2008; 8:1400–1458.

Grossman JS, Meyer MI, Wang YC, Mulligan GJ, Kobayashi R, Helfman DM. The use of antibodies to the polypyrimidine tract binding protein (PTB) to analyze the protein components that assemble on alternatively spliced pre-mRNAs that use distant branch points. RNA (New York, N.Y.). 1998; 4(6):613–625. https://doi.org/10.1017/s1355838298971448

Grosso AR, Martins S, Carmo-Fonseca M. The emerging role of splicing factors in cancer. EMBO Rep. 2008; 9(11):1087–1093. DOI:10.1038/embor.2008.189

Guitard E, Barlat I, Maurier F, Schweighoffer F, Tocque B. Sam68 is a Ras-GAP-associated protein in mitosis, Biochem. Biophys. Res. Commun. 1998;245:562 – 566.

Hong W, Resnick RJ, Rakowski C, Shalloway D, Taylor SJ, Blobel GA. Physical and functional interaction between the transcriptional cofactor CBP and the KH domain protein Sam68. Mol. Cancer Res. 2002; 1:48–55.

Huot MÉ, Vogel G, Zabarauskas A, Ngo CT, Coulombe-Huntington J, Majewski J, Richard S. The Sam68 STAR RNA-binding protein regulates mTOR alternative splicing during adipogenesis. Molecular cell. 2012;46(2):187–199. https://doi.org/10.1016/j.molcel.2012.02.007

Iijima T, Wu K, Witte H, Hanno-Iijima Y, Glatter T, Richard S, Scheiffele P. SAM68 regulates neuronal activity-dependent alternative splicing of neurexin-1. Cell. 2011; 147(7):1601–1614. https://doi.org/10.1016/j.cell.2011.11.028

Ishidate T, Yoshihara S, Kawasaki Y, Roy BC, Toyoshima K, Akiyama T. Identification of a novel nuclear localization signal in Sam68, FEBS Lett.1997; 409:237 – 241.

Itoh H, Kakuta T, Genda G, Sakonju I, Takase K. Canine serum alkaline phosphatase isoenzymes detected by polyacrylamide gel disk electrophoresis. J. Vet. Med. Sci. 2002; 64:35-39.

Jan E, Motzny CK, Graves LE, Goodwin EB. The STAR protein, GLD-1, is a translational regulator of sexual identity in Caenorhabditis elegans. EMBO J. 1999; 18:258–269. //doi.org/10.1093/emboj/18.1.258

Jones AR, Schedl T. Mutations in gld-1, a female germ cell-specific tumor suppressor gene in Caenorhabditis elegans, affect a conserved domain also found in Src-associated protein Sam68. Genes Dev. 1995; 9:1491–1504.

Kalvik TV and Arnesen T. Protein N-terminal acetyltransferases in cancer. Oncogene 2013; 32:269–276.

Kaplan EL and Meier P. Nonparametric Estimation from Incomplete Observations. J. Am. Stat. Assoc. 1958; 53:457–481.

Khanmohammadi A, Aghaie A, Vahedi E, et al. Electrochemical biosensors for the detection of lung cancer biomarkers: A review. Talanta. 2020; 206:120251. DOI:10.1016/j.talanta.2019.120251

Kim CJ, Nishi K, Isono T, Okuyama Y, Tambe Y, Okada Y, Inoue H. Cyclin D1b variant promotes cell invasiveness independent of binding to CDK4 in human bladder cancer cells. Molecular carcinogenesis. 2009; 48(10):953–964. https://doi.org/10.1002/mc.20547

Klippel S, Wieczorek M, Schumann M, Krause E, Marg B, Seidel T, Meyer T, Knapp EW, Freund C. Multivalent binding of formin-binding protein 21 (FBP21)-tandem-WW domains fosters protein recognition in the pre-spliceosome. J Biol Chem 2011; 286: 38478–87.

Knudsen K. E. The cyclin D1b splice variant: an old oncogene learns new tricks. Cell division. 2006; 1:15. https://doi.org/10.1186/1747-1028-1-15

Koedoot E, Wolters L, van de Water B, Dévédec S. Splicing regulatory factors in breast cancer hallmarks and disease progression. Oncotarget. 2019; 10(57): 6021–6037. https://doi.org/10.18632/oncotarget.27215.

Kota V. et al, SUMO Modification of the RNA-Binding Protein La Regulates Cell Proliferation and STAT3 Protein Stability. Mol. Cell Biol. 2018; 38: https://doi.org/10.1128/MCB.00129-17.

Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through the utilization of emerging technologies. Nat Clin Pract Oncol. 2008; 5(10):588–599. DOI:10.1038/ncponc1187

Kumari S. et al, Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. PLoS One. 2012; 50411, https://doi.org/10.1371/journal.pone.0050411.

Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970; 227(5259):680-685.

Lang V, Mege D, Semichon M, Gary-Gouy H, Bismuth GA, A dual participation of ZAP-70 and Src protein tyrosine kinases is required for TCR-induced tyrosine phosphorylation of Sam68 in Jurkat T cells, Eur. J. Immunol. 1997; 27:3360 – 3367.

Latour S, Veillette A. Proximal protein tyrosine kinases in immunoreceptor signaling, Curr. Opin. Immunol. 2001; 13:299 – 306.

Lawe DC, Hahn C, Wong AJ. The Nck SH2/SH3 adaptor protein is present in the nucleus with the nuclear protein Sam68, Oncogenes 1997; 14:223 – 231.

Li B and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011; 12:323. https://doi.org/10.1186/1471-2105-12-323.

Li T. et al, A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods 2017; 14:61–64.

Li Y, Ren Z, Peng Y. et al. Classification of glioma based on prognostic alternative splicing. BMC Med Genomics 2019; 12:165. https://doi.org/10.1186/s12920-019-0603-7

Lin Q, Taylor SJ, Shalloway D. Specificity and Determinants of Sam68 RNA Binding IMPLICATIONS FOR THE BIOLOGICAL FUNCTION OF K HOMOLOGY DOMAINS. J. Biol. Chem. 1997; 272:27274–27280. DOI: 10.1074/jbc.272.43.27274

Liu K, Li L, Nisson P. E, Gruber C, Jessee J, Cohen S. N. Neoplastic Transformation and Tumorigenesis Associated with Sam68 Protein Deficiency in Cultured Murine Fibroblasts. J. Biol. Chem. 2000; 275:40195–40201. DOI: 10.1074/jbc.M006194200

Liu S, Cheng C. Alternative RNA splicing and cancer. Wiley Interdiscip Rev RNA. 2013; 4(5):547–566. DOI:10.1002/wrna.1178

Liu X, Wu J, Zhang D, et al, Identification of Potential Key Genes Associated with the Pathogenesis and Prognosis of Gastric Cancer Based on Integrated Bioinformatics Analysis. Front Genet. 2018; 9:265. Published 2018 Jul 17. DOI:10.3389/fgene.2018.00265

Liu, Z, Zhang, G, Chen, Z. et al. Prussian blue-doped nanogold microspheres for enzyme-free electrocatalytic immunoassay of p53 protein. Microchim Acta 2014; 181:581–588. https://doi.org/10.1007/s00604-013-1149-6

Liyasova MS, Ma K, Lipkowitz S. Molecular pathways: cbl proteins in tumorigenesis and antitumor immunity-opportunities for cancer treatment. Clin. Cancer Res. 2015; 21:1789–1794.

Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. Biochem J. 2009; 417(1):15–27. DOI:10.1042/BJ20081501

Lopez-Bigas N, Audit B, Ouzounis C, et al. Are splicing mutations the most frequent cause of hereditary disease? FEBS Lett. 2005; 28:579(9):1900-3. DOI: 10.1016/j.febslet.2005.02.047.

Lu L, Seenivasan R, Wang YC, Yu JH, Gunasekaran S. An Electrochemical Immunosensor for Rapid and Sensitive Detection of Mycotoxins Fumonisin B1 and Deoxynivalenol. Electrochimica. Acta 2016; 213:89-97. https://doi.org/10.1016/j.electacta.2016.07.096.

Lukong KE, Richard S. Sam68, the KH domain-containing superSTAR. Biochim Biophys Acta. 2003; 1653(2):73–86. doi:10.1016/j.bbcan.2003.09.001

Maa MC, Leu TH, Trandel BJ, Chang JH, Parsons SJ, A protein that is related to GTPase activating protein-associated p62 complexes with phospholipase Cg, Mol. Cell. Biol. 1994; 14:5466 – 5473

Macias MJ, Wiesner S, Sudol M, WW and SH3 domains, two different scaffolds to recognize proline-rich ligands, FEBS Lett. 2002; 513:30 – 37.

Mardis ER. A decade's perspective on DNA sequencing technology. Nature 2011; 470:198-203.

Maroni P, Citterio L, Piccoletti R, Bendinelli P. Sam68 and ERKs regulate leptin-induced expression of OB-Rb mRNA in C2C12 myotubes. Molecular and cellular endocrinology. 2009; 309(1-2):26–31. https://doi.org/10.1016/j.mce.2009.05.021

Matter N, Herrlich P, Konig, H. Signal-dependent regulation of splicing via phosphorylation of Sam68. Nature 2002; 420:691–695.

McBride AE, Silver PA. State of the arg: protein methylation at arginine comes of age. Cell. 2001; 106(1):5–8. doi:10.1016/s0092-8674(01)00423-8.

McLaren M, Asai K, Cochrane A. A novel function for Sam68: enhancement of HIV-1 RNA 3' end processing. RNA (New York, N.Y.). 2004; 10(7):1119–1129. https://doi.org/10.1261/rna.5263904

Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010; 11:31-46.

Meyer NH, Tripsianes K, Vincendeau M, Madl T, Kateb F, Brack-Werner R, & Sattler M. Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qua1 domain. The Journal of biological chemistry. 2010; 285(37):28893–28901. https://doi.org/10.1074/jbc.M110.126185

MMP-3-induced EMT and genomic instability. Nature. 2005; 436(7047):123-7.

Mok SC. et al, A gene signature predictive for outcome in advanced ovarian cancer identifies a survival factor: microfibril-associated glycoprotein 2. Cancer Cell 2009; 16:521–532.

Moritz S, Lehmann S, Faissner A, von Holst A. An induction gene trap screen in neural stem cells reveals an instructive function of the niche and identifies the splicing regulator Sam68 as a tenascin-C-regulated target gene. Stem cells (Dayton, Ohio). 2008; 26(9):2321–2331. https://doi.org/10.1634/stemcells.2007-1095

Mosmann T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. J. Immunol. Methods. 1983; 67:55–59.

Najib SV. Sanchez-Margalet, Sam68 associates with the SH3 domains of Grb2 recruiting GAP to the Grb2 – SOS complex in insulin receptor signaling, J. Cell. Biochem. 2002; 86:99 – 106.

Najib S, Martin-Romero C, Gonzalez-Yanes C, Sanchez-Margalet V. Role of Sam68 as an adaptor protein in signal transduction. Cell Mol. Life Sci. 2005; 62:36–43.

Naomi Li, Stéphane Richard, Sam68 functions as a transcriptional coactivator of the p53 tumor suppressor, Nucleic Acids Research. 2016; 44(18):8726–8741, https://doi.org/10.1093/nar/gkw582

NCI Biomarker 2009Available from http://www.cancer.gov/dictionary/?searchTxt=biomarkerAccessed Sep 24, 2010

Nimse SB, Sonawane MD, Song KS, Kim T. Biomarker detection technologies and future directions. Analyst. 2016; 141(3):740–755. DOI:10.1039/c5an01790d

Olshavsky NA, Comstock CE, Schiewer MJ, et al. Identification of ASF/SF2 as a critical, allele-specific effector of the cyclin D1b oncogene. Cancer Res. 2010; 70(10):3975–3984. DOI:10.1158/0008-5472.CAN-09-3468

Oltean S.and Bates DO. Hallmarks of alternative splicing in cancer. Oncogene. 2014; 1–8. DOI:10.1038/onc.2013.533

Pagliarini V, Pelosi L, Bustamante MB, Nobili A, Berardinelli MG, D'Amelio M, Musarò A, Sette C. SAM68 is a physiological regulator of SMN2 splicing in spinal muscular atrophy. The Journal of cell biology. 2015; 211(1):77–90. https://doi.org/10.1083/jcb.201502059

Parikh JR, Klinger B, Xia Y, Marto JA, Blüthgen N. Discovering causal signaling pathways through gene-expression patterns. Nucleic acids research, 2010; 38:W109–W117. https://doi.org/10.1093/nar/gkq424.

Paronetto MP, Achsel T, Massiello A, Chalfant CE, Sette C. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. J. Cell Biol. 2007; 176:929–939.

Paronetto MP, Cappellari M, Busà R, Pedrotti S, Vitali R, Comstock C, Hyslop T, Knudsen KE, Sette C. Alternative splicing of the cyclin D1 proto-oncogene is regulated by the RNA-binding protein Sam68. Cancer research. 2010; 70(1):229–239. https://doi.org/10.1158/0008-5472.CAN-09-2788

Paronetto MP, Zalfa F, Botti F, Geremia R, Bagni C, Sette C. The nuclear RNA-binding protein Sam68 translocates to the cytoplasm and associates with the polysomes in mouse spermatocytes. Mol. Biol. Cell. 2006; 17:14–24.

Pedrotti S, Bielli P, Paronetto MP, Ciccosanti F, Fimia GM, Stamm S, Manley JL, Sette C. The splicing regulator Sam68 binds to a novel exonic splicing silencer and functions in SMN2 alternative splicing in spinal muscular atrophy. The EMBO journal. 2010; 29(7):1235–1247. https://doi.org/10.1038/emboj.2010.19

Peng Y and Croce CM. The role of MicroRNAs in human cancer. Signal Transduct. Target Ter. 2016; 1:15004. https://doi.org/10.1038/sigtrans.2015.4.

Peri S. et al, Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res. 2004; 32:D497–501.

Pilotte J, Larocque D, Richard S. Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKING apoptotic inducer. Genes Dev. 2001; 15:845–858. DOI: 10.1101/gad.860301

R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2009. http://www.R-project.org [ISBN 3-900051-07-0]

Radisky DC, Levy DD, Littlepage LE, et al. Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. Nature. 2005; 436(7047):123–127. doi:10.1038/nature03688

Rain JC, Rafi Z, Rhani Z, Legrain P, Krämer A. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. RNA. 1998; 4(5):551–565. doi:10.1017/s1355838298980335

Rajan P, Gaughan L, Dalgliesh C, El-Sherif A, Robson C.N, Leung H.Y, Elliott D.J. The RNA-binding and adaptor protein Sam68 modulates signal-dependent splicing and transcriptional activity of the androgen receptor. J. Pathol. 2008; 215:67–77.

Ray D. et al, A compendium of RNA-binding motifs for decoding gene regulation. Nature 2013; 499:172–177.

Razick S, Magklaras G, Donaldson IM. iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics. 2008; 9:405. https://doi.org/10.1186/1471-2105-9-405 (2008).

Reynier F. et al, Importance of correlation between gene expression levels: application to the type I interferon signature in rheumatoid arthritis. PLoS One. 2011; 6: e24828, https://doi.org/10.1371/journal.pone.0024828.

Richard S, Vogel G, Huot ME, Guo T, Muller WJ, Lukong KE. Sam68 haploinsufficiency delays onset of mammary tumorigenesis and metastasis. Oncogene. 2008; 27(4):548–556. DOI:10.1038/sj.onc.1210652

Richard S, Yu D, Blumer KJ, Hausladen D, Olszowy DW, Connelly PA, Shaw AS. Association of p62, a multi-functional SH2- and SH3-binding protein, with src-family tyrosine kinases, Grb2, and phospholipase Cg-1. Mol. Cell. Biol. 1995; 15:186 – 197.

Ronkainen NJ, Halsall HB, Heineman WR. Electrochemical biosensors. Chem. Soc. Rev. 2010; 39:1747–1763.

Rosenberger S, De-Castro Arce J, Langbein L, Steenbergen RD, Rösl F. Alternative splicing of human papillomavirus type-16 E6/E6* early mRNA is coupled to EGF signaling via Erk1/2 activation. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107(15):7006–7011. https://doi.org/10.1073/pnas.1002620107

Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, Moro-Sibilot D, Brichon PY, Lantuejoul S, Hainaut P, Laffaire J, de Reynies A, et al,Ectopic activation 13 of germline and placental genes identifies aggressive metastasis-prone lung cancers. Sci. Transl. Med. 2013. 5,186ra166.https://doi.org/10.1126/scitranslmed.3005723.

Sambrook J. Molecular Cloning: a Laboratory Manual.4th edition. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sánchez-Jiménez F, Sánchez-Margalet V. Role of Sam68 in post-transcriptional gene regulation. Int J Mol Sci. 2013;14(12):23402–23419. doi:10.3390/ijms141223402

Sawyers CL. The cancer biomarker problem. Nature. 2008; 452(7187):548–552. DOI:10.1038/nature06913

Scotti MM, Swanson MS. RNA mis-splicing in disease. Nat Rev Genet. 2016, 17(1):19–32. DOI:10.1038/nrg.2015.3

Seeler JS. and Dejean A. SUMO and the robustness of cancer. Nat. Rev. Cancer 2017; 184–197.

Sette C. Alternative splicing programs in prostate cancer. International Journal of cell biology, 2013; 458727. https://doi.org/10.1155/2013/458727

Sharma S, Liao W, Zhou X, et al. Exon 11 skipping of E-cadherin RNA downregulates its expression in head and neck cancer cells. Mol Cancer Ther. 2011; 10(9):1751-9. DOI: 10.1158/1535-7163.MCT-11-0248.

Sherr CJ. Cancer cell cycles. Science. 1996; 274:1672–1677.

Shi H, Wei SH, Leu YW, et al, Triple analysis of the cancer epigenome: an integrated microarray system for assessing gene expression, DNA methylation, and histone acetylation. Cancer Res. 2003; 63:2164-71.

Shoaie, N., Daneshpour, M., Azimzadeh, M., Mahshid, S., Khoshfetrat, S. M., Jahanpeyma, F., Foruzandeh, M. Electrochemical sensors and biosensors based on the use of polyaniline and its nanocomposites: a review on recent advances. Microchimica Acta 2019: 186(7). doi:10.1007/s00604-019-3588-1

Shultz JC, Goehe RW, Murudkar CS, et al. SRSF1 regulates the alternative splicing of caspase 9 via a novel intronic splicing enhancer affecting the chemotherapeutic sensitivity of non-small cell lung cancer cells. Mol Cancer Res. 2011; 9(7):889–900. DOI:10.1158/1541-7786.MCR-11-0061

Sienel W, Dango S, Ehrhardt P, Eggeling S, Kirschbaum A, Passlick B. The future in diagnosis and staging of lung cancer. Molecular techniques. Respiration 2006;73: 575–80.

So AY, Sookram R, Chaudhuri AA, et al, Dual mechanisms by which miR-125b represses IRF4 to induce myeloid and B-cell leukemias. Blood 2014; 124:1502-12.

So AY, Zhao JL, Baltimore D. The Yin and Yang of microRNAs: leukemia and immunity. Immunol Rev 2013;253:129-45.

Srivastava A, Gupta VB. Methods for the determination of limit of detection and limit of quantitation of the analytical methods. Chron. Young Sci. 2011; 2:21-25. https://doi.org/10.4103/2229-5186.79345.

Stallcup MR, Role of protein methylation in chromatin remodeling and transcriptional regulation. Oncogenes. 2001; 20:3014 – 3020.

Stark C. et al, BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006; 34:D535–539.

Stirewalt DL. et al, Identification of genes with abnormal expression changes in acute myeloid leukemia. Genes Chromosomes Cancer. 2008; 47:8–20.

Stockley J. et al, The RNA-binding protein Sam68 regulates expression and transcription function of the androgen receptor splice variant. Nat. Publ. Gr. 2015; 1–13.

Stockley J, Markert E, Zhou Y, Robson CN, Elliott DJ, Lindberg J, Leung HY, Rajan P. The RNA-binding protein Sam68 regulates expression and transcription function of the androgen receptor splice variant AR-V7. Scientific reports. 2015; 5:13426. https://doi.org/10.1038/srep13426

Sumantran VN, Ealovega MW, Nuñez G, Clarke MF, Wicha MS. Overexpression of Bcl-XS sensitizes MCF-7 cells to chemotherapy-induced apoptosis. Cancer research. 1995; 55(12):2507–2510.

Sveen A, Kilpinen S, Ruusulehto A. et al. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene 2016; 35:2413–2427.

Taylor SJ, Shalloway D. An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. Nature. 1994; 368(6474):867–871. DOI: 10.1038/368867a0.

Taylor SJ, Anafi M, Pawson T, Shalloway D. Functional interaction between c-src and its mitotic target, Sam68, J. Biol. Chem. 1995; 270:10120 – 10124.

Todaro M, Gaggianesi M, Catalano V, et al. CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. Cell Stem Cell. 2014; 14(3):342–356. DOI:10.1016/j.stem.2014.01.009

Tomlins SA, Laxman B, Varambally S, et al, Role of the TMPRSS2-ERG gene fusion in prostate cancer. Neoplasia. 2008; 10:177–188.

Tripathi S. et al, Meta and Orthogonal Integration of Influenza "OMICs" Data Define a Role for UBR4 in Virus Budding. Cell HostMicrobe. 2015; 18:723–735.

Trub T, Frantz JD, Miyazaki M, Band H, Shoelson SE. The role of a lymphoid-restricted, Grb2-like SH3-SH2-SH3 protein in T cell receptor signaling, J. Biol. Chem. 1997; 272:894 – 902.

Turei D, Korcsmaros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat. Methods. 2016;13: 966–967 (2016).

Valacca C, Bonomi S, Buratti E, Pedrotti S, Baralle FE, Sette C, Ghigna C, Biamonti G. Sam68 regulates EMT through alternative splicing-activated nonsense-mediated mRNA decay of the SF2/ASF proto-oncogene. The Journal of cell biology. 2010; 191(1):87–99. https://doi.org/10.1083/jcb.201001073

Venables JP. Aberrant and alternative splicing in cancer. Cancer Res. 2004; 64(21):7647–7654. DOI:10.1158/0008-5472.CAN-04-1910

Vogel LB, Fujita DJ. p70 phosphorylation and binding to p56lck is an early event in interleukin-2 induced onset of cell cycle progression in T-lymphocytes, J. Biol. Chem. 1995; 270:2506 – 2511.

Wan L, Yu W, Shen E, Sun W, Liu Y, Kong J, et al. SRSF6-regulated alternative splicing that promotes tumor progression offers a therapy target for colorectal cancer. Gut. 2017. DOI: 10.1136/gutjnl-2017-314983.

Wang Q, Xu T, Tong Y, et al. Prognostic Potential of Alternative Splicing Markers in Endometrial Cancer. Mol Ther Nucleic Acids. 2019; 18:1039–1048. DOI:10.1016/j.omtn.2019.10.027

Wang Y, Chen D, Qian H, et al, The splicing factor RBM4 controls apoptosis, proliferation, and migration to suppress tumor progression. Cancer Cell. 2014; 26(3):374–389. DOI:10.1016/j.ccr.2014.07.010

Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of GO terms. Bioinformatics. 2007; 23:1274–1281.

Wang K. et al, MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 2010; 38:e178, https://doi.org/10.1093/nar/gkq622.

Wang LL, Richard S, Shaw AS. P62 Association with RNA is regulated by tyrosine phosphorylation. J. Biol. Chem. 1995; 270:2010–2013.

Wang Y, Liu J, Huang BO, Xu YM, Li J, Huang LF, Lin J, Zhang J, Min QH, Yang WM, Wang XZ. Mechanism of alternative splicing and its regulation. Biomedical reports. 2015; 3(2):152–158. https://doi.org/10.3892/br.2014.407

Weigelt B, Lo AT, Park CC, Gray JW, Bissell MJ. HER2 signaling pathway activation and the response of breast cancer cells to HER2-targeting agents is dependent strongly on the 3D microenvironment. Breast cancer research and treatment. 2010; 122(1):35–43. https://doi.org/10.1007/s10549-009-0502-2

Weinstein JN. et al, The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. 2013; 45:1113–1120.

Weng A, Thomas SM, Rickles RJ, Taylor JA, Brauer BA, Seidel-Dugan C, Michael WM, Dreyfuss G, Brugge JS. Identification of Src, Fyn, and Lyn SH3-binding proteins: implications for a function of SH3 domains, Mol. Cell. Biol. 1994; 14:4509 – 4521.

Wu J, Zhou L, Tonissen K, Tee R, Artzt K. The quaking I-5 (QKI5) has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm, J. Biol. Chem. 1999; 274:29202 – 29210.

Xenarios I. et al, DIP: the database of interacting proteins. Nucleic Acids Res. 2000; 28:289–291.

Yanaihara N, Caplen N, Bowman E, et al, Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006; 9(3):189–198. DOI:10.1016/j.ccr.2006.01.025

Yang Q, Zhao J, Zhang W, Chen D, Wang Y. Aberrant alternative splicing in breast cancer. Journal of molecular cell biology. 2019; 11(10):920–929. https://doi.org/10.1093/jmcb/mjz033

Yang Y. et al, Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. Nat. Commun. 2014; 5:3231. https://doi.org/10.1038/ncomms4231 (2014).

Yang Y. et al, Protein SUMOylation modification and its associations with disease. Open Biol. 2017. https://doi.org/10.1098/rsob.170167.

Younis I, Dittmar K, Wang W, Foley SW, Berg MG, Hu KY, et al, Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. Elife 2013; e00780. DOI: 10.7554/eLife.00780

Yu M, Hong W, Ruan S, et al, Genome-Wide Profiling of Prognostic Alternative Splicing Pattern in Pancreatic Cancer. Front Oncol. 2019; 9:773. DOI:10.3389/fonc.2019.00773

Yu G. et al, GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 2010; 26:976–978.

Zaffran S, Astier M, Gratecos D, Semeriva M. The held-out wings (how) Drosophila gene encodes a putative RNA-binding protein involved in the control of muscular and cardiac activity. Development 1997; 124:2087–2098.

Zanzoni A. et al, MINT: a Molecular INTeraction database. FEBS Lett. 2002; 513:135–140.

Zeng Y, Bao J, Zhao Y, Huo D, Chen M, Qi Y, Yang M, Fa H, Hou C. A sandwich-type electrochemical immunoassay for ultrasensitive detection of non-small cell lung cancer biomarker CYFRA21-1. Bioelectrochemistry 2018; 120:183-189. https://doi.org/10.1016/j.bioelechem.2017.11.003.

Zhong Z, Shan J, Zhang Z, Qing Y. The Signal-Enhanced Label-Free Immunosensor Based on Assembly of Prussian Blue-SiO2 Nanocomposite for Amperometric Measurement of Neuron-Specific Enolase. Electroanalysis. 2010; 22:7. https://doi.org/10.1002/elan.201000221.

Zhu L, Xu L, Jia N, et al. Electrochemical immunoassay for carcinoembryonic antigen using gold nanoparticle-graphene composite modified glassy carbon electrode. Talanta. 2013; 116:809–815. DOI:10.1016/j.talanta.2013.07.069

Zong Z, Li H, Yi C, Ying H, Zhu Z, Wang H. Genome-Wide Profiling of Prognostic Alternative Splicing Signature in Colorectal Cancer. Front Oncol. 2018; 8:537. Published 2018 Nov 20. DOI:10.3389/fonc.2018.00537

Zorn AM, Krieg PA. The KH domain protein encoded by quaking functions as a dimer and is essential for notochord development in Xenopus embryos. Genes Dev. 1997; 11:2176–2190. DOI: 10.1101/gad.11.17.2176

Appendix

Table A1. Culture medium for bacteria

Media	Components	Concentration (%)	pH
Luria broth (LB)	Tryptone	1.0	7.2
	Yeast extract	0.5	
	NaCl	0.5	
2xTY	Tryptone	1.6	7.2
	Yeast extract	1.0	
	NaCl	0.5	

Table A2. List of buffers and solutions

Buffers/solutions	Composition
TAE- Tris Acetate EDTA buffer (50X)	24.2 g Tris base, 5.7 mL acetic acid, 10 mL of 0.5 M EDTA.
Phosphate Buffer Saline (PBS)	0.137 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4.
PBST	PBS containing 0.1 % Tween-20
Binding buffer (Sam68 purification)	1XPBS buffer,150mM NaCl, PMSF, pH 7.4
Elution buffer (Sam68 purification)	100 mM Tris-Cl, 300MM NaCl, 0.2mM EDTA 20mM reduced glutathione, pH 8
Trypsin- EDTA	0.05% Trypsin, 0.53mM EDTA in PBS
RIPA buffer	50mM Tris-Hcl, 150mM NaCl, 1% Triton X100, 0.1% SDS, 1mM PMSF, 50mM sodium fluoride, 1mM sodium orthovandate

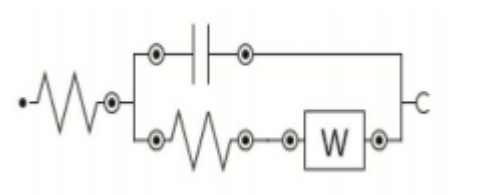
Buffers/solutions for SDS-PAGE

30% Acrylamide-bisacrylamide solution (100mL)	29.2 g Acrylamide, 0.8 g bisacrylamide
0.5 M Tris HCl, pH 6.8 (100 mL)	6.06 g of Tris base, pH adjusted to 6.8 with 2N HCl
1.5 M Tris HCl, pH 8.8 (100 mL)	18.18 g of Tris base, pH adjusted to 8.8 with 2N HCl
10 % Ammonium persulfate (APS)	0.1 g of APS in 1 mL distilled H_2O
Sample loading buffer (6X)	50 mM Tris HCl of pH 6.8, 2% SDS, 10% glycerol, 1% β-mercaptoethanol, 0.1% Bromophenol blue
Staining solution	50% Methanol, 10% acetic acid, 40% H_2O, 0.25% commasive brilliant blue R250
Destaining solution	30% Methanol, 10% acetic acid, 60% H_2O
12% Separating gel (5 mL)	1.6 mL H_2O, 2 mL 30% Acrylamide-bisacrylamide solution, 1.3 mL 1.5 M Tris HCl, pH 8.8, 0.05 mL of 10% SDS, 0.05 mL of 10% APS, 0.002 mL TEMED
5 % Stacking gel (2 mL)	1.4 mL H_2O, 0.33 mL 30% Acrylamide-bisacrylamide solution, 0.25 mL 1.5 M Tris HCl, pH 6.8, 0.02 mL of 10% SDS, 0.02 mL of 10% APS, 0.002 mL TEMED

Buffers/solutions for Western Blot

Transfer Buffer	25mM Tris Base, 39mM Glycine, 20% methanol
Blocking solution	5% BSA in 0.1% PBST (PBS buffer with Tween 20.

Figure A1. <u>Randle's equivalent circuit.</u>



This is simple Randle's equivalent circuit with NOVA software for EIS data of the current study after deducing the EIS data and the component values using non-linear least square curve fitting technique.

Publications

1. Sumithra B., Saxena U. & Das A. B. Alternative splicing within the Wnt signaling pathway: role in cancer development. Cell Oncol. 39:1-13, (2016).

2. Sumithra B., Saxena U. & Das A. B. A comprehensive study on the genome-wide coexpression network of Sam68/Sam68 reveals its cancer and patient-specific association. Sci. Rep. 9, 11083 (2019).

3. Sumithra B., V.S.P.K. Sankara Aditya Jayanthi., Hari Chandana Manne., Rashmika Gunda., Urmila Saxena., Asim Bikas Das. Antibody-based biosensor to detect oncogenic splicing factor Sam68 for the diagnosis of lung cancer. Biotechnology letters. DOI: 10.1007/s10529-020-02951-9 (2020).

Conferences and Workshops

1. Sumithra.B, Asim Bikas Das. Transcriptional Regulation and Protein-Protein Interaction Networks of Human Oncogenic Splicing Factor Sam68*, 6th International Conference on Stem Cells and Cancer: Proliferation, Differentiation and Apoptosis (ICSCC 2015), Pune India. *** Best Poster award**

2. Sumithra.B, Asim Bikas Das. Oncogenic Sam68 links Signaling Pathways and Alternative Splicing, 9th Indo Global Summit on Cancer Therapy November 02-04, 2015 Hyderabad, India.

3. Participated in 7th IEEE International Conference on Technology for Education (T4E) 2015, NIT Warangal

4. Two-day national seminar on "Research Trends in Animal Biotechnology" from 27-28th November 2014 at Kakatiya University.

5. Presented a poster in the "9th Indo global summit on cancer therapy" conference at Hyderabad from 2-4th November 2015.

6. Two-day national workshop on "Research Trends in Animal Biotechnology" from 21 -22nd January 2016 at NIT Warangal.

7. "Bioinformatics Workshop on High Throughput Biological Data Analysis" from 23rd to 26th February 2016 at Andhra University.