Genetic approaches to discover novel oncogenes in human cancer

By

Mr. Prajish Sundaram Iyer

[LIFE09201204001]

Tata Memorial Centre, Navi Mumbai

A thesis submitted to the

Board of Studies in Life Sciences

in partial fulfilment of requirements

for the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

November, 2017

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Mr. Prajish Sundaram Iyer entitled "Genetic approaches to discover novel oncogenes in human cancer "and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

S. N. Dalal 27/11/17

Chairman – Dr. Sorab N. Dalal Date:

27/11/17

Guide/Convener – Dr. Amit Dutt Date:

27/11/2017

External Examiner – Prof. Sudhir Krishna Date:

27/11/2017

Member 1 – Dr. Prasanna Venkatraman Date:

27/11/17

Member 2 – Dr. Shailesh V. Shrikhande Date:

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.
I hereby certify that I have read this thesis prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

27/11/17

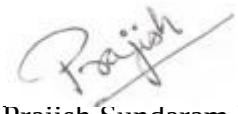
Date: 27th November 2017
Place: Navi Mumbai

Dr. Amit Dutt
Guide

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.



Navi Mumbai — Prajish Sundaram Iyer

Date 27.11.2017

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Navi Mumbai

Prajish Sundaram Iyer

List of Publications arising from the thesis

Journal

Iyer P, Barreto SG, Sahoo B, Chandrani P, Ramadwar MR, Shrikhande SV, Dutt A. Non-typhoidal Salmonella DNA traces in gallbladder cancer. Infectious agents and cancer. 2016; 11:12. (**Thesis Work**)

Chapters in books and lecture Notes : None

Conferences

1. **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo, P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Abstract 4397: Deciphering the Diversity of Somatic Alterations and *Salmonella* Infection in Gallbladder Cancer by Whole Exome Sequencing (April 1st – 5th, 2017) Proceedings of the 108th Annual Meeting of the American Association for Cancer Research, Washington D.C**(Poster Presentation).**
2. **P.Iyer**,N.Gardi,M.Ranjan,P.Chandrani,P.Upadhyay,M.R.Ramadwar,S.V.Shrikhande, A.Dutt AbstractK002: Discovery of somatic *ERBB2* alterations in human gallbladder cancer New Ideas in Cancer – Challenging Dogmas (26th- 28th Feb 2016), European Journal of Cancer,10.1016/S0959-8049(16)31953-0 **(Poster Presentation)**
3. **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo,P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Deciphering the Diversity of Somatic Alterations and *Salmonella* Infection in Gallbladder Cancer by Whole Exome Sequencing (April 8th – 10th, 2016) 35th Annual Convention of Indian Association for Cancer Research **(Oral Presentation under award category)**
4. **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo, P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Landscape of somatic alterations and identification of non-typhoidal *Salmonella* traces in gallbladder cancer (January 29th – 30th 2016) MOSCON conference. **(Poster Presentation)**

5. **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo, P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Deciphering the Diversity of Somatic Alterations and *Salmonella* Infection in Gallbladder Cancer by Whole Exome Sequencing (October 3rd - 5th, 2016) 2016 NextGen Genomics and Bioinformatics **(Nature reviews Best Poster award)**
6. **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo, P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Deciphering the Diversity of Somatic Alterations and *Salmonella* Infection in Gallbladder Cancer by Whole Exome Sequencing (December 7th – 11th, 2016) India International Science Festival (Poster Presentation)-**Selected among 8000 posters in the Life science category**
7. Chandrani P*, Kulkarni V*, **Iyer P**, Upadhyay P, Chaubal R, Das P, Mulherkar R, Singh R, and Dutt A. HPVDetector : A tool to detect HPV and their integration sites using next generation sequencing data (November 17th- 19th , 2014.) 2014 NextGen Genomics and Bioinformatics **(Nature reviews Best Poster award)**
8. P Upadhyay**,** P Chandrani, R Chaubal, N Oak, M Shetty, K Karve, R Prasad, **P Iyer** and A.Dutt "Integrated Genomic Characterization of Head & Neck Cancer Cell Lines Derived from Indian Patients" at The Global Cancer Genomics Consortium Second Annual Symposium: Genomics Medicine in Cancer Research, 2012, ACTREC, India.**(Best Poster award)**
9. R Chaubal, **P Iyer**, V.Kulkarni, I Mittra, A.Dutt A Next Generation Sequencing approach to determine traces of human genetic material in NIH3T3 cells transformed by chromatin & DNA fragments extracted from human cancer patients at The Global Cancer Genomics Consortium Second Annual Symposium: Genomics Medicine in Cancer Research, 2012, ACTREC, India **(Best Poster award)**

Others

1. Chandrani P, Prabhash K, Prasad R, Sethunath V, Ranjan M, **Iyer P**, Aich J, Dhamne H, Iyer DN, Upadhyay P, Mohanty B, Chandna P, Kumar R, Joshi A, Noronha V, Patil V, Ramaswamy A, Karpe A, Thorat R, Chaudhari P, Ingle A, Choughule A, Dutt A. Drug-sensitive FGFR3 mutations in lung adenocarcinoma. Annals of oncology : official journal of the European Society for Medical Oncology. 2017; 28(3):597-603.

2. Upadhyay P, Gardi N, Desai S, Sahoo B, Singh A, Togar T, **Iyer P**, Prasad R, Chandrani P, Gupta S, Dutt A. TMC-SNPdb: an Indian germline variant database derived from whole exome sequences. Database : the journal of biological databases and curation. 2016; 2016.

3. Chandrani P, Upadhyay P, **Iyer P**, Tanna M, Shetty M, Raghuram GV, Oak N, Singh A, Chaubal R, Ramteke M, Gupta S, Dutt A. Integrated genomics approach to identify biologically relevant alterations in fewer samples. BMC genomics. 2015; 16:936.

4. Chandrani P, Kulkarni V, **Iyer P**, Upadhyay P, Chaubal R, Das P, Mulherkar R, Singh R, Dutt A. NGS-based approach to determine the presence of HPV and their sites of integration in human cancer genome. British journal of cancer. 2015; 112(12):1958-65.

5. Choughule A, Sharma R, Trivedi V, Thavamani A, Noronha V, Joshi A, Desai S, Chandrani P, **Sundaram P**, Utture S, Jambhekar N, Gupta S, Aich J, Prabhash K, Dutt A. Coexistence of KRAS mutation with mutant but not wild-type EGFR predicts response to tyrosine-kinase inhibitors in human lung cancer. British journal of cancer. 2014; 111(11):2203-4.

Prajish Sundaram Iyer

ACKNOWLEDGEMENTS

One of the obivious transistion from my master's program was to do a Ph.D. Though it was not my obvious choice to do Ph.D in India, but after I joined Dutt laboratory at ACTREC, it was totally a different ball game. In this journey of five years, there are so many experiences that cannot be shared in a single piece of acknowledgment. Today, I would like to thank all of you on this occasion.

Firstly, I would like to express my deepest and earnest gratitude for my mentor, Dr. Amit Dutt, for his constant guidance, motivation and support in all aspects of my career throughout my tenure. It was his aggression and infectious enthusiasm for the science which made me doing science as a passion. His constructive, sharp and extremely critical supervision has always encouraged me to perform interdisciplinary research in cancer and develop deep scientific temperament. Throughout my tenure, he was a sharp critic but a caring mentor. His words would echo in my body language which transformed me to a great extent in my scientific pursuits. His throat cutting remarks has made me stronger to face the brutal world of science.

I am highly privileged to you for allowing me to actively participating in such an endeavour. I am thankful to doctoral committee members, Dr. Sorab N. Dalal (chairperson), Dr. Prasanna Venkataraman and Dr Shailesh Shrikhande for their continuous critical advice and constructive suggestion during doctoral committee meetings. Advices made during DC meetings were invaluable to my research work and has allowed me to progress in focused manner. I am also thankful to Dr. Rita Mulherkar (ex-chairperson), for their advice in my thesis work. I would like to thank Dr Harsha Gowda for providing gallbladder cancer cell lines at a crucial period of my tenure.

It was privileged to utilize the excellent infrastructure and facilities at ACTREC and Tata MemorialCentre for which I would like to thank Dr. Rajendra Badwe (Director, TMC), Dr. Shubhada Chiplunkar(Director, ACTREC), Dr. Sudeep Gupta (Deputy Director, ACTREC), Dr. Rajiv Sarin (Ex-Director,ACTREC), Dr. Surekha Zingde (Ex-Deputy Director, ACTREC). I would also like to thank ACTREC for providing me regular fellowship, Wellcome Trust/DBT India Alliance, Department of Biotechnology, India and Tata Memorial Centre for funding my research work. I am appreciative to Homi Bhabha National Institute (HBNI) and Tata Memorial Centre (TMC) - Sam Mistry foundation for proving financial support to present my work at an international meeting.

I am grateful to my clinical collaborators Dr. Shailesh Shrikhande, Dr. Mukta Ramadwar, Mr Anand Deshpande, Dr Vikram Chaudhary, Ms.Manisha Kulkarni for their dedication and active support for providing the gallbladder tissue samples and well annotated clinical information of patients.

I would like to express my deepest appreciation to Dr. Shilpee Dutt for her valuable advice during lab meetings as well as doctoral committee meetings, which helped me widen my research interests.

I am exceptionally grateful to all the patients for agreeing to be a part of my thesis work and helping me to conduct research.

I would like to thank the various ACTREC facilities staffs for their assistance. I am thankful to Mr. Dandekar sir for his dexterous assistance in common instrument facility. Additionally, help from staff in sequencing, microscopy, flow cytometry, sequencing, IT, library, steno-pool, administration, account, canteen, and engineering and security personals. It's due to their efforts I could get all the facilities running smoothly and organized manner. A special thank you to Mr. Naresh, Sharda Mam, Anand sir, Tanuja, Jayraj, M.A. Sharma, Chitra, Aruna Mam, Karan sir, Arun Shetty for always being ready to help in administrative work.

I express my humble thank you to all the past and present members of Shilpee Laboratory for providing such a nurturing atmosphere and contributing to my personal and professional growth. A special thanks to Jacinth, Anagha, Sameer, Ekjot, Jyoti, Priya, Shailesh, Tejal, Swati, Samreen, Neha, Aloka, Saket, Poorvaja for creating a friendly atmosphere and being a part of my scientific and non-scientific stuff. Special thanks to Sameer for experimental work. Special thanks to Smita mam for supporting and encouraging me at all times.

The Dutt laboratory members have immensely contributed to my personal and professional growth by providing an amazing fertile ground and creating an interdisciplinary research ecosystem in the lab. I am thankful to Malika, Nilesh, Rohan, Bikram, Trupti, Pratik, Mukul, Ratnam, Asim, Sanket, Prachi, Vidya, Mayur for their constant support in my work and motivation during difficult times. Special thanks to Malika for working really hard on the project and turning the project into a goldstone. Special thanks to Rohan, Nilesh, Bikram and Pratik for helping me with computational work. I am especially thankful to Pawan for constant support during my ups and downs. Thanks to Trupti for helping me during this time of thesis. Special thanks to Ratnam and Renu for being supportive backbone of lab by managing the IRB and administrative stuffs. I would also like to thank Kanishka, Ankita, Sunil, Reshma, Sharan, Rashmi, Renu, Vaibhav, Deepak Iyer, Kunal for being great friends and assisting me at various occasions during my tenure. Special thanks to Sunil, for the conversations that we used to share in central park. I really indebted to your constant support and meaningful interactions in and out of the lab. I am thankful to Post-doctoral fellows at Dutt Laboratory, Dr. Kuldeep, Dr. Jyoti, Dr.Hemant, Dr. Manoj for training me in several experimental stuffs, constant advice, feedback in my research work.

I am thankful to members of lab, Bhaskar and Neelima for their timely help for my experiments. Special thanks to Asim for timely help for my cell culture experiments. Special thanks to newbie Hitesh for taking out time to check the draft. Special thanks to Sanket for helping me with cell line whole exome sequencing. I am thankful to Deepak Amburle, Deepak Chauhan, Mr. Rane for being providing endless technical support and being exceedingly friendly. Special thanks to Dhananjay sir and Shailesh sir for cheering me up and making me feel like a king. Special thanks to Deepak for accompanying me for tea, helping me with media, trypsin and all the non-scientific discussions during all this tenure.

I am thankful to my Ph.D batch mates, Gopal, Bhavik, Saujanya, Pratik, Arunabha for being supportive in all phases of tenure. We all shared several memorable moments together in an out of ACTREC campus which I will remember lifetime. Special thanks to my labmate Mukul for being a caring and a tolerant labmate. I would remember the fun times we have shared together in the lab.

Special thanks to all company representatives who have been my friends and helped me during difficult times.

Last but not the least, I am immensely thankful to my parents who have unconditionally helped me in all my endeavours and being a strong pillar in my life. Special thanks to my papa for always cheering me up with astrological advices. And my moma, praying always for me. I am very thankful to my brother Prashant anna who always encouraged me in all parts of my tenure.

Lastly, I don't believe much in god. But what I know is hardwork, perseverance, honesty and belief in oneself are all different forms I would see god. And only when the time is right these things work. Waiting for the right time……

Thank you

TABLE OF CONTENTS

Homi Bhabha National Institute
SYNOPSIS OF Ph.D. THESIS

Name of the Student: Prajish sundaram Iyer

1. **Name of the Constituent Institution:** Tata Memorial Centre, ACTREC
2. **Enrolment No. and Date of Enrolment:** LIFE09201204001, 31st July 2012
3. **Title of the Thesis:** Genetic approaches to discover novel oncogenes in human cancer
4. **Board of Studies:** Life Science

1. INTRODUCTION

Cancer is a disease of the genome. Recent developments in high-throughput genetic mutation profiling facilitate to do a comprehensive analysis of the cancer genome using next-generation sequencing technologies[1]. Large-scale profiling projects have revealed a landscape of the cancer genome that includes a diverse variety of cancer genomic alterations such as point mutations, copy number variation, and translocations. Cancer genome and transcriptome sequencing have revealed additional clinically relevant novel gene fusions in solid tumors[2]. The developments in the technologies have helped to characterize exomes, genomes, and epigenomes of various cancers. Despite developments in next-generation sequencing technologies not much is understood regarding the rare cancer types such as biliary tract cancers due to its low prevalence in the western countries[3].

Biliary tract cancers are a group of heterogeneous cancers that arise either in intra-or extrahepatic bile ducts or the gallbladder. These cancers are presented at late advanced stages hence have a poor prognosis. Among the biliary tract cancers, gallbladder cancer is one of the most aggressive cancer with poor prognosis. Surgical resection has been preferred as an

option for resectable gallbladder cancer and offers a potential cure. The major risk factor for gallbladder cancer is gallstones however there are other minor risk factors such as female gender, obesity, *Salmonella* infections, cholangitis and gallbladder polyps[4]. The median survival is less than a year and is less than 5%[4]. The adjuvant chemotherapy given to gallbladder cancer patients consists of gemcitabine and platinum-based therapy. There is enough evidence that suggests the resistance of the cancer type to cytotoxic agents [5, 6]. Hence there is an unmet need to identify potential targeted therapies in gallbladder cancer. In India, highest incidence has been reported in Delhi and Bhopal in women (6.6 and 5.2 per 1,00,000 respectively) which is far more higher than south India(0.6 – Chennai and 0.8 – Bangalore)[7]. Comprehensive genomic characterization of gallbladder cancer in India is still dismal despite high prevalence. There are few candidate gene-based studies from India that have identified few targets but these targets are still not in clinical practice [8-10]. There are few reports from China and the west to characterize the gallbladder cancer genome using next-generation sequencing technologies [11-13]. However, the molecular mechanism of the cancer is still poorly understood. Hence more sequencing studies with larger number of samples are required to identify candidate targets in gallbladder cancer.

Mutations in the *EGFR* family members (*EGFR, ERBB2, ERBB3*) have been recently known to be altered at a frequency of 10-15% in gallbladder cancers and the signaling pathway has been shown to be altered in the pathogenesis of gallbladder cancer[11, 14] Some of these alterations in the *EGFR* family have already been shown to be sensitive to tyrosine kinase inhibitors in NSCLC and breast cancers[15, 16]. The EGFR receptor family consists of EGFR/ErbB1, ErbB2 (HER2), ErbB3 (HER3) and ERbB4 (HER4). All receptors except HER2 bind specific ligands via extracellular domain. Upon ligand binding, these receptors use HER2 as a preferential dimerization partner. Homo/Heterodimerization of receptors results in phosphorylation of residues in the intracellular domain resulting in activation of several signaling pathways such as Ras/Raf/MAPK and the PI3K-Akt pathways[17]. Reports

suggest that though EGF does not bind HER2, implying that EGFR may be involved in HER2 signal transduction. Oncogenic transformation by EGFR or HER2 requires high receptor expression while moderate levels are sufficient to cause transformation. Also, reports suggest that down-regulation of normally expressed EGFR suppressed the ability of overexpressed HER2 suggesting HER2 requires EGFR for sustained signaling and transforming potential [18]. This synergistic activity of EGFR-HER2 heterodimerization may be particularly significant as these receptors are upregulated concomitantly in breast and other tumors[19] [20-22]. Hence it is pivotal to study EGFR signaling in gallbladder cancer from the Indian gallbladder cancer perspective.

Pathogenic infections have been associated with cancer worldwide. About 18-20% of the malignancies have been attributed to infections[23]. Among the infection-related cancers, cancers of the stomach, cervix, and liver detain the highest incidence figures and are largely attributable to Helicobacter pylori, human papillomaviruses and hepatitis B & C viruses respectively[24]. In the case of viruses as carcinogens, the critical part of the virus is integrated into the cancer cell resulting in the expression of viral oncogenes that disrupt cell cycle check-points, inhibit apoptosis and contribute to cell immortalization[25]. Other organisms such as *H.pylori*, *Fusobacterium* the chronic persistent infection leads to inflammation which in turn lead to the release of chemokines, cytokines which can result in deregulation of the immune system and promote neovascularization[26, 27]. Among the various risk factors for gallbladder cancer infections with enteric organisms like *Salmonella* typhi are of core importance. The presence of gallstones and the chronic typhoid carrier state might co-operate in the pathogenesis of gallbladder carcinoma, however, the cause and effect relationship is still needed to be ascertained. There is increasing evidence that products of degradation of bile salts by intestinal bacteria may contribute to tumorigenic process however exact causal role needs to be determined[28, 29]. A Recent report by Scanu et al provided a mechanistic role of chronic *Salmonella* infection in host triggering cell transformation

pathways [30]. Most of the studies have focussed on association studies with only the typhoidal *Salmonella* species while no systematic studies have been done to find the association of non-typhoidal species in gallbladder cancer. Few studies have shown the presence of *H.pylori* species in gallbladder samples by PCR-based methods however mechanistic studies are still dismal[31]. However large epidemiological studies and better detection methods at a higher resolution are needed to understand the role of *H.pylori/S.typhi* in gallbladder cancer.

2. OBJECTIVES

The objective of this research proposal is to identify novel oncogenic mutations and pathogenic sequences in gallbladder cancer using genomic approaches. The identification of such oncogenic mutations could be a useful step towards the development of novel targeted therapies. We intend to accomplish this objective as follows:

1) Apply next-generation sequencing analysis of gallbladder cancer to identify pathogenic sequences in gallbladder cancer using computational subtraction method.

2) Apply next-generation sequencing analysis of gallbladder cancer to identify genes whose somatic genomic alterations suggest the properties of driver oncogenes. In a more directed approach, we will sequence exome, from gallbladder tumors of Indian origin cases.

3) We will test candidate oncogenes identified in Objective 2 by gain-of-function assays for cellular transformation and activation of known oncogenic signaling pathways.

Objective 1- Apply next-generation sequencing analysis of gallbladder cancer to identify pathogenic sequences in gallbladder cancer using computational subtraction method.

Specific objective 1a: Detection of *Salmonella* sequences from exome sequencing data.

Epidemiological findings support and indicate the association of *Salmonella* with gallbladder cancer. However, the reports exist only for the typhoidal *Salmonella*, while no reports exist for the association of non-typhoidal *Salmonella* with gall bladder cancer which has been associated only with a systemic illness that triggers an inflammatory response. So we propose to identify the presence of *Salmonella* sequences in gallbladder exome sequencing data using HPVDetector [32] with the addition of *Salmonella* genome as a reference genome in addition to HPV genome.

Whole exome data for these 26 samples were analyzed to detect *Salmonella* traces using HPVDetector pipeline, modified to include additional genome sequence of 6 common *Salmonella* isolates. The computational approach, in brief, subtracts all reads that align with the human genome and aligns remaining reads to HPV and *Salmonella* reference database from NCBI. While HPV16 was detected in 1 gallbladder sample, *Salmonella* isolates were found across multiple samples: *S. typhi* Ty2 (3 samples), *S. typhi* CT18 (6 samples), *S. typhimurium* LT2 (10 samples), *S. choleraesuis* SCB67 (5 samples), *S. paratyphi* TCC (3 samples), and *S. paratyphi* SPB7 (7 samples). In total, *Salmonella* reads were found in 19 of 26 gallbladder tissues (tumor as well as adjacent normal tissues).Typhoidal *Salmonella* species were present in 11 of 26 gallbladder cancer samples, consistent with as known earlier. In addition, we present the first evidence to support the association of even non-typhoidal *Salmonella* species in 12 of 26 gallbladder cancer with 6 samples co-infected with typhoidal and non-typhoidal isolates. To test the specificity of our assay we re-analysed the whole exome data by taking the reverse of the exome data and did not find any spurious *Salmonella* reads. To test the sensitivity of the assay, we downsampled our raw fastq data from 100X to 1X of one of the sample using Downsample Sam (http://broadinstitute.github.io/picard/), the function of Picard Toolkit. Distinct *Salmonella* reads were detected at as low as 10X whole exome coverage that increased linearly.

Specific objective 1b: Validation of *Salmonella* sequences identified from exome sequencing data.

Confirmation of the true identity of *Salmonella* sequences identified using HPVDetector by PCR amplification of read sequences from tumor samples and further to be confirmed by Sanger sequencing.

Further, we confirmed the presence of *Salmonella* sequences using PCR by amplifying 150bp read sequence from 4 samples and Sanger sequencing to validate the true identity of sequences discovered by reference modified HPVDetector.

2) Apply next-generation sequencing analysis of gallbladder cancer to identify genes whose somatic genomic alterations suggest the properties of driver oncogenes. In a more directed approach, we will sequence exome, from gallbladder tumors of Indian origin cases.

Specific objective 2a -Exome sequencing of gallbladder tumor samples

Sample collection - We collected 26 fresh frozen gallbladder tumor samples for whole exome sequencing and 98 FFPE samples for extended validation. We extracted DNA from fresh frozen samples and processed for whole exome sequencing. Out of 98 FFPE blocks, 27 FFPE blocks were suitable for our study.

Whole exome sequencing and analysis-

To investigate the somatic mutation spectrum of Indian gall bladder cancer genome we have analyzed 17 tumors (10 tumor-matched normal paired and 7 unpaired tumors using whole exome sequencing approach. The average coverage for sequencing these samples was around >100X which was suitable for variant calling. Using various steps of filtering in bioinformatics pipeline, we identify 383 somatic alterations across 17 tumors, which includes an average 112 synonymous, 245 missense, 8 nonsense, 8 indels and 8 splice site changes. The average mutation rate considering the paired tumors is about 7.7 mutations/Mb. We

further extended the analysis by comparing our study with COSMIC (Catalogue of somatic alterations in cancer – Gallbladder cancer) and recent exome sequencing in gall bladder cancer.[11] We identified 18 genes that were common in our study and these studies. We found TP53 (35.2%), *ERBB2* (17.6%), *SF3B1* (17.6%), *ATM* (17.6%) and *AKAP11* (17.6%) mutations in more than two samples. We validated some of the alterations identified in *TP53*, *ERBB2*, *ERBB3*, *SMAD4* and *CTNNB1* In the recent exome study ERBB pathway related genes were significantly mutated[11], we extended the discovery of three different activating mutations of *ERBB2* and single mutation in *ERBB3* in our study to an independent validation sample set of 27 FFPE (Formalin fixed paraffin embedded tissues) tumor-samples. We validated 1/3 of the *ERBB2* alterations identified by exome sequencing in the independent set of samples by Sanger sequencing. Out of the two kinase domain mutations of *ERBB2*, V777L was recurrently mutated in 6 out of 44 samples(13% overall mutation frequency) while I767M was found only in a single sample. *ERBB2* (V777L and I767M) has been shown to be activating in *ERBB2* amplification negative breast cancer cell lines and NIH3T3 cell line by colony forming and 3D matrigel assays. A recent report with 9 gall bladder cancer patients identified one patient with the *ERBB2* V777L mutation who showed a mixed response to lapatinib[14]. We identified a C-terminal novel alteration in *ERBB3* (*ERBB3* R1127H) which is not reported in the literature and it further warrants functional validation. To gain insight into the mutation spectrum of gallbladder cancer cell lines, we performed whole exome sequencing of 5 gallbladder cancer cell lines(OCUG1, NOZ, G415, TGBC2TKB, and SNU308) at an average coverage of >100X. Using several steps of filtering in exome sequencing pipeline we identified a total of 2154 alterations comprising of 1930 missense mutations, 65 nonsense mutations, 70 splice site mutations, 83 silent mutations, 4 start codon SNP and 2 Non-stop mutations. We did not observe any hotspot alteration in *ERBB2* in any of the cell lines. We identified *KRAS* alterations *KRAS* G12V in the NOZ cell line and *KRAS* G13D in the G415 cells as opposed to primary tumors where we did not observe *KRAS*

alterations. We identified a polymorphism in *ERBB2* I655V in SNU308 which is a non-activating alteration in *ERBB2*. We validated some of the variants identified in these cell lines by Sanger sequencing to confirm true positive variant discovery by whole exome sequencing.

Specific objective 2b-Functional validation in gallbladder cancer cell lines

To investigate the phosphorylation status of ERBB family of proteins in gallbladder cancer cell lines we used a Phospho-RTK array (R&D systems) which would identify the phosphorylation status of 49 RTKs in an array format spotted in duplicates on a nitrocellulose membrane using a pan anti-phospho-tyrosine antibody conjugated to horseradish peroxidase. Out of the four cell lines analyzed (OCUG1, G415, NOZ, and TGBC2TKB), we observed hyperphosphorylation of EGFR in all gallbladder cancer cell lines while the mild amount of phosphorylation of HER2 was observed in two of the cell lines. On treating these cell lines with an inhibitor such as BIBW-2992(a known ERBB2 and EGFR inhibitor) OCUG1 was found to be highly sensitive to BIBW-2992 than other cell lines in MTT based experiments. Further, we checked downstream components of ERBB2 in OCUG1, p-MAPK levels decreased with increased concentration of the drug. Soft agar colony formation also decreased with increased concentration of drug with maximum inhibition observed at 1 µM and 10µM. Also, wound healing assay in gallbladder cell lines (OCUG1, G415) indicated that cell migration was inhibited in presence of the inhibitor as compared to the control.

We performed experiments in the presence and absence of EGF and treatment with the BIBW-2992. On treatment with BIBW-2992, in presence of EGF, there was the complete abolishment of p-MAPK levels indicating the cell proliferation was efficiently inhibited by BIBW-2992. Also, we checked p-EGFR and p-HER2 levels in the treated cells, we observed that there was the complete abolishment of phosphorylation of EGFR and HER2 in the treated cell lines. We performed stable knockdown of ERBB2 in OCUG1, G415, TGBC2TKB, and NOZ cells with five shRNA constructs. Efficient knockdown was

observed with sh1, sh3, sh4 and sh5 as analyzed by western blot analysis. Soft agar colony formation assay of the knockdown clones in the three cell lines (OCUG1, G415, and TGBC2TKB) indicated that colony formation decreased in the shRNA clones as compared to scrambled control with the strongest inhibition observed in the sh1 and sh3 clone. However, NOZ cell line did not show much difference in soft agar colony formation assay. Growth curve analysis of shRNA clones in the three cells (OCUG1, G415, and TGBC2TKB) indicated that cell proliferation was affected in a time-dependent manner. However, NOZ cells did not show any much difference in the growth pattern in the knockdown clones. Invasive behavior of gallbladder cancer cells was reduced in the knockdown clone (sh1 clone) of OCUG1, G415, and NOZ as compared to the scrambled control indicated by Transwell cell Invasion assay.

3. CONCLUSION

Gall bladder cancer has the highest incidence among the biliary tract cancers. Despite this high incidence rate, coupled with a comparable mortality rate, the genomic causality underlying this disease remains unexplored. Using a highly sensitive methodology that resolves the genome of the disease at base pair resolution(whole exome sequencing), we set out to identify somatic aberrations (mutations and copy number) and infections using computational subtraction methods which may play a causal role in disease pathogenesis.

We specifically identified recurrent, actionable *HER2* alterations as well as copy number changes in EGFR which we show to be sensitive to a pan-HER2 inhibitor. We further observed a differential response of gallbladder cancer cell lines to the pan-HER2 inhibitor, which was primarily based on the presence of different *KRAS* mutations (codon 12 and codon 13 alterations). Similar observations in colorectal cancer have been reported wherein patients with *KRAS* (G13D) mutations respond better to anti-EGFR therapy than *KRAS* (G12V) mutations. These findings may have a clinical relevance in gallbladder cancer and allow

patient stratification and could preclude gallbladder cancer patients from anti-EGFR therapy on the basis of *KRAS* mutational status.

Further studies with larger number samples would be required to have greater insights into the mutation spectrum of Indian gallbladder cancer genome, and confirm these findings. Thus, our discovery introduces a hitherto unknown modality of targetted therapeutic intervention in this disease, which may change the current therapeutic regimen in gallbladder cancer, and introduce scope for precision medicine in the clinics for this dreaded disease.

Additionally, using next-generation sequencing, we identified the presence of DNA sequences of infectious agents (non-typhoidal *Salmonella*) in gallbladder cancer patient tissue, which may be associated with disease progression. Our study identifies a new association of non-typhoidal *Salmonella* with gallbladder cancer. We propose a hypothesis that the presence of non-typhoidal *Salmonella* species in our study along with typhoidal species, provides the inflammatory stimulus required for carcinogenesis. Our study extends the current scope of treatment and provides a basis for treating the non-typhoidal species, along with typhoidal species, for reducing chronic infection due to *Salmonella* in gallbladder cancer. Further, we observe co-occurrence of *TP53* alterations and *Salmonella* infections in gallbladder cancer patients. Detection of the *Salmonella* bacteria using molecular approaches may allow better management of the disease in the current treatment regimen for gallbladder cancer. Further studies would be required to attribute causality of the disease to *Salmonella* infections in gallbladder cancer.

Taken together, identification of *EGFR* family alterations and *Salmonella* infections in gallbladder cancer may allow better treatment and management of the disease.

References

1. Qu, K., et al., Meta-signature of mutated genes in gallbladder cancer: evidence based high throughput screening assays. Ann Transl Med, 2016. 4(11): p. 229.
2. Vogelstein, B., et al., Cancer genome landscapes. Science, 2013. 339(6127): p. 1546-58.

3. Chan, E. and J. Berlin, Biliary tract cancers: understudied and poorly understood. J Clin Oncol, 2015. 33(16): p. 1845-8.

4. Hundal, R. and E.A. Shaffer, Gallbladder cancer: epidemiology and outcome. Clin Epidemiol, 2014. 6: p. 99-109.

5. Vikhanskaya, F., et al., Mechanism of resistance to cisplatin in a human ovarian-carcinoma cell line selected for resistance to doxorubicin: possible role of p53. Int J Cancer, 1997. 72(1): p. 155-9.

6. Lowe, S.W., et al., p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. Cell, 1993. 74(6): p. 957-67.

7. Kapoor, V.K. and A.J. McMichael, Gallbladder cancer: an 'Indian' disease. Natl Med J India, 2003. 16(4): p. 209-13.

8. Srivastava, K., et al., Candidate gene studies in gallbladder cancer: a systematic review and meta-analysis. Mutat Res, 2011. 728(1-2): p. 67-79.

9. Kumari, N., et al., Mutation profiling in gallbladder cancer in Indian population. Indian J Pathol Microbiol, 2014. 57(1): p. 9-12.

10. Kimura, Y., et al., Frequent beta-catenin alteration in gallbladder carcinomas. J Exp Clin Cancer Res, 2003. 22(2): p. 321-8.

11. Li, M., et al., Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. Nat Genet, 2014. 46(8): p. 872-6.

12. Javle, M., et al., Biliary cancer: Utility of next-generation sequencing for clinical management. Cancer, 2016.

13. Javle, M., et al., Molecular characterization of gallbladder cancer using somatic mutation profiling. Hum Pathol, 2014. 45(4): p. 701-8.

14. Javle, M., et al., HER2/neu-directed therapy for biliary tract cancer. J Hematol Oncol, 2015. 8: p. 58.

15. Lynch, T.J., et al., Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med, 2004. 350(21): p. 2129-39.

16. Kancha, R.K., et al., Differential sensitivity of ERBB2 kinase domain mutations towards lapatinib. PLoS One, 2011. 6(10): p. e26760.

17. Fiszman, G.L. and M.A. Jasnis, Molecular Mechanisms of Trastuzumab Resistance in HER2 Overexpressing Breast Cancer. Int J Breast Cancer, 2011. 2011: p. 352182.

18. Zhou, X. and Y.M. Agazie, The signaling and transformation potency of the overexpressed HER2 protein is dependent on the normally-expressed EGFR. Cell Signal, 2012. 24(1): p. 140-50.

19. Brennan, P.J., et al., HER2/neu: mechanisms of dimerization/oligomerization. Oncogene, 2000. 19(53): p. 6093-101.

20. Lee, J.W., et al., Somatic mutations of ERBB2 kinase domain in gastric, colorectal, and breast carcinomas. Clin Cancer Res, 2006. 12(1): p. 57-61.

21. Cancer Genome Atlas, N., Comprehensive molecular portraits of human breast tumours. Nature, 2012. 490(7418): p. 61-70.

22. Stephens, P.J., et al., The landscape of cancer genes and mutational processes in breast cancer. Nature, 2012. 486(7403): p. 400-4.

23. de Martel, C. and S. Franceschi, Infections and cancer: established associations and new hypotheses. Crit Rev Oncol Hematol, 2009. 70(3): p. 183-94.

24. Pagano, J.S., et al., Infectious agents and cancer: criteria for a causal relation. Semin Cancer Biol, 2004. 14(6): p. 453-71.

25. zur Hausen, H., Oncogenic DNA viruses. Oncogene, 2001. 20(54): p. 7820-3.

26. Ljungh, A. and T. Wadstrom, The role of microorganisms in biliary tract disease. Curr Gastroenterol Rep, 2002. 4(2): p. 167-71.

27. Vandeven, N. and P. Nghiem, Pathogen-driven cancers and emerging immune therapeutic strategies. Cancer Immunol Res, 2014. 2(1): p. 9-14.

28. Aries, V., et al., Bacteria and the aetiology of cancer of the large bowel. Gut, 1969. 10(5): p. 334-5.

29. Samaras, V., et al., Chronic bacterial and parasitic infections and cancer: a review. J Infect Dev Ctries, 2010. 4(5): p. 267-81.

30. Scanu, T., et al., *Salmonella* Manipulation of Host Signaling Pathways Provokes Cellular Transformation Associated with Gallbladder Carcinoma. Cell Host Microbe, 2015. 17(6): p. 763-74.

31. de Martel, C., et al., Helicobacter species in cancers of the gallbladder and extrahepatic biliary tract. Br J Cancer, 2009. 100(1): p. 194-9.

32. Chandrani, P., et al., NGS-based approach to determine the presence of HPV and their sites of integration in human cancer genome. Br J Cancer, 2015. 112(12): p. 1958-65.

Publications

a.Published

- **Prajish Iyer**, Savio George Barreto, Bikram Sahoo, Pratik Chandrani, Mukta R Ramadwar, Shailesh V Shrikhande, Amit Dutt(2016) ***'****Non-typhoidal Salmonella DNA traces in gallbladder cancer"* BMC Infectious Agents and cancer(**Thesis Work**)

- P. Chandrani, K. Prabhash, A. Choughule, R. Prasad, V. Sethunath, M. Ranjan, **P. Iyer,** J. Aich, H. Dhamne, D. N. Iyer, P. Upadhyay, B. Mohanty, P. Chandna, R. Kumar, A. Joshi, V. Noronha, V. Patil, A. Ramaswamy, A. Karpe, R. Thorat, P. Chaudhari, A. Ingle, A. Dutt(2016) 'Drug Sensitive FGFR3 mutations in lung adenocarcinoma Ann Onc

- Chandrani P*, Kulkarni V*, ***Iyer P***, Upadhyay P, Chaubal R, Das P, Mulherkar R, Singh R, and Dutt A. (2015). ***"****NGS Based Approach to Determine the Presence of HPV and Their Sites of Integration in Human Cancer Genome."* British Journal of Cancer

- Pratik Chandrani*, Pawan Upadhyay*, ***Prajish Iyer***, Mayur Tanna, Madhur Shetty, Gorantala Venkata Raghuram, Ninad Oak, Ankita Singh, Rohan Chaubal, Manoj Ramteke, Sudeep Gupta and Amit Dutt(2015) *"Integrated genomics approach to identify biologically relevant alterations in fewer samples."* BMC Genomics

- A Choughule, R Sharma, V Trivedi, A Thavamani, V Noronha, A Joshi, S Desai, P Chandrani, ***P Sundaram***, S Utture, N Jambhekar, S Gupta, J Aich, K Prabhash, A Dutt.(2014)" *Coexistence of KRAS mutation with mutant but not wild-type EGFR predicts response to tyrosine-kinase inhibitors in human lung cancer* British Journal of Cancer

- Pawan Upadhyay, Nilesh Gardi, Sanket Desai, Bikram Sahoo, Ankita Singh, Trupti Togar, **Prajish Iyer,** Ratnam Prasad, Pratik Chandrani, Sudeep Gupta, Amit Dutt (2016) TMC-SNPdb: an Indian germline database derived from whole exome sequences Database (Oxford)

b. Accepted: NA

c. Communicated: NA

d. Other Publications: NA

Conference abstracts

- **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo, P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Abstract 4397: Deciphering the Diversity of Somatic Alterations and *Salmonella* Infection in Gallbladder Cancer by Whole Exome Sequencing (April 1st – 5th, 2017) Proceedings of the 108th Annual Meeting of the American Association for Cancer Research, Washington D.C.
- **P.Iyer**,N.Gardi,M.Ranjan,P.Chandrani,P.Upadhyay,M.R.Ramadwar,S.V.Shrikhande, A.Dutt AbstractK002: Discovery of somatic *ERBB2* alterations in human gallbladder cancer New Ideas in Cancer – Challenging Dogmas (26th- 28th Feb 2016), European Journal of Cancer,10.1016/S0959-8049(16)31953-0 (Poster Presentation)
- **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo,P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Deciphering the Diversity of Somatic Alterations and *Salmonella* Infection in Gallbladder Cancer by Whole Exome Sequencing (April 8th – 10th, 2016) 35th Annual Convention of Indian Association for Cancer Research (Oral Presentation under award category)
- **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo, P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Landscape of somatic alterations and identification of non-typhoidal *Salmonella* traces in gallbladder cancer (January 29th – 30th 2016) MOSCON conference. (Poster Presentation)
- **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo, P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Deciphering the Diversity of Somatic Alterations and *Salmonella* Infection in Gallbladder Cancer by Whole Exome Sequencing (October 3rd - 5th, 2016) 2016 NextGen Genomics and Bioinformatics **(Nature reviews Best Poster award)**
- **P.Iyer**, N.Gardi, M.Ranjan, Bikram Sahoo, P.Chandrani, P.Upadhyay, M.R.Ramadwar, S.V.Shrikhande, A.Dutt Deciphering the Diversity of Somatic Alterations and *Salmonella* Infection in Gallbladder Cancer by Whole Exome Sequencing (December 7th – 11th, 2016)

India International Science Festival (Poster Presentation)-**Selected among 8000 posters in the Life science category**

- Chandrani P*, Kulkarni V*, **Iyer P**, Upadhyay P, Chaubal R, Das P, Mulherkar R, Singh R, and Dutt A. HPVDetector : A tool to detect HPV and their integration sites using next generation sequencing data (November 17th- 19th , 2014.) 2014 NextGen Genomics and Bioinformatics **(Nature reviews Best Poster award)**
- P Upadhyay**,** P Chandrani, R Chaubal, N Oak, M Shetty, K Karve, R Prasad, **P Iyer** and A.Dutt "Integrated Genomic Characterization of Head & Neck Cancer Cell Lines Derived from Indian Patients" at The Global Cancer Genomics Consortium Second Annual Symposium: Genomics Medicine in Cancer Research, 2012, ACTREC, India.**(Best Poster award)**

- P Upadhyay, P Chandrani, R Chaubal, N Oak, M Shetty, K Karve, R Prasad, **P Iyer** and A.Dutt "Integrated Genomic Characterization of Head & Neck Cancer Cell Lines Derived from Indian Patients" at The Global Cancer Genomics Consortium Second Annual Symposium: Genomics Medicine in Cancer Research, 2012, ACTREC, India.**(Best Poster award)**
- R Chaubal, **P Iyer**, V.Kulkarni, I Mittra, A.Dutt A Next Generation Sequencing approach to determine traces of human genetic material in NIH3T3 cells transformed by chromatin & DNA fragments extracted from human cancer patients at The Global Cancer Genomics Consortium Second Annual Symposium: Genomics Medicine in Cancer Research, 2012, ACTREC, India **(Best Poster award)**

Signature of student

Date 16/4/17

Doctoral Committee:

S. No.	Name	Designation	Signature	Date
1.	Dr. Sorab N Dalal	Chairperson		17/4/17
2.	Dr. Amit Dutt	Guide & Convener		18/4/17
3.	Dr. Prasanna Venkatraman	Member		16/4/17
4.	Dr Shailesh Shrikhande	Member		16/4/17

Forwarded through:

18/4/17

Dr. S.V. Chiplunkar

Director, ACTREC

Chairperson, Academic & Training Programme

ACTREC

Dr. S. V. Chiplunkar
Director
Advanced Centre for Treatment, Research & Education in Cancer (ACTREC)
Tata Memorial Centre
Khanghar, Navi Mumbai 410210.

To

Dean-Academic

PROF. K. S. SHARMA
DIRECTOR (ACADEMICS)

Dr. K. Sharma,Director, Academics,

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ACTREC	Advanced Centre for Treatment Research and Education in Cancer
COSMIC	Catalogue of Somatic Mutations in Cancer
dbSNP	Single Nucleotide Polymorphism Database
DNA	Deoxyribonucleic acid
EGFR	Epidermal Growth Factor Receptor
FFPE	Formalin-Fixed, Paraffin-Embedded
HBNI	Homi Bhabha National Institute
HPV	Human Papillomavirus
ICGC	International Cancer Genome Consortium
IRB	Institutional Review Board
KRAS	Kirsten rat sarcoma viral oncogene homolog
mL	Millilitre
TCGA	The Cancer Genome Atlas
TMC	Tata Memorial Centre
TMH	Tata Memorial Hospital
TMC-SNPdb	Tata Memorial Centre - Single Nucleotide Polymorphism Database
TSCC	Tongue Squamous Cell Carcinoma
NGS	Next-Generation Sequencing
SNP	Single Nucleotide Polymorphism
GBC	Gallbladder cancer
LDL	Low Density Lipoprotein
ApoE	Apolipoprotein E
ApoB	Apolipoprotein B
ERBB2	erbb receptor tyrosine kinase 2
HER2	Human epidermal growth factor receptor 2
APBDJ	Anomalous junction of the pancreaticobiliary duct
CDKN2A	Cyclin dependent kinase inhibitor 2A
BMI	Body mass index
CASP8	Caspase 8
PTGS2	Prostaglandin endopreoxide synthase 2
TLR	Toll like receptor
SHBG	Sex Hormone binding globulin
NAT2	N-acetyl trasnferase 2
XRCC1	X-Ray Repair Cross Complementing 1
IDH1	Isocitrate dehydrogenase 1
COX2	Cyclooxygenase 2
FHIT	Fragile histidine triad

SUMMARY

Gallbladder cancer is a rare neoplasm. In India, gallbladder cancer is a major problem in the northern part of the country with its highest incidence of 22/1,00,000 women and risk factors such as gallstones, female gender, and genetic alterations. Genome-wide studies are far in dismal. There is an unmet need to understand the genomic landscape of Indian gallbladder cancer genome. I interrogated the coding region of gallbladder cancer genome of 27 samples (10 paired and 7 unpaired tumors) using whole exome sequencing at an average coverage of 100X and above. First, I analyzed the exome sequencing data for identifying *Salmonella* sequences as well as the presence of 143 HPV types using computational subtraction based HPVDetector tool. I found an interesting association of typhoidal *Salmonella* strains in 11 of 26 gallbladder cancer samples and non-typhoidal *Salmonella* species in 12 of 26 samples, 6 samples were co-infected with both. I observed co-occurrence of *TP53* alterations in 4 of 16 *Salmonella* positive samples while I did not observe *TP53* alterations in *Salmonella* negative samples.

Secondly, my analysis of the whole exome data led to the identification of 383 somatic alterations across 17 tumors, which includes an average 112 synonymous, 245 missense, 8 nonsense, 8 indels and 8 splice site changes. I found recurrent alterations in *TP53*, *CTNNB1*, *SF3B1*, *ATM*, *AKAP11* and other genes by exome sequencing analysis. Of specific mention, my work has led to the discovery of a recurrent activating *ERBB2* V777L mutation in 6 of 44 gallbladder cancer samples with an overall mutation frequency of 13%; along with *KRAS* G12V and G13D mutation in 2 of 4 gallbladder cancer cell lines. I demonstrated that treatment of these cells with either *ERBB2*-specific or *EGFR*-specific shRNA or with irreversible *EGFR* inhibitor BIBW-2992 inhibits transformation and survival along with migration and invasion characteristics of gallbladder cancer cells with wild-type *KRAS* or those harboring *KRAS*(G13D) but not *KRAS*(G12V) mutation.

In overall, I present the first landscape of somatic alterations in Indian gallbladder cancer genome and identification of non-typhoidal *Salmonella* species along with co-occurrence of *TP53* alterations that could aid in the treatment of gallbladder cancer. More importantly, my study implicates *ERBB2* as a novel therapeutic target in gallbladder cancer, and puts forward the first evidence that the presence of *KRAS* G12V but not *KRAS* G13D mutation may preclude patients to respond to anti-EGFR treatment in gallbladder cancer, similar to the clinical algorithm commonly practiced to stratify patients for anti-EGFR treatment in colorectal cancer.

1. CHAPTER1: INTRODUCTION AND REVIEW OF LITERATURE

1.1 Human cancer and genomics

For the past two decades, we have witnessed a tremendous advancement in understanding of the pathogenesis of cancer. The process of carcinogenesis arises through a multistep, mutagenic process whereby cancer cells acquire common properties such as unlimited growth potential, self-sufficiency in growth signals, and resistance to antiproliferative cues and apoptotic cues[2]. Many of these traits have been bought by series of accumulating genetic alterations that involve gain-of-function mutations, amplification, and/or overexpression of key oncogenes together with the loss of function mutation, deletion and/or epigenetic silencing of key tumor suppressors[3]. 60% of cancer deaths are constituted by malignancies of five organs i.e. lung, liver, stomach, head & neck and colon worldwide [4]. India also matches the global pattern of these cancer types however there is a higher proportion of head & neck and cervical cancer in India (GLOBOCAN, 2012; http://globocan.iarc.fr). Chemotherapy, surgery, and radiation are the most common conventional treatment options available to the patients. However, with increasing resistance to conventional therapies, there is unmet need to identify molecular targets that could help in designing better treatment strategies for patients.

The development of technologies in analyzing nucleic acids together with advanced computational approaches has facilitated the study of cancer in a way which was previously not possible[5]. Cancer is a disease of the genome characterized by a diversity of genetic and epigenetic alterations[6]. The early efforts in the cancer genome analysis have helped in identification of new targets for cancer therapy and new insights into the relationship between specific genetic mutations and their clinical response as well as new approaches for diagnosis[5, 7]. The rapid pace of development of sequencing technologies such as next-generation sequencing technologies (NGS) has impacted the field of cancer genomics while dramatically reducing the cost of data production[6]. These developments have further

motivated largescale coordinated cancer genomic efforts (TCGA, ICGC) to perform comprehensive profiling of tumors and enable genome-informed personalized cancer medicine[8].

1.2 Genomics of rare cancer types

The large-scale genome characterization efforts have been focused on most common cancer types such as brain, lung, head and neck, breast and so on. Very few genomic efforts have been concentrated on rare cancer types. One of the rare cancer types is a group of cancers of the biliary tract that arise from the biliary epithelium. The biliary tract cancers are further classified into three major types as intrahepatic (intrahepatic cholangiocarcinoma), (extrahepatic cholangiocarcinoma) and gallbladder carcinoma. These cancers are generally very aggressive in nature. Patients present their cases in later stages and systematic chemotherapeutic regimens generally have dismal response rates. Hence, the treatment strategies are often palliative in nature[9]. Due to the rarity of these malignancies worldwide except for few regions, the treatment strategies for these cancers has been identical. With recent developments in the molecular techniques and NGS, it has been demonstrated that each tumor type has a unique genomic landscape[10]. Among the tumor types of the biliary tract, gallbladder cancer is one the most common and aggressive biliary tract cancer. The genomic landscape of gallbladder cancer is not well characterized[11]. As a result, identification of molecular targets may be important for genomics-guided precision medicine approaches as well as biomarker-driven clinical trial design.

1.3 Gallbladder Cancer

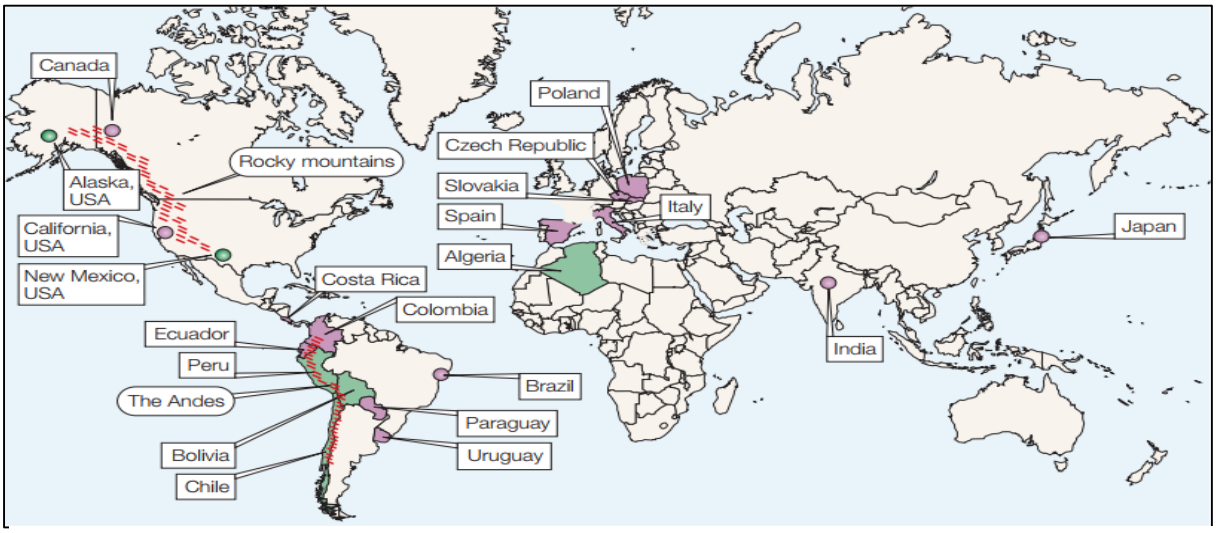
1.3.1 Definition and Epidemiology of GBC

Gallbladder cancer is one of the most common malignancies of the biliary tract and is ranked fifth among the gastrointestinal cancers worldwide. Gallbladder cancer (GBC) is female gender biased and mostly affects at advanced ages[12, 13]. GBC is regarded as highly lethal

diseases of the biliary tract with 5-year survival estimates less than 5%. The global occurrence of gallbladder cancer varies with different regions and ethnicities, reaching epidemic levels in some regions and ethnicities. The basis of this variability could be attributed to different geographical conditions, environmental exposures and genetic predisposition to carcinogenesis[1]. GBC develops over a period of 5 to 15 years with metaplasia to dysplasia, carcinoma-in-situ and then, invasive cancer. The prognosis of GBC is dismal and surgical resection is a current curative option for patients with GBC. However, less than 10% of the patients are presented at the resectable stage, while 50% of patients have lymph node metastasis[14]. Epidemiologically, mortality rates of gallbladder cancer are higher in countries with higher incidence.

1.3.2 Epidemiology of gallbladder cancer

Worldwide, higher rates of GBC are observed in Mapuche Indians of the Chile followed by North and South American Indians. Mortality rates are high in American Indians: 15.5 per 100,000 women vs. 7.5 per 100,000 in men from La Paz, Bolivia and 11.3 per 100,000 in women vs. 4 per 100,000 in men from New Mexico. There are other high-risk areas include Eastern Europe (14/100,000 in Poland), Northern India (22.5/100,000 in Delhi), South Pakistan (11.3/100,000), Israel (5/100,000) and Japan (7/100,000). The incidence in China,



I-Figure 1: Overall worldwide variation in incidence of gallbladder cancer.
(Adapted from Wistuba et al., 2004 Nature reviews Cancer) The areas shown in green show very high incidences while the areas shown in purple indicate moderately high incidences [1]

especially in Shanghai, have doubled over the years[15]. GBC is relatively low in the United States and Mediterranean countries (UK, France, and Norway)[1, 16]. In the United States, Hispanic women and men have a higher incidence of GBC than non-Hispanic men[1].

A retrospective study carried out in North Central India during 2007-2008 identified gallbladder cancer to be at a fourth position after head and neck, breast and cervical cancer. Within the Indian population highest incidence has been reported in northern cities(3.7 per 100 000 for male and 8.9 per 100 000 for female and in Bhopal it is 1.6, 2.5 per 100 000 for male and female, respectively) as compared to southern cities (eg in Chennai, the incidence is 0.5 per 100 000 for male and 0.8 per 100 000 for 100 000 for female and in Bangalore, incidence for male is 0.6 per 100 000 and for female it is 0.7 per 100 000 population) female and in Bangalore, (incidence for male is 0.6 per 100 000 and for female it is 0.7 per 100 000 population)[12, 17].

1.3.3 Unmet need to treat gallbladder cancer in India

Gallbladder cancer is very common in the northern and north-eastern states of India. The mean survival rate with advanced stages of cancer is 6 months with a 5-year survival rate of less than 5%[1]. Since early diagnosis of the cancer is difficult, most of the gallbladder cancers (95%) are detected at advanced stages where curative resection is not possible. Of the remaining 5% who have stage I or II diseases, cholecystectomy is performed for symptomatic gallstones. Chemotherapy and radiotherapy are given for unresectable cancers, however, the survival frequencies are low in such cases. Few studies from India have shown the association of ABC transporter genes like *ABCB4*, *ABCB11*, *CYP7A1*, *ApoB*, *ApoE* and LDL receptor polymorphisms in gallstone diseases is also implicated in gallbladder cancer[18]. Gallstone disease with typhoidal infections is an important risk factor for gallbladder cancer, is also common in northern India[19]. However, secondary prevention by prophylactic cholecystectomy is controversial, as there is no evidence to support it[20]. There are different studies in India investigating the role of pesticides, trace elements, bacteria in bile, bile

composition, chronic typhoid carriage, hormonal factors, and genetic factors like *KRAS* alterations in the causation of gallbladder cancer[21-25]. However, these studies are limited by the fewer number of samples and systematic genome-wide studies are dismal. Lack of systematic clinical trials in India investigating the role of target therapies in gallbladder cancer. Hence there is an unmet need to study the cancer type in a systematic and comprehensive at a genome-wide scale.

1.3.4 Application of next generation sequencing in gallbladder cancer

The recent developments in sequencing technologies have helped in molecular characterization of several rare cancer types. Biliary tract cancers are one of the rare cancer types which are comprised of intrahepatic cholangiocarcinoma, extrahepatic cholangiocarcinoma, and the gallbladder carcinoma. Genomic profiling of gallbladder cancers using mass spectrometry and targeted sequencing technologies identified mutations in *IDH1* and *TP53* as the most recurrently altered genes in gallbladder cancer[26]. Another comprehensive study of 57 samples by whole exome sequencing and ultra-deep sequencing of cancer-related genes identified mutations in TP53(47%), *KRAS*(7.8%), *ERBB2*(9.8%), *ERBB3*(11.8%) and also the authors identified ERBB pathway as the most recurrently mutated pathway in gallbladder cancer affecting up to 36.8% of GBC samples. Further using multivariate analysis the authors show that the cases with ERBB pathway alterations have the worse prognosis[27]. Another study using ion torrent based amplicon sequencing of 46 genes in 9 gallbladder cancer samples identified one patient with activating *ERBB2* alteration and rest of the other samples with ERBB2 amplification. Patient with the *ERBB2* mutation had a mixed response to the pan-*HER2* inhibitor[28]. Using targeted sequencing of 236 cancer-related genes of 9 gallbladder cancer patients, recurrent mutations were observed in *TP53*, *ARID1A*, and *KRAS*. Transcriptome sequencing of 8 gallbladder cancer patients and 3 normal samples identified 519 genes to be differentially expressed and identified liver X receptors and farnesoid receptors to be top canonical pathways to be deregulated in gall bladder

cancer[29]. Another RNA-seq study of 3 tumors and adjacent normal samples identified 161 differentially expressed genes and the authors observed enrichment of genes related to pathways such as cell cycle, enzyme modulators, and pathways in cancer[30]. Despite the higher prevalence in India, no genome-wide studies have been done using next-generation sequencing technologies.

1.3.5 The landscape of known genomic alterations in gallbladder cancer.

The most common alteration reported in gallbladder cancer which occurs earlier in the dysplasia to carcinoma sequence is *p53* alterations. The most common alterations are in exon 5 and exon 8. Most of the p53 alterations are missense alterations that increase the stability of the protein. The frequency of alterations reported is above 50%. Loss of heterozygosity of p53 occurred earlier and more frequently than protein overexpression[31]. The frequency of *KRAS* alterations is quite variable in gallbladder cancer ranging from 39-59%. Most of the *KRAS* alterations have been reported in codon 12. Higher frequency of *KRAS* alterations has been reported in patients having the anomalous junction of the pancreaticobiliary duct(APBDJ) suggesting reflux of pancreatic juice might contribute to carcinogenic process[1, 31, 32]. Inactivation of *CDKN2A* has been observed in half of the GBC cases that occur by a combination of mutations, deletion, and abnormal hypermethylation. Increased expression of *CDK4* and *cyclin D1* detected by immunostaining in 41-60% of samples has been noted in the progression of gallbladder cancer[33]. Cyclooxygenase 2 (*COX2*) overexpression has been observed to occur earlier in the cascade of gallbladder carcinogenesis[34]. Loss of heterozygosity and SNPs have been observed in *DCC* (deleted in colorectal carcinoma) gene in gallbladder cancer and are considered as an early event in the cascade of gallbladder carcinogenesis[35]. Reduction of *FHIT* (Fragile Histidine Triad) expression has been observed in the progression of gallbladder cancer from dysplasia to invasive carcinoma[36]. LOH of 3p and 9p has been related to the progression of gallbladder cancer. Also, increasing LOH proportions has been observed on chromosomes 3p, 9q, 8p, and 22q in normal,

dysplasia and malignant tissue[37]. Mismatch repair gene alterations are frequently reported in gallbladder cancer. High-frequency microsatellite instability (MIN-H) has been reported in early and late gallbladder cancers[31]. *HER2* alterations have been reported in Chinese and Japanese population. Overexpression of *HER2* has been reported in 30-60% of GBC cases and gene amplification is found in 70% of the cases[38]. In a mouse model system (BK5-ERBB2 mice model) overexpression of *HER2* in the basal layer of the biliary tract, epithelium leads to the development of gallbladder carcinoma by 3 months of age. However, the mouse gallbladder tumors were different from human tumors characterized by adenoma precursors and papillary structures that filled the gallbladder lumen[39]. Expression levels of *HER2* varied depending on the increasing grade of the tumor.

1.3.6 Targeted therapy in gallbladder cancer

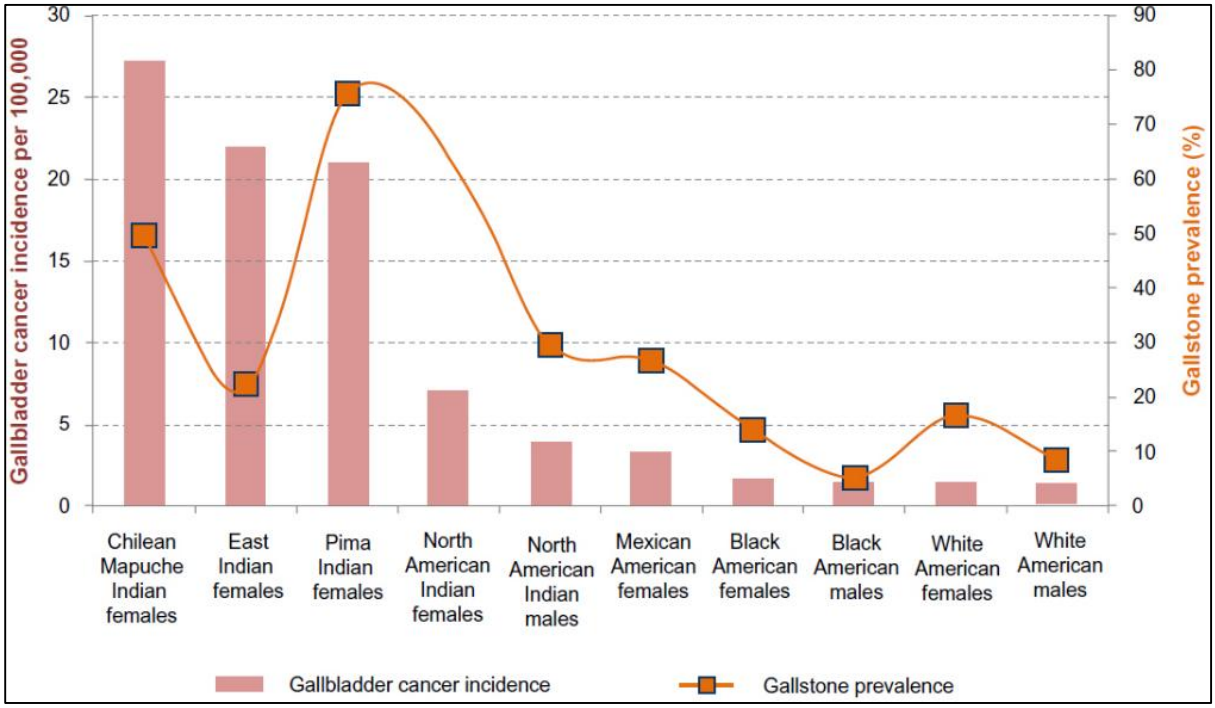
The conventional mode of treatment for gallbladder cancer is surgery for resectable cancers and there are gemcitabine and fluoropyrimidine-based chemotherapeutic regimens for unresectable cancers. Few reports have evaluated the effect of targeted therapies along with the conventional treatment. Some studies suggest the benefit from blockade of *EGFR* by oral tyrosine kinase inhibitor erlotinib or anti-EGFR monoclonal antibody cetuximab [40, 41]. Another study evaluated the benefit of the efficacy of bevacizumab, a monoclonal antibody targeting VEGF along with erlotinib in a phase II trial. Overall, there was a partial response among nine patients with six patients sustained beyond three weeks, and with an overall median response of 8.4 months. The other half of the patients had a stable disease[42]. Likewise, other inhibitors such as sunitinib and sorafenib have their modest benefit in biliary tract cancers[43, 44]. Hence, targeted therapy has shown some promise in gallbladder cancer, however with better screening of patients with alterations the response rates may be improved in the treatment of gallbladder cancer.

1.3.6 Risk factors for gallbladder cancer

The major risk factors for gallbladder cancer are chronic inflammation, geographical and ethnic variation, increasing age, female gender, low socioeconomic status, genetic predisposition, infections, low cholecystectomy cases and exposure to certain chemicals. Some of the risk factors are discussed below:

1.3.6.1 Gallstones and cholecystitis

Gallstones are the major risk factor for gallbladder cancer accounting for 60-90% cases in different regions of the world. A very common phenomenon is chronic inflammation due to gallstone irritation of the gallbladder wall, which is observed in 85% of the cases. Patients with gallstone have a higher incidence of gallbladder cancer[15]. The association between with cholelithiasis and cancer may explain why female gender, multiparity and increased body mass indices (also a risk for gallstone formation) are associated with developing carcinoma of the gallbladder. The size of the gallstones (>3cm) and duration of stones in the gallbladder have a stronger association with a pathogenesis that culminates in cancer [45, 46]. Though gallstones are associated risk factor, which is likely facultative rather than causative prophylactic cholecystectomy is not favored for clinically silent gallstones except for large stones and elderly patients with cholethiasis[47]. The decrease in incidence and mortality of gallbladder cancer began decades before the introduction of laparoscopic cholecystectomy and apparently stabilized in the past decade. There is no temporal relationship known to exist between laparoscopic cholecystectomy rate and the incidence and mortality rates of gallbladder cancer.



I-Figure 2: Incidence of gallbladder cancer and gallstones in different ethnicities. (Adapted from Hundal et al 2014 Clinical Epidemiology). There is a co-incidence of gallstones presence and incidence of gallbladder cancer in different ethnicities[48]

.

1.3.6.2 Porcelain gallbladder

Chronic inflammation can lead to calcification of the gallbladder known as porcelain gallbladder. The gallbladder wall becomes fragile and acquires bluish color hence the term porcelain is used. Even though the frequency is uncommon(less than 1%), tends to occur in older women of the sixth decade. The porcelain gallbladder is frequently (average 25% 12-61%) associated with gallbladder cancer in most but not in all reports. Only those with stippled calcification are pre-malignant while complete calcification is less likely to be associated with carcinoma[49]. Therefore, gallbladders with stippled calcification or multiple punctate calcifications in the glandular spaces of the mucosa must be removed prophylactically[48].

1.3.6.3 Age and Gender

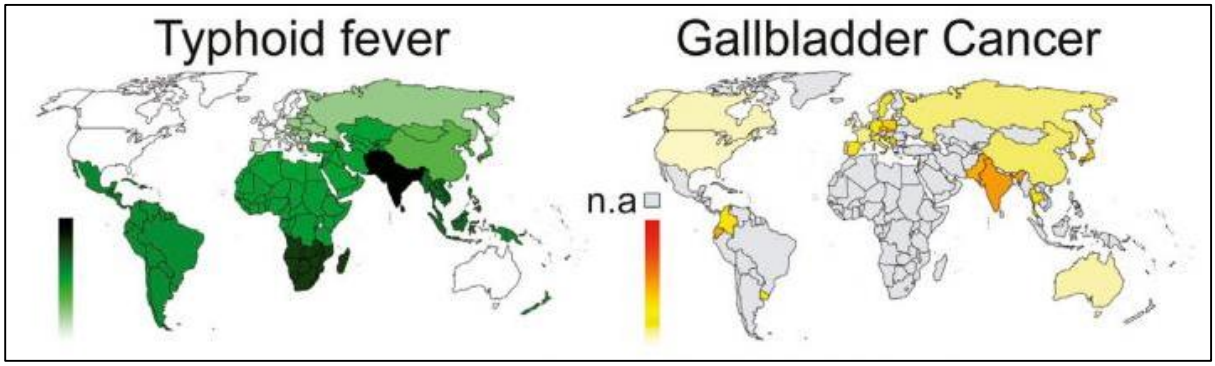
The rates of gallbladder cancer tend to increase with increasing age. The disease is more common among elderly women and more than 90% of patients are above the age of 50 years.

1.3.6.4 Diet and Obesity

It has been observed that there is an increased risk of gallbladder cancer with consumption of high-calorie diet, high carbohydrate, and increased oily foods. For every 5-point increase in BMI, there is an increase in relative risk of developing gallbladder cancer in women by 1.59 and 1.09 in men. High consumption of red chili pepper has been found to be increasing the risk of development of gallbladder cancer[50]. Intake of red meat is also found as a risk factor for gallbladder cancer while inverse correlation exists with intake of vegetables, vitamin E, vitamin C and fiber with gallbladder carcinogenesis[15]. In an epidemiological study, the authors estimated the contents of aflatoxins B1, B2 D1 and D2 in red chilies used by Bolivian and Peru population suffering from GBC. Aflatoxin is a liver carcinogen that is associated with the proliferation of bile duct epithelium in humans and animals. The levels of aflatoxins were high in GBC patients which suggest a possible association of aflatoxin with gallbladder cancer[1].

1.3.6.5 Bacterial infections

Several reports suggest that chronic infection of the gallbladder with *Salmonella typhi* increases the risk of developing gallbladder cancer[51]. The infection of *Salmonella* is via fecal-oral route through contaminated food and water. Though the bacteria is cleared by neutrophils and macrophages, few bacteria reach the gallbladder and establish a carrier state in the gallbladder. The bacteria produce toxins which may be involved in persistent infection of the gallbladder leading to cancer. Epidemiological findings indicate that those who become carriers of *S.typhi* have 8.47 times increased the risk of developing gallbladder cancer than those who have acute typhoid and cleared the infection[52]. Few reports also indicate that infection with Helicobacter pylori may lead to gallbladder cancer[53].



I-Figure 3: Worldwide incidence of typhoid and gallbladder cancer.
(Adopted from Scanu et al, 2015 Cell host, and Microbe) There is a close association between the incidence of typhoid fever and gallbladder cancer.

1.3.6.6 Genetic polymorphisms in gallbladder cancer

A large number of GWAS (Genome-wide association studies) studies have identified the association of commonly occurring polymorphisms with gallbladder cancer. Several reports suggest the association of SNPs in DNA repair genes with increased risk of gallbladder cancer. *OGG1* is a DNA repair enzyme in humans. SNPs identified in patients with gallstones in this gene (*OGG1* Cys/Cys genotype) had lower DNA repair activity and conferred a higher risk of gallbladder cancer[18]. Similarly, another SNPs identified in another DNA repair enzyme *FEN1* (Flap endonuclease 1) also increased the risk of gallbladder cancer. A significant association was observed between body mass index (BMI) and *CYP1A1* rs2606345 SNP on GBC risk, with non-obese (BMI<23 kg/m^2) carriers of the T allele having a 3.3-fold risk (95% CI=1.8–6.1). Polymorphisms in Wnt signaling genes such as *APC* rs11954856, *GLI-1* rs2228226, and *AXIN-2* rs4791171 were found to be associated with poor survival in advanced GBC patients[54]. Recent reports suggest there is a strong association of SNPs in *ABCB1* and *ABCB4* with an increased risk of gallbladder cancer[55]. Other SNPs that have been reported so far from different studies are *CYP7A1,* Apolipoprotein X-baI, *CCR5 Delta32, XRCC1, ABCG8,* Cholecystokinin receptor A, *NAT2, SHBG, TLR, CASP8, PTGS2* showed a significant association with increased risk of gallbladder cancer[56].

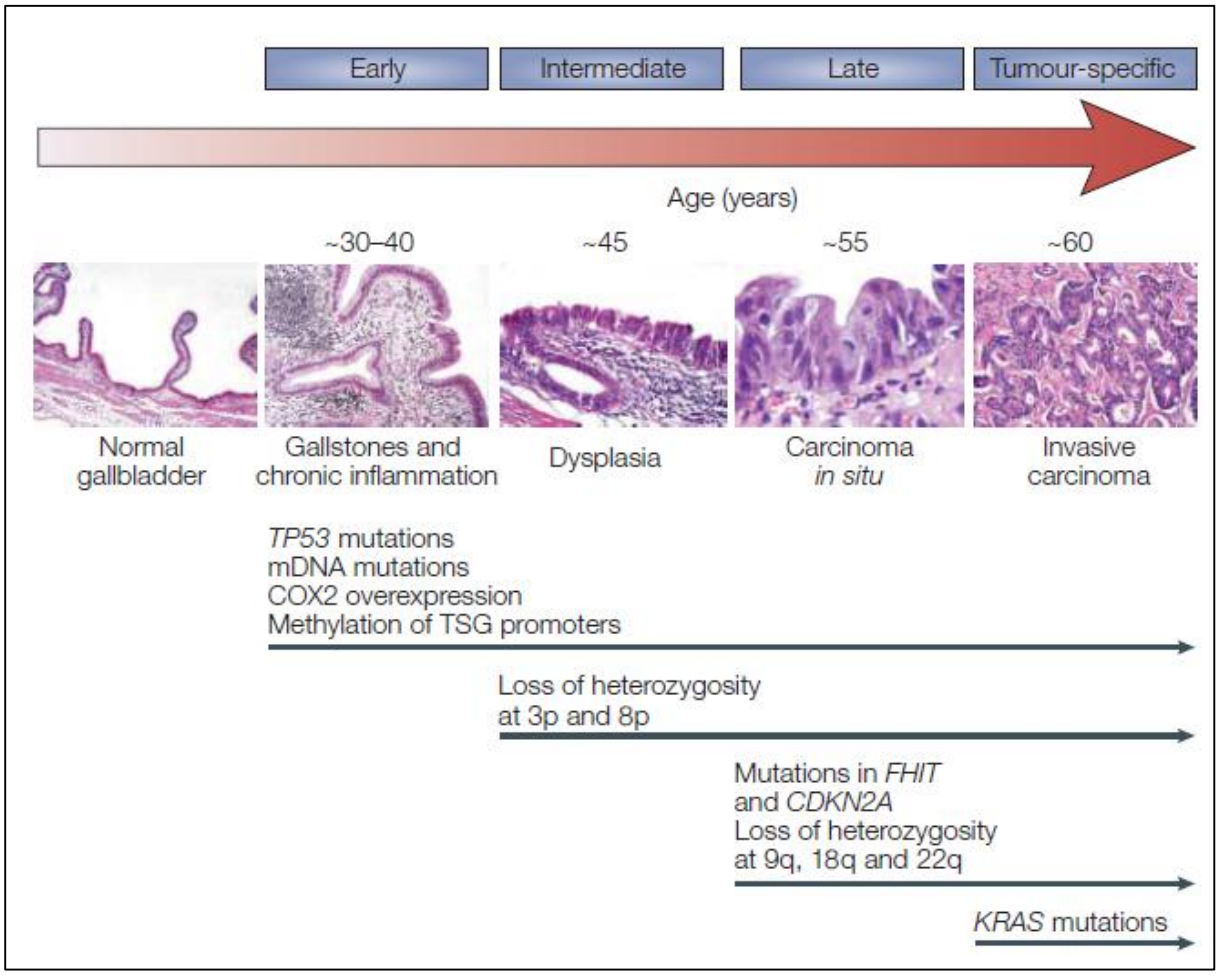
1.3.6.7 Environmental effects

Though there is no conclusive evidence to support the association of environmental pollutants with gallbladder cancer. Few reports suggest the higher biliary concentration of cadmium, chromium, and lead were found in the bile of cancer patients than patients with gallstones. Increased expression of metallothionein has been observed in GBC patients which may play a role in gallbladder carcinogenesis.

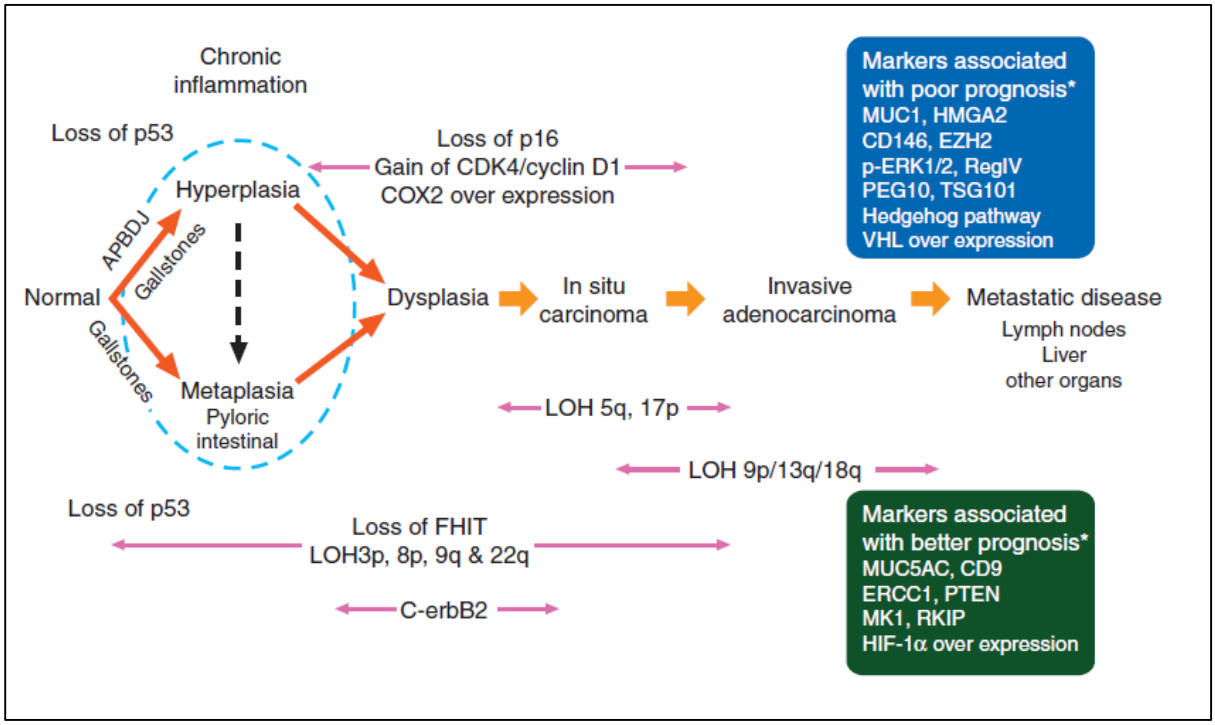
1.4 Molecular pathology of gallbladder cancer

There are proposed two pathways in the pathogenesis of gallbladder pathogenesis viz dysplasia-carcinoma sequence and adenoma-invasive carcinoma. In the first model, normal epithelium changes to dysplasia due to chronic irritation or inflammation thus progressing to carcinoma in situ and eventually leading to invasive cancer. In the second model, gallbladder polyp is formed by the initial glandular proliferation of the epithelium, malignant transformation occurs within this initially formed an initial benign mass[31]. In addition to the previously proposed models, the updated model "gallbladder carcinogenesis and dissemination model incorporates the course of disease after the development of invasive carcinoma and addition of new molecular markers that could be used for diagnosis or response to therapy[31]. Chronic inflammation plays an important role in gallbladder cancer. There is a reduction in expression of inflammatory markers such as *COX2*, *EGFR a*nd other markers at advanced stages of cancer. The protective influence of the expression of markers in the advanced stages is still needed to be ascertained. Most dysplasias and carcinoma in situ are observed after cholecystectomy when the entire lesion is removed and evidence suggest that progression could occur from precursor lesions to infiltrating carcinoma. About 90% of the gallbladder cancers are adenocarcinoma, of which 10-37% of carcinomas cannot be identified with certainty on gross examination since the macroscopic findings are similar to chronic cholecystitis[57]. The majority of the carcinomas originate in the fundus (60%), body (30%) and neck (10%)[58]. Most of the gallbladder cancers are well to moderately differentiated adenocarcinomas. Few previously reported histological studies have identified

papillary neoplasms in 0.4% of the cholecystectomies and 6-7% of the invasive carcinomas. Papillary neoplasms are associated with favorable prognosis as compared to non-tumoral counterparts[57, 58]. The favorable prognosis is attributed to their outward growth forming a polyp or mass and delayed invasion into the gallbladder wall. A staging system was proposed in which Stage I is limited to the mucosa; Stage II is limited to the muscular layer; Stage III is limited to the perimuscular layer; Stage IV is limited to the lymph nodes and Stage V has hepatic or other distant metastasis[13].



I-Figure 4: Dysplasia to invasive carcinoma model of GBC involving sequential histopathological and molecular changes associated with gallstones and inflammation. (Adapted from Wistuba et al, 2004 Nature reviews cancer) The multistage process of the pathogenesis of gallbladder cancer involving series of histological changes from normal epithelium to invasive carcinoma with different molecular changes at different stages[1].



I-Figure 1: Gallbladder carcinogenesis and dissemination model.
(Adapted from S.G Barreto et al, 2014 Ann Oncol) Gallbladder carcinogenesis and dissemination model The multi-step process of carcinogenesis of gallbladder involving a series of changes from normal epithelium to invasive carcinoma via two pathways metaplasia/hyperplasia as well as dysplasia and metastasis to lymph nodes, liver and other distant organs[31]

1.5 Research objective

Rationale

Gallbladder cancer is an aggressive disease with poor prognosis. Currently, surgery is the only curative mode of treatment for the disease. Chemotherapy and radiotherapy are used as adjuvants; however, there is very little effect on survival of the patients. Moreover, molecular targets have not been identified yet. My research proposal takes the advantage of combining genomic approaches followed by functional validation of the alterations discovered to build up a molecular framework of therapeutically relevant alterations thereby benefit the patients with the deadly disease. The objectives of the thesis are listed below:

Objectives

1) Apply next-generation sequencing analysis of gallbladder cancer to identify pathogenic sequences in gallbladder cancer using computational subtraction method.

2) Apply next-generation sequencing analysis of gallbladder cancer to identify genes whose somatic genomic alterations suggest the properties of driver oncogenes. In a more directed approach, we will sequence exome, from gallbladder tumors of Indian origin cases.

3) We will test candidate oncogenes identified in Aim 2 by gain-of-function assays for cellular transformation and activation of known oncogenic signaling pathways.

CHAPTER 2: NON TYPHOIDAL *SALMONELLA* TRACES IN GALLBLADDER CANCER (an excerpt; as published in *BMC Infectious Agents and Cancer (2016);11:12*

Abstract

Background

We earlier proposed a genetic model for gallbladder carcinogenesis and its dissemination cascade. However, the association between gallbladder cancer and 'inflammatory stimulus' to drive the initial cascade in the model remained unclear. A recent study suggested infection with *Salmonella* can lead to changes in the host signalling pathways in gallbladder cancer.

Findings

We examined the whole exomes of 26 primary gallbladder tumour and paired normal samples for presence of 143 HPV (Human papillomavirus) types along with 6 common *Salmonella* serotypes (*S. typhi* Ty2, *S. typhi* CT18, *S. typhimurium* LT2, *S. choleraesuis* SCB67, S. *paratyphi* TCC, and *S. paratyphi* SPB7) using a computational subtraction pipeline based on the HPVDetector, we recently described. Based on our evaluation of 26 whole exome gallbladder primary tumors and matched normal samples: association of typhoidal *Salmonella* species were found in 11 of 26 gallbladder cancer samples, and non-typhoidal *Salmonella* species in 12 of 26 gallbladder cancer, with 6 samples were found co-infected with both.

Conclusions

We present the first evidence to support the association of non-typhoidal *Salmonella* species along with typhoidal strains in gallbladder cancer. *Salmonella* infection in the chronic carrier state fits the role of the 'inflammatory stimulus' in the genetic model for gallbladder carcinogenesis that may play a role in gallbladder cancer analogous to *Helicobacter pylori* in gastric cancer.

2.1 Introduction

2.1.1 Infections in cancer

Worldwide, it has been estimated that 20% of cancers are attributed to infectious agents. It's known that viral and bacterial pathogens have been postulated to play an important role in the development of cancer [59, 60]. Among the pathogens, viruses account for the majority of malignancies from a universal perspective. There are seven oncogenic viruses (hepatitis B and C (HBV and HCV), human papillomavirus (HPV), Epstein-Barr virus (EBV) and human T cell lymphoma virus 1(HTLV-1), Merkel cell polyomavirus, Kaposi's sarcoma virus, one oncogenic bacterium (*Helicobacter pylori*) play a tumorigenic role in the development of cancer[61]. The vital portion of the viral genome can be found in a cancer cell resulting in the expression of viral genes disrupt cell cycle, inhibit apoptosis thus contributing to the cellular process of immortalization. In contrast pathogens like bacteria (*H.pylori, O.viverrini, and S.typhi)* produce a chronic inflammatory state that leads to the production of cytokines, prostaglandins which can result in deregulation of the immune system and neovascularization[60]. All these pathogens have been associated with the carcinogenic process, however, the molecular mechanism has not been elucidated so far.

2.1.2 Infections in gallbladder cancer

Gallbladder cancer is one of the most common cancers of the biliary tract. One of the major risk factors is infections with *Salmonella* bacteria[51]. Several epidemiological studies from India, especially from the northern part of the country have reported a chronic carriage of typhoid and gallbladder cancer[62]. Typhoid is caused by *Salmonella enterica* serovar Typhi. *Salmonella* typhi enters the bloodstream through contaminated food or water. After crossing the intestinal epithelial barrier, *Salmonella* is phagocytosed and are systematically spread to produce acute disease. About 3-5% of typhoid patients become chronic carriers, with gallbladder providing the niche for its persistence[63]. Chronic carriers are symptomatic and have an approximately 8-fold risk of developing gallbladder cancer than the non-carriers[64].

Few reports suggest the co-operative relationship between gallstones and *Salmonella* typhoidal infections in the carcinogenesis of gallbladder cancer, however, the cause and effect relationship is not ascertained[65]. The proposed mechanism of tumorigenicity by *Salmonella* bacteria is the production of b-glucuronidase that result in deconjugation of toxins and bile acids which in turn lead to carcinogenesis[66]. Other evidence indicates that products of degradation of bile acids by the bacteria may contribute to tumorigenesis[67, 68]. Recent evidence shows that *Salmonella* enterica infection induces transformation in predisposed murine gallbladder organoids, fibroblasts with *TP53* mutations and *c-MYC* amplification by activation of MAPK and AKT pathways[19]. The typhoidal *Salmonella* was strongly associated with gallbladder cancer however non-typhoidal *Salmonella species (S. Typhimurium and S. Choleraesius)* that elicits a stronger immune response is linked with the systemic illness (gastroenteritis) have as yet not associated with gallbladder cancer[69]. We proposed a gallbladder cancer carcinogenesis model based on current understanding of tumor biology[31]. However, the lacunae in the model is the driving force behind inflammation-related changes is not ascertained. Here we examined the exomes of primary gallbladder tumor and paired normal samples for the presence of 6 common *Salmonella* serotypes with available genome information (S. *typhi* Ty2, S. *typhi* CT18, S. *typhimurium* LT2, S. *choleraesuis* SCB67, S. *paratyphi* TCC, and S. *paratyphi* SPB7) using a computational subtraction pipeline based on the HPVDetector tool.

2.2 Materials and Methods

2.2.1 Patient information:

Twenty six fresh frozen primary tumor and matched normal tissues were obtained from the tissue repository of Tata Memorial Hospital (TMH). The Institutional Review Board (IRB) and the Ethics Committee (EC) of Tata Memorial Centre (TMC) - Advanced Centre for Treatment, Research and Education in Cancer (ACTREC) (Mumbai, India) approved the

project (#104). Since this was a retrospective analysis, the IRB and the EC waived the need for an informed consent. Patients were randomly selected based on the availability of fresh frozen tissues. The patient characteristics including age, gender, gallstone status and histopathology were recorded.

2.2.2 PCR analysis for *Salmonella* isolates

The PCR method used for *Salmonella* detection has been previously described [12]. Nested PCR was carried out in a 25µl volume containing 10 µl KAPA 2X ready mix master-mix (Kapa Biosystems catalog no-KK1024), 10pmol primer and 100 ng of genomic DNA. Following the first round of PCR (94 °C for 1 min, 55 °C for 1 min 15 s,72 °C for 3 min – 40 cycles) with ST1 and ST2 primers,5 µl PCR product was used as template for nested PCR using ST3 and ST4 primers (94 °C for 1 min, 68 °C for 1 min 15 s, 72 °C for 3 min – 40 cycles). We also performed validation of *Salmonella* sequences using read specific primers. The PCR conditions - 94 °C for 1 min, 59 °C for 30s, 72 °C for 45 s – for 30 cycles.

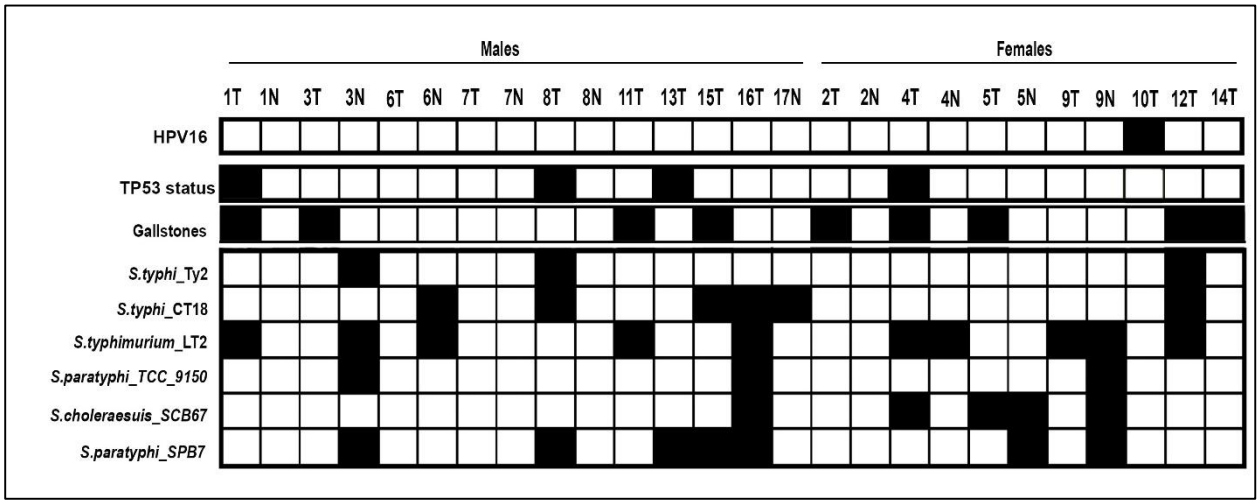
2.2.3 Sequencing and analysis

Exome capture and library preparation were performed using Agilent Sure select in-solution (low-input capture500 ng) target enrichment technology. Genomic DNA was sheared and size selected (150–200 bp) and ligated to adaptors and run on Illumina Nextseq 500 platform to generate 150 bp paired-end reads at a coverage of 100X and above. To detect *Salmonella* traces, the HPVDetector pipeline was used, as described previously [11]. Briefly, reads were aligned against six known *Salmonella* species genomes in addition of the HPVDetector dataset of 143 HPV types, as downloaded from the National Centre for Biotechnology Information (NCBI), using BWA (Burrows wheeler algorithm) aligner (v0.6.2). All reference sequences were annotated and concatenated to compose multi-fasta sequences using bio-perl modules. The alignment files were parsed using UNIX shell program to detect the types of *Salmonella* represented by at least one read that aligned to a particular *Salmonella* type with high confidence.

2.3 Results

2.3.1 HPVDetector pipeline identifies *Salmonella* sequence present in gallbladder cancer samples

We performed PCR based analysis of 26 gallbladder tumor and paired normal samples to detect the presence of *Salmonella* DNA using pan primers, as described earlier [12]. None of the gallbladder samples were found to be positive for *Salmonella* (data not shown). As a next step, whole exome data for these 26 samples (generated in-house, manuscript in make) were analyzed to detect *Salmonella* traces using HPVDetector pipeline, modified to include additional genome sequence of 6 common *Salmonella* isolates. The computational approach, in brief, subtracts all reads that align to human genome and aligns remaining reads to HPV



II-Figure 1: Profiling the occurrence of 143 HPV types and 6 *Salmonella* isolates across 26 gallbladder cancer patients.

Heat map representation of 6 *Salmonella* isolates (in row) found across 26 gallbladder samples -- 17 tumours and 9 matched normal (in the column) are shown. Solid boxes indicate the presence of reads from *Salmonella* genome in the corresponding gallbladder sample. The samples (column) have been grouped based on gender as shown above the heat map. Solid boxes in the first row indicate the presence of HPV16 sequence. Solid boxes in the second row indicate the presence of a mutation in *TP53*. Reads of HPV16 were found in 1 of 26 samples (10 T). 9 of 17 gallbladder tumour samples were associated with gallstones as shown by solid boxes in the second row. Typhoidal *Salmonella* isolates were found in 11 of 26 gallbladder cancer samples, non-typhoidal *Salmonella* isolates were found in 12 of 26 gallbladder cancer, with 6 of 26 samples co-infected with both.

and *Salmonella* reference database from NCBI. While HPV16 was detected in 1 gallbladder sample, *Salmonella* isolates were found across multiple samples: S.*typhi* Ty2 (3samples),

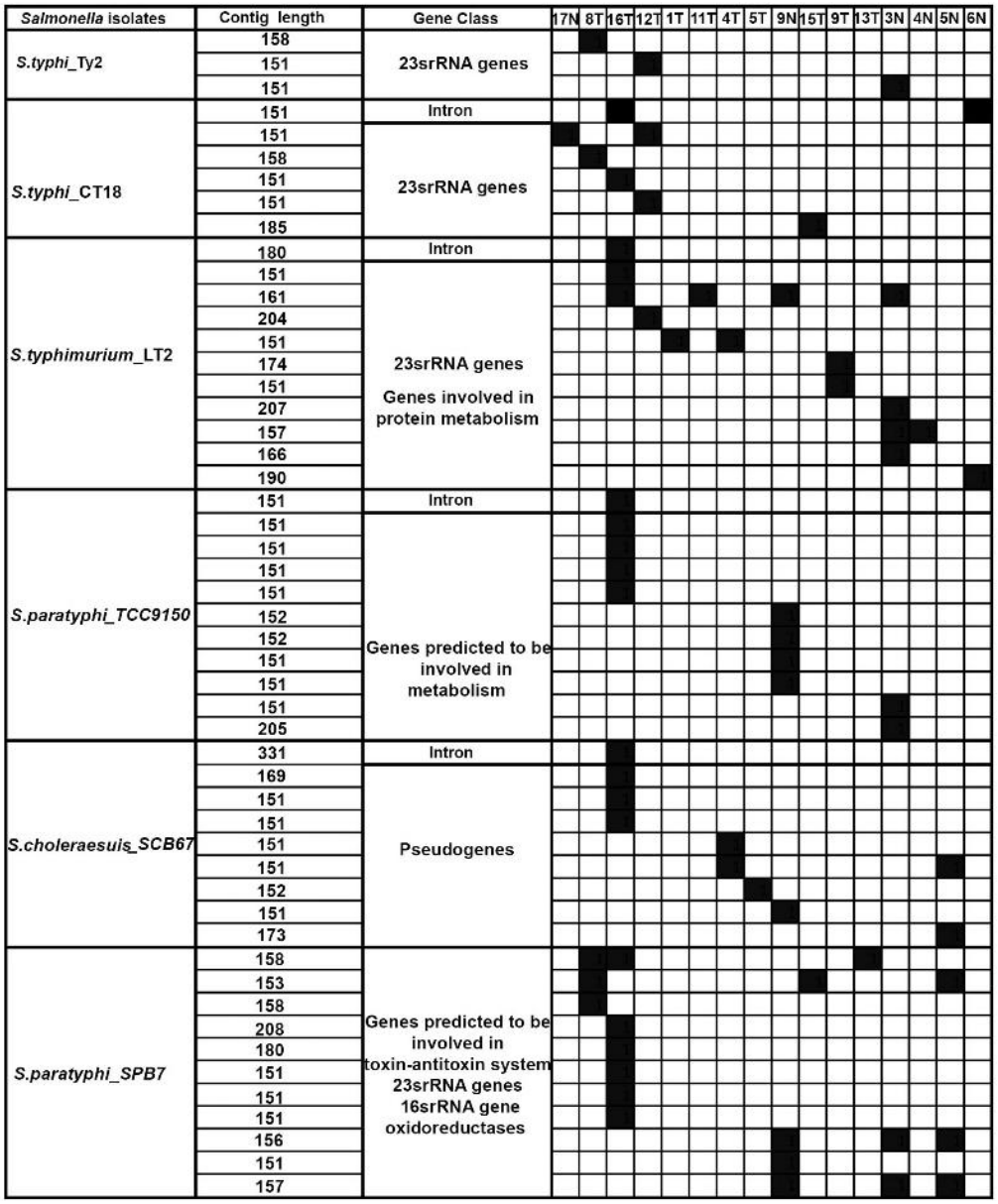
S.*typhi* CT18 (6 samples), S. *typhimurium* LT2 (10 samples), S.*choleraesuis* SCB67 (5 samples), S. *paratyphi* TCC (3 samples), and S. *paratyphi* SPB7 (7 samples). In total, *Salmonella* reads were found in 19 of 26 gallbladder tissues (tumor as well as adjacent normal tissues). Interestingly, 10 of 19 samples were co-infected with multiple isolates independent of gender or gallstone status (**Figure 1**).

2.3.2 Annotation of the *Salmonella* reads found in gallbladder cancer samples

A variable number of overlapping reads of variant lengths for each isolate were assembled into contigs based on Clustal X2 multiple alignment. The unique stretch of contigs generated were annotated based on gene annotation database of *Salmonella* isolates from NCBI (National center for biotechnology information) database. 114 reads of multiple *Salmonella* isolates were found in 19 of 26 samples analyzed.47 of 114 reads of *Salmonella* ORF (open reading frame) were identified as encoding for bacterium genes known to be involved in metabolism and those related to the toxin-antitoxin system. Rest of the reads aligned to the *Salmonella* ribosomal genes, understandably due to their relatively higher abundance (**Figure 2**)

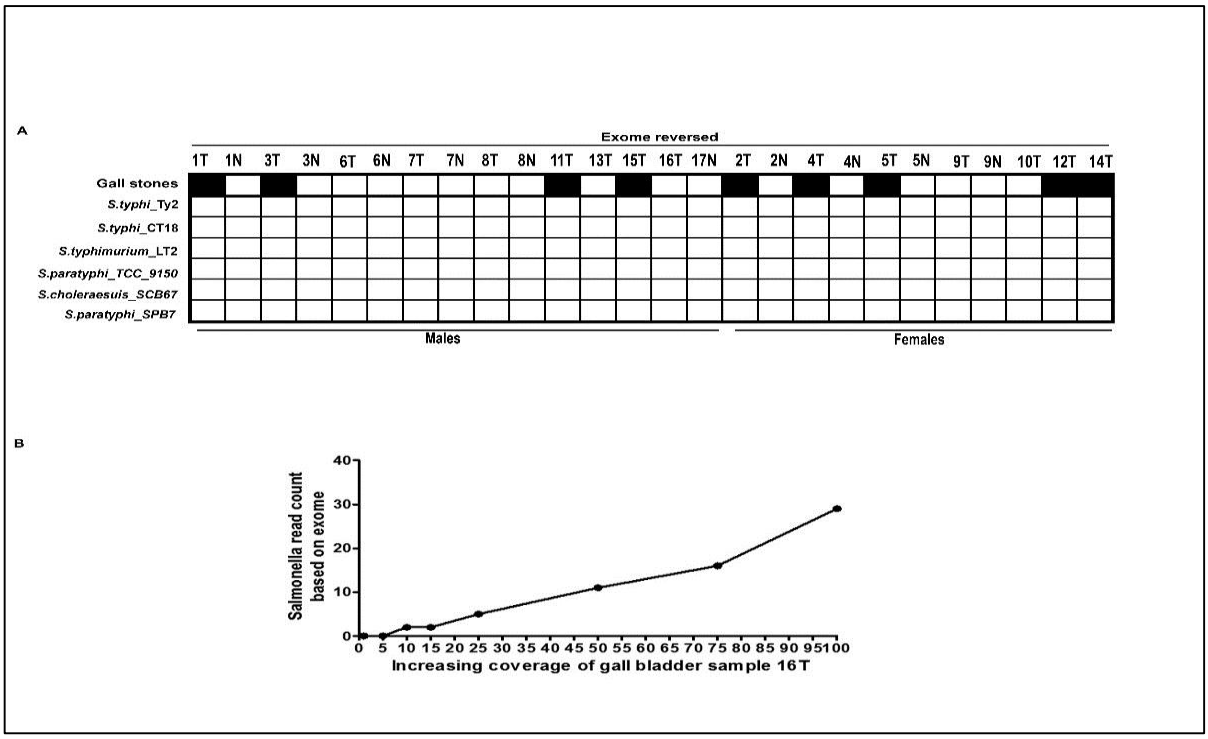
2.3.3 HPVDetector pipeline is specific and highly sensitive to detect true *Salmonella* traces

To assess the specificity of our assay, we re-analyzed whole exome data of all samples by taking their reverse (not complement) to simulate random sequence, but retaining composition of nucleotides and genome complexity, using an in-house perl script, as described earlier. We found no spurious *Salmonella* reads when the primary tumor whole exome sequence was reversed in any of the 26 samples, suggesting the computational pipeline used was specific to detect *Salmonella* traces. To test the sensitivity of our assay, raw FASTQ file of a primary tumor sample 16 T that was found positive for *Salmonella* reads was downsampled to 1X, 5X, 10X, 15X, 25X, 50X, 75X and 100X coverage using Picard Toolkit's downsampleSam function (http://broadinstitute.github.io/picard/), as described earlier. The resulting FASTQ files were analysed for detection of *Salmonella* reads using the HPVDetector pipeline(**Figure 3**).



II-Figure 2: Detailed annotation of read sequences of different *Salmonella* species identified across gallbladder cancer patient samples.

Abundance and annotation of *Salmonella* reads found across the 16 of 26 gall bladder cancer samples. Heat map representation of individual *Salmonella* reads (in rows) identified from 6 different isolates found across the 16 gall bladder cancer samples (in column) is shown. Variable length and number of overlapping reads, each of 150 bp obtained from paired end Illumina sequence for each isolate, were assembled into contigs based on Clustal *X*2 multiple alignment. The unique total length of contigs generated is shown in second column reflecting the total length of the gene covered in the study. The contigs generated were annotated based on gene annotation database of *Salmonella* isolates from NCBI database. A representative general class for all genes identified is shown in the third column.

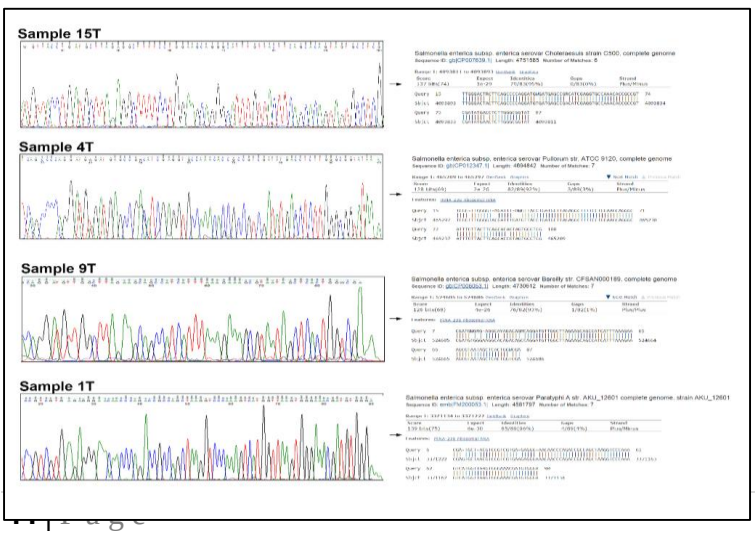


II-Figure 3: Specificity and sensitivity for detection of *Salmonella* reads in whole exome sequencing of gallbladder samples.

(A) Specificity for detection of *Salmonella* reads in whole exome sequence of gallbladder samples. Exome sequenced reads were reversed (not complement) to maintain the genome complexity and used an input file to detect random *Salmonella* reads. No *Salmonella* reads were found in the samples with reversed whole exome sequence. **(B)** Sensitivity for detection of *Salmonella* reads in gallbladder samples as a function of increasing genome sequence coverage. Gallbladder tumor sample 16 T with the highest number of *Salmonella* reads was down-sampled to 1x, 5x, 10x, 15x, 25x, 50x, 75x and 100x. *Salmonella* reads were counted (black line) and plotted against increasing coverage of the genome on the x-axis.

2.3.4 Sanger validation of *Salmonella* reads identified in gallbladder cancer samples

We have attempted to validate the presence of *Salmonella* read sequences identified by HPVDetector in 4 of 16 *Salmonella* positive samples using Sanger sequencing.



II-Figure 4: Sanger validation of *Salmonella* read sequences in GBC samples.
Individual read sequences were PCR amplified and Sanger sequencing trace of individual read sequence with their blast output is represented in the figure.

2.4 Discussion

We examined *Salmonella* and HPV DNA sequences in gallbladder tumors and paired normal, the high incidence of *Salmonella* sequence found in 16 of 26 samples analyzed in the study suggests a possible role of *Salmonella* infection in gallbladder cancer analogous to *Helicobacter* pylori in gastric cancer and *Fusobacterium* in colon cancer[70, 71]. We demonstrate the presence of typhoidal *Salmonella* species in 11 of 26 gallbladder cancer samples, consistent with as known earlier. In addition, we present the first evidence to support the association of even non-typhoidal *Salmonella* species in 12 of 26 gallbladder cancer, with 6 of 26 samples co-infected with typhoidal as well as non-typhoidal *Salmonella* isolates. Systemic inflammation is known to be associated with a poor prognosis in gallbladder cancer[72]. Owing to the ability of *Salmonella* infection to stimulate a host response, it is likely that these bacteria are able to provide the continued 'inflammatory stimulus' necessary for carcinogenesis. Recent reports suggest that *Salmonella* infections promote malignant transformation in genetically predisposed mice, murine gall bladder organoids and fibroblasts with *TP53* mutations[19]. We observed 4 of 16 *Salmonella* positive samples harbored *TP53* mutations while we did not observe *TP53* mutations in *Salmonella* negative samples. *Salmonella* isolates in the chronic carrier state thus fits the role of the 'inflammatory stimulus' in the genetic model for gallbladder carcinogenesis and its dissemination cascade, which may trigger transformation through chronic inflammation, but not for maintenance of tumorigenesis[31].The focus of treatment in typhoid-endemic countries such as India has historically been solely on eliminating typhoidal *Salmonella* species often underestimating the contribution of the non-typhoidal isolates that show an inherent higher resistance to the standard antibiotics[69] resulting in their ability to lead to chronic carrier state in humans. The presence of non-typhoidal *Salmonella* species in our study highlights that in typhoid as well as gallbladder cancer endemic countries such as India and other similar countries, efforts must be directed not only at treating typhoid fever, but

also diagnosing and appropriately managing non-typhoidal *Salmonella* species. This simple approach could reduce the chronic carrier state of these species in humans, which by our hypothesis may be contributing to the inflammatory stimulus driving gallbladder carcinogenesis. Thus, this simple strategy may help reduce in the incidence of gallbladder cancer. While this study validates and extends the association of *Salmonella* with gallbladder carcinoma, further study is required to establish the causality of infection to the disease.

II-Table 1: Primer sequences for Salmonella validation

Primer	5'-3'
OAD 1352 _Stymurium_9T_F	TCGACCAGTGAGCTATTACGC
OAD1353 _Stymurium_9T_R	GCCAGCTAAGGTCCCAAAGT
OAD1354_Stymurium_4T_F	TCTCTCAAGCGCCTTGGTAT
OAD1355_Stymurium_4T_R	CGAGGCACTACTGTGCTGAA
OAD1359_Stymurium_1T_F	CCCACATCGTTTCCCACTTA
OAD1360_Stymurium_1T_R	GCGAATTCCGGAGAATGTTA
OAD1361_Styphi_15T_F	CCAGCTCGCGTACCTCTTTA
OAD1362_Styphi_15T_R	ATACCGCCCAAGAGGTCATA

3. CHAPTER 3:

ERBB2 and *KRAS* Alterations Mediate Response to EGFR Inhibitors in early stage Gallbladder Cancer

Abstract

The uncommonness of gallbladder cancer in the developed world has contributed to the generally poor understanding of the disease. The development of new and effective treatment, therefore, has been and continues to be a major public health imperative. We report mutation analysis of 44 predominantly early-staged gallbladder tumors and 5-gallbladder cancer cell lines by a combination of whole exome and directed sequencing. We discover recurrent activating *ERBB2* (V777L) somatic mutation in 6 of 44 gallbladder primary tumors with an overall mutation frequency of 13%; along with *KRAS* activating mutations in 3 cancer samples. Consistent with whole exome findings, phospho-proteomic array profile of 49-tyrosine kinase revealed constitutive phosphorylation of ERBB2 and EGFR that were found to heterodimerize. We demonstrate that treatment with *ERBB2*-specific, *EGFR*-specific shRNA or with a covalent EGFR family inhibitor BIBW-2992 inhibits transformation, survival, migration and invasion characteristics of gallbladder cancer cells harboring wild-type or *KRAS* (G13D) but not *KRAS* (G12V) mutation. In summary, our studies implicate *ERBB2* as an important therapeutic target in early-stage gallbladder cancer. We also present the first evidence that the presence of *KRAS* (G12V), but not *KRAS* (G13D) mutation, may preclude gallbladder cancer patients to respond to anti-EGFR treatment, similar to the clinical algorithm commonly practiced to opt for anti-EGFR treatment in colorectal cancer.

3.1 Introduction

Genomically matched therapies targeting activated tyrosine kinases have shown promise across multiple cancer types [73]. The success of tyrosine kinase inhibitors (TKIs) such as imatinib, a BCR-ABL fusion protein inhibitor [74]; vemurafenib, a RAF inhibitor [75]; lapatinib, an inhibitor of ERBB2 [76]; erlotinib and crizotinib, inhibitors of EGFR and ALK, respectively [77-79]; and, others have provided a powerful validation for precision cancer medicine. Although these treatments offer great promise, selective genomic profiling of tumors tend to impede broader implementation of genome-based cancer care [80]. For example, an inadequacy to account for multiple relevant genetic alterations likely resulted in comparable outcomes in a recently performed randomized trial where multiple cancer type patients were profiled for selected driver alterations and randomized to receive genomically-

matched versus conventional therapy [81]. Such important clinical studies underscore the need for convergence of information for multiple genetic alterations to ensure the success of future clinical trial designs, with specific emphasis for consideration of co-occurring alterations that could potentially render tumors unlikely to benefit from genomically-matched treatments. Some prototypical examples include *KRAS*, *NRAS*, and *BRAF* mutations in colorectal cancers or secondary *EGFR* mutations in lung cancer against anti- EGFR targeted therapies [82].

The EGFR family of receptor tyrosine kinases (RTK) consists of *EGFR*, *HER2*, *HER3* and *HER4* (human EGFR-related- 2, -3, and -4). A ligand-bound EGFR family member forms a homo- or hetero-dimer to activate the PI3K-AKT-mTOR or RAS-RAF-MAPK downstream signaling pathway to evade apoptosis and enhance cell proliferation [83-85]. Interestingly, of all EGFR family members, HER2 lacks a ligand binding domain and forms preferred partner for other members to heterodimerize even in the absence of ligand [83]. Deregulation of EGFR family RTK-signaling network endows tumor cells with attributes to sustain their malignant behavior and survival, as is frequently observed in breast cancer, lung cancer, pancreatic cancer, head and neck cancer and colorectal cancer [86]. Interfering with the EGFR pathway thus forms the basis for the development of targeted anticancer therapies such as RTK-targeted antibodies (Cetuximab and Herceptin) and small-molecule inhibitors of RTK kinase (Erlotinib, Lapatinib, Afatinib, etc.) that have shown dramatic clinical response [86]. In such responses, however, the co-occurrence of a *KRAS* mutation – a downstream component of the pathway-- preclude patients from anti-EGFR treatment in colorectal cancer, wherein *KRAS* codon 12, but not codon 13 mutations are associated with poor outcomes [87-89], underscoring their prognostic impact.

Gallbladder cancer, the most common malignancy of biliary tract, is a rare form of cancer in the world where chemotherapy and other palliative treatments have little effect on overall survival of patients [14, 90]. The poor understanding of gallbladder cancer due to its uncommonness in the western world but high prevalence in Chile and the Indian subcontinent lends itself to the need for further research [1, 31, 91]. The 5-year survival rate of an early stage T1 gallbladder carcinoma is nearly 100%, however, of advanced stage, T3/T4 is less than 15% [90, 92]. A hope for longer-term survival has specifically been promising for an early staged T2 carcinomas with an intermediate 5-year survival [93]. Literature suggests *HER2* overexpression in 12–15 % of advanced stage gallbladder cancers with a favorable

response to HER2 directed therapy [26, 28, 38, 94]. Moreover, three recent studies analyzed whole exome sequence of advanced stage gallbladder tumors among non-Indian populations with consistent findings [10, 27, 95]. In order to understand the landscape of somatic alterations among a clinically distinct early staged pT1/pT2 Indian gallbladder cancer patients, we performed whole exome sequencing of 17 early staged tumor-normal paired gallbladder samples, 5 gallbladder cancer cell lines followed by validation in 27 additional tumor samples. Here, we report novel somatic mutations of *ERBB2* in gallbladder cancer, and its therapeutic implication in presence and absence of *KRAS* (G12V) and (G13D) mutations

3.2 Materials and Methods

3.2.1Patient Information:

A total of 27 fresh frozen samples (10 tumor-normal paired and 7 orphan tumors) were utilized for whole exome sequencing. An additional set of 27 FFPE samples were utilized as a validation set. Tumor-normal paired samples were collected at Tata Memorial Hospital and Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Mumbai. Sample set and (ACTREC-TMC) Internal Review Board (IRB) --IRB Project Number # 104-- approved study protocols. Formalin-fixed paraffin-embedded tissue blocks were collected from the tissue repository of Tata memorial hospital (TMH-TTR) in compliance with the guidelines. These tissues were examined for tumor content and the tumor content was in the range of 40-90%. Patient samples and characteristics are provided in Table 1.

3.2.2 **DNA extraction**

Genomic DNA was extracted from fresh frozen samples by using Qiagen Blood and Cell culture DNA kit. The extracted DNA yield and quality were assessed using Nanodrop ND2000 (Thermo scientific). The extracted DNA (about 1µg) from the fresh-frozen tissue specimens were sent to Genotypic Technology Pvt Ltd, Bangalore for exome sequencing. Genomic DNA from FFPE blocks was extracted using Qiagen QiAmp DNA FFPE Tissue kit as per manufacturer instructions. The extracted DNA yield and quality were assessed using Nanodrop ND2000 (Thermo scientific). These samples were further checked for integrity by PCR amplification of GAPDH (96bp). These samples were used for extended Sanger

validation of identified variants in exome sequencing. The primers used for validation are listed in Table 5.

3.2.4 Exome sequencing capture, library construction, and sequencing

Exome capture and sequencing were performed as described previously [96]. Briefly, Agilent Sure select in-solution (low-input capture-500ng) were used to capture ~62Mb region of human genome comprising of ~201,121 exons representing ~20,974 gene sequences, including 5'UTR, 3'UTR, microRNAs and other non-coding RNA. Sequencing was run with 150bp paired-end reads to achieve coverage of 100X and was performed according to Illumina standard protocol.

3.2.5 Exome analysis pipeline and somatic mutation calling

The variant analysis was performed as described previously [96, 97] and detailed in supplementary material and methods. MutSigCV v2.0 [98] and IntOgen [99] were used for identification of the significantly mutated gene and p value ≤ 0.05 was considered as the threshold for significance. The variants were excluded if they were present in exclusively in dbSNP, TMC-SNPdb or both. Also, we removed variants that were identified in all three databases – COSMIC (v68) [100], dbSNP (v142) [101] and TMC-SNPdb database [102]. The annotated cancer-associated variants were annotated using Oncotator (v1.1.6.0) [103] and restricted our further analysis to only coding variants. Intogen (https://www.intogen.org/search) was used to calculate the significance of frequently mutated gene in our cohort. Since our dataset was inherently not suitable for above tools due to a limited number of tumor samples (n=17), we have also performed extensive functional prediction tool based analysis for non-synonymous variants using nine different tools as described earlier [97]. Total number of identified somatic substitutions in exome sequencing were extracted from MutSigCV output and were processed to calculate the number and frequency distribution of various transitions and transversions.

3.2.6 Copy number analysis from Exome sequencing data

Control-FREEC [104] was used for copy number analysis from BAM files of variant calling analysis. Genes with Segments-of-Gain-Or-Loss (SGOL) score ≥ 4 were defined as amplified genes and ≤ -2 as deleted genes by cghMCR package of R (http://bioconductor.org/packages/release-/bioc/html/cghMCR.html). The validation of somatic copy number changes was performed as described previously [96].

3.2.7 Cell culture and reagents

Human GBC cell lines (OCUG1, SNU308, TGBC2TKB, NOZ, and G415) obtained as a kind gift from Dr. Akhilesh Pandey (IOB, Bangalore) were cultured in DMEM media containing 10% FBS, 100 units/ml penicillin, and 100mg/ml streptomycin and amphotericin. All cell lines were incubated at 37^{o}C with 5% CO_2 . The cell lines were authenticated by DNA short tandem repeat (STR) profiling using Promega Geneprint 10 system in conjunction with GeneMarker HID software tool. All cell lines were made mycoplasma free if necessary with EZKill Mycoplasma removal reagent (HiMedia).

3.2.8 Soft Agar assay

All experiments were performed in triplicates as described earlier [105]. Briefly, anchorage-independent growth was assessed for the knockdown clones of *ERBB2* and *EGFR* along with respective scrambled control. About 1ml of 2X DMEM supplemented with 20% FBS containing (1ml of 1.6% agar) to obtain 0.8% agar was added to the six-well plate as bottom agar and was allowed to solidify. Next, 5 * 103 cells were supplemented with 1ml of 2X DMEM containing 0.8% agar to obtain 0.4% agar and were added to the bottom agar as top agar. The cells were incubated for 2 weeks at 37^{o}C and 5% CO2 .Colonies were counted under a microscope with a magnification of 10X.

3.2.9 Virus production

293FT cells were seeded in 6 well plates one day before transfection and each of the lentiviral constructs along with packaging plasmids -pPAX helper vector and pVSVG were transfected using Lipofectamine 3000 reagent (Invitrogen) as per manufacturer's protocol. The viral soup was collected 48 and 72 hrs post-transfection, passed through the 0.45µM filter and stored at 4^{O}C. Respective cells for transduction were seeded one day before infection in a six-well plate and allowed to grow to reach 50-60% confluency. One ml of the virus soup (1:1 dilution) and 8ug/ml of polybrene (Sigma) was added to cells and incubated for six hours. Cells were selected with puromycin (Sigma) (2µg/ml) selection for 2 days as further described earlier [96].

3.2.10 Growth Curve

Growth curve assay was performed on a 24 well plate format with a cell density of 20000cells/well. Cell growth was assessed post 48hr and 96hr by counting the cells using a hemocytometer and was recorded. Cell proliferation was calculated as percentage proliferation normalized to scrambled control. All the experiments were performed in triplicates.

3.2.11 MTT assay

1000 cells per well were seeded in 96 well plate followed by incubation with the drug for 72 hours and six replicate per concentration and subsequently incubated with MTT (0.5 mg/ml) for 4 hours and then MTT assay was performed and data was acquired at 570nm using Microplate reader. Percentage cell viability was calculated against vehicle treated.

3.2.12 Western blotting

Cells were lysed in RIPA buffer and protein concentration was estimated using BCA (MP Biomedical) method. 50 µg protein was separated on 10% SDS-PAGE gel, the transfer was verified using Ponceau S (Sigma), transferred to nitrocellulose membrane and blocked in Tris-buffered saline containing 5% BSA (Sigma) and 0.05% Tween-20(Sigma). The primary antibody against Total HER2 (sc-33684 Dilution 1:500), Total EGFR (1005) (sc-03 Dilution 1:500), Total ERK2(C-14) (sc- 154 Dilution 1:500) and β-Actin(I-19)-R (sc-1616-R Dilution 1:3000) were obtained from Santa Cruz biotechnology. The primary antibodies Phospho-HER2 (Tyr1248) (AP0152 Dilution 1:500) from Abclonal and Phospho-p44/42 (T202/Y204) MAPK (#4370) Dilution 1:1000), Phospho-*EGFR* (Y1068) (#2234 Dilution 1:500) were obtained from Cell signaling technology respectively. Thiazolyl blue tetrazolium bromide (MTT, TC191) was obtained from Hi-Media.

3.2.13 Receptor tyrosine kinase proteome array

The relative amount of 49 tyrosine kinases were evaluated using Proteome Profiler Human Phospho- RTK array kit (ARY001B – Proteome Profiler, R&D systems) and the protocols were followed as per manufacturer's recommendation. Briefly, cells were harvested, washed with 1X PBS and lysed after which 400µg of protein was mixed with a buffer and incubated

with pre-blocked nitrocellulose membrane at 4^{o}C. Subsequently, the membranes were probed using detection antibodies and probed using streptavidin-HRP, after which signals were developed using the chemi-reagents provided with the kit. The Pixel density of each spot in the array in duplicate was quantified using Image J macro-Protein array analyzer plug in. The average pixel density of the duplicate spots for each of the kinases was subtracted from the negative density and was plotted, as detailed earlier [106].

3.2.14 Invasion assay

Invasion ability of the cells was assessed in Transwell system using cell culture inserts for 24 well plates with 8µm pores (BD Biosciences, NJ). The upper side of the cell culture insert was coated with Matrigel (BD Biosciences, San Jose, CA). GBC cells were seeded at a density of $2 * 10^4$ on the upper side of the coated Matrigel in presence of serum-free DMEM. Complete DMEM media with 10% FBS was added to the lower side of the insert and were incubated at 37^{o}C in 5% CO_2 incubator for 12-14hrs. Post incubation the non-migratory cells on the lower side of the cell culture insert were removed using a cotton swab. The transwell chambers were fixed and stained with 0.1% crystal violet. The invasion ability was estimated by counting the cells that have migrated to the lower side of the cell culture insert. Cells in visual field with a magnification of 20X were counted in each Transwell chamber in triplicates.

3.2.15 Wound healing assay

Confluent monolayers in 6 well plate are subjected to scratch with a sterile pipette tip. After this, cells are washed with 1X PBS to remove debris and subsequently incubated with media. Cell migration at the wound surface was measured during a period of 20h under an inverted .microscope. The quantification of cell migration was done using Cell Profiler wound healing pipeline for three independent wounds in 3 independent experiments.

3.2.16 Survival analysis

Survival analysis was assessed using statistical package SPSS statistics v20. Overall survival could be calculated from the date of registration and the endpoint was taken as the date of

death with censoring implied at last date of follow-up. Correlation analysis was assessed using descriptive statistics of SPSS statistics v20.

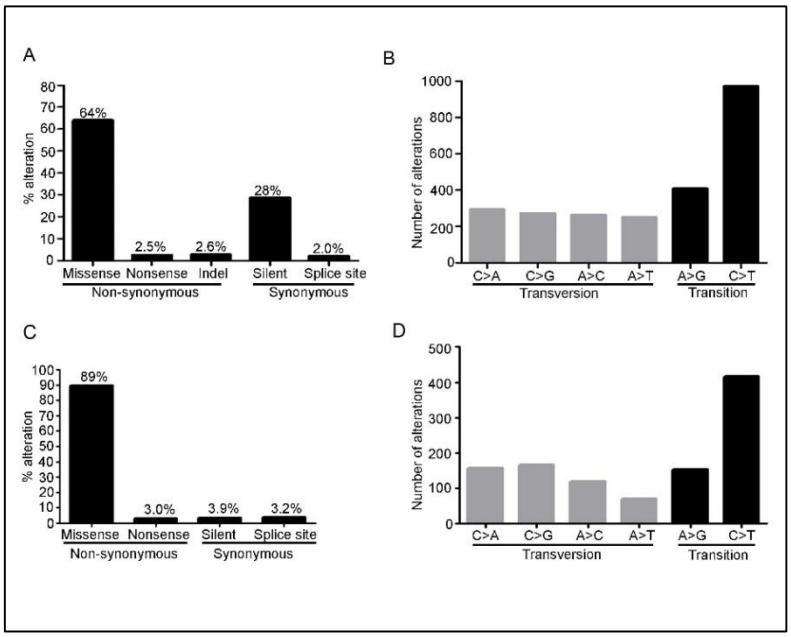
3.3Results

3.3.1 Integrated genomics and proteomics approach identify aberrant alterations in members of the *EGFR* family in gallbladder cancer

We performed whole-exome sequencing on paired tumor and germline DNA samples from 17 patients with gallbladder cancer and 5 gallbladder cancer cell lines (Table 1 and 2). We achieved >100-fold mean sequence coverage of targeted exonic regions. The average non-synonymous mutation rate was found to be 7.7 mutations per megabase (Table 3), which is significantly higher than as reported for other populations [27]. The nucleotide mutation pattern was observed to be enriched for C>T transition followed by A>G transition (Figure 1), consistent with previous reports [27]. A total of 5060 somatic variants found across 17 tumors consisted of 3239 missense, 1449 silent, 131 nonsense, 135 indels and 106 splice site mutations. Somatic mutations in genes previously reported to be altered in gallbladder cancer, including recurrent mutations in *TP53* (35.2%), *ERBB2, SF3B1, ATM*, and *AKAP11* at 17.6% each were found to be mutated at comparable frequencies [27] (Figure 2). For validation of a few *TP53, ERBB2, ERBB3, SMAD4* and *CTNNB1* mutations, sanger-based sequencing was carried out in a subset of patients (Figure 3). Among set novel alterations, we observed significant somatic mutations in chromatin modifier genes such as *SF3B1*, *ATRX*, *CREBBP* and *EZH2* that are known to play a significant role in other cancer types [107-110]. In addition, we also found two tumor samples that harbored known activating kinase domain mutations in *ERBB2*, (V777L) and (I767M) [111, 112]; while two samples harbored *EGFR* (I1005V) and *ERBB3* (R112H) mutation. We identified 5 more samples with *ERBB2* mutations harboring (V777L) mutations in an additional set of 27 gallbladder cancer samples (Figure 2B). Interestingly, copy number analysis using cghMCR software identified *EGFR* amplification with a highest Segment Gain Or Loss (SGOL) score of 18 (Figure 2A), as reported earlier [10, 28, 113]. In overall, we observed genomic amplification in *EGFR, CDK4, MDM4, CCND1, CCNE1, MYC, STK11* and *BRD3*, and deletion in *FHIT*, *SMAD4*, *TRIM33* and *APC*.

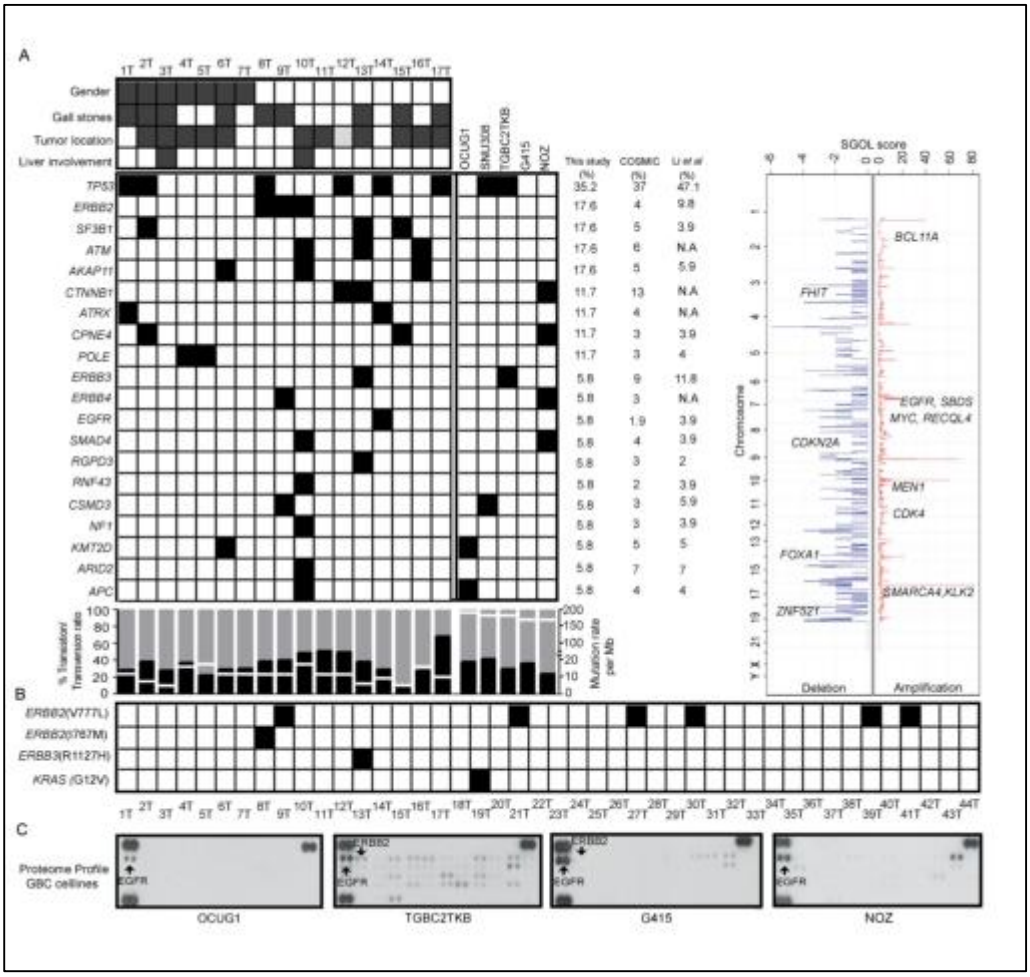
Next, to correlate differential activation of signaling molecules with their genomic alterations, we performed phospho-proteomic profile of four gallbladder cell lines for 49 receptor tyrosine kinases using a phospho-RTK array. Consistent with whole exome findings, we

observed varying levels of EGFR and ERBB2 constitutive phosphorylation in all gallbladder cancer cell lines based on their phospho-proteome (Figure 2C) and follow up validation by western blot analysis (Figure 4). Interestingly, the whole exome data analysis and Sanger sequencing-based validation also revealed that 1 of 44 gall bladder patients and NOZ cells harbor *KRAS* (G12V) mutation; G415 cells harbor *KRAS* (G13D) mutant allele; while OCUG1 and SNU308 cells were wild-type for *KRAS* (Figure 3)These four cell lines thus represent diverse gallbladder cancer sub-classes based on their *KRAS* mutant allele status [27]. Of note, *KRAS* mutations are known to predict plural clinical outcome in response to EGFR inhibitors in colorectal and lung cancer along with other mutations[114, 115].



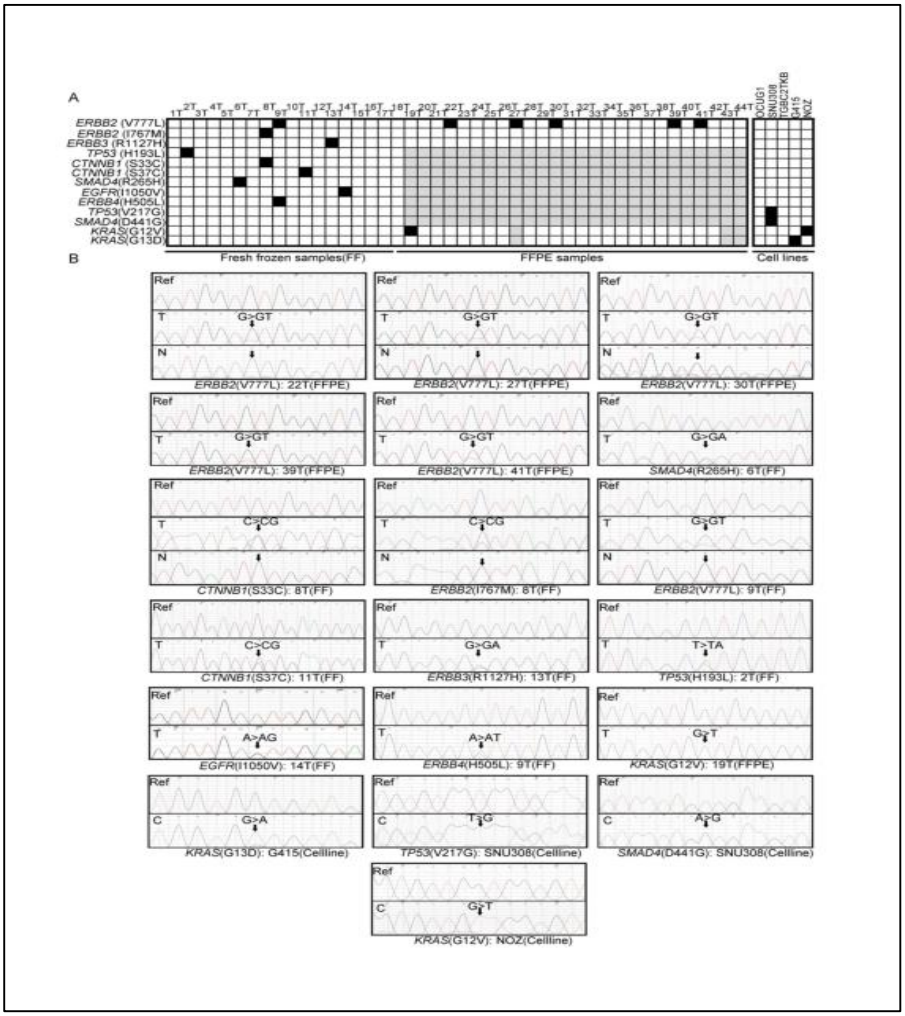
III-Figure 1: Distribution and features of variants identified in whole exome sequence of tumor samples and cell lines

A) The distribution of variants in different classes (missense, indel, silent and splice site alterations) is represented in the bar graph. Each bar represents the percentage frequency of the class in the exome sequencing data. **B)** The distribution of substitution in somatic variants. Each bar represents the percentage frequency of the substitution in the exome sequencing data



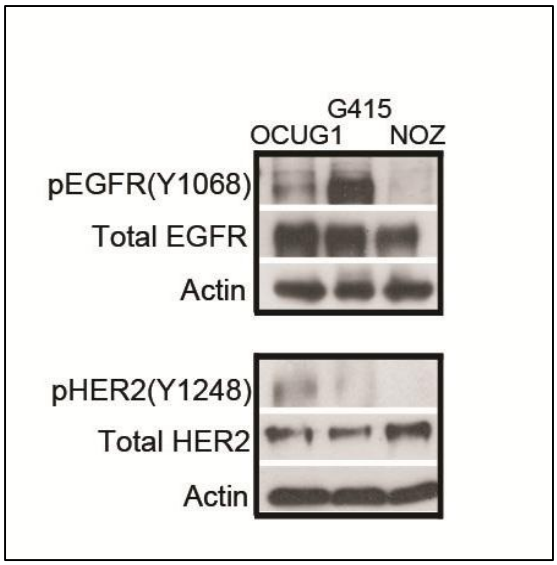
III-Figure 2: Integrated genome analysis of 17 primary tumor samples, 5 cell lines and tyrosine kinase protein array of 5 gallbladder cancer cell lines

A) The heat map represents the altered genes that were common in the COSMIC (Catalogue of somatic alterations in gallbladder cancer) and exome sequencing study in gallbladder cancer (Li *et al., 2014* Nature genetics) with our study in the top panel of the figure. The right panel of the heat map indicates the frequency of mutations and it's across each gene in our study and COSMIC &. Li *et al., 2014* study. The distribution of mutations is represented by mutation/Mb. Above the heat map is a representation of clinicopathological features of samples. (Sex – Grey indicates females and white are males, Gallstones – Grey indicates present and white is absent, Liver involvement- Grey indicates positive and white indicates negative, Tumor location- Grey indicates neck and white indicate body.) Below the heat map is the percentage ratio of transition and transversion across the tumor samples B) Copy number alterations – The plot represents chromosomal regions with copy number loss/gains. The x-axis is represented by a score of segment gain or segment loss (SGOL score) while the y-axis represents the chromosomal positions. Copy number gain is indicated by red with positive SGOL score while copy number loss is indicated by blue with a negative SGOL score. Cancer-associated genes are annotated in their representative amplified regions. C) RTK phosphorylation in gallbladder cancer cell lines Protein lysates of gallbladder cancer cell lines were incubated on RTK arrays and phosphorylation status was determined by subsequent incubation with anti-phosphotyrosine horseradish peroxidase. Each RTK is spotted in duplicate and the pair of dots in each corner of the membrane are positive controls and negative control. The representative arrays are shown in the cell lines (OCUG1, TGBC2TKB, G415, and NOZ). The positive signals for phosphorylation of ERBB2 and EGFR are indicated by arrows in respective cell lines.



III-Figure 3: Validation of alterations in primary tumors & cell lines identified by whole exome seq

A)Heat map representation of alterations identified by whole exome sequencing in the discovery set and its validation in the additional validation set of samples. Black filled box indicates the samples in which the corresponding mutations are validated by Sanger sequencing B) Sanger sequencing chromatogram of the alterations identified by whole exome sequencing in respective samples. The position of the aberration is indicated by a black arrowhead



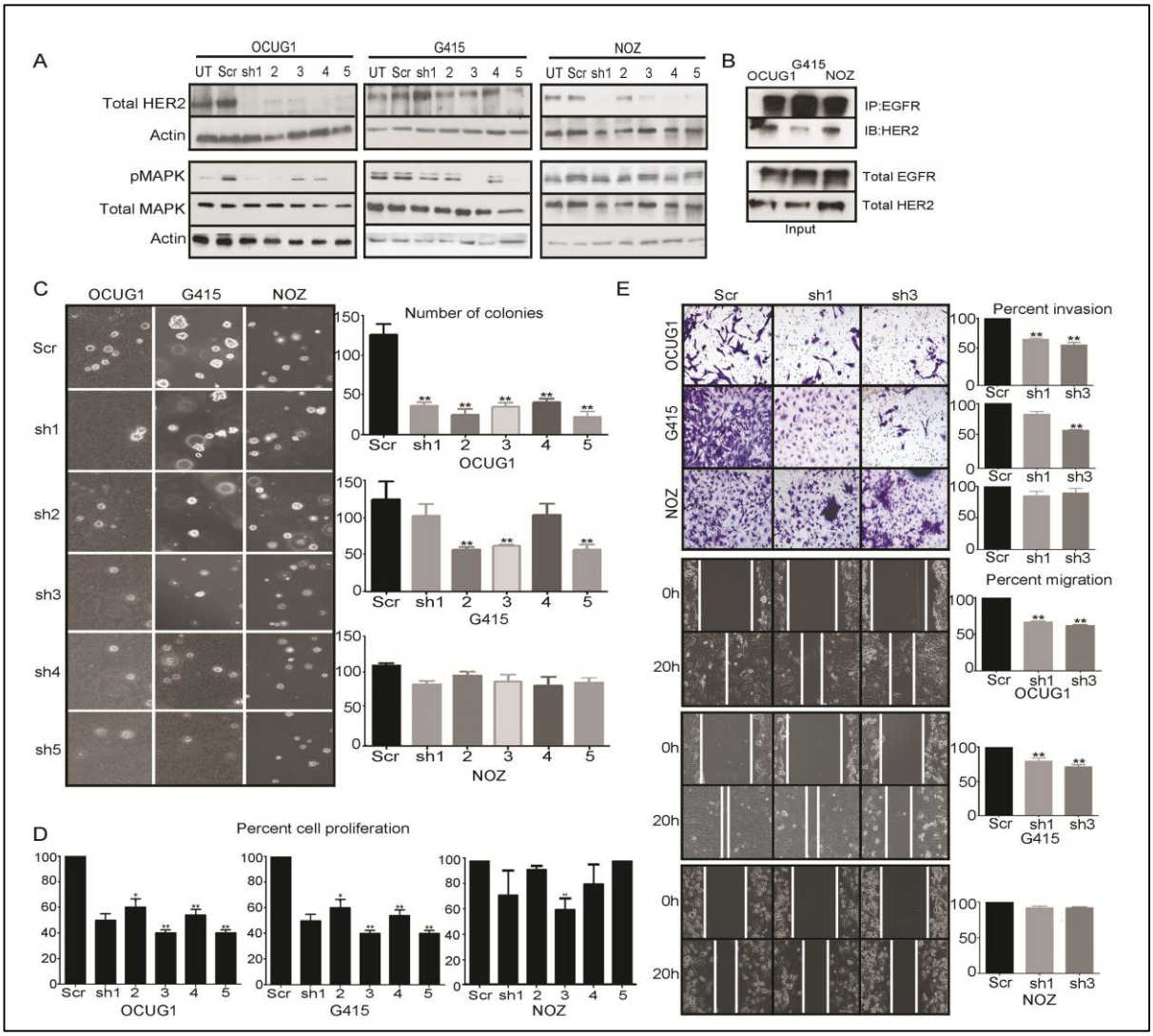
III-Figure 4: Validation of tyrosine kinase array using western blot analysis in GBC cell lines

Phosphorylation of ERBB2 and *EGFR* were examined using western blot analysis with indicated antibodies as shown in GBC cell lines. Actin was used as a loading control

3.3.2 *ERBB2* and *EGFR* are essential for the survival of gallbladder cancer cells not harboring *KRAS G12V* mutant allele

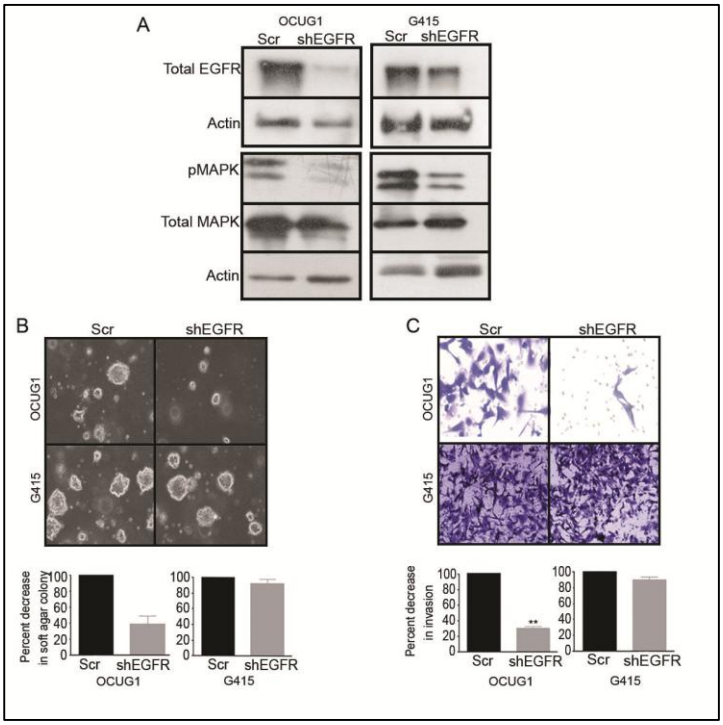
To determine the significance of *EGFR* and *ERBB2* constitutive phosphorylation and *KRAS* mutant alleles in gallbladder cancer cells, we set out to establish whether expression of *ERBB2* is required for gallbladder tumor cell survival. We tested a series of five shRNA constructs in three gallbladder tumor cell lines expressing *ERBB2* with wild-type *KRAS* in OCUG1 cells, along with G415 and NOZ cells harboring the *KRAS* (G13D) and *KRAS* (G12V) mutant alleles, respectively. We identified three shRNA constructs that efficiently knocked down expression of *ERBB2* and inhibited the constitutive phosphorylation of MAPK in OCUG1 and G415 cells but not in NOZ cells (Figure 6A), consistent with drug-sensitive outcome described in colorectal cancer wherein cells harboring wild-type *KRAS* or mutant *KRAS* (G13D) allele are sensitive to *EGFR* inhibitor but not those harboring mutant *KRAS* (G12V) mutant allele [116]. This suggests that *KRAS* (G13D) but not *KRAS* (G12V) still requires upstream *EGFR* signaling in gallbladder cancer cells, similar to as established in colorectal cancer[117]. Next, we used these cells to demonstrate that knockdown of *ERBB2* inhibited anchorage-independent growth, cell survival, cell invasion and migration efficiently in OCUG1 and G415 cells but not in NOZ cells (Figure 5C-E). Furthermore, as unlike other *EGFR* family members, *ERBB2* does not require ligand binding for dimerization but can be activated by heterodimerization [118], we asked if *EGFR* mediates the activation of downstream signaling pathways. We performed co-immunoprecipitation of *EGFR* and *ERBB2* to establish that *ERBB2* interacts with *EGFR* in gallbladder cells (Figure 5B), possibly similar to *ERBB3* as shown earlier in gallbladder cells [27]. Moreover, to test if *ERBB2* requires *EGFR* also for sustained signaling and transforming potential, we knocked down the expression of *EGFR* in OCUG1 and G415 cells. The knockdown of *EGFR* inhibited anchorage-independent growth, cell survival, cell invasion and migration in OCUG1 but not in G415 cells, similar to *ERBB2* knockdown (Figure 6). Taken together, this

suggests that *ERBB2* requires *EGFR* or other members of the family possibly to dimerize for activation, such that down-regulation of *EGFR* and potentially other members suppress the functionality of *ERBB2*, as has been previously reported in breast cancer [119].



III-Figure 5: Effect of shERBB2 on *ERBB2* expression, colony formation, cell invasion, migration and proliferation in gallbladder cancer cell lines

A) Western blot analysis of *ERBB2* knockdown constructs (sh1-sh5) with respect to scrambled (Scr) and untransfected control in gallbladder cancer cells (OCUG1, G415, and NOZ). Phosphorylation of MAPK was examined using western blot analysis with indicated antibodies as shown. Actin was used as a loading control. **B)** Anti-*EGFR* (IP) immunoprecipitates were analyzed by western blot (WB) using anti-*EGFR* and anti-*HER2* antibodies. **C)** Anchorage-independent growth of gallbladder cancer cell lines (OCUG1, G415, and NOZ) was evaluated with different shRNA constructs (sh1-sh5). The graph represents the reduction in a number of colonies (±SD) relative to scrambled (Scr) control cells. *P< 0.05 vs scrambled control (Scr) **D)** Gallbladder cancer cells (OCUG1, G415, NOZ) were seeded in a 12 well plate to perform growth curve assay with 5 different knockdown clones (sh1 to sh5) relative to Scr (scrambled control). The graph represents the percentage reduction in cell proliferation (± SD) relative to scrambled control cells *P<0.05 vs scrambled control. **E)** Gallbladder cancer cells (OCUG1, G415, NOZ) were seeded in Matrigel-coated Transwell chambers with knockdown clone (sh1,sh3) and scrambled control to perform invasion assay. The graph represents the percentage reduction in invasion (± SD) relative to scrambled control untreated cells. *P<0.05 vs scrambled control (Panel E, top). Gallbladder cancer cells (OCUG1, G415, NOZ) were seeded in a six-well plate with sh1, sh3 clone along with a scrambled control for wound migration assay. The graph represents the percentage reduction in migration (± SD) relative to scrambled control untreated cells. *P<0.05 vs scrambled control (Panel E, bottom).

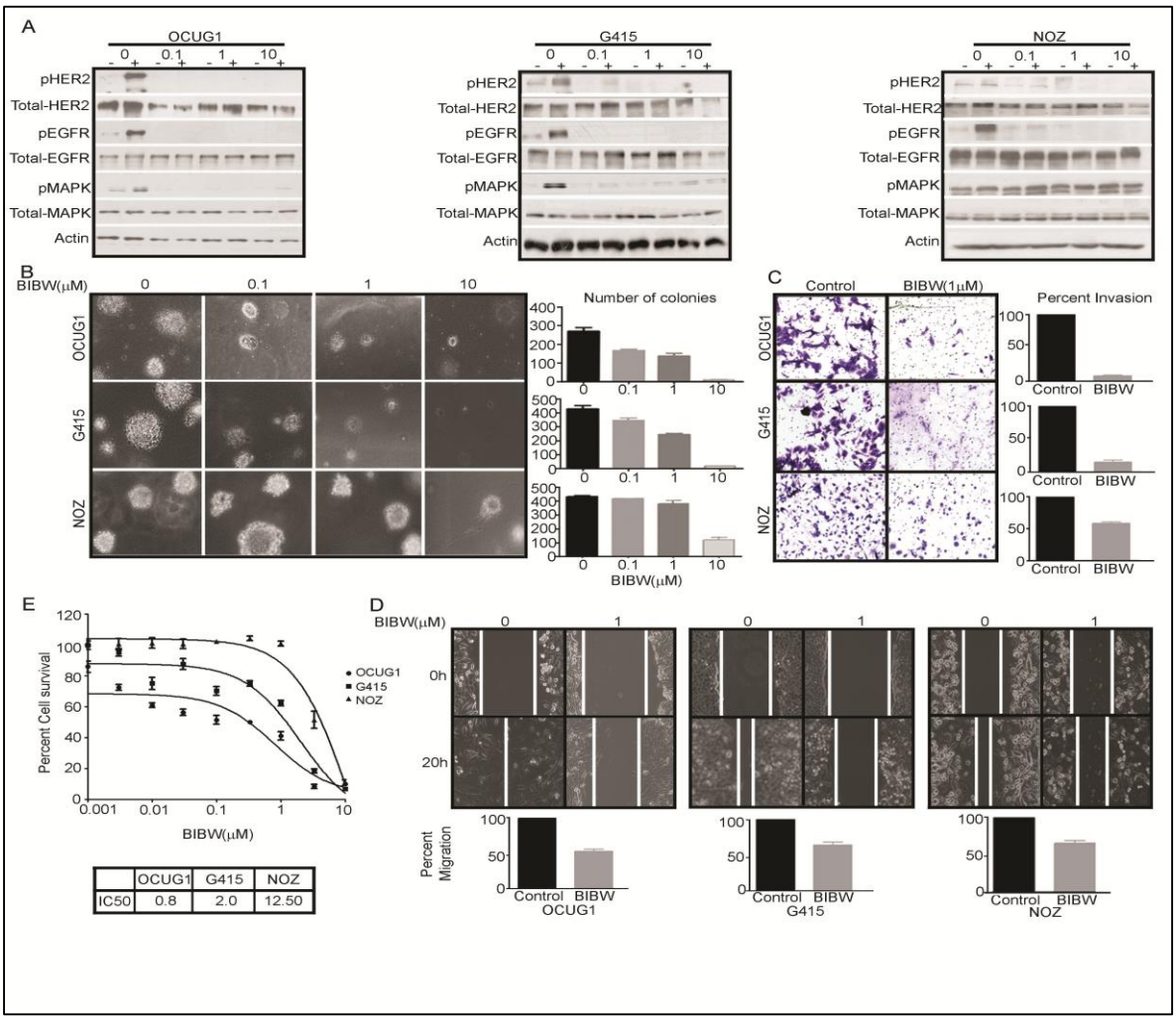


III-Figure 6: Effect of shEGFR on EGFR expression, colony formation, invasion &proliferation

A) Western blot analysis of EGFR knockdown construct shEGFR with respect to scrambled control in gallbladder cancer cells (OCUG1, G415). **B, C)** Anchorage-independent growth, cell invasion potential of gallbladder cancer cell lines (OCUG1, G415) were evaluated with shRNA construct (sh-EGFR) with respect to scrambled control. The graph represents the percentage reduction in cell invasion, number of colonies (±SD) relative to scrambled control. *P< 0.05 vs scrambled control

3.3.3 Irreversible *EGFR* inhibitors block proliferation and survival of gallbladder cancer cell lines

Next, we investigated whether inhibition of kinase activity of EGFR family receptor tyrosine kinases would be effective against gallbladder cancer cell lines. Treatment of the OCUG1 and G415 cells with BIBW-2992, but not reversible EGFR inhibitor gefitinib (data not shown), similarly abolished phosphorylation of MAPK in OCUG1 cells, which was constitutively phosphorylated in the untreated gallbladder cell lines compared to the NOZ cells and resulted in a marked decrease in colony formation in soft agar and cell survival in liquid culture, with IC50s of 0.8 uM in OCUG1 and 2.0 uM in G415 cells, whereas no effect was observed on NOZ cells harboring *KRAS G12V* mutant allele (Figure 7).

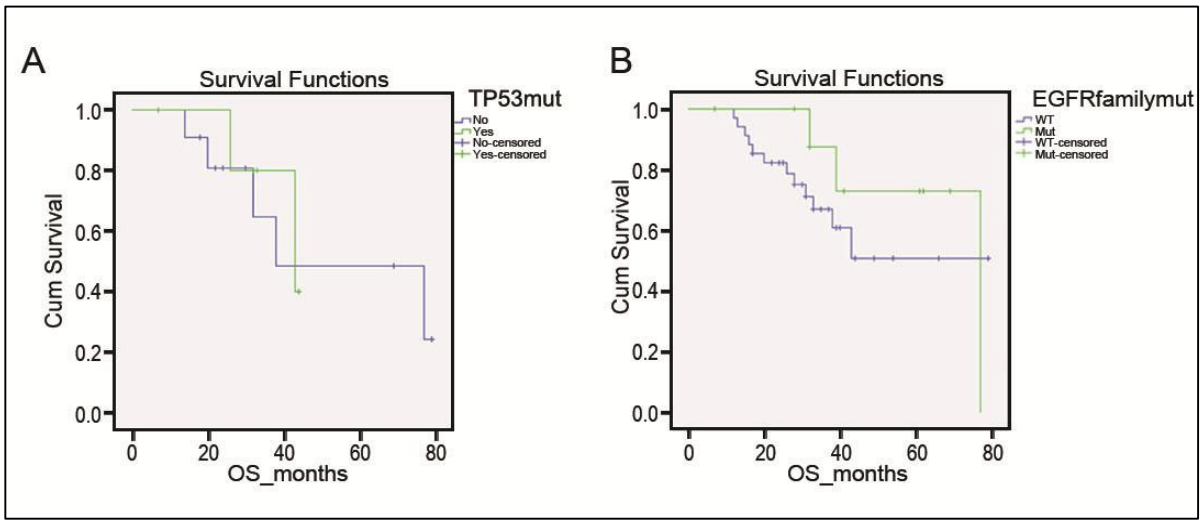


III-Figure 7: Effects of BIBW on EGFR-HER2 pathway in gallbladder cell lines

Biochemical and phenotypic effects of BIBW on the EGFR-HER2 pathway in gallbladder cell lines **A)** Cells were serum starved overnight and post serum starvation cell lines were treated with BIBW2992 for 12-14hrs. Cells were then induced with EGF(20ng/ml)[EGF(+), solid bars; EGF(-) bars with hatched lines] for 5 minutes. Phosphorylation of ERBB2 and EGFR along with phosphorylation of MAPK and AKT were examined using western blot analysis with indicated antibodies as shown. Actin was used as a loading control.**B)** Anchorage-independent growth of gallbladder cancer cell lines (OCUG1, G415, NOZ) was evaluated in presence of the inhibitor at different concentrations(0-10μM). The graph represents the reduction in a number of colonies (±SD) relative control untreated cells. *P< 0.05 vs control (μM). **C)** Gallbladder cancer cells (OCUG1, G415, NOZ) were seeded in Matrigel-coated Transwell chambers in the presence (1μM) or absence (Control) of BIBW (Panels-A-D) to perform invasion assay. The graph represents the percentage reduction in invasion (± SD) relative to control untreated cells. *P<0.05 vs control (0μM). **D)** Migration ability of the gallbladder cancer cells was evaluated in the presence (1μM) and absence of the inhibitor (Control) (A- G415 and B- NOZ). The graph represents the percentage reduction in migration (± SD) relative control untreated cells. *P< 0.05 vs control (μM).**E)** Cell proliferation was determined using MTT assay. Gallbladder cancer cell lines (OCUG, G415, NOZ) were treated with BIBW2992 at different doses in the range of 0-10μM. Data represent the mean (percent survival ± standard deviation SD) of the three independent experiments, each performed in 6 replicates and are presented with respect to control cells (DMSO, untreated control).

3.3.4 Clinical correlation of *TP53* and *EGFR* family mutations in GBC patients

In overall, the patient cohort represents a good subset of fairly early stage disease who received experienced and good quality radical surgery in a tertiary referral center. The Kaplan-Meier survival analysis with respect to *TP53* mutation status revealed an overall survival of 40 months (n=6; 95% CI: 34.1-70.7) compared to 52 months (n=11; HR: 0.8, 95% CI: 1.1-4.4 P= 0.799) among patients with wild-type *TP53* (Figure 8A). Thus, the overall survival of patients with mutations decreased as compared to patients with wild-type *TP53*. This observation is consistent with previous reports wherein *TP53* mutations are associated are known to be associated with worse prognosis in various cancer types [120, 121]. Interestingly, overall survival with respect to *EGFR* family mutation status was 65 months (n=9; 95% CI: 49.5-.82.1) compared to 54 months (n=35; 95% CI: 42.1-65.2) among patients with wild-type *EGFR* family genes (Figure 8B). This observation is in contrast to a recent report, wherein *ERBB2* mutations are associated with lower survival [27].



III-Figure 8: Overall survival of GBC patients with respect to *TP53* and *EGFR* mutations

A) Kaplan-Meier survival analysis curve calculated for gallbladder cancer patients with respect to TP53 status (N=17). **B)** Kaplan Meier survival analysis curve calculated for gallbladder cancer patients with respect to EGFR family mutation status (N=43). OS (Overall survival is given in months)

Discussion

This study represents the first genomic landscape of Indian gallbladder cancer that reveals somatic mutations in *TP53, ERBB2, ATM, AKAP11, SMAD4,* and *CTNNB1* similar to as reported in Chinese and Caucasian population. Our mutation pattern analysis revealed an enrichment for C>T transition followed by A>G transition, a signature which suggests an underlying chronic inflammation leading to GC to AT polyclonal transition [122], as reported earlier [27, 123]. We also observed significant somatic mutations in chromatin modifier genes such as *SF3B1, ATRX*, *CREBBP* and *EZH2* that have not been reported earlier in gallbladder cancer, indicating potential therapeutic options among the Indian population. Analyzing the potential effects of somatic alterations on survival of gallbladder cancer patients, we observed a trend among patients with wild-type *TP53* to survive longer than patients with *TP53* mutations, which is known to predict failure of chemotherapy in several cancer types [124] and is consistent with previous reports observed in gallbladder cancer [125].

Comparison of whole exome sequencing analysis of gallbladder cancer samples of Japanese origin with our study suggests four genes (TP53, ARID2, EGFR and SMAD4) to be significantly altered, with frequency as shown in III-Additional Supporting Table-1 [10]. Of note, consistent with a recent report that described alterations in *ERBB2* and *ERBB3* at a frequency of 9.8% and 11.8% respectively among Chinese gallbladder cancer [27], we found recurrent activating *ERBB2* (V777L) mutation in 6 of 44 gallbladder cancer samples with an overall mutation frequency of 13%, in addition to *ERBB3* (R112H) and *EGFR* (I1005V) mutation occurring at 2%, each in our sample set. The (V777L) alteration has been shown to be sensitive to lapatinib in biliary tract cancer, breast cancer cell lines and other isogenic systems overexpressing the alteration [28, 111, 112]. Functional studies performed using gallbladder cell lines establish that *ERBB2* and *EGFR* are essential for the survival of gallbladder cancer cells. Given that *ERBB2* lacks the ligand binding domain, the co-immunoprecipitation experiments suggest that ERBB2 dimerize with EGFR, and possibly with other members, to constitutively activate the pathway. Interestingly, genetic or pharmacological ablation of *ERBB2* and *EGFR* function, using EGFR small-molecule irreversible inhibitor BIBW-2992, diminishes the survival, anchorage-independent growth, migration and invasion characteristics of gallbladder cancer cell lines, suggesting members of the EGFR family as an effective therapeutic target.

Furthermore, while *KRAS* mutations in gallbladder cancer have been reported to occur at a frequency from 3% to 30 % [126], some co-occurring with activating *ERBB3* mutation [30], we observed *KRAS* (G12V) and (G13D) mutation in 1 of 44 primary gallbladder tumors and

2 of 5 gallbladder cancer cell lines that are known to be associated with differential clinical outcome in response to anti-*EGFR* therapy in colorectal cancer [127-129]. The biological characteristics of *KRAS* mutation are known to vary by cancer types as those found in pancreatic and non-small cell lung cancers are predominantly at codon 12, while in colorectal and gallbladder mutations appears to be in codon 12 and codon 13 [130]. Moreover, the clinical response among patients and studies with isogenic colon cell line indicate *KRAS* (G13D) mutation as sensitive but (G12V) as resistant to anti-*EGFR* therapy suggesting codon 13 mutations are still dependent on inductive upstream *EGFR* signaling and exhibit weaker *in vitro* transforming activity than codon 12 mutations [89, 128].

In summary, besides suggesting adoption of anti-EGFR therapy as a therapeutic option in gallbladder cancer based on *ERBB2* alteration, we present the first evidence that presence of *KRAS* (G12V) but not *KRAS* (G13D) mutation may preclude such patients to respond to the treatment, similar to the clinical algorithm commonly practiced based on *EGFR* alteration in colorectal cancer. Due to smaller sample size, this study remains underpowered to reach statistical significance for patients harboring alterations in *ERBB2* mutations. However, due to a low prevalence rate of the disease, target accrual in clinical trials has been a bottleneck in gallbladder cancer. This study thus forms the basis to test the efficacy of ERBB2 inhibitors in gallbladder cancer and would help inform design a randomized clinical trial by considering the inclusion of gallbladder patients under basket clinical trials such as the NCI–Molecular Analysis for Therapy Choice (NCI-MATCH) trial that are genomically matched [131].

Tables

III-Table 1: Clinical characteristics of the primary tumor samples

Clinical Characteristic	Variable	Frequency (N=44)
Age	Median range	53(20-75)
Sex	Male	17
	Female	27
Ethnicity	Asian (Indian)	44
Histological type	Adenocarcinoma	33
	Adenosquamous	4
	Squamous	3
	No information	4
Degree of Differentiation		
	Poor	10
	Moderate	28
	Well	2
	No information	4
Lymphovascular emboli		
	Yes	3
	No	40
	No info	1
Perineural invasion		
	Yes	4
	No	39
	No info	1
Gallstones	Yes	11
	No	32
	No info	1

III -Table 2: Exome sequencing quality control and statistics of primary tumor samples

Sample ID	Sample type	Total Reads	Mapped Reads	Mapped reads in pair	Duplication Rate (%)	Mapping %	Coverage
AD0752	Normal	63858844	35885515	34086308	11.5	56.2	168
AD0759	Normal	43455624	27369663	26143858	17.6	63.0	129
AD0754	Normal	54998184	37809365	36492238	22.7	68.7	177
AD0755	Normal	66115458	47741170	46444766	21.3	72.0	224
AD0756	Normal	39594984	23065796	21858806	16.7	58.3	108
AD0757	Normal	51254660	34364520	32857726	18.0	67.0	161
AD0437	Normal	44676668	29362140	28289852	25.5	65.7	139
AD0439	Normal	49994928	29952529	28671136	28.1	59.0	142
AD0746	Normal	37563776	21449942	20284354	17.3	57.1	100
AD0435	Normal	43581276	28263519	27285190	23.1	64.9	134
AD0739	Tumor	37795668	25657903	24820778	24.9	67.9	122
AD0758	Tumor	59032310	36867711	35288198	19.4	62.5	173
AD0741	Tumor	46034048	28780365	27600442	24.0	62.5	136
AD0742	Tumor	27047422	17293941	16669644	30.0	63.9	82
AD0743	Tumor	27047423	17293841	16669744	20.0	63.0	114
AD0744	Tumor	46434392	27105282	26073070	12.2	58.4	129
AD0438	Tumor	48018464	30157789	28970210	25.6	62.8	143
AD0440	Tumor	39564826	24359457	23292676	25.2	61.6	115
AD0748	Tumor	54159317	32433869	31016282	10.0	59.9	152
AD0745	Tumor	63778254	41274673	39646496	28.7	64.7	195
AD0750	Tumor	48550174	28391632	27054752	10.0	58.5	133
AD0747	Tumor	74760204	48372303	46588150	14.9	64.0	227
AD0740	Tumor	37795669	25657907	24820778	24.9	67.88583	128
AD0736	Tumor	54826754	35184594	35057946	12.1	62.1	167
AD0738	Tumor	36430489	25438691	25703748	25.2	71.0	127
AD0477	Tumor	46034058	28780375	27600442	24	63.519735	136
AD0761	Tumor	43581277	28263519	27285190	23.11	64.924357	134
AD1019	Cell line	68889377	68328077	68185712	26.9	65.9	220
AD1020	Cell line	73569111	73125971	73032062	26.69	66.9	235
AD1021	Cell line	64521902	63920376	63796570	26.37	67.9	205
AD1022	Cell line	62013432	61500302	61396508	26.05	68.9	198
AD1023	Cell line	63793578	62868021	62689988	26.5	69.9	204

III-Table 3 : Statistics of alterations sample-wise in exome sequencing of primary tumor samples

		Unpaired samples							Paired samples									
Tumor		**10T**	**11T**	**12T**	**13T**	**14T**	**15T**	**16T**	**1T**	**2T**	**3T**	**4T**	**5T**	**6T**	**7T**	**8T**	**9T**	**18T**
Overall distribu	**Total number of variants**	10242	6601	18193	14198	8079	21303	31940	10253	19491	6048	2566	9995	14709	33874	29240	54578	62559
	3UTR entries	360	213	312	229	307	122	423	381	524	216	142	384	265	1254	988	2086	10089
	5Flank entries	0	0	0	0	0	0	1	0	1	0	0	1	1	2	1	1	2
	5UTR entries	156	89	170	113	138	82	216	264	242	101	69	223	199	552	484	888	1863
	IGR entries	1362	758	6415	5343	1002	10698	14151	1854	6476	1180	472	1570	6002	4728	4396	7594	10821
	Intron entries	5630	3263	8128	5845	4248	7957	12436	5357	9270	3006	968	5563	5560	18204	14612	31722	28375
	De_novo_Start_InF000rame entries	0	1	2	0	3	0	2	0	4	2	2	1	0	7	7	5	15
	De_novo_Start_OutOfFrame entries	7	5	6	3	6	1	6	4	3	2	1	2	3	15	16	20	33
	Frame_Shift_Del entries	18	23	18	6	17	6	13	32	19	5	7	16	12	42	46	69	68
	Frame_Shift_Ins entries	10	3	3	7	7	0	9	27	15	10	6	19	11	27	36	47	60
	In_Frame_Del entries	3	3	2	0	1	0	3	1	3	2	1	1	3	2	2	8	10
	In_Frame_Ins entries	0	2	2	0	1	0	1	2	2	0	0	0	0	1	2	4	6
	lincRNA entries	155	57	677	503	96	1167	1512	156	435	126	56	173	552	468	408	739	968
	RNA entries	454	279	681	557	303	894	1281	462	810	272	128	448	653	1243	1096	2040	1922
	Start_Codon_Del entries	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0
	Start_Codon_Ins entries	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	2	1
	Start_Codon_SNP entries	2	3	0	0	1	1	1	1	0	0	0	0	0	9	8	5	9
	Stop_Codon_Del entries	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	2
	Stop_Codon_Ins entries	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1	1
	Missense mutation entries	1156	1019	970	894	1057	236	1107	916	903	636	388	780	873	3579	3448	4652	4157
	Nonsense mutation entries	16	22	15	9	23	7	27	18	24	19	13	13	62	52	48	51	40
	Nonstop entries	2	2	2	0	0	1	1	3	2	0	0	0	2	7	9	6	6
	Silent entries	875	825	747	668	821	114	710	729	723	446	295	764	481	3561	3519	4467	3955
	Splice entries	36	34	43	20	48	17	40	44	35	25	17	34	30	119	112	169	156
NOVEL, DBSNP, COSMIC and MyLAB distribution	**Total number of variants**	10242	6601	18193	14198	8079	21303	31940	10253	19491	6048	2566	9995	14709	33874	29240	54578	62559
	Novel entries	1624	962	1242	982	1085	1002	1755	633	825	541	262	228	1456	1061	805	1449	1770
	Exclusive Cosmic entries	25	23	20	25	15	8	31	17	21	15	14	7	102	26	19	28	20
	Exclusive DBsnp entries	2537	1766	9731	8353	1883	15707	20712	490	4205	244	75	137	4830	1616	1244	2352	7633
	Exclusive MyLAB entries	281	123	210	122	225	210	318	618	878	611	307	571	468	776	559	1616	955
	Cosmic+DBsnp common entries	89	124	60	80	75	18	67	18	14	12	19	6	31	69	63	86	103
	Cosmic+MyLAB common entries	8	3	5	2	5	3	19	39	33	48	12	25	20	46	19	73	47
	DBsnp+MyLAB common entries	5262	3177	6537	4300	4345	4309	8687	7536	12726	4180	1538	8081	7345	26353	22533	43886	47405
	Cosmic+DBsnp+MyLAB common entries	416	423	388	334	446	46	351	902	789	397	339	940	457	3927	3998	5088	4626
	Total Cosmic entries	538	573	473	441	541	75	468	976	857	472	384	978	610	4068	4099	5275	4796
COSMIC and Novel distribution	**Total number of variants**	1746	1112	1327	1089	1180	1031	1872	707	893	616	307	266	1609	1202	906	1636	1940
	3UTR entries	55	32	40	33	37	22	58	18	29	13	7	4	52	39	24	41	373
	5Flank entries	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5UTR entries	22	9	15	19	15	13	24	5	12	10	5	5	21	10	15	19	89
	IGR entries	170	112	259	213	132	277	431	93	131	56	25	29	212	126	91	144	207
	Intron entries	876	471	554	396	574	427	728	264	390	233	92	100	605	536	370	770	734
	De_novo_Start_InF000rame entries	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	De_novo_Start_OutOfFrame entries	0	1	1	2	1	1	0	1	0	0	0	0	3	2	1	0	1
	Frame_Shift_Del entries	5	10	6	2	3	5	5	4	3	0	2	2	2	8	5	7	8
	Frame_Shift_Ins entries	4	1	1	1	3	0	4	2	0	2	0	1	3	3	2	5	12
	In_Frame_Del entries	2	1	1	0	0	0	3	1	1	1	0	0	1	0	0	2	3
	In_Frame_Ins entries	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	lincRNA entries	24	7	19	11	11	18	32	10	8	9	4	8	23	9	7	23	27
	RNA entries	55	33	60	31	38	44	55	24	39	15	10	9	52	44	30	49	44
	Start_Codon_Del entries	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Start_Codon_Ins entries	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0
	Start_Codon_SNP entries	0	1	0	0	1	0	0	0	0	0	0	0	0	2	2	0	1
	Stop_Codon_Del entries	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Stop_Codon_Ins entries	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Missense mutation entries	352	277	247	247	214	155	364	185	195	171	101	63	456	267	229	376	275
	Nonsense mutation entries	5	9	2	5	11	6	16	1	8	8	3	2	49	6	7	5	1
	Nonstop entries	2	0	0	0	0	1	0	0	1	0	0	0	2	0	0	1	1
	Silent entries	163	139	113	123	130	57	144	88	74	94	52	42	116	127	111	185	154
	Splice entries	11	6	9	5	9	5	8	10	2	4	6	0	12	23	11	9	10

III-Table 4: Statistics of alterations sample-wise in exome sequencing of cell lines

Sample Name	OCUG1	G415	NOZ	SNU308	TGBC2TKB
Total number of variants	142977	135771	133874	129221	126770
Novel entries	2629	3023	2350	2140	2213
Exclusive Cosmic entries	61	56	57	56	48
Exclusive DBsnp entries	7916	8094	8820	7244	6762
Exclusive MyLAB entries	907	999	825	864	796
Cosmic+DBsnp common entries	252	249	226	256	190
Cosmic+MyLAB common entries	55	77	61	60	53
DBsnp+MyLAB common entries	118486	111331	109581	106767	105472
Cosmic+DBsnp+MyLAB common entri	12671	11942	11954	11834	11236
Total Cosmic entries	13039	12324	12298	12206	11527
Total number of variants	142977	135771	133874	129221	126770
3UTR entries	19761	18618	18294	18306	17976
5Flank entries	4128	4200	3799	3933	3886
5UTR entries	3173	3134	3074	3054	3083
IGR entries	21631	21058	21954	19401	19025
Intron entries	65074	60516	58556	57015	56625
De_novo_Start_InF000rame entries	22	18	19	19	27
De_novo_Start_OutOfFrame entries	62	54	47	60	53
Frame_Shift_Del entries	164	141	152	160	157
Frame_Shift_Ins entries	129	113	106	119	109
In_Frame_Del entries	13	15	13	12	14
In_Frame_Ins entries	12	11	12	12	13
lincRNA entries	2394	2479	2512	2252	2062
RNA entries	5705	5535	5503	5320	4922
Start_Codon_Del entries	2	2	0	1	1
Start_Codon_Ins entries	3	3	3	3	1
Start_Codon_SNP entries	10	11	12	10	11
Stop_Codon_Del entries	3	2	1	1	3
Stop_Codon_Ins entries	3	2	1	3	1
Missense mutation entries	9709	9377	9454	9224	9004
Nonsense mutation entries	100	99	91	97	109
Nonstop entries	10	10	15	10	13
Silent entries	10485	10013	9882	9848	9316
Splice entries	384	360	374	361	359
Total number of COSMIC+NOVEL vari	15668	15347	14648	14346	13740
3UTR entries	559	603	482	501	476
5Flank entries	268	303	231	265	239
5UTR entries	165	190	176	166	173
IGR entries	553	597	596	467	454
Intron entries	2026	2140	1741	1690	1691
De_novo_Start_InF000rame entries	2	4	1	1	1
De_novo_Start_OutOfFrame entries	7	4	4	3	2
Frame_Shift_Del entries	94	75	85	95	83
Frame_Shift_Ins entries	58	54	52	55	54
In_Frame_Del entries	4	6	5	5	4
In_Frame_Ins entries	1	1	3	1	1
lincRNA entries	54	56	52	50	44
RNA entries	274	295	276	241	204
Start_Codon_Del entries	2	2	0	1	1
Start_Codon_Ins entries	0	0	0	0	0
Start_Codon_SNP entries	4	4	3	5	3
Stop_Codon_Del entries	3	1	1	0	2
Stop_Codon_Ins entries	1	0	0	0	0
Missense mutation entries	5427	5214	5211	5057	4886
Nonsense mutation entries	49	51	47	47	55
Nonstop entries	4	2	4	3	5
Silent entries	5929	5570	5497	5518	5185
Splice entries	184	175	181	175	177

III-Table 5: Primers used for validation of mutations

Primer	(5'-3')
OAD1089 _CTNNb1 F_S37C_Forward	TTTGATGGAGTTGGACATGG
OAD1090 _CTNNb1 R_S37C_Reverse	CCTCAGGATTGCCTTTACCA
OAD1091_CTNNB1F_S33C_Forward	AGCTGATTTGATGGAGTTGGA
OAD1092_CTNNB1R_S33C_Reverse	CACTCAGAGAAGGAGCTGTGG
OAD_1093_MLL3F_C1114R_Forward	TCATAGCCCACAGGGAAGAG
OAD_1094_MLL3R_C1114R_Reverse	GATTGCTAGCATCGTGCAAC
OAD_1095_SMAD4F_R265H_Forward	TGAAATGGATGTTCAGGTAGGA
OAD_1096_SMAD4R_R265H_Reverse	TCAATGGCTTCTGTCCTGTG
OAD1097_ERBB2F_I767M_Forward	ATCCCTGATGGGGAGAATGT
OAD1098_ERBB2R_I767M_Reverse	GGGTCCTTCCTGTCCTCCTA
OAD1098_ERBB2F_V777L_Forward	GAGGCTGTGTGGTGTTTGG
OAD1099_ERBB2R_V777L_Reverse	CGTGGATGTCAGGCAGATG
OAD1401_ERBB2_I655V_Forward	ACCCCAAACTAGCCCTCAAT
OAD1402_ERBB2_I655V_Reverse	AGGGGGTGGTGGGTCAGT
OAD_583_TP53_H193L Forward	AGGCCCTTAGCCTCTGTAAG
OAD_584_TP53_H193L Reverse	TGCAGCTGTGGGTTGATTCC
OAD1105_ERBB3F_R1127H_Forward	GAGTCATCAGAGGGGCATGT
OAD1106_ERBB3R_R1127H_Reverse	GGGAATGGTAGGCGCTATCT
OAD1407_TP53_V217G Forward	CTTAACCCCTCCTCCCAGAG
OAD1408_TP53_V217G Reverse	CATGAGCGCTGCTCAGATAG
OAD1409_KRAS_G12V Forward	TTGGATCATATTCGTCCACAA
OAD1410_KRAS_G12V Reverse	TTTTCATGATTGAATTTTGTAAGG
OAD 1411_CTNNB1_R582W Forward	AGGAGAATGCCCTGTTTGTT
OAD 1412_CTNNB1_R582W Reverse	TTATGGTCCCTAATTTTCTGAAATG
OAD 1413_SMAD4_D441G Forward	CGGATTACCCAAGACAGAGC
OAD 1414_SMAD4_D441G Reverse	CTAGGAGCAAGGCAGCAAAC
OAD1401_ERBB2_I655V_Forward	ACCCCAAACTAGCCCTCAAT
OAD1402_ERBB2_I655V_Reverse	AGGGGGTGGTGGGTCAGT
OAD1545_EGFR _I1050V_Forward	ATACCCTCCATGAGGCACAC
OAD1546_EGFR_ I1050V_Reverse	CCAGGAGTCACGCTTTGAAC

III- Additional supporting data:

3.1 Comparison of significantly altered genes from whole exome sequencing of gallbladder cancer samples

The table compares the frequency of alterations common in this study and Japanese whole exome sequencing analysis with 29 samples. N.A designates frequency not available.

Gene name	This study (% Freq)	Nakamura et al, 2015 (% Freq)
TP53	35.3	41.3
ARID2	5.8	17.9
EGFR	5.8	14.3
SMAD4	5.8	3.4
ERBB2	17.6	N.A
SF3B1	17.6	N.A
ATM	17.6	N.A
AKAP11	17.6	N.A
CTNNB1	11.8	N.A
ATRX	11.8	N.A
CPNE4	11.8	N.A
POLE	11.8	N.A
ERBB3	5.8	N.A
ERBB4	5.8	N.A
RGPD3	5.8	N.A
RNF43	5.8	N.A
CSMD3	5.8	N.A
NF1	5.8	N.A
KMT2D	5.8	N.A
APC	5.8	N.A

4.CHAPTER 4: SUMMARY AND CONCLUSION

Gallbladder cancer is an aggressive disease with poor prognosis. The disease has been least studied from the functional genomics perspective due to the rarity of the disease worldwide. The cancer is treated mostly by surgery without lymph node involvement[49]. Currently, gem-cis(gemcitabine-cisplatin) or gem-ox(gemcitabine oxaliplatin) based treatment is given to advanced cases without any targeted therapy which offers a very modest survival benefit[12]. Few reports suggest the efficacy of the combination of cetuximab and gemcitabine in gallbladder cancer, thus demonstrating some promise for targeted therapy in gallbladder cancer[41]. Hence, there is an unmet need for better therapeutically relevant targets in GBC especially in a country like India, where the cancer is highly prevalent in the northern part of the country. Another aspect of gallbladder cancer is an association of infections with gallbladder cancer. Epidemiological studies suggest a strong association of typhoidal *Salmonella* infections with gallbladder cancer, however, the other group of *Salmonella* - non-typhoidal *Salmonella* was not associated with gallbladder cancer [51, 69]. These reports suggest an only association, no causal relationship has been established. It has been proposed that gallbladder infections with *Salmonella* trigger host manipulation pathways under predisposed of condition thus leading to cancer [19]. These factors led to the origin of my thesis. The major focus of my thesis is to identify therapeutically relevant alterations in gallbladder cancer and understanding the association of infections in gallbladder cancer using integrated approaches.

4.1 Mutant KRAS predicts response to EGFR inhibitors in gallbladder cancer cell lines.

Mutations in *KRAS* are predictive biomarkers of anti-EGFR therapy in colorectal cancer [88]. Similarly, *KRAS* alterations are also known in gallbladder cancer[132]. However, not much has been studied neither at the genomic or functional level to understand the overall biology of gallbladder cancer. The disease has been associated with worse outcomes. There are very few genome-wide studies regarding gallbladder cancer from the Chinese and Caucasian[26,

27] and from India, there are only candidate gene-based studies[18]. Hence, using a sensitive methodology such as whole exome sequencing, I set out to identify somatic aberrations (mutations and copy number) in primary tumors and tyrosine kinase proteome profiler to check the expression of activated tyrosine kinases in gallbladder cancer cell lines.

Using the whole exome sequencing approach, I identified alterations in known cancer-associated genes such as *TP53*, *ERBB2*, *SMAD4*, and *β-catenin* which are already known in the literature[22] for gallbladder cancer. Further, I specifically identified recurrent, actionable *ERBB2* alterations as well as copy number changes in *EGFR* which we show to be sensitive to *EGFR* family inhibitor. Using proteome profiler in cell lines, I observed higher phosphorylation of EGFR as compared to other tyrosine kinases in gallbladder cancer cell lines. Also, using biochemical approaches, I further observed a differential response of gallbladder cancer cell lines to the *EGFR* family inhibitor, which was primarily based on the presence of different *KRAS* mutations (codon 12 and codon 13 alterations). Similar observations in colorectal cancer have been reported wherein patients with *KRAS* (G13D) mutations respond better to anti-*EGFR* therapy than *KRAS* (G12V) mutations[133]. These findings in gallbladder cancer cell lines may have a clinical relevance in gallbladder cancer and allow patient stratification and could preclude gallbladder cancer patients from anti-EGFR therapy on the basis of *KRAS* mutational status. Our findings may change the current treatment regimen in gallbladder cancer and introduce the scope of the genome-guided precision medicine in the clinics for this disease.

4.2 Association of non-typhoidal *Salmonella* with gallbladder cancer

Epidemiological reports suggest a strong association of chronic typhoidal *Salmonella* infections with gallbladder cancer[51]. However, non-typhoidal *Salmonella* infections have been associated with systemic illness such as gastrointestinal infections[69]. Using a modified HPVDetector tool for computational subtraction, I identified the presence of non-typhoidal DNA sequences in gallbladder patients using whole exome sequencing data, which may be associated with disease progression. Further, I observed co-occurrence of *TP53*

alteration with *Salmonella* infections in gallbladder cancer patients. Recent reports suggest *Salmonella* infections induce manipulation of host metabolic pathways in *TP53* mutant and *c-Myc* amplified predisposed mice, murine gallbladder organoids, and fibroblasts[19]. Infections with typhoidal and the non-typhoidal *Salmonella* provides the necessary inflammatory stimulus for gallbladder cancer[123]. These findings expand the current understanding of infections of *Salmonella* with gallbladder cancer. Also, the findings may suggest better management of the disease by detection of *Salmonella* using molecular approaches.

In overall, my thesis puts forward the first landscape of somatic alterations in Indian gallbladder cancer genome and leads to the identification of non-typhoidal *Salmonella* species along with co-occurrence of *TP53* alterations that could potentially aid in better management of gallbladder cancer. More importantly, my study implicates *ERBB2* as a novel therapeutic target in gallbladder cancer and puts forward the first evidence that the presence of *KRAS* G12V but not *KRAS* G13D mutation may preclude patients to respond to anti-EGFR treatment in gallbladder cancer, as reported in colorectal cancer.

CHAPTER 5: BIBLIOGRAPHY

1. Wistuba, II and A.F. Gazdar, *Gallbladder cancer: lessons from a rare tumour.* Nat Rev Cancer, 2004. **4**(9): p. 695-706.
2. Luo, J., N.L. Solimini, and S.J. Elledge, *Principles of cancer therapy: oncogene and non-oncogene addiction.* Cell, 2009. **136**(5): p. 823-37.
3. Macconaill, L.E. and L.A. Garraway, *Clinical implications of the cancer genome.* J Clin Oncol, 2010. **28**(35): p. 5219-28.
4. Mortality, G.B.D. and C. Causes of Death, *Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013.* Lancet, 2015. **385**(9963): p. 117-71.
5. Chin, L., et al., *Making sense of cancer genomic data.* Genes Dev, 2011. **25**(6): p. 534-55.
6. Mardis, E.R. and R.K. Wilson, *Cancer genome sequencing: a review.* Hum Mol Genet, 2009. **18**(R2): p. R163-8.
7. Garraway, L.A. and E.S. Lander, *Lessons from the cancer genome.* Cell, 2013. **153**(1): p. 17-37.
8. International Cancer Genome, C., et al., *International network of cancer genome projects.* Nature, 2010. **464**(7291): p. 993-8.
9. Sicklick, J.K., et al., *Genomics of gallbladder cancer: the case for biomarker-driven clinical trial design.* Cancer Metastasis Rev, 2016. **35**(2): p. 263-75.
10. Nakamura, H., et al., *Genomic spectra of biliary tract cancer.* Nat Genet, 2015. **47**(9): p. 1003-10.
11. Shaffer, E.A., *Gallbladder cancer: the basics.* Gastroenterol Hepatol (N Y), 2008. **4**(10): p. 737-41.
12. Kapoor, V.K. and A.J. McMichael, *Gallbladder cancer: an 'Indian' disease.* Natl Med J India, 2003. **16**(4): p. 209-13.
13. Lazcano-Ponce, E.C., et al., *Epidemiology and molecular pathology of gallbladder cancer.* CA Cancer J Clin, 2001. **51**(6): p. 349-64.
14. Rakic, M., et al., *Gallbladder cancer.* Hepatobiliary Surg Nutr, 2014. **3**(5): p. 221-6.
15. Stinton, L.M. and E.A. Shaffer, *Epidemiology of gallbladder disease: cholelithiasis and cancer.* Gut Liver, 2012. **6**(2): p. 172-87.
16. Randi, G., S. Franceschi, and C. La Vecchia, *Gallbladder cancer worldwide: geographical distribution and risk factors.* Int J Cancer, 2006. **118**(7): p. 1591-602.
17. Barbhuiya, M.A., et al., *Gallbladder cancer incidence in Gwalior district of India: Five-year trend based on the registry of a regional cancer center.* Indian J Cancer, 2015. **52**(3): p. 430-7.
18. Srivastava, K., et al., *Candidate gene studies in gallbladder cancer: a systematic review and meta-analysis.* Mutat Res, 2011. **728**(1-2): p. 67-79.
19. Scanu, T., et al., *Salmonella Manipulation of Host Signaling Pathways Provokes Cellular Transformation Associated with Gallbladder Carcinoma.* Cell Host Microbe, 2015. **17**(6): p. 763-74.
20. Sheth, S., A. Bedford, and S. Chopra, *Primary gallbladder cancer: recognition of risk factors and the role of prophylactic cholecystectomy.* Am J Gastroenterol, 2000. **95**(6): p. 1402-10.
21. Basu, S., et al., *Heavy and trace metals in carcinoma of the gallbladder.* World J Surg, 2013. **37**(11): p. 2641-6.
22. Kumari, N., et al., *Mutation profiling in gallbladder cancer in Indian population.* Indian J Pathol Microbiol, 2014. **57**(1): p. 9-12.
23. Saranga Bharathi, R., et al., *Female sex hormone receptors in gallbladder cancer.* J Gastrointest Cancer, 2015. **46**(2): p. 143-8.
24. Sharma, V., et al., *Role of bile bacteria in gallbladder carcinoma.* Hepatogastroenterology, 2007. **54**(78): p. 1622-5.
25. Shukla, V.K., et al., *Organochlorine pesticides in carcinoma of the gallbladder: a case-control study.* Eur J Cancer Prev, 2001. **10**(2): p. 153-6.
26. Javle, M., et al., *Molecular characterization of gallbladder cancer using somatic mutation profiling.* Hum Pathol, 2014. **45**(4): p. 701-8.
27. Li, M., et al., *Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway.* Nat Genet, 2014. **46**(8): p. 872-6.

28. Javle, M., et al., *HER2/neu-directed therapy for biliary tract cancer.* J Hematol Oncol, 2015. **8**: p. 58.
29. Zuo, M., et al., *RNA sequencing-based analysis of gallbladder cancer reveals the importance of the liver X receptor and lipid metabolism in gallbladder cancer.* Oncotarget, 2016. **7**(23): p. 35302-12.
30. Gu, X., et al., *RNA sequencing reveals differentially expressed genes as potential diagnostic and prognostic indicators of gallbladder carcinoma.* Oncotarget, 2015. **6**(24): p. 20661-71.
31. Barreto, S.G., A. Dutt, and A. Chaudhary, *A genetic model for gallbladder carcinogenesis and its dissemination.* Ann Oncol, 2014. **25**(6): p. 1086-97.
32. Hanada, K., et al., *Gene mutations of K-ras in gallbladder mucosae and gallbladder carcinoma with an anomalous junction of the pancreaticobiliary duct.* Am J Gastroenterol, 1999. **94**(6): p. 1638-42.
33. Feng, Z., et al., *The risk factor of gallbladder cancer: hyperplasia of mucous epithelium caused by gallstones associates with p16/CyclinD1/CDK4 pathway.* Exp Mol Pathol, 2011. **91**(2): p. 569-77.
34. Legan, M., et al., *Expression of cyclooxygenase-2 is associated with p53 accumulation in premalignant and malignant gallbladder lesions.* World J Gastroenterol, 2006. **12**(21): p. 3425-9.
35. Wistuba, II, et al., *Allele-specific mutations involved in the pathogenesis of endemic gallbladder carcinoma in Chile.* Cancer Res, 1995. **55**(12): p. 2511-5.
36. Koda, M., et al., *Expression of Fhit, Mlh1, and P53 protein in human gallbladder carcinoma.* Cancer Lett, 2003. **199**(2): p. 131-8.
37. Wistuba, II, et al., *High resolution chromosome 3p, 8p, 9q and 22q allelotyping analysis in the pathogenesis of gallbladder carcinoma.* Br J Cancer, 2002. **87**(4): p. 432-40.
38. Roa, I., et al., *Overexpression of the HER2/neu Gene: A New Therapeutic Possibility for Patients With Advanced Gallbladder Cancer.* Gastrointest Cancer Res, 2014. **7**(2): p. 42-8.
39. Wu, Q., et al., *Therapeutic effect of rapamycin on gallbladder cancer in a transgenic mouse model.* Cancer Res, 2007. **67**(8): p. 3794-800.
40. Lee, J., et al., *Gemcitabine and oxaliplatin with or without erlotinib in advanced biliary-tract cancer: a multicentre, open-label, randomised, phase 3 study.* Lancet Oncol, 2012. **13**(2): p. 181-8.
41. Gruenberger, B., et al., *Cetuximab, gemcitabine, and oxaliplatin in patients with unresectable advanced or metastatic biliary tract cancer: a phase 2 study.* Lancet Oncol, 2010. **11**(12): p. 1142-8.
42. Lubner, S.J., et al., *Report of a multicenter phase II trial testing a combination of biweekly bevacizumab and daily erlotinib in patients with unresectable biliary cancer: a phase II Consortium study.* J Clin Oncol, 2010. **28**(21): p. 3491-7.
43. El-Khoueiry, A.B., et al., *SWOG 0514: a phase II study of sorafenib in patients with unresectable or metastatic gallbladder carcinoma and cholangiocarcinoma.* Invest New Drugs, 2012. **30**(4): p. 1646-51.
44. Yi, J.H., et al., *A phase II study of sunitinib as a second-line treatment in advanced biliary tract carcinoma: a multicentre, multinational study.* Eur J Cancer, 2012. **48**(2): p. 196-201.
45. Andrea, C. and A. Enzo, *Cholesterol Gallstones Larger Than 3 cm Appear to Be Associated With Gallbladder Cancer: Identification of a High Risk Group of Patients That Could Benefit From Preventive Cholecystectomy.* Ann Surg, 2016. **263**(3): p. e56.
46. Csendes, A., et al., *Number and size of stones in patients with asymptomatic and symptomatic gallstones and gallbladder carcinoma: a prospective study of 592 cases.* J Gastrointest Surg, 2000. **4**(5): p. 481-5.
47. Shrikhande, S.V., et al., *Cholelithiasis in gallbladder cancer: coincidence, cofactor, or cause!* Eur J Surg Oncol, 2010. **36**(6): p. 514-9.
48. Hundal, R. and E.A. Shaffer, *Gallbladder cancer: epidemiology and outcome.* Clin Epidemiol, 2014. **6**: p. 99-109.
49. Kanthan, R., et al., *Gallbladder Cancer in the 21st Century.* J Oncol, 2015. **2015**: p. 967472.

50. Serra, I., et al., *Association of chili pepper consumption, low socioeconomic status and longstanding gallstones with gallbladder cancer in a Chilean population.* Int J Cancer, 2002. **102**(4): p. 407-11.
51. Kumar, S., S. Kumar, and S. Kumar, *Infection as a risk factor for gallbladder cancer.* J Surg Oncol, 2006. **93**(8): p. 633-9.
52. Samaras, V., et al., *Chronic bacterial and parasitic infections and cancer: a review.* J Infect Dev Ctries, 2010. **4**(5): p. 267-81.
53. de Martel, C., et al., *Helicobacter species in cancers of the gallbladder and extrahepatic biliary tract.* Br J Cancer, 2009. **100**(1): p. 194-9.
54. Yadav, A., et al., *Association of Wnt signaling pathway genetic variants in gallbladder cancer susceptibility and survival.* Tumour Biol, 2016. **37**(6): p. 8083-95.
55. Mhatre, S., et al., *Common genetic variation and risk of gallbladder cancer in India: a case-control genome-wide association study.* Lancet Oncol, 2017. **18**(4): p. 535-544.
56. Liebe, R., et al., *Modifiable Factors and Genetic Predisposition Associated with Gallbladder Cancer. A Concise Review.* J Gastrointestin Liver Dis, 2015. **24**(3): p. 339-48.
57. Lai, C.H. and W.Y. Lau, *Gallbladder cancer--a comprehensive review.* Surgeon, 2008. **6**(2): p. 101-10.
58. Gore, R.M. and R.P. Shelhamer, *Biliary tract neoplasms: diagnosis and staging.* Cancer Imaging, 2007. **7 Spec No A**: p. S15-23.
59. de Martel, C. and S. Franceschi, *Infections and cancer: established associations and new hypotheses.* Crit Rev Oncol Hematol, 2009. **70**(3): p. 183-94.
60. Selgrad, M., et al., *The role of viral and bacterial pathogens in gastrointestinal cancer.* J Cell Physiol, 2008. **216**(2): p. 378-88.
61. Vandeven, N. and P. Nghiem, *Pathogen-driven cancers and emerging immune therapeutic strategies.* Cancer Immunol Res, 2014. **2**(1): p. 9-14.
62. Shukla, V.K., et al., *Carcinoma of the gallbladder--is it a sequel of typhoid?* Dig Dis Sci, 2000. **45**(5): p. 900-3.
63. Mager, D.L., *Bacteria and cancer: cause, coincidence or cure? A review.* J Transl Med, 2006. **4**: p. 14.
64. Caygill, C.P., et al., *The association between typhoid carriage, typhoid infection and subsequent cancer at a number of sites.* Eur J Cancer Prev, 1995. **4**(2): p. 187-93.
65. Dutta, U., et al., *Typhoid carriers among patients with gallstones are at increased risk for carcinoma of the gallbladder.* Am J Gastroenterol, 2000. **95**(3): p. 784-7.
66. Kinoshita, N. and H.V. Gelboin, *beta-Glucuronidase catalyzed hydrolysis of benzo(a)pyrene-3-glucuronide and binding to DNA.* Science, 1978. **199**(4326): p. 307-9.
67. Chipman, J.K., *Bile as a source of potential reactive metabolites.* Toxicology, 1982. **25**(2-3): p. 99-111.
68. Aries, V., et al., *Bacteria and the aetiology of cancer of the large bowel.* Gut, 1969. **10**(5): p. 334-5.
69. Crump, J.A., et al., *Epidemiology, Clinical Presentation, Laboratory Diagnosis, Antimicrobial Resistance, and Antimicrobial Management of Invasive Salmonella Infections.* Clin Microbiol Rev, 2015. **28**(4): p. 901-37.
70. Kostic, A.D., et al., *Genomic analysis identifies association of Fusobacterium with colorectal carcinoma.* Genome Res, 2012. **22**(2): p. 292-8.
71. Marshall, B.J. and J.R. Warren, *Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration.* Lancet, 1984. **1**(8390): p. 1311-5.
72. Wu, X.S., et al., *Evaluation of two inflammation-based prognostic scores in patients with resectable gallbladder carcinoma.* Ann Surg Oncol, 2014. **21**(2): p. 449-57.
73. Krause, D.S. and R.A. Van Etten, *Tyrosine kinases as targets for cancer therapy.* N Engl J Med, 2005. **353**(2): p. 172-87.
74. Druker, B.J., et al., *Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome.* N Engl J Med, 2001. **344**(14): p. 1038-42.
75. Flaherty, K.T., et al., *Inhibition of mutated, activated BRAF in metastatic melanoma.* N Engl J Med, 2010. **363**(9): p. 809-19.

76. Rusnak, D.W., et al., *The effects of the novel, reversible epidermal growth factor receptor/ErbB-2 tyrosine kinase inhibitor, GW2016, on the growth of human normal and tumor-derived cell lines in vitro and in vivo.* Mol Cancer Ther, 2001. **1**(2): p. 85-94.
77. Lynch, T.J., et al., *Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib.* N Engl J Med, 2004. **350**(21): p. 2129-39.
78. Christensen, J.G., et al., *Cytoreductive antitumor activity of PF-2341066, a novel inhibitor of anaplastic lymphoma kinase and c-Met, in experimental models of anaplastic large-cell lymphoma.* Mol Cancer Ther, 2007. **6**(12 Pt 1): p. 3314-22.
79. Paez, J.G., et al., *EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy.* Science, 2004. **304**(5676): p. 1497-500.
80. McGranahan, N. and C. Swanton, *Biological and therapeutic impact of intratumor heterogeneity in cancer evolution.* Cancer Cell, 2015. **27**(1): p. 15-26.
81. Le Tourneau, C., et al., *Randomised proof-of-concept phase II trial comparing targeted therapy based on tumour molecular profiling vs conventional therapy in patients with refractory cancer: results of the feasibility part of the SHIVA trial.* Br J Cancer, 2014. **111**(1): p. 17-24.
82. De Roock, W., et al., *Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis.* Lancet Oncol, 2010. **11**(8): p. 753-62.
83. Spivak-Kroizman, T., et al., *Heterodimerization of c-erbB2 with different epidermal growth factor receptor mutants elicits stimulatory or inhibitory responses.* J Biol Chem, 1992. **267**(12): p. 8056-63.
84. Graus-Porta, D., et al., *ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling.* EMBO J, 1997. **16**(7): p. 1647-55.
85. Garner, A.P., et al., *An antibody that locks HER3 in the inactive conformation inhibits tumor growth driven by HER2 or neuregulin.* Cancer Res, 2013. **73**(19): p. 6024-35.
86. Tebbutt, N., M.W. Pedersen, and T.G. Johns, *Targeting the ERBB family in cancer: couples therapy.* Nat Rev Cancer, 2013. **13**(9): p. 663-73.
87. Imamura, Y., et al., *Specific mutations in KRAS codons 12 and 13, and patient prognosis in 1075 BRAF wild-type colorectal cancers.* Clin Cancer Res, 2012. **18**(17): p. 4753-63.
88. Yoon, H.H., et al., *KRAS codon 12 and 13 mutations in relation to disease-free survival in BRAF-wild-type stage III colon cancers from an adjuvant chemotherapy trial (N0147 alliance).* Clin Cancer Res, 2014. **20**(11): p. 3033-43.
89. Morelli, M.P. and S. Kopetz, *Hurdles and complexities of codon 13 KRAS mutations.* J Clin Oncol, 2012. **30**(29): p. 3565-7.
90. Zhu, A.X., et al., *Current management of gallbladder carcinoma.* Oncologist, 2010. **15**(2): p. 168-81.
91. Shrikhande, S., et al., *Cholelithiasis in gallbladder cancer: coincidence, cofactor, or cause!* Eur J Surg Oncol, 2010. **36**(6): p. 514-519.
92. Dai, M., Y. Fong, and A. Lowy, *Treatment of T3 gallbladder cancer.* J Gastrointest Surg, 2009. **13**(11): p. 2040-2.
93. Miller, G. and W.R. Jarnagin, *Gallbladder carcinoma.* Eur J Surg Oncol, 2008. **34**(3): p. 306-12.
94. Wolff, A.C., et al., *Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update.* J Clin Oncol, 2013. **31**(31): p. 3997-4013.
95. Jiao, Y., et al., *Exome sequencing identifies frequent inactivating mutations in BAP1, ARID1A and PBRM1 in intrahepatic cholangiocarcinomas.* Nat Genet, 2013. **45**(12): p. 1470-1473.
96. Upadhyay, P., et al., *Notch pathway activation is essential for maintenance of stem-like cells in early tongue cancer.* Oncotarget, 2016. **7**(31): p. 50437-50449.
97. Chandrani, P., et al., *Drug-sensitive FGFR3 mutations in lung adenocarcinoma.* Ann Oncol, 2017. **28**(3): p. 597-603.
98. Lawrence, M.S., et al., *Mutational heterogeneity in cancer and the search for new cancer-associated genes.* Nature, 2013. **499**(7457): p. 214-8.

99. Gonzalez-Perez, A., et al., *IntOGen-mutations identifies cancer drivers across tumor types.* Nat Methods, 2013. **10**(11): p. 1081-2.
100. Forbes, S.A., et al., *The Catalogue of Somatic Mutations in Cancer (COSMIC).* Curr Protoc Hum Genet, 2008. **Chapter 10**: p. Unit 10 11.
101. Sherry, S.T., et al., *dbSNP: the NCBI database of genetic variation.* Nucleic Acids Res, 2001. **29**(1): p. 308-11.
102. Upadhyay, P., et al., *TMC-SNPdb: an Indian germline variant database derived from whole exome sequences.* Database (Oxford), 2016. **2016**.
103. Ramos, A.H., et al., *Oncotator: cancer variant annotation tool.* Hum Mutat, 2015. **36**(4): p. E2423-9.
104. Boeva, V., et al., *Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data.* Bioinformatics, 2012. **28**(3): p. 423-5.
105. Chandrani, P., et al., *Integrated genomics approach to identify biologically relevant alterations in fewer samples.* BMC Genomics, 2015. **16**: p. 936.
106. Godbole, M., et al., *Progesterone suppresses the invasion and migration of breast cancer cells irrespective of their progesterone receptor status - a short report.* Cell Oncol (Dordr), 2017. **40**(4): p. 411-417.
107. Mullighan, C.G., et al., *CREBBP mutations in relapsed acute lymphoblastic leukaemia.* Nature, 2011. **471**(7337): p. 235-9.
108. Bodor, C., et al., *EZH2 Y641 mutations in follicular lymphoma.* Leukemia, 2011. **25**(4): p. 726-9.
109. Lovejoy, C.A., et al., *Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway.* PLoS Genet, 2012. **8**(7): p. e1002772.
110. Yoshida, K. and S. Ogawa, *Splicing factor mutations and cancer.* Wiley Interdiscip Rev RNA, 2014. **5**(4): p. 445-59.
111. Kancha, R.K., et al., *Differential sensitivity of ERBB2 kinase domain mutations towards lapatinib.* PLoS One, 2011. **6**(10): p. e26760.
112. Bose, R., et al., *Activating HER2 mutations in HER2 gene amplification negative breast cancer.* Cancer Discov, 2013. **3**(2): p. 224-37.
113. Nakazawa, K., et al., *Amplification and overexpression of c-erbB-2, epidermal growth factor receptor, and c-met in biliary tract cancers.* J Pathol, 2005. **206**(3): p. 356-65.
114. Tural, D., et al., *Association KRAS G13D tumor mutated outcome in patients with chemotherapy refractory metastatic colorectal cancer treated with cetuximab.* Hepatogastroenterology, 2013. **60**(125): p. 1035-40.
115. Choughule, A., et al., *Coexistence of KRAS mutation with mutant but not wild-type EGFR predicts response to tyrosine-kinase inhibitors in human lung cancer.* Br J Cancer, 2014. **111**(11): p. 2203-4.
116. Osumi, H., et al., *Cetuximab treatment for metastatic colorectal cancer with KRAS p.G13D mutations improves progression-free survival.* Mol Clin Oncol, 2015. **3**(5): p. 1053-1057.
117. Kumar, S.S., et al., *KRAS G13D Mutation and Sensitivity to Cetuximab or Panitumumab in a Colorectal Cancer Cell Line Model.* Gastrointest Cancer Res, 2014. **7**(1): p. 23-6.
118. Linggi, B. and G. Carpenter, *ErbB receptors: new insights on mechanisms and biology.* Trends Cell Biol, 2006. **16**(12): p. 649-56.
119. Zhou, X. and Y.M. Agazie, *The signaling and transformation potency of the overexpressed HER2 protein is dependent on the normally-expressed EGFR.* Cell Signal, 2012. **24**(1): p. 140-50.
120. Kandioler-Eckersberger, D., et al., *TP53 mutation and p53 overexpression for prediction of response to neoadjuvant treatment in breast cancer patients.* Clin Cancer Res, 2000. **6**(1): p. 50-6.
121. Kandioler-Eckersberger, D., et al., *The TP53 genotype but not immunohistochemical result is predictive of response to cisplatin-based neoadjuvant therapy in stage III non-small cell lung cancer.* J Thorac Cardiovasc Surg, 1999. **117**(4): p. 744-50.
122. Yanagisawa, N., et al., *Sporadic TP53 transition mutations in chronic cholecystitis are possibly linked to gallbladder carcinogenesis.* Anticancer Res, 2010. **30**(11): p. 4443-9.

123. Iyer, P., et al., *Non-typhoidal Salmonella DNA traces in gallbladder cancer.* Infect Agent Cancer, 2016. **11**: p. 12.
124. Petitjean, A., et al., *TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes.* Oncogene, 2007. **26**(15): p. 2157-65.
125. Doval, D.C., et al., *Expression of epidermal growth factor receptor, p53, Bcl2, vascular endothelial growth factor, cyclooxygenase-2, cyclin D1, human epidermal receptor-2 and Ki-67: Association with clinicopathological profiles and outcomes in gallbladder carcinoma.* J Carcinog, 2014. **13**: p. 10.
126. Muller, B.G., X. De Aretxabala, and M. Gonzalez Domingo, *A review of recent data in the treatment of gallbladder cancer: what we know, what we do, and what should be done.* Am Soc Clin Oncol Educ Book, 2014: p. e165-70.
127. Guerrero, S., et al., *K-ras codon 12 mutation induces higher level of resistance to apoptosis and predisposition to anchorage-independent growth than codon 13 mutation or proto-oncogene overexpression.* Cancer Res, 2000. **60**(23): p. 6750-6.
128. De Roock, W., et al., *Association of KRAS p.G13D mutation with outcome in patients with chemotherapy-refractory metastatic colorectal cancer treated with cetuximab.* JAMA, 2010. **304**(16): p. 1812-20.
129. Li, W., et al., *Colorectal carcinomas with KRAS codon 12 mutation are associated with more advanced tumor stages.* BMC Cancer, 2015. **15**: p. 340.
130. Prior, I.A., P.D. Lewis, and C. Mattos, *A comprehensive survey of Ras mutations in cancer.* Cancer Res, 2012. **72**(10): p. 2457-67.
131. Do, K., G. O'Sullivan Coyne, and A.P. Chen, *An overview of the NCI precision medicine trials-NCI MATCH and MPACT.* Chin Clin Oncol, 2015. **4**(3): p. 31.
132. Roa, J.C., et al., *[K-ras gene mutation in gallbladder carcinoma].* Rev Med Chil, 2004. **132**(8): p. 955-60.
133. Chen, J., et al., *Association between KRAS codon 13 mutations and clinical response to anti-EGFR treatment in patients with metastatic colorectal cancer: results from a meta-analysis.* Cancer Chemother Pharmacol, 2013. **71**(1): p. 265-72.

79 | P a g e

6.CHAPTER 6: APPENDIX

(Attached as a google drive link due to huge size of the document)

6.1 APPENDIX I: List of variants identified in whole exome sequencing of 27 samples

6.2 APPENDIX 2: List of variants identified in whole exome sequencing of 5 gallbladder cancer cell lines.

6.3 APPENDIX 3: Detailed annotation table of read sequences of different salmonella species identified across gallbladder cancer patient samples

7. CHAPTER 7: REPRINTS OF PUBLICATIONS