MOLECULAR TARGETS FOR CANCER THERAPEUTICS: INSIGHTS FROM GENOMIC ABERRATIONS AND PROTEIN INTERACTIONS

VIDHI MALIK

DEPARTMENT OF BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY

INDIAN INSTITUTE OF TECHNOLOGY DELHI

SEPTEMBER 2020

MOLECULAR TARGETS FOR CANCER THERAPEUTICS: INSIGHTS FROM GENOMIC ABERRATIONS AND PROTEIN INTERACTIONS

by

VIDHI MALIK

DEPARTMENT OF BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

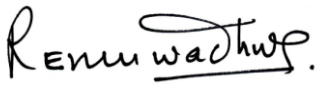
SEPTEMBER 2020

This thesis is dedicated to the memory of my father, for his love, endless support, encouragement & sacrifices

CERTIFICATE

This is to certify that the thesis entitled **'Molecular Targets for Cancer Therapeutics: Insights from Genomic Aberrations and Protein Interactions'** being submitted by **Ms. Vidhi Malik** to the Indian Institute of Technology Delhi for the award of the degree of '**Doctor of Philosophy**', is a record of the bonafide research work carried out by her, which has been prepared under our supervision in conformity with the rules and regulations of the Indian Institute of Technology Delhi. The research reports and the results presented in this thesis have not been submitted for any degree or diploma in any other University or Institute.

Prof. D. Sundar
Head, Department of Biochemical Engineering and Biotechnology
Indian Institute of Technology Delhi



Dr. Renu Wadhwa
Research Group Leader, Cellular and Molecular Biotechnology Research Institute
National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

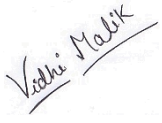
ACKNOWLEDGEMENTS

Firstly, I would like to express my gratitude to my supervisor Prof. D. Sundar for his constant support and guidance for my Ph.D. research. I sincerely appreciate the freedom he has given to me to pursue various projects without objection. He is a person with great organization and management skills and always aims for perfection. Working under his supervision always motivated me to work with my full potential. I really cannot thank him enough for his sensitivity and generous support that helped me going through tough times of life. I am also very grateful to my co-supervisor Dr. Renu Wadhwa from AIST, Japan for her scientific advice and knowledge and many insightful discussions and suggestions. I would also like to thank Dr. Sunil Kaul from AIST, Japan for providing me an opportunity to perform *wet lab* experiments and for his valuable advice and encouragement to give my best.

My sincere thankfulness to the rest of my thesis committee: Prof. Shilpi Sharma, Prof. Ritu Kulshreshtha, and Prof. Ashok Kumar Patel (Kusuma School of Biological Sciences, IITD), for not only their insightful comments and encouragement, but also for the questions which motivated me to widen my research from various perspectives.

I thank my fellow lab-mates for the stimulating discussions, for tiring working hours, the sleepless nights we were working together before deadlines, and for all the fun we have had in the last few years. I would especially like to thank my seniors Ms. Shayoni Dutta and Mr. Shashank P. Katiyar for their help and guidance, I have learnt a lot from them. I would also like to mention sincere efforts of Mr. Yogesh Kalakoti, Mr. Vipul Kumar and Mr. R. Navaneethan for helping me in some of my projects. I am very grateful to Ms. Anjani Kumari, Ms. Shayoni Dutta, Ms. Kamya Singh and Mr. Moolchand Sigar for providing me emotional support and strength and ensuring that I never lost hope. I appreciate the contribution of all the people who shared the lab with me at some point of time during this long journey, for keeping the lab environment full of energy, love and fun. It was a pleasure to meet all these people.

Lastly, but most importantly, I would like to thank my family, my parents, sisters, brother and my in-laws for all their love and support and for always believing in my capabilities. Your faith in me has always fueled my motivations. I would specifically like to mention the support and efforts of my parents, Mr. Pradeep Malik and Ms. Sushma Malik, and my husband Mr. Siddharth Teotia, for always ensuring that I never doubt about my abilities. They always inspired me to aspire for knowledge, health and peace.



Vidhi Malik

ABSTRACT

Cancer is a major health concern in both developed and developing countries. Lack of knowledge of underlying mechanisms leading to cancer, is one of the reasons responsible for absence of effective medication with least side effects. Adding to complexity of the situation, any two cancer patients will not respond in a similar manner to same treatment due to different molecular signature of aberration that lead to cancer. This is a major motivation driving the research in the area of personalized medicine. Large numbers of cancer-genomics projects like The Cancer Genome Atlas (TCGA) and Genomics of Drug Sensitivity in Cancer (GDSC) provide a wealth of information at multi-omics level. Machine learning-based models trained on multi-omics data and clinical health record of thousands of patients has a potential to a) identify all possible set of molecular aberration signature responsible for each cancer type and b) correlate these aberration signature to drug-response prediction. In this thesis, I have attempted to leverage the technology of deep- and machine learning to identify genomic aberrations that could serve as potential therapeutic target or prognostic/diagnostic biomarkers that could aid in prediction of survival-prognosis and drug response of patients in clinical settings. In addition to use of available data to bridge the gap between research and clinical practices, our laboratory group is actively involved in encouraging the use of compounds, extracted from natural sources like 'Ashwagandha' and 'honeybee propolis', in cancer treatment by studying the underlying mechanism of its anti-cancer activity specific to cancer cells only. Withanolides (secondary metabolites of Ashwagandha) and CAPE (bioactive component of honeybee propolis) have a very impressive anti-cancer pharmacological profile due to their reported properties like anti-metastatic, anti-apoptotic, anti-stress, anti-oxidant, anti-inflammatory and anti-angiogenic activity. In light of the reported activities of these natural compounds and pathways dysregulation reported in cancer cells with respect to their normal counterparts, an attempt has been made in this thesis to link this information to identify possible pathways affected by these compounds. The identification of pathways were followed by exploring the possible target protein to find the action mechanism behind specific activity of these natural compounds using bioinformatics and experimental approaches. This kind of basic research on unveiling the interactions and mode of action of these natural compounds with their target proteins and their resultant effect on pathways in cancer cell and normal cells, will aid in promoting the use of natural compounds in clinical practices in the long run.

सार

विकसित और विकासशील दोनों देशों में कैंसर एक चिंताजनक स्वास्थ्य समस्या है। कैंसर के बुनियादी कारणों में ज्ञान की कमी का होना ही कम दुष्प्रभाव वाली किसी सफल दवाई के ना बना पाने का असली कारण है। कैंसर का उपचार और भी जटिल इसलिए हो जाता है ,क्योंकि कोई भी दो कैंसर पीड़ित व्यक्ति समान चिकित्सा के प्रति एक जैसी प्रतिक्रिया प्रदर्शित नहीं करते, क्योंकि आणविक स्तर पर कैंसर पैदा करने वाले कारक भिन्न भिन्न होते हैं। यह व्यक्तिगत चिकित्सा के क्षेत्र में अनुसंधान को प्रेरित करने का मुख्य कारण है। बड़ी संख्या में कैंसर जिनोमिक्स परियोजनाएं जैसे कि द कैंसर जीनोम एटलस (टी.सी.जी.ए.), और जिनोमिक्स ऑफ ड्रग सेंसटिविटी इन कैंसर (जी.डी.एस.सी.) मल्टी-ओमिक्स स्तर पर इस क्षेत्र में काफी जानकारी प्रदान कर रहें है। मशीन लर्निंग आधारित मॉडल जो हजारों रोगियों के नैदानिक स्वास्थ्य रिकॉर्ड और मल्टी ओमिक्स डाटा पर प्रशिक्षित है, के लिए संभावित है कि (क) वह प्रत्येक प्रकार के होने वाले कैंसर के लिए जिम्मेदार आणविक स्तर पर होने वाले परिवर्तन को पहचान सके (ख) और इन परिवर्तनों के अध्ययन से कैंसर रोगी पर किसी दवाई की प्रतिक्रिया को पूर्व सूचित कर सकें। इस थीसिस में मैंने जिनोमिक स्तर पर परिवर्तनों को पहचानने के लिए डीप और मशीन लर्निंग तकनीक का प्रयोग किया है। यह संभावित है कि इनमें से कुछ जिनोमिक परिवर्तनों का उपयोग चिकित्सकीय लक्ष्य एवं निदान बायो मार्कर की तरह किया जा सके, तथा मशीन लर्निंग आधारित पूर्वसूचना प्रदान करने वाला मॉडल रोगियों के जीवित रहने की दर व उन पर दवाई के होने वाले असर के अनुमान में मदद कर सकता है। अनुसंधान और नैदानिक प्रथाओं के बीच की दूरी को खत्म करने के लिए उपलब्ध आंकड़ों के उपयोग के अलावा, हमारी प्रयोगशाला की टीम ,प्राकृतिक स्रोतों से निकाले गए योगीको, जैसे कि अश्वगंधा और मधुमक्खी प्रोपोलिस के उपयोग को कैंसर के इलाज में सक्रिय रुप से बढ़ावा दे रही है। हमारी प्रयोगशाला की टीम इन प्राकृतिक यौगिकों की कैंसर विरोधी गतिविधि के अंतर्निहित तंत्र और कैंसर कोशिकाओं के प्रति उनकी विशेषता का अध्ययन कर रही है। विथनॉलिड्स (अश्वगंधा के सेकेंडरी मेटाबोलाइट) और सी.ए.पी.इ. (मधुमक्खी के प्रोपोलिस का बायो एक्टिव घटक) में एंटी-मेटास्टेटिक, एंटी-एपोप्टोटिक, एंटी-स्ट्रेस, एंटी-ऑक्सीडेंट, एंटी-इंफ्लेमेटरी जैसे उनके रिपोर्ट किए गए गुणों के कारण कैंसर रोधी औषधीय प्रोफ़ाइल है। इस थीसिस में इन यौगिकों से संभावित प्रभावित पाथवेज की पहचान हेतु कैंसर कोशिकाओं में घोषित पाथवेज में आनियंत्रण और इन प्राकृतिक योगिकों की प्रतिवेदित एंटी कैंसर प्रतिक्रियाओं को जोड़ने का प्रयास किया गया है। इन यौगिकों से प्रभावित जैविक पाथवेज की पहचान के बाद इनके संभावित लक्ष्य प्रोटीन की पहचान करने के लिए जैव सूचना विज्ञान और प्रायोगिक दृष्टिकोणों का उपयोग करके इन प्राकृतिक यौगिकों की विशिष्ट गतिविधि के पीछे की क्रियाविधि को खोजा गया है। इस प्रकार के मूलभूत शोध जो इन प्राकृतिक यौगिकों के काम करने के तरीके एवं इनके टारगेट प्रोटींस के साथ तालमेल का अनावरण करती हैं,और कैंसर कोशिका और सामान्य कोशिकाओं के पाथवेज पर उनके परिणामी प्रभाव की व्याख्या करती हैं। भविष्य में इस प्रकार के शोधों के परिणाम स्वरूप नैदानिक प्रथाओं में प्राकृतिक यौगिकों के प्रयोग को बढ़ावा मिलेगा।

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABL1/ Abl	Abelson murine leukemia viral oncogene homolog 1
ASH-WEX	Water extract of Ashwagandha leaves
BRCA	Breast Invasive Carcinoma
CAPE	Caffeic Acid Phenethyl Ester
CCLE	Cancer Cell Line Encyclopedia
CNV	Copy Number Variation
FDA	U.S. Food and Drug Administration
GDSC	Genomics of Drug Sensitivity in Cancer
IKK	IκB kinase
IVS	Inverse virtual screening
LUAD	Lund Adenocarcinoma
MD	Molecular dynamics
MM	Multiple Myeloma
MMPs	Matrix metalloproteinases
NFκB	Nuclear factor kappa light chain enhancer of activated B cells
NSCLC	Non-small cell lung cancer
RNA-Seq	RNA Sequencing
ROC	Receiver operating characteristic
TCGA	The Cancer Genome Atlas
TEG	Triethylene glycol
TD-10	Triethyleneglycol dimethacrylate
TD-11	Tetraethyleneglycol dimethacrylate
VEGF-A	Vascular endothelial growth factor A
WHO	World Health Organization
Wi-A	Withaferin A
Wi-N	Withanone