

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

© Indian Institute of Technology Delhi (IITD), New Delhi, 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

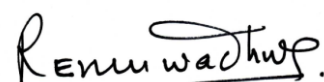
This is to certify that the thesis entitled '**Molecular Targets for Cancer Therapeutics: Insights from Genomic Aberrations and Protein Interactions**' being submitted by **Ms. Vidhi Malik** to the Indian Institute of Technology Delhi for the award of the degree of '**Doctor of Philosophy**', is a record 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

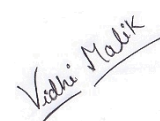
ACKNOWLEDGEMENTS

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. Kamyia 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 सेंसिटिविटी इन कैंसर (जी.डी.एस.सी.) मल्टी-ओमिक्स स्तर पर इस क्षेत्र में काफी जानकारी प्रदान कर रहे हैं। 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 रूप 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 प्रोफाइल है। इस थीसिस में इन यौगिकों से संभावित प्रभावित पाथवेज की पहचान हेतु कैंसर कोशिकाओं में घोषित पाथवेज में आनियंत्रण और इन प्राकृतिक यौगिकों 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

Certificate	i
Acknowledgements	iii
Abstract.....	v
संर.....	vii
List of Figures.....	xvii
List of Tables	xxi
List of Abbreviations	xxv
Chapter 1 Introduction.....	1
1.1 Background	1
1.1.1 Anti-cancer potential of medicinal herbs.....	2
1.1.2 Secondary metabolites from natural sources shortlisted for this study	3
1.1.2.1 Active compounds extracted from leaves of Ashwagandha	3
1.1.2.2 CAPE, bioactive compound from propolis of honeybee hives	4
1.1.3 Approaches for identification of target proteins to study the underlying mechanism of anti-cancer activity of natural compounds	5
1.1.3.1 Literature evidences for specific activity of target protein in cancer biology to unveil mechanism of counter-activity offered by natural compounds.	6
1.1.3.2 Inverse virtual screening (IVS) of natural compounds with a database of target proteins.....	6
1.1.3.3 Transcriptome analysis for identification of dysregulated pathways in cancer	7
1.1.4 Study the action mechanism of anti-cancer activity of natural compounds	9
1.1.5 Multi-omics data analysis of cancer cell lines.....	11
1.1.6 Application of multi-omics cancer data analysis.....	11
1.1.6.1 Predicting genomic aberration signature	12
1.1.6.2 Survival prognosis	12
1.1.6.3 Drug response prediction.....	12
1.2 Motivation.....	13
1.3 Definition of problem.....	14
1.4 Objectives.....	15

Chapter 2 Molecular mechanism of anti-cancer potential of active components of Ashwagandha.....	17
2.1 Introduction.....	17
2.2 Effect of Wi-A on NFκB pathway activation.....	18
2.2.1 Computational Methods	19
2.2.1.1 Molecular docking of Wi-A with IKKβ/NEMO association domain	19
2.2.1.2 Molecular docking of Wi-A with NFκB dimer (p50-p65 dimer) DNA binding domain.....	20
2.2.1.3 MD simulation studies revealing the effect of Wi-A on IKKβ/NEMO association domain and DNA binding domain of NFκB dimer	21
2.2.1.4 Analysis of molecular docking and MD simulations results	22
2.2.2 Results and Discussion.....	23
2.2.2.1 Effect of Wi-A on IKK complex assembly or its conformation	23
2.2.2.2 Effect of Wi-A on DNA binding activity of NFκB dimer	27
2.2.3 Experimental evidences for effect of Wi-A on NFκB pathway	29
2.2.4 Conclusion	30
2.3 Mechanism of anti-metastatic property of withanolides by targeting EMT biomarker, vimentin.....	31
2.3.1 Background	31
2.3.2 Computational Methods	32
2.3.2.1 Molecular interaction study of Wi-A and 3βmWi-A with monomeric and multimeric forms of vimentin.....	32
2.3.2.2 Analysis of molecular docking and MD simulation results	33
2.3.3 Results and Discussion.....	33
2.3.4 Experimental evidences for anti-metastatic activity of Wi-A and 3βmWi-A.....	38
2.3.5 Conclusion	39
2.4 Effect of Wi-A on RAD51 activity to unveil its underlying action mechanism on alternate lengthening of telomeres (ALT) cell lines	39
2.4.1 Background	39
2.4.2 Computational methods	40
2.4.3 Results and Discussion.....	41
2.4.4 Conclusion	44

2.5 Study the mechanism of anti-metastasis and anti-angiogenic activities of TEG derivatives - TD-10 and TD-11	45
2.5.1 Background	45
2.5.2 Computational Methods	45
2.5.2.1 Molecular docking to check the effect of TD-10 and TD-11 on MMPs (MMP2, MMP7 and MMP9) family of protein.....	45
2.5.2.2 Molecular docking to check the effect of TD-10 and TD-11 on VEGF-A protein and VEGF-A-VEGFR-1 complex	46
2.5.2.3 MD simulation of protein-ligand complexes	46
2.5.3 Results and Discussion.....	47
2.5.3.1 Effect of TD-10 and TD-11 on catalytic site of MMP2, MMP7 and MMP9	47
2.5.3.2 Effect of TD-10 and TD-11 on VEGF-A-VEGFR-1 interaction	51
2.5.3.3 Experimental evidences for anti-metastatic activity of TD-10 and TD-11	52
2.5.4 Conclusion	53
Chapter 3 Exploring the action mechanism of natural compounds as ATP competitive inhibitors of protein kinases	55
3.1 Background	55
3.2 IVS of Wi-A with protein kinases.....	56
3.3 Exploring the effect of Wi-A and Wi-N combination for dual inhibition of ABL1 protein.....	58
3.3.1 Background	58
3.3.2 Methods	60
3.3.3 Results and Discussion.....	60
3.3.3.1 Dual targeting of Abl with imatinib and asciminib	61
3.3.3.2 Dual targeting of Abl with Wi-A as catalytic inhibitor and Wi-N as allosteric inhibitor	63
3.3.3.3 Dual targeting of Abl with Wi-A as both catalytic and allosteric inhibitor	65
3.3.3.4 Dual targeting of Abl with imatinib as catalytic inhibitor and Wi-A as allosteric inhibitor	65
3.3.4 Conclusion	66
3.4 Study the effect of natural compounds for different insertion mutation at exon 20 of EGFR protein	67
3.4.1 Background	67

3.4.2 Computational methods	68
3.4.3 Results and Discussion.....	69
3.4.3.1 Effect of erlotinib on EGFR exon 20 insertion mutations	70
3.4.3.2 Effect of poziotinib on EGFR exon 20 insertion mutations.....	75
3.4.3.3 Effect of TAS-6417 on EGFR exon 20 insertion mutations	77
3.4.3.4 Effect of CAPE on EGFR exon 20 insertion mutations	80
3.4.3.5 Effect of Wi-A on EGFR exon 20 insertion mutations.....	82
3.4.3.6 Effect of Wi-N on EGFR exon 20 insertion mutations.....	85
3.4.4 Conclusion	87
3.5 Study the effect of natural compounds on BRAF wild type and BRAF V600E mutant protein activity	88
3.5.1 Background	88
3.5.2 Computational Methods	89
3.5.3 Results and Discussion.....	90
3.5.3.1 Structural insight into the effect of VEM and BGB283 on BRAF wild type and BRAF V600E mutant on ATP binding site	91
3.5.3.2 Structural insight into the effect of natural compounds on ATP binding site of BRAF wild type protein	94
3.5.3.3 Structural insight into the effect of natural compounds on ATP binding site of BRAF V600E mutant protein	96
3.5.4 Conclusion	97
Chapter 4 Transcriptome analysis for identification of dysregulated pathways in different subtypes of breast cancer	99
4.1 Background	99
4.2 Computational Methods.....	100
4.2.1 Data retrieval.....	100
4.2.2 Preprocessing of RNASeq reads	102
4.2.2.1 Conversion of sra files downloaded from DNAnexus to fastq format files	102
4.2.2.2 Quality check using FastQC	103
4.2.2.3 Trimming of low quality reads	103
4.2.2.4 Trimming of reads to even length.....	103
4.2.2.5 Estimation of insert size	103

4.2.3 Mapping of RNASeq reads on reference genome	104
4.2.4 Transcriptome reconstruction	104
4.2.5 Reference annotation based transcript (RABT) assembly	105
4.2.6 Transcript sequence extraction using gffread	105
4.2.7 Differential expression analysis	105
4.2.8 Pathway analysis	106
4.3 Results	107
4.3.1 Dataset retrieval and preprocessing of initial RNASeq reads.....	107
4.3.2 Reference based mapping of reads using TopHat.....	109
4.3.3 Differential expression analysis	110
4.3.4 Pathway analysis	110
4.4 Conclusion	116
Chapter 5 Survival-prognosis prediction model for cancer patients	117
5.1 Background	117
5.2 Multi-omics data integration for identification of molecular signature of aberration responsible for prediction of survival of lung adenocarcinoma patients	117
5.2.1 Introduction.....	117
5.2.2 Computational Methods.....	118
5.2.2.1 Dataset preprocessing	118
5.2.2.2 Feature selection.....	119
5.2.2.3 Multi-omics integration based binary classification models for survival prediction	119
5.2.3 Results and Discussion.....	120
5.2.3.1 Selection of the optimal number and combination of features for improving survival prediction	120
5.2.3.2 Neural network-based pattern recognition classifier	121
5.2.3.3 External validation of multi-omics and single-omics based neural network models.....	123
5.2.4 Conclusion	124
5.3 Identification of CNV aberration signature responsible for prediction of survival and prognosis of multiple myeloma patients of indigenous population	125

5.3.1 Introduction	125
5.3.2 Methods	126
5.3.2.1 Autoencoder based feature selection for DNN based survival prediction of patients.....	126
5.3.2.2 NCA based feature selection for neural-network based pattern recognition model for prediction of survival and prognosis of patients	127
5.3.3 Results	128
5.3.3.1 Stacked Autoencoder based DNN for classification of patients based on their survival and prognosis chances	128
5.3.3.2 Neural network-based pattern recognition model trained on features shortlisted by NCA for prediction of survival and prognosis of patients	130
5.3.3.3 Biological significance of identified gene signature of aberration as diagnostic and prognostic biomarker	132
5.3.4 Conclusion	132
5.4 Multi-omics data integration for identification of molecular signature of aberration responsible for prediction of survival of patients and drug response estimation in breast cancer	133
5.4.1 Introduction	133
5.4.2 Computational Methods	134
5.4.2.1 Datasets	134
5.4.2.2 Preprocessing of TCGA breast cancer patients data.....	135
5.4.2.3 Preprocessing of breast cancer cell line data obtained from GDSC	135
5.4.2.4 Constructing representative gene sets.....	135
5.4.2.5 Multi-omics feature integration	136
5.4.2.6 Survival prediction	136
5.4.2.7 Drug response prediction.....	136
5.4.3 Results and Discussion	137
5.4.3.1 Multi-omics integration improves survival prediction in BRCA patients	138
5.4.3.2 Multi-omics signature predicts drug response in BRCA cell lines	140
5.4.3.3 Biological significance of identified gene signature of aberration as diagnostic and prognostic biomarker	142
5.4.4 Conclusion	144
Chapter 6 Conclusion of the thesis	145

References	163
Appendices.....	175
List of publications from the thesis	183
Resume of author	185

LIST OF FIGURES

Figure 1.1-1 Structure of Withanolides, Wi-A and Wi-N.	3
Figure 1.1-2 Structure of TEG Derivatives, TD-10 and TD-11	4
Figure 1.1-3: Structure of CAPE	5
Figure 1.1-4: Diagrammatic representation of IVS pipeline	7
Figure 1.1-5: Methodology for transcriptome analysis of human cancer cell lines.....	8
Figure 2.2-1 Binding of Wi-A with IKK β -NEMO complex and its effect on structural stability.....	24
Figure 2.2-2 Effect of Wi-A on IKK complex.....	25
Figure 2.2-3 Binding of Wi-A with NF κ B dimer (p50-p65) complex and its effect on structural stability.....	28
Figure 2.2-4 Immunofluorescence microscopy finding for effect of Wi-A on IKK complex activity.....	30
Figure 2.3-1 RMSD Fluctuations of vimentin protein and vimentin-ligand complexes with time.	34
Figure 2.3-2 Interactions formed by Wi-A and 3 β mWi-A with vimentin monomer, dimer and tetramer.	36
Figure 2.3-3 MM-GBSA binding energy of vimentin monomer, dimer and tetramer with Wi-A (grey) and 3 β mWi-A (Orange).	37
Figure 2.3-4 . Experimental evidences for anti-metastatic effect of Wi-A and 3 β mWi-A.....	38
Figure 2.4-1 Effect of Wi-A on RAD51 nucleofilament assembly.	42
Figure 2.5-1 Interaction of TD-10 and TD-11 with different MMPs.	48

Figure 2.5-2 Interactions formed by TD-10 (A) and TD-11 (B) at catalytic site of MMP7 protein.	50
Figure 2.5-3 Interaction of TD-10 and TD-11 with VEGFA protein dimer.....	51
Figure 2.5-4 Interaction of TD-10 and TD-11 with VEGFA-VEGFR-1 protein complex dimer.	52
Figure 2.5-5 Line diagram showing the possible mechanism of mortalin-dependent anti-metastasis activity of the TEG derivatives TD-10 and TD-11.	53
Figure 3.3-1 Abl-Imatinib _{2(cat)} -Asciminib _(allos) Complex comparison with Abl-ATP-Myr complex.....	62
Figure 3.3-2 Effect of natural compounds on catalytic and allosteric site of Abl protein dimer.	63
Figure 3.4-1 Superimposition of active and inactive conformation of EGFR-ATP complex.....	69
Figure 3.4-2 Interaction of erlotinib at ATP binding site of EGFR exon 20 insertion mutants-erlotinib and EGFR exon 20 insertion mutants-erlotinib-ATP complexes.....	71
Figure 3.4-3 Protein RMSD plots of different EGFR exon 20 insertion mutant-ligand complexes.	74
Figure 3.4-4 Interaction of poziotinib at ATP binding site of EGFR exon 20 insertion mutants-poziotinib and EGFR exon 20 insertion mutants-poziotinib-ATP complexes.	76
Figure 3.4-5 Interaction of TAS-6417 at ATP binding site of EGFR exon 20 insertion mutants-TAS-6417 and EGFR exon 20 insertion mutants-TAS-6417-ATP complexes.	78
Figure 3.4-6 Interaction of CAPE at ATP binding site of EGFR exon 20 insertion mutants-CAPE and EGFR exon 20 insertion mutants-CAPE-ATP complexes.	81
Figure 3.4-7 Interaction of Wi-A at ATP binding site of EGFR exon 20 insertion mutants-Wi-A and EGFR exon 20 insertion mutants-Wi-A-ATP complexes.....	84

Figure 3.4-8 Interaction of Wi-N at ATP binding site of EGFR exon 20 insertion mutants-Wi-N and EGFR exon 20 insertion mutants-Wi-N-ATP complexes.....	86
Figure 3.5-1 Interaction of VEM and BGB283 at ATP binding site of BRAF wild type and V600E mutant.	92
Figure 3.5-2 Interaction of CAPE, Wi-A and Wi-N at ATP binding site of BRAF wild type protein.	95
Figure 3.5-3 Interaction of CAPE, Wi-A and Wi-N at ATP binding site of BRAF V600E mutant protein.	97
Figure 4.2-1 Methodology adopted for transcriptome analysis of breast cancer cell lines and normal breast cell lines.....	102
Figure 4.3-1 Comparison of DEGs obtained for basal and non-malignant subtype and mapping of genes dysregulated in basal subtype by DAVID functional annotation tool.....	113
Figure 4.3-2 Cell cycle pathway map for basal subtype gene set.....	114
Figure 4.3-3 ECM receptor interaction pathway map for basal subtype geneset.	115
Figure 5.2-1 Diagrammatic view of different model generated for survival prediction of LUAD patients.	120
Figure 5.2-2 Performance of multi-omics integration based neural network model of LUAD patient’s survival prediction.....	122
Figure 5.3-1 Flowchart representing architecture of DNN1, DNN2 and neural network-based classifier trained on CNV data.....	129
Figure 5.3-2 Performance of of DNN1, DNN2 and neural network-based classifier trained on CNV data of MM patients.	130
Figure 5.3-3 Confusion matrix of neural network-based pattern recognition model trained on features selected using NCA.....	131
Figure 5.4-1 Performance of multi-omics integration based neural network survival prediction model of BRCA patients.....	138

Figure 5.4-2 Violin plot showing the distribution of IC50 values of sensitive (blue) and resistant (orange) cell lines for first 21 drugs.141

Figure 5.4-3 Performance of drug response model for BRCA cell lines..... 141

LIST OF TABLES

Table 2.2-1 Properties of Wi-A in complex with IKK β /NEMO association domain and NF κ B dimer.	26
Table 2.3-1 Details of interactions formed by withanolides at vimentin protein binding site along with calculated binding energy of protein-ligand complexes and their docking scores.....	35
Table 2.4-1 List of interactions formed by ATP and Wi-A at RAD51 protomer-protomer interface and their binding energy at ATPase domain.....	43
Table 2.5-1 Interactions formed by TD-10 and TD-11 with MMP class of proteins under study.....	49
Table 3.2-1 Top scoring hits for targets of Wi-A obtained from IVS.	56
Table 3.3-1 Binding energies obtained by separate docking of inhibitors at each sites of Abl protein.	61
Table 3.3-2 Interactions formed by inhibitors at catalytic and allosteric site of Abl-Imatinib _{2(cat)} -Asciminib _(allos) complex (treated as control), Abl-Wi-A _{2(cat)} -Wi-N _(allos) complex and Abl-Wi-A _{2(cat)} -Wi-A _(allos) complex..	64
Table 3.4-1 Influence of compounds on different structural properties of EGFR exon 20 insertion mutants as ATP competitive inhibitors.....	72
Table 3.4-2 Interactions formed by erlotinib and ATP with chain A of different EGFR exon 20 insertion mutant-erlotinib-ATP complexes selected for this study.....	73
Table 3.4-3 Interactions formed by poziotinib and ATP with chain A of different EGFR exon 20 insertion mutant-poziotinib-ATP complexes selected for this study.	75
Table 3.4-4 Interactions formed by TAS-6417 and ATP with chain A of different EGFR exon 20 insertion mutant-TAS-6417-ATP complexes selected for this study.	79
Table 3.4-5 Interactions formed by CAPE and ATP with chain A of different EGFR exon 20 insertion mutant-CAPE-ATP complexes selected for this study.	82

Table 3.4-6 Interactions formed by Wi-A and ATP with chain A of different EGFR exon 20 insertion mutant-Wi-A-ATP complexes selected for this study.	83
Table 3.4-7 Interactions formed by Wi-N and ATP with chain A of different EGFR exon 20 insertion mutant-Wi-N-ATP complexes selected for this study.	85
Table 3.5-1 Structural properties of BRAF wild type-Inhibitor dimer complexes and BRAF V600E mutant-Inhibitor dimer complexes.	91
Table 3.5-2 Interactions formed by inhibitors with both chain B and C of wild type BRAF protein dimer.	93
Table 3.5-3 Interactions formed by inhibitors with both chain B and C of BRAF V600E mutant protein dimer.	94
Table 4.2-1 Details of SRA runs selected for the study.	100
Table 4.3-1 Various parameter of reads belong to each run, such as length of reads before and after trimming, calculated insert size and standard deviations and mate-pair-dist.	107
Table 4.3-2 Mapping percentage of reads obtained for each run, categorized on the basis of cell type.	109
Table 4.3-3 Number of differentially expressed transcripts found in pairwise testing of five conditions of breast cancer.	110
Table 4.3-4 Pathways mapped by DAVID involving upregulated genes in different breast cancer subtypes.	111
Table 4.3-5 Pathways mapped by DAVID involving downregulated genes in different breast cancer subtypes.	112
Table 4.3-6 Pathways reported by global pathway analysis of DEGs of different breast cancer subtypes.	113
Table 5.2-1 Description of the total number of features and number of features selected using a feature selection algorithm for each dataset, that were used for survival prediction.	121

Table 5.2-2 Prediction accuracy of multi-omics based integration models for survival prediction of LUAD cancer patients.	123
Table 5.2-3 External validation prediction accuracy of our multi-omics-based integration neural net classification model trained on group I and group II data when provided with single-omics data and clinical features as input	123
Table 5.2-4 Prediction accuracy of single-omics based neural network classifier trained on features selected for group I and external validation performance of each model when provided with single-omics data and clinical features as input.	124
Table 5.3-1 Performance of neural network-based classifier	132
Table 5.3-2 Biological significance of identified gene CNV signature of aberration	132
Table 5.4-1 Description of the total number of features for each dataset that were used for multi-omics data integration type for building survival and drug response models.	137
Table 5.4-2 Performance of neural network-based classifier for survival prediction of BRCA patients	139
Table 5.4-3 External validation prediction accuracy of our multi-omics-based integration neural net classification model for survival prediction of BRCA patients when provided with single-omics and five-omics data along with clinical features as input	139
Table 5.4-4 Biological significance of genes that aid in prediction of both survival and drug response of BRCA patients.....	142
Table 5.4-5 KEGG pathways mapped by identified signature of aberration for survival and drug response prediction model.	143

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	I κ B kinase
IVS	Inverse virtual screening
LUAD	Lung Adenocarcinoma
MD	Molecular dynamics
MM	Multiple Myeloma
MMPs	Matrix metalloproteinases
NF κ B	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
