Probing of biomarkers predictive of pan-cancer drug sensitivity and resistance and drug repurposing

Thesis submitted to University of Hyderabad, India in partial fulfilment for the award of degree of Doctor of Philosophy in Biochemistry

By

Santosh Kumar
16LBPH04

Department of Biochemistry

School of Life Sciences

University of Hyderabad

Hyderabad-500046

Telangana, India

January, 2023

University of Hyderabad
(A Central University established in 1974 by Act of Parliament)
Hyderabad -500046, India

Certificate

(For Ph.D. Dissertation)

This is to certify that this thesis entitled **"Probing of biomarkers predictive of pan-cancer drug sensitivity and resistance and drug repurposing"** submitted by **Mr. Santosh Kumar** bearing registration number **16LBPH04** in partial fulfilment of the requirement for the award of Doctor of Philosophy in the Department of Biochemistry, School of Life Sciences, is a bonafide work carried out by him under my supervision and guidance.

This thesis is free from plagiarism and has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma.

Part of this thesis has been presented in the following conferences:

- Santosh Kumar, Seema Mishra . **"Probing of predictive markers of drug sensitivity & resistance pan-cancer"** Poster presented in the symposium on Accelerating Biology Digitizing Life, organized by the Centre for Development of Advanced Computing (CDAC). Pune, India, during 09-11 January 2018.
- Santosh Kumar, Seema Mishra . **"Identification of biomarkers predictive of pan-cancer drug sensitivity and resistance and their role"** Poster presented in the 17th International Conference on Bioinformatics, organized by the Jawaharlal Nehru University (JNU), New Delhi, India, during 26-28 September 2018.
- Santosh Kumar, Seema Mishra . **"Probing of biomarkers predictive of pan-cancer drug sensitivity & resistance"** Poster presented in the national seminar on Biomolecular interactions in development and diseases, organized by the Department of Biochemistry, UoH, Hyderabad, India, during 26-28 September 2019.

Published in the following journal:

- **Santosh Kumar,** Seema Mishra. *MALAT1* as master regulator of biomarkers predictive of pan-cancer multi-drug resistance in the context of recalcitrant NRAS signaling pathway identified using systems-oriented approach. *Sci Rep* 12, 7540 (**2022**).

Other papers:

- Seema Mishra, **Santosh Kumar,** Choudhuri, K.S.R. *et al.* Structural exploration with AlphaFold2-generated STAT3α structure reveals selective elements in STAT3α-GRIM-19 interactions involved in negative regulation. *Sci Rep* 11, 23145 **(2021)**.

Further the student has passed the following courses towards fulfilment of coursework requirement for Ph.D.

Course code	Name	Credits	Pass/Fail
BC 801	Analytical Techniques	4	Passed
BC 802	Research ethics, Data analysis and Biostatistics	3	Passed
BC 803	Lab seminar and Records	5	Passed

Supervisor

Dr. SEEMA MISHRA
Assistant Professor
Deptt. of Biochemistry
School of Life Sciences
University of Hyderabad
HYDERABAD-500 046. INDIA

Head
Dept. of Biochemistry

HEAD
Dept. of Biochemistry
SCHOOL OF LIFE SCIENCES
UNIVERSITY OF HYDERABAD
HYDERABAD-500 046.

Dean
School of Life Sciences

6/11/2023

DEAN
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

University of Hyderabad
(A Central University established in 1974 by Act of Parliament)
Hyderabad -500046, India

Declaration

I, **Santosh Kumar**, hereby declare that this thesis entitled "**Probing of biomarkers predictive of pan-cancer drug sensitivity and resistance and drug repurposing**" has been carried out by me under the guidance and supervision of **Dr. Seema Mishra,** is an original and independent research work. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma.

Santosh Kumar

Santosh Kumar
(16LBPH04)

Date: 05/01/2023

Seema

Supervisor

Acknowledgments

I express my deepest sense of gratitude to my supervisor, Dr. Seema Mishra, for giving me an opportunity to work under her able guidance, providing me the lab facilities and constant help and support throughout my PhD work. It gives me immense pleasure to admit that Dr. Seema Mishra, my mentor with integrity and perseverance, has been a constant source of inspiration throughout my doctoral discourse.

I would like to thank my doctoral committee members, Prof. B. Senthilkumaran and Dr. N. Prakash Prabhu, for their valuable suggestions, guidance and monitoring my work. My special thanks to Prof. Gutti Ravi Kumar for his kind of constant support and encouragement throughout my PhD.

I would also thank all the faculty members of Department of Biochemistry and School of Life Sciences for their support during the course of my Ph.D. work.

I would like to thank the former Head of Department of Biochemistry Prof. N. Siva Kumar and Prof. Mrinal Kanti Bhattacharya and the Present head, Prof. Krishnaveni Mishra.

I would like to thank former and present Dean, School of Life Sciences Prof. P. Reddanna and Prof. Kolluru V A Ramaiah, Prof. S. Dayananda and Prof. N. Siva Kumar for allowing me to use the facilities of the school.

I wish to thank all my former and present labmates A. Saleembhasha, Kesaban Sankar Roy Choudhury, Akash Ghosh and Praveena, Euphinia, Ayang, Aanchal and Raghunandan respectively, for their timely help and support. I wish them all success and happiness in their upcoming journeys.

I thank all my lab M.Sc. students Sagara Gurusinghe, Abdul Jasir, Aditya, Soumi, Udita, Sumathi, Sathish, Shubhashree for creating a cheerful work atmosphere. I wish them all success and happiness in their lives

I thank all my teachers since my childhood, whose guidance and encouragement at each step was instrumental in shaping my career.

I acknowledge University of Hyderabad for providing BBL (non-NET) fellowship, and CSIR, New Delhi for providing JRF and SRF fellowship in the form of financial assistance during my PhD.

I would like to take this opportunity to thank all my friends, Deepak Kashyap, Swati Singh, Sheetal Uikey, Swati Dahariya, Chhaya Rani Kispotta, Ravikash Maurya, and Ravish Gupta. I wish them all success and happiness in their lives. My special thanks to Deepak Kashyap for helping me in my Pre-PhD presentation and thesis writing.

I also thank my PhD batchmate friends Neera Yadav, Ranay Yadav, Arpita Prusty and Deepak Saini. I wish them all success and happiness in their lives.

I would like to thank my seniors Pankaj, Ashish and Anita for their kind of help and support.

I would like to thank the non-teaching staff of the Department of Biochemistry Chary and Prabhu for helping in all the documentation work during PhD and I also thank non-teaching staff of Dean Office, School of Life Sciences.

I fall short of words to thank my family members, especially my sisters, Swati Singh, Archana Singh, Sheetal Singh, and Vandana Singh and brother in laws, Virendra Patel and Manoj Patel for their blessings, inseparable support, prayers and love.

I thank my cousins Alok Singh, Arvind Singh, Indrawati, and Soni for being there with me all through the course of my Ph.D and for their love, support and encouragement.

My floral wishes and lots of love to all my nephews and niece, Shiwangi, Shiv Balak, and Neha….

Finally, I express my gratitude and thank my father Sri. Ram Jatan Singh, my mother Smt. Lakshmina Devi and almighty God for their celestial blessings and providing me the moral strength to lead the life against all the odds. I owe all my success to them.

Santosh Kumar

Dedicated to my parents

Table of content

Abbreviation

WHO	:	World Health Organization
DNA	:	Deoxyribonucleic acid
UV	:	Ultraviolet
BCR	:	Breakpoint cluster region protein
TP53	:	Tumor protein
RB1	:	Retinoblastoma protein
CML	:	Chronic Myeloid Leukaemia
ABC	:	ATP- binding cassette transporter family
FEN1	:	Flap endonuclease
FANCG	:	FA Complementation Group G
RAD23B	:	RAD23 Homolog B
BCL-2	:	B-cell lymphoma 2
Tβ4	:	Thymosin β4
AML	:	Acute myeloid leukemia
RAF	:	Rapidly accelerated fibrosarcoma
HRAS	:	Harvey rat sarcoma viral oncogene homolog
KRAS	:	Kirsten rat sarcoma viral oncogene homolog
NRAS	:	Neuroblastoma rat sarcoma viral oncogene homolog
GTP	:	Guanosine triphosphate
GDP	:	Guanosine diphosphate
PI3K	:	Phosphoinositide 3-kinases
ERK	:	Extracellular signal regulated kinase
MAPK	:	Mitogen-activated protein kinase
LncRNA	:	Long Non-coding RNA
GDSC	:	Genomics of Drug Sensitivity in Cancer
TCGA	:	The Cancer Genome Atlas
COSMIC	:	Catalogue of Somatic Mutations in Cancer
ANOVA	:	Analysis of variance

CCLE	:	Cancer cell line encyclopaedia
GA	:	Genetic algorithm
KNN	:	K-nearest neighbours
KEGG	:	Kyoto encyclopedia of genes and genomes
GO	:	Gene ontology
TF	:	Transcription factor
TG	:	Target gene
GEO	:	Gene expression omnibus
MINT	:	Molecular Interaction database
STRING	:	Search tool for the retrieval of interacting genes/proteins
ORTI	:	Open-access repository transcription factor interactions
HTRI	:	Human transcriptional regulation interaction
TRED	:	Transcriptional regulatory element database
TRRD	:	Transcriptional regulatory region database
RCSB	:	Research collaboratory for structural bioinformatics
PDB	:	Protein data bank
NMR	:	Nuclear magnetic resonance
wwPDB	:	Worldwide Protein data bank
MeV	:	MultiExperiment Viewer
CMS	:	Comparative marker selection
NGS	:	Next-generation sequencing
ChIP	:	Chromatin immunoprecipitation
IC50	:	Inhibitory concentration
ALL	:	Acute lymphocytic leukemia
BLCA	:	Bladder urothelial carcinoma
DBLC	:	Lymphoid neoplasm diffuse large B-cell lymphoma
LIHC	:	Liver hepatocellular carcinoma
LUAD	:	Lung adenocarcinoma
LUSC	:	Lung squamous cell carcinoma

MB	:	Medulloblastoma
MM	:	Multiple myeloma
NB	:	Neuroblastoma
SCLC	:	Small-cell lung cancer
THCA	:	Thyroid carcinoma
SKCM	:	Skin cutaneous melanoma
DEGs	:	Differentially express genes
GDSC	:	Genomics of Drug Sensitivity in Cancer
CNS	:	Central nervous System
SKCM	:	Skin Cutaneous Melanoma
BLCA	:	Urothelial Bladder Carcinoma
LUSC	:	Lung Squamous Cell Carcinoma
THCA	:	Thyroid Cancer
LUAD	:	Lung Adenocarcinoma
LIHC	:	Liver Hepatocellular Carcinoma
GO	:	Gene Ontology
KEEG	:	Kyoto Encyclopedia of Genes and Genomes
TYR	:	Tyrosinase
PMEL	:	Premelanosome Protein
MLANA	:	Melanoma Antigen Recognized by T cells
EDNRB	:	Endothelin Receptor Type B
FN1	:	Fibronectin 1
TIMP1	:	Tissue Inhibitor of Metalloproteinases 1
CD44	:	Cluster of Differentiation 44
SPARC	:	Secreted Protein Acidic and Cysteine Rich
SNAI2	:	Snail Family Transcriptional Repressor 2
MMP1	:	Matrix Metallopeptidase 1
MMP14	:	Matrix Metallopeptidase 14
TIMP3	:	Tissue Inhibitor of Metalloproteinases 3

VEGFC	:	Vascular Endothelial Growth Factor C
GCN	:	Gene co-expression network
PPI	:	Protein-protein network
PVT1	:	Plasmacytoma Variant Translocation 1
SNHG11	:	Small Nucleolar RNA Host Gene 11
MIR22HG	:	MIR22 Host Gene
TP73-AS1	:	TP73- Antisense RNA1
ALDH1A1	:	Aldehyde Dehydrogenase Family Member A1
HOTAIR1	:	HOX Transcript Antisense RNA1
ER	:	Estrogen receptor
MALAT1	:	Metastasis Associated Lung Adenocarcinoma Transcript 1
EGR1	:	Early Growth Response 1
AR	:	Androgen Receptor
YBX1	:	Y-Box Binding Protein 1
UTR	:	Untranslated Region
CDS	:	Coding Sequences
HA	:	Hyaluronic acid
HABD	:	Hyaluronic acid binding domain
NSCLC	:	Non-small cell lung cancer

List of figures and tables

Chapter-1

Chapter-3

Chapter-4

Chapter-5

Chapter-6

Chapter-7

Chapter-8

List of tables

Chapter-3

Chapter-4

Chapter-5

Chapter-6

Chapter-7

Chapter 1

Introduction and review of literature

1.1 Cancer

Cancer is one of the most dreaded disease caused when normal cells transform into tumour cells leading to abnormal cell growth or division through a multi-stage process. It especially arises from a pre-cancerous lesion to form a malignant tumour ***(Roy & Saikia 2016; WHO*)**. Cancer cells continue unregulated proliferation instead of responding appropriately to the signals which control normal cell behaviour, and they simultaneously invade surrounding normal tissues and consequently migrating to other body parts (metastasize) through the blood vascular system ***(Cooper, 2000)***. Cancer can arise as a result of abnormal proliferation of cells from different parts of the body, and based on the originating cell types, there are more than a hundred distinct types of cancer, which generally differ in their behaviour and responses to applied treatments. Tumors are classified into benign and malignant tumors based on their characteristics. Benign tumors remain confined in their primary location without invading the surrounding normal tissues of the body and also do not spread to distant body part. Benign tumors are known to grow slowly and have distinct borders. However, malignant tumors have characteristics of both invading the surrounding normal tissues and spreading throughout the body (metastasis) through the circulatory or lymphatic systems. Malignant tumors are classified as cancer because they are significantly more harmful than benign tumors due to their capacity to infiltrate and metastasis. While benign tumors can be removed surgically, it is tough to treat malignant tumors due to their frequent relapse and spreading to distant body sites (***Cooper, 2000; Kumar et al., 2015; Patel, 2020).***

1.2 How cancer arises?

Cancer development and progression is a complex process which arises due to the aggregation of numerous genetic alterations, suggesting that cancer is a genetic disease which involves a host of functional and genetic abnormalities ***(He et al. 2007; Semi & Yamada 2015; Vogelstein & Kinzler 2004).*** These abnormalities can include various genetic and epigenetic modifications

which induces chromosomal instability leading to the initiation and promotion of cancer development (***Semi & Yamada 2015***). For instance, overall changes in degrees of DNA methylation (***Feinberg & Vogelstein 1983***), and also site-specific DNA hyper- methylation at promoters of certain genes are one of the most frequently analysed epigenetic alterations associated with increased cancer frequency (***Feinberg & Tycko 2004; Ushijima 2005).*** Alterations in the pattern of histone modification, which includes acetylation, methylation and phosphorylation plays a significant role in tumorigenesis ***(Nowacka-Zawisza & Wiśnik 2017)***, as well as in the development of genomic mutations (***He et al. 2007***) and other insults that can be conductive to the expression or suppression of target genes in tumors. Genomic and epigenetic abnormalities increases oncogenic signals that alters the regulation of downstream target genes transcriptionally, thereby resulting in changes in the transcriptional regulatory networks. The transcriptional changes caused by oncogenic signals could be a secondary effect of the genetic and epigenetic alterations (***Semi & Yamada 2015***).

1.2.1 Risk factors associated with genetic and epigenetic changes in cancer

A cancer risk factor can be anything that increases the feasibility of growth of cancer in human body. Cancer risk factors may incorporate exposure to chemical carcinogens, or other substances and life style etc. Risk factors associated with cancer also include things like age and family history which people cannot control. Family history of some cancers can be an indication of a possible inherited form of cancer (**13**). These risk factors mainly grouped into two mutually exclusive modules: intrinsic and non-intrinsic risk factors (***Wu et al. 2018).***

i. **Intrinsic cancer risk factors:** It is an inevitable natural mutations that arises because of random errors during DNA replication and confers specific attribute to the human being. Intrinsic risk of cancer occurs in all dividing cells due to the basal mutation rate. Intrinsic risk factors are unmodifiable and unavoidable.

ii. **Non-intrinsic cancer risk factors:** Owing to their versatile mechanism, non-intrinsic factors include two groups, as exogenous and endogenous risk factors.

a) **Exogenous factors** are chemical carcinogens (mutagen), xenobiotic, viruses and lifestyle associated factors (e.g. smoking, nutrient intake, hormone therapy and physical activity) which are exogenous (extrinsic) to the host. Tobacco smoke for lung cancer, UV radiation for skin cancer, and viruses for cervical and liver cancer have been identified as exogenous cancer risk factors (***Wu et al., 2018***). These factors are modifiable non-intrinsic factors.

b) **Endogenous factors** are known as partially modifiable factors and associated with the features of an individual (e.g., immune and DNA damage response, hormone levels) and impact on the control of cell growth and genomic integrity.

Exogenous (environment) and endogenous (hereditary/genetic) risk factor direct to complex endogenous activity such as ageing, inflammation and obesity, these processes also influence the steroid hormones level in an individual, which could have a role in breast cancer.

1.3 Worldwide cancer statistics

Cancer is a serious worldwide public health issue and one of the leading cause of morbidity and mortality in the world. With about 10 million deaths reported in 2020 (**Fig.1B**)**,** one among the six deaths was from cancer. The most frequent cancer incidence reported globally are; breast, lung, colon, rectum and prostate cancers ***(WHO; Ferlay et al. 2020; Sung et al. 2021***). Almost 1.3 million new cases and 8.5 lakh deaths were reported in India in the year 2020 (**Fig. 1E**)**.**

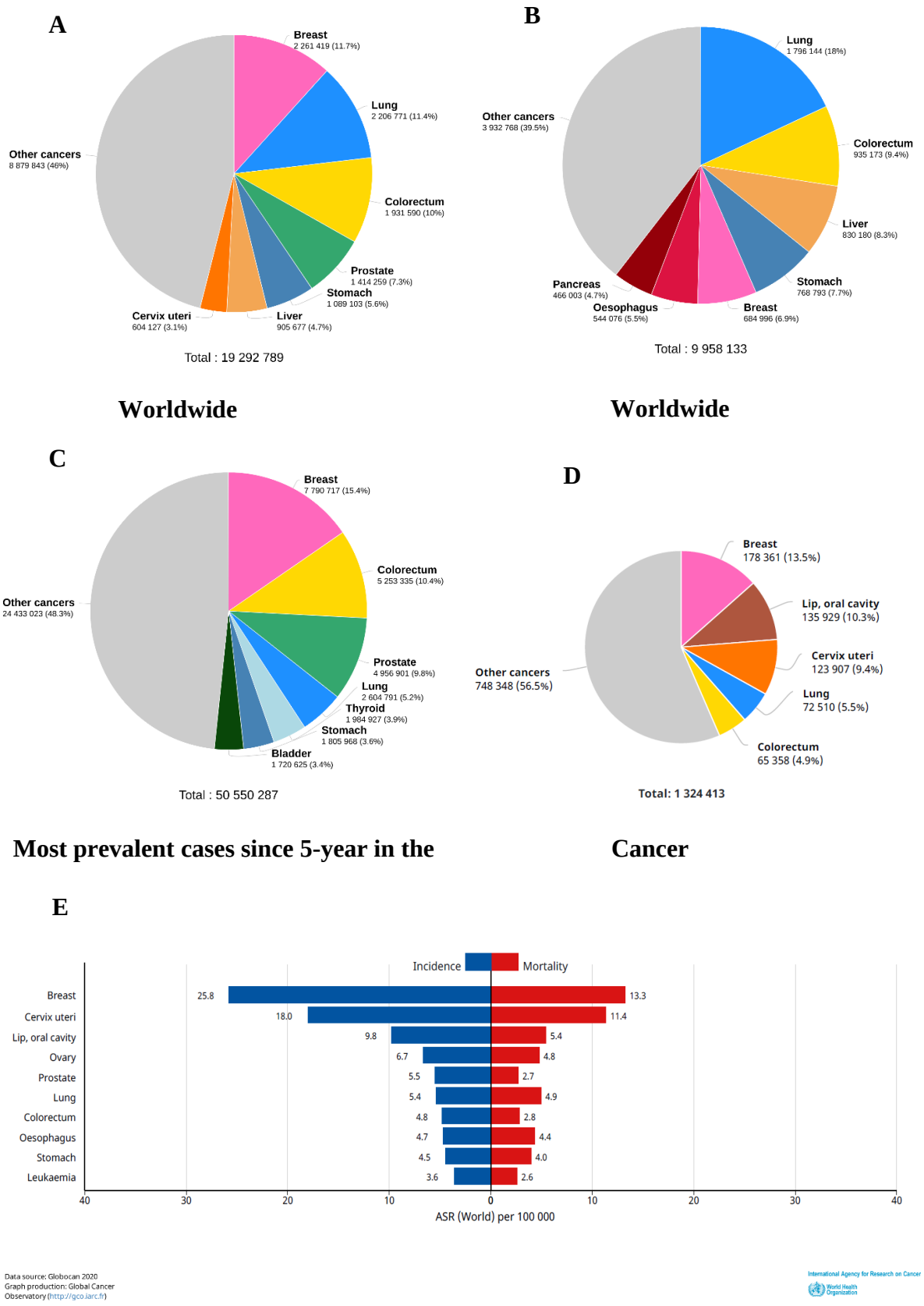


Figure 1: Global cancer statistics. A) Worldwide cancer incidence, B) Worldwide death due to cancer, C) Most prevalent cancer worldwide, D) Cancer incidence in India, E) Top 10 cancer incidence and mortality in India.

1.4 History and origin of cancer

A Greek physician Hippocrates (also known as "Father of Medicine" 460-370 BC) for the first time called cancer as a disease of uncontrolled cell division. He coined the terms 'carcinoma' (meaning crab in Greek) and 'carcinos' and to narrate cancer-forming and non-cancer forming tumors, respectively. The description of disease was named after the crab whose finger-like projections resembled the spreading from a cancer called to mind imitated the shape of a crab. A Roman physician, Celsus (28-50 BC), translated the crab (Greek term) into cancer. Later Galen (130-200 AD), another Greek physician stated tumors by using the word *oncos* (Greek for swelling). Today, the terminology coined by Galens (oncos) is widely used to designate cancer specialists (oncologists) while the malignant tumors are still designated by the crab (carcinos) analogy proposed by Hippocrates and Celsus **(18).**Though cancer has been a cause of great pain for humanity since time immemorial, its incidences have substantially increased in recent times because of factors like; an ageing population, rise in number of exogenous carcinogens, risky health behaviours, etc (***Faguet, 2015***).

1.5 Cancer treatment strategies

Cancer is one of the most dominating disease in the world and in the past, many therapies have come into existence for cancer treatment. Currently, there are various methods being used for cancer therapy. The kinds of treatment that a patient receive depend on what type of cancer they contains and its stage. The most common cancer therapies are chemotherapy, targeted, surgery, and radiation. Other therapies include hormonal therapy, immunotherapy, laser, etc (**19-21**).

i. **Chemotherapy:** In chemotherapy, drugs may be given orally or intravenously to kill cancerous cells. Two different drugs can be given together at the same time or one after the other at different time points. Chemotherapy is being used to cure cancer by shrinking it to stop or slow down its growth and thereby preventing cancer from spreading. Chemotherapy

is being used to treat a variety of cancers worldwide. For some cancer patients, chemotherapy may be the only option for cancer treatment that they can receive. But most often, patients also receive other treatments along with chemotherapy. The types of therapy that patients need its depends on the type of cancer they have, if the cancer has spread or migrated, and if they have other health issues **(19-21**).

ii. **Targeted therapy:** This is a target-specific therapy for some cancers, where most cancer patients carry a target for a certain drug, so they can be treated with that drug. Targeted therapy is a method of cancer treatment that could use either small-molecule drugs or monoclonal antibodies that targets proteins to stop the process of growth, division, and spread of cancer by triggering cancer cells to undergo cell death on their own or kill cancer cells directly in the body. This therapy has also become the foundation of precision medicine. As in the case of targeted therapy, the specific targets within the cancer cells are attacked, and little to no harm is caused to the normal cell. These targeted protein molecules play a pivotal role in the growth and survival of cancer. Using these targets, the drug molecules paralyzed the spreading of cancer cells (**19-21**).

iii. **Surgery:** It is a commonly used therapy for a variety of cancers. This therapy works best for solid tumors that are contained in one area. Surgeons remove out the bulk of cancerous cells (tumors) and also some of the adjacent normal tissue from the patient's body through the surgical operation. Sometimes, surgery is also done to relieve the consequences or side effects caused by a tumor. Surgery is not applicable for blood cancer such as leukemia or for metastasized cancers (**19-21**).

iv. **Radiation therapy:** It is also known as radiotherapy and is being used for cancer treatment that applies a heavy dose (high frequency) of radiation to kill or slows down the growth of cancerous cells by damaging their DNA and thus shrink tumors. Cancer cells stop growing

or die whose DNA is damaged from where it cannot be repaired, they are broken down and eliminated by the body. Mostly x-rays or radioactive seeds have been used in radiotherapy to destroy the cancer cells. Cancer cells grow and divide quicker in compare to healthy cells in the human body. Radiotherapy destroys or kills cancer cells more than normal healthy cells because radiation is more harmful or susceptible to fast-growing/dividing cells. This type of therapy arrests the growth and division of cancer cells and then directs to cell death instead of killing cancer cells right away. This type of therapy may take days or weeks before causing enough damage in the DNA of cancer cells to die. Radiation therapy is categorized into two major groups: External beam is the most frequent form of radiotherapy that uses X-ray radiation or particles projected at the tumor tissue from the outside of the body to kill cancer cells. Internal beam radiotherapy delivers radiation inside the body via radioactive seeds (pills or liquid) placed within or near the tumor through a vein (intravenous).

v. **Hormonal therapy:** It is a type of treatment used to treat majorly those cancers which are fuelled by hormones and are also called endocrine therapy. Surgery and drugs help in the stoppage or slowing down the cancer growth, are being used to stop the natural endocrine hormones from functioning on the organs which acquires hormones to grow. The surgery involves the removal of hormone-making organs, like ovaries and testes. Endocrine therapy is primarily used to treat cancer, where it decreases the chances of relapse of cancer and blocks or slows down its growth and survival, and eases cancer symptoms that mostly used to bring down or restrict symptoms of men's prostate cancer who are incapable of having surgery or radiotherapy. Endocrine therapy is being used to treat mainly prostate, ovarian, and breast cancers which uses steroid hormones to grow. Along with other cancer treatments, hormone therapy is one of the most frequently used therapy.

vi. **Immunotherapy:** Immunotherapy is another type of treatment to cure cancer that boosts the immune system to fight against cancer. Immunotherapy basically depends on the ability of the human body to fight against infection and other diseases. It uses immune cells, e.g., white blood cells and tissues of the lymph system in the body, to promote a stronger immune system to work in a more powerful manner or attacking way to fight against cancer. Immunotherapy is used to stop or slow down the cancer cell's growth, preventing from metastasizing of cancer cells to distant parts of the human body and remove the cancer cells by boosting the ability of the immune system. Immunotherapy is a form of biological therapy that uses substances (immune cells) made in living organisms to cure cancer. Various type of immunotherapies is being used for cancer treatment which includes; immune checkpoint inhibitor, immune system modulators, T-cell transfer therapy and monoclonal antibody. Immunotherapy can be given orally (pills or capsules), intravenously (directly into a vein), topical (through cream rub on the skin) and intravesical (directly injected into the bladder). Even though the immune system is well designed to stop or reduce the growth of cancer, however, the cancer cells find ways to avoid destruction by the immune system through different genetic modifications, thereby decreasing their visibility, having inconspicuous surface proteins, or by changing the biochemistry of the normal cells around them (**19-21**).

Despite these advances in treatment and also being a promising option to cure cancer, currently, about 90% of chemotherapy failures happen during the invasion and migration of cancers and these failures are majorly due to the occurrence of drug resistance in cancer cells (***Mansoori et al. 2017; Longley & Johnston 2005***).

1.6 Drug resistance in cancer

Drug resistance is a widely-known incident which develops due to the inability of anticancer drugs to cure cancer because of restricted effectiveness (***Holohan et al. 2013***). The concept of drug resistance was initially observed in microbes when bacteria were observed to show resistance to some particular antibiotics, but later, a similar type of mechanisms have been found in several other diseases, including cancer ***(Housman et al. 2014)***. Furthermore, various major cancer treatments, including chemotherapy, surgery, radiotherapy, immunotherapy and a combination of therapy, are being used as promising cancer treatments as selective therapies based on the stronger laws and principles of biology and molecular genetics in the tumor development (***Urruticoechea et al. 2010; Baskar et al. 2012; Damin & Lazzaron 2014; Khalil et al. 2016***). A large number of malignant tumor cells become resistant to the drug in the chemotherapy and later the administration of a certain drug. So, drug resistance in the field of cancer remains a major problem and is also responsible for most relapses and death due to cancer (***Mansoori et al. 2017***). There is a diverse range of possible factors and mechanisms involved in cancer drug resistance, including genetic mutations, epigenetic changes, alteration in drug metabolism, increased rate of drug efflux, and several other cellulars and molecular mechanisms (***Wang et al. 2019; Holohan et al. 2013***).

1.6.1 Type of drug resistance

Resistance to chemotherapeutic treatment is mainly categorized into two broad groups, which are the following. (**Fig. 2**).

a) **Intrinsic drug resistance (IDR):** Drug resistance primarily or naturally present before receiving chemotherapy and which is mediated by pre-existing elements in the bulk of malignant tumor cells that make the therapy ineffective is referred to as intrinsic drug resistance (***Holohan et al. 2013***). Intrinsic resistance, also defined as innate resistance, arises because of naturally present (first-line) mutation, tumors heterogeneity, and activation of various intrinsic pathways against anticancer drugs, and this type of drug

resistance generally exist in cancer before treatment due to mutation in drug target genes having a crucial role in tumor growth or apoptosis (***Wang et al. 2019***). For example, Snail and Slug suppressed p53-mediated apoptosis in ovarian cancer to induce radioresistance and chemoresistance ***(Kurrey et al. 2009***).

b) **Acquired drug resistance (ADR):** ADR may arise during or post chemotherapeutic drug treatment of cancer cells that were sensitive in the beginning, as well as through many other adaptive responses (***Holohan et al. 2013***). It can be because of secondary proto-oncogene activation, altered expression of drug targets, or mutations in target protein and tumor microenvironment changes in the latter part of treatment (***Wang et al. 2019***). Acquired resistance appears in cancer when an advance mutation in drug targets alters their molecular structures; as an example, gatekeeper mutation in the oncogenic kinase domain of BCR-ABL1 (T315) developed imatinib (STI-571) resistance in chronic myeloid leukaemia (CML) patients (***Gorre et al. 2001***), combined loss of function of TP53 and RB1 induces enzalutamide resistance and increase cellular plasticity in prostate cancer (***Mu et al. 2017***). Although, there are various other mechanisms which can promote drug resistance in human cancer, and it could be intrinsic or acquired resistance.

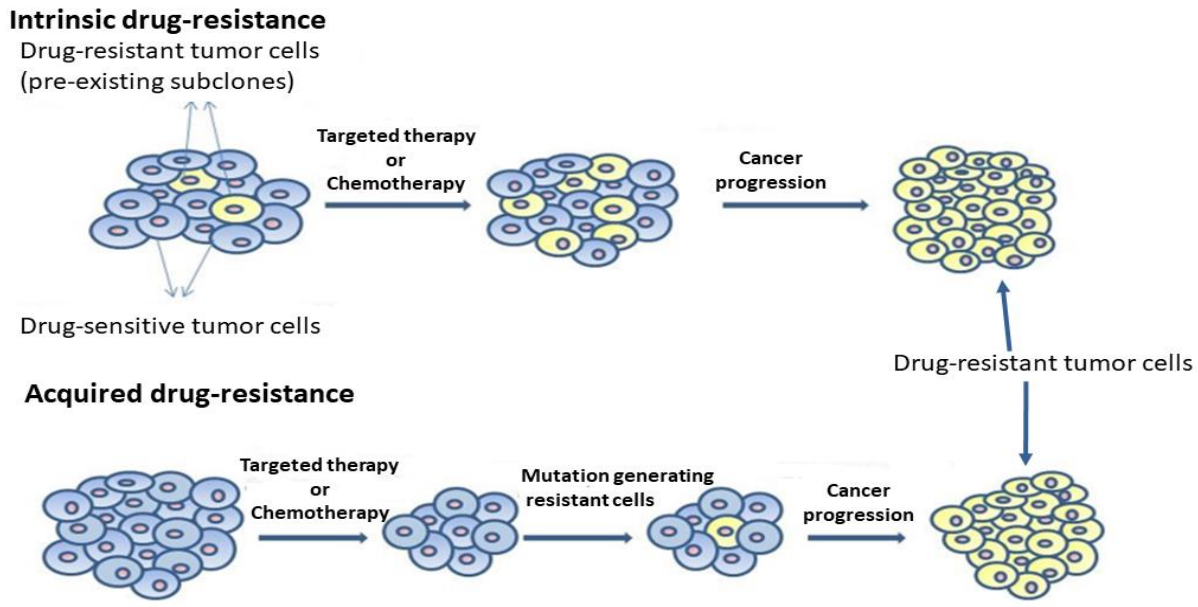


Figure 2: Drug resistance mechanism in cancer. Intrinsic drug resistance and acquired drug resistance. (*Dai et al., 2020*)

1.6.2 Mechanisms induce drug- resistance in cancer

i. **Drug Efflux:** It is one of the most widely known processes to induce drug resistance in cancer by enhancing drug efflux, which involves the reduction in drug accumulation. Transmembrane transporter proteins from the ATP-binding cassette (ABC) transporter family are known to license drug efflux and are crucial regulators at the plasma membranes of non-tumor cells. Apart from human cells, ABC transporter family proteins are also present in other phyla, where they play a key role in the transport of a variety of substances over cell membranes (***Housman et al. 2014***). For example, ABCC2 and ABCC3 are transporter proteins which transport a variety of chemotherapeutic agents, such as etoposide, cisplatin and doxorubicin, and their overexpression induces multidrug resistance in cancer (***Folmer et al. 2007; Balaji et al. 2016***).

ii. **DNA damage repair:** A variety of anticancer drugs induce DNA damage, and it could be either directly or indirectly that causes cancer drug resistance, such as platinum-based drugs and topoisomerase inhibitors, respectively. Upon DNA damage by drugs, the cells respond either by the function of repair or cell death. Therefore, the efficiency of DNA-damaging

drugs extensively depends on the capacity of DNA damage repair by cancer cells (***Holohan et al. 2013***). For example, many genes, such as FEN1, FANCG and RAD23B, upregulated in human colon cancer resistant to 5-FU, involved in DNA repair (***De Angelis et al. 2004***). A tumor suppressor protein p53 expression induced by 5-FU treatment in response to DNA damage leads to either repair or induced cell death (***De Angelis et al. 2006***). A mutation in tumor suppressor protein P53 induced drug resistance by disrupting DNA-damage-induced cell cycle arrest (***Fan et al., 1994).***

iii. **Cell death inhibition:** As the apoptosis and autophagy are two central regulatory mechanisms that cause cell death. Wherein these activities are hostile to each other, and both of them contribute to programmed cell death. Apoptosis is known to have two well-established pathways: an intrinsic pathway where caspase-9, Akt and B-cell lymphoma 2 (BCL-2) protein family members play a key role, facilitated by the mitochondria activation, whereas the presence of death receptors on the cellular surface facilitates extrinsic pathway. Both intrinsic and extrinsic pathways finally merge and guide to apoptosis through the activation of downstream protein caspase-3 (***Housman et al. 2014***). Mutations, amplifications, overexpression and chromosomal translocation of these protein-coding genes have been widely associated with a diverse group of malignant tumors and chemotherapy and targeted therapy resistance (***Holohan et al. 2013***). Earlier studies demonstrate that BCL-2 overexpression induces resistance to the cytotoxic chemotherapeutic agent in human small-cell lung cancer (***Sartorius & Krammer et al. 2002***). In contrast, autophagy is a lysosomal degradation process to maintain cellular biosynthesis, where cellular organelles and protein degradation take place (***Holohan et al. 2013***). Autophagy arises in an acidic pH of lysosome due to phagolysosomal death. Drugs such as chloroquine and its derivatives increase the pH of lysosome to inactivate its

digestive enzymes to avert this process and play an important role in inhibiting autophagy-dependent resistance to chemotherapy (***Sasaki et al. 2010***).

iv. **Epigenetic Alterations caused drug resistance:** Epigenetic alterations are an emerging and important mechanism that contributes to cancer drug resistance during chemotherapy. Growing evidence of epigenetic modifications engaged in the evolution of cancer drug resistance brought people's attention to it, which includes the rise in drug efflux, increased DNA repair, and altered cell death ***(Wang et al. 2019).*** Epigenetic modifications include alterations related to DNA methylation, histone modification through acetylation or methylation, chromatin remodeling, and non-coding RNA. For instance, the demethylation of an oncogene at the promoter region of DNA would induce gene expression that caused drug resistance in cancer. A previous study suggests that a G-actin monomer binding protein, thymosin β4 (Tβ4), aberrantly expressed due to demethylation and active modification of histone H3 at the promoter region, is responsible for antiangiogenic therapy resistance by the acquisition of characteristics like stem cell in a hepatocellular carcinoma cell line (***Ohata et al. 2017***).

v. **Tumor cell heterogeneity in cancer:** Apart from the drug resistance development in cancer cells by the various vital mechanisms discussed above, heterogeneity of the tumor cell population is another aspect that may cause therapeutic resistance in cancer by extending cancer relapse. Studies reveal that within the heterogeneous population of cancer cells, a fraction of cells possess stem cell-like characteristics that are generally drug resistant. Along with that, a small proportion of some adult malignant cells also have the potency to feature drug resistance.

In cancer treatment, a drug kills only those cancer cells which are drug-sensitive, and drug-resistant cancer cells remain alive and may expand cancer. A few of these drug-resistant cancer cells may migrate via the vascular system and be able to form a new tumor in a distant part of

the body. However, the heterogeneous population of cancer cells can be seen while in circulation or even in solid tumors (***Housman et al. 2014***). For example, an early study determined two coexisting dominant clones of acute myeloid leukemia (AML), where one clone of AML was sensitive to the drug while the other clone was resistant. So, there is a possibility that relapse of this AML disease in patients later in drug treatment may be the result of the cancer cell growth due to the presence of drug-resistant clones (***Parkin et al. 2013***).

In order to conquer drug resistance, a large number of cancer genomic biomarkers have been identified in cancer, which is strongly associated with the effectiveness of an anticancer drug in cancer cell lines ***(Garnett et al. 2012).*** High throughput screening of anticancer drugs against established cancer cell lines for therapeutic drug sensitivity and resistance patterns anticipate an approach to pinpoint proper cancer subtypes and key biomarkers that may direct to the initial phase of clinical trials for a variety of novel therapeutic compounds to undergo for drug development. To reveal clinically relevant gene-drug interactions, a large number of new anticancer drug molecules have been used in screening at a massive amount against a broad range of human cancer cell lines (***Iorio et al. 2016; O'Driscoll & Clynes 2006***). Due to the limitation of an imperfect understanding of the landscape of driver genes in cancer, earlier screening of drugs was laborious work. But now, it is possible to view drug effectiveness in such models through the lens of clinically meaningful oncogenic alterations. Such kind of studies on gene/protein-drug associations is key in identifying and rectifying the complication of acquired drug resistance in cancer and in proposing novel therapeutic gene/protein biomarkers.

Based on these findings of biomarkers in different types of drug-resistant cancers, we aim to study and identify predictive biomarkers in mutant NRAS pan-cancer systems, for which no common biomarkers have been identified.

1.7 RAS and NRAS (Neuroblastoma RAS viral oncogene homolog) protein

An intensive search for the key genes found to be frequently involved in cancer drug resistance led us to the RAF-RAS family of genes. Among the proto-oncogenes, RAS proto-oncogenes (*HRAS, KRAS* and *NRAS*) are a family of GDP/GTP-regulated switches that play a significant role in controlling the activity of various key signaling pathways required for survival and cell growth (***Houben et al. 2004; Irahara et al. 2010;***). RAS proto-oncogenes are frequently expressed in human cancer and remain constitutively activated due to point mutations, while mutated RAS family genes are present in 20% of human cancers and widely contribute to tumor growth, programmed cell death, invasion, and induce the formation of new blood vessels and also involved in inducing the drug resistance (***Downward 2003; Irahara et al. 2010***). In human cancer, *KRAS* accounts for about 85% of all *RAS* mutations, *NRAS* cover about 15% and *HRAS* holds for less than 1% of mutations, and RAS family genes mutations are highly tumor-specific (***Downward 2003***). *KRAS* is reported to be highly mutated in lung, pancreatic, endometrial, colorectal, biliary tract, cervical and colon cancer, while the highest incidence of *NRAS* mutation is found in myeloid leukaemia, melanoma, bladder, neuroblastoma and thyroid cancer, etc. (***Schubbert et al. 2007; Lau & Haigis 2009***). The most frequent oncogenic mutation in *RAS* family genes (including *NRAS*) occurs at codons G12, G13 and Q61 (***Schubbert et al. 2007***).

1.7.1 NRAS signaling pathway

NRAS proteins bind with GTP to initiate the signal by activating various downstream "effector" pathways, such as RAF→MEK→ERK and PI3K→AKT cascades (**Fig. 3**) (***Bertoli et al. 2019; Irahara et al. 2010)***. Protein kinases encoded by the RAF family genes mediate cellular responses to growth signals and are regulated by activated RAS (***Houben et al. 2004***).

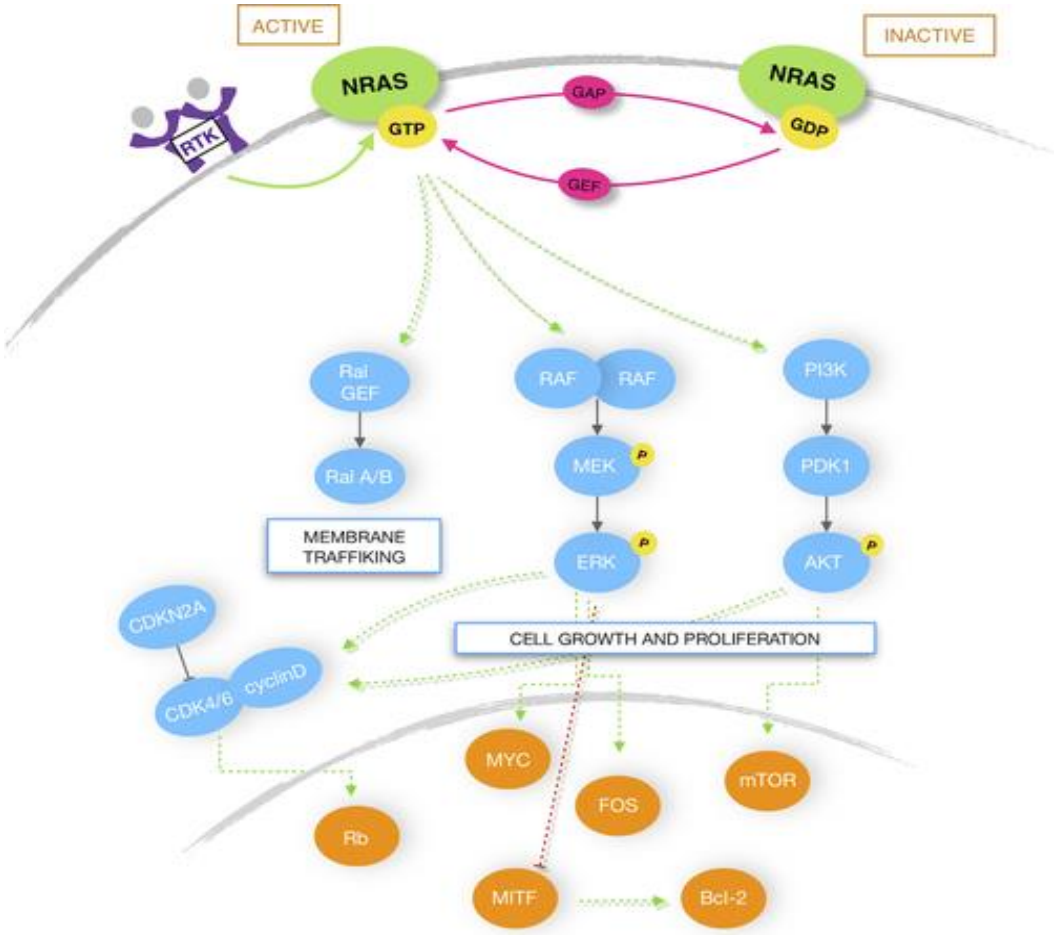


Figure 3: Schematic representation of NRAS signaling pathway. NRAS protein is activated (NRAS-GTP) by guanine nucleotide exchange factors and inactivated (NRAS-GDP) by GTPase-activating proteins. NRAS gene/protein mutation impairs the GTPase activity and remains in a constitutively activated state. Active NRAS activates downstream molecules, in turn, Ral/GEF, MEK/ERK and PI3K/AKT pathways *(Modified from Bertoli et al., 2019).*

A previous study suggests that mutations in *NRAS* or *BRAF* are highly associated with the significantly declined survival rate of patients with metastatic cancer (***Houben et al. 2004).*** While *NRAS* secondary mutation is linked with the mechanisms directly involved in cancer drug resistance (***Le et al. 2013). Le et al.*** found that acquired Vemurafenib-resistance mediated by a secondary mutation in NRAS in melanoma cells harbouring BRAF mutation, where PB04 inhibited ERK1/2 phosphorylation. Experiments with *NRAS* mutant cells showed apoptotic stress, which suppresses apoptosis and also induces drug resistance in growing cancer (***Wang et al. 2013; Haigis et al. 2008; Le et al. 2013***). RAS-targeted therapy has long been elusive (***Healy et al. 2021***). Because of the immensely high affinity of NRAS toward nucleotides like

GTP and GDP and the high intracellular concentrations of GTP, mutated NRAS remains constitutively active; hence it's very difficult to directly target mutant NRAS protein. Hence, the development of drugs for mutant NRAS is largely unsuccessful, and currently, there is no targeted therapy that has been approved for NRAS- mutant protein in cancer (***Johnson & Puzanov 2015*).** Some drugs have been developed with the potential to treat NRAS-mutant cancers, such as the MEK inhibitor binimetinib was used for NRAS-mutant melanoma, went under phase III trial but due to no difference in overall survival, did not get approval for NRAS-mutant melanoma treatment ***(Dummer et al. 2017; Garcia-Alvarez et al. 2021),*** LXH254 is a pan-RAF inhibitor which has antitumor activity in preclinical NRAS-mutant models and completed Phase I clinical trial study by February 2022 in patients with solid tumors harboring MAPK pathway alterations ***(https://clinicaltrials.gov/ct2/show/NCT02607813).*** At present, to the best of our understanding, there is no targeted drug therapy that has yet been approved for the cancers harbouring *NRAS* mutation, and several therapeutic inhibitors are currently under investigation. Even after years of extensive research, several candidates under investigation, and vast knowledge of signaling and associated drug-kinase interactions, not a single targeted therapy has been found to be supportive for NRAS-mutant cancers (***Boespflug et al. 2017; Garcia-Alvarez et al. 2021***).

It is widely known that the binding of RAS effector proteins to the RAS-GTP complex initiates the signal by activating a variety of downstream pathways that act as an effector, such as the MAPK and PI3K signaling cascades (***Rajalingam et al. 2007***). Hence, in order to screen for druggable targets, we proposed our hypothesis that the genes/proteins other than *NRAS* associated with the MAPK signaling pathway and are in direct or indirect linked with *NRAS* might be serve as promising targets. Further, understanding the regulatory environment of such druggable targets would be crucial to circumvent the effects of refractory mutant *NRAS* in cancer drug resistance.

Towards this, to understand the regulation of biomarkers in mutant *NRAS*-harbouring drug-resistant pan-cancer systems. Further, we focused on the identification of newly emerging key regulators, such as long non-coding RNAs (lncRNAs), to pinpoint key master regulators of selected coding biomarkers genes, apart from the omnipresent proteins. LncRNAs are known as new molecular players in cancer, acting as key regulators of coding gene expression. LncRNAs may directly or indirectly regulate pan-cancer drug sensitivity and resistance through their actions on such predictive biomarker targets.

This study aimed to probe the possible functional roles of predictive coding biomarkers as well as their regulatory mechanisms in the drug-resistant pan-cancer system by employing microarray data and drug response (sensitivity) data from the updated database Genomics of Drug Sensitivity in Cancer (GDSC) and The Cancer Genome Atlas (TCGA). We also constructed various biological networks such as gene co-expression, protein-protein interaction, and regulatory networks and analyzed the network using methods both qualitatively and quantitatively ***(Mishra, 2014)*** to pinpoint probable biomarkers.

Further, comprehensive studies on the regulation of these druggable targets by lncRNAs at the mRNA level. This provides a new insights into their regulatory pattern and mechanisms of these lncRNAs. These insights are highly expected to help in improving the pan-cancer drug sensitivity to these selected drugs and are also useful in drug repurposing studies utilizing our chosen target.

1.8 References

1. Roy PS, Saikia BJ. (2016) Cancer and cure: A critical analysis. *Indian J Cancer*. 53(3):441-442. doi: 10.4103/0019-509X.200658. PMID: 28244479.
2. WHO, https://www.who.int/news-room/fact-sheets/detail/cancer. (Accessed on April 2022).
3. Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. The Development and Causes of Cancer. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9963/
4. Kumar D, Sharma S, Verma S, et al. (2015) Molecular Signalling Saga in Tumour Biology. *Journal of Tumor*, 3(2): 309-313
5. Patel A. (2020) Benign vs Malignant Tumors. *JAMA Oncol*. 6(9):1488. doi:10.1001/jamaoncol.2020.2592
6. He, M., Rosen, J., Mangiameli, D., Libutti, S.K. (2007). Cancer Development and Progression. In: Mocellin, S. (eds) Microarray Technology and Cancer Gene Profiling. Advances in Experimental Medicine and Biology, vol 593. *Springer*, New York, NY. https://doi.org/10.1007/978-0-387-39978-2_12
7. Semi K, Yamada Y. (2015) Induced pluripotent stem cell technology for dissecting the cancer epigenome. *Cancer Sci*. 106(10):1251-6. doi: 10.1111/cas.12758. PMID: 26224327; PMCID: PMC4638022.
8. Vogelstein B, Kinzler KW. (2004) Cancer genes and the pathways they control. *Nat Med.* 10(8):789-99. doi: 10.1038/nm1087. PMID: 15286780.
9. Feinberg AP, Vogelstein B. (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature.* 301(5895):89-92. doi: 10.1038/301089a0. PMID: 6185846.
10. Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer*. 4(2):143-53. doi: 10.1038/nrc1279. PMID: 14732866.
11. Ushijima T. (2005) Detection and interpretation of altered methylation patterns in cancer cells. *Nat Rev Cancer*. 5(3):223-31. doi: 10.1038/nrc1571. PMID: 15719030.
12. Nowacka-Zawisza M, Wiśnik E. (2017) DNA methylation and histone modifications as epigenetic regulation in prostate cancer (Review). *Oncol Rep.* 38(5):2587-2596. doi: 10.3892/or.2017.5972. PMID: 29048620.
13. National institute of health, National Cancer Institute website. https://www.cancer.gov/about-cancer/causes-prevention/risk. (Accessed on April 2022).
14. Wu S, Zhu W, Thompson P, Hannun YA. (2018) Evaluating intrinsic and non-intrinsic cancer risk factors. *Nat Commun.* 9(1):3490. doi: 10.1038/s41467-018-05467-z. PMID: 30154431; PMCID: PMC6113228.
15. Ferlay J, Ervik M, Lam F, et al. (2020) Global Cancer Observatory: Cancer Today. *Lyon: International Agency for Research on Cancer*; (https://gco.iarc.fr/today, accessed April 2022).
16. Sung H, Ferlay J, Siegel RL, et al. (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 71(3):209-249. doi: 10.3322/caac.21660. PMID: 33538338.

17. Faguet GB. (2015) A brief history of cancer: age-old milestones underlying our current knowledge database. Int J Cancer. 136(9):2022-36. doi: 10.1002/ijc.29134. PMID: 25113657.
18. The history of cancer: Understanding what cancer is: Ancient times to present, American cancer society, revised 2018, https://www.cancer.org/cancer/cancer-basics/history-of-cancer/what-is-cancer.html#written_by. Accessed on May 2022).
19. American Cancer Society website. Treatment types. www.cancer.org/treatment/treatments-and-side-effects/treatment-types.html. Accessed on May, 2022.
20. https://medlineplus.gov/ency/patientinstructions/000901.htm. Accessed on May 2022).
21. National institute of health, National Cancer Institute website. Types of cancer treatment https://www.cancer.gov/about-cancer/treatment/types. Accessed on May 2022).
22. Mansoori B, Mohammadi A, Davudian S, et al. (2017) The Different Mechanisms of Cancer Drug Resistance: *A Brief Review*. *Adv Pharm Bull*. 7(3):339-348. doi: 10.15171/apb.2017.041. PMID: 29071215; PMCID: PMC5651054.
23. Longley DB, Johnston PG. (2005) Molecular mechanisms of drug resistance. *J Pathol*. 205(2):275-92. doi: 10.1002/path.1706. PMID: 15641020.
24. Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. (2013) Cancer drug re-sistance: an evolving paradigm. *Nat Rev Cancer*. (10):714-26. doi: 10.1038/nrc3599. PMID: 24060863.
25. Housman G, Byler S, Heerboth S, et al. (2014) Drug resistance in cancer: an overview. *Cancers (Basel)*. 6(3):1769-92. doi: 10.3390/cancers6031769. PMID: 25198391; PMCID: PMC4190567.
26. Urruticoechea A, Alemany R, Balart J, et al. (2010) Recent advances in cancer therapy: an overview. *Curr Pharm Des.* 16(1):3-10. PMID: 20214614.
27. Baskar R, Lee KA, Yeo R, Yeoh KW. (2012) Cancer and radiation therapy: current advances and future directions. *Int J Med Sci.* 9(3):193-9. PMID: 22408567; PMCID: PMC3298009.
28. Damin DC, Lazzaron AR. (2014) Evolving treatment strategies for colorectal cancer: a critical review of current therapeutic options. *World J Gastroenterol.* 20(4):877-87. PMID: 24574762; PMCID: PMC3921541.
29. Khalil DN, Smith EL, Brentjens RJ, Wolchok JD. (2016) The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. *Nat Rev Clin Oncol.* 13(5):273-90. 13(6):394. PMID: 26977780; PMCID: PMC5551685.
30. Wang X, Zhang H, Chen X. (2019) Drug resistance and combating drug resistance in cancer. *Cancer Drug Resist*. 2:141-160. PMID: 34322663; PMCID: PMC8315569.
31. Dai Z, Gu XY, Xiang SY, et al. (2020) Research and application of single-cell sequencing in tumor heterogeneity and drug resistance of circulating tumor cells. *Biomark Res*. PMID: 33292625; PMCID: PMC7653877.
32. Kurrey NK, Jalgaonkar SP, Joglekar AV, et al. (2009) Snail and slug mediate radioresistance and chemoresistance by antagonizing p53-mediated apoptosis and acquiring a stem-like phenotype in ovarian cancer cells. *Stem Cells.* 27(9):2059-68. PMID: 19544473.
33. Gorre ME, Mohammed M, Ellwood K, et al. (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science. 293(5531):876-80. PMID: 11423618.

34. Mu P, Zhang Z, Benelli M, et al.(2017) SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. Science. 355(6320):84-88. PMID: 28059768; PMCID: PMC5247742.
35. Folmer Y, Schneider M, Blum HE, Hafkemeyer P. (2007) Reversal of drug resistance of hepatocellular carcinoma cells by adenoviral delivery of anti-ABCC2 antisense constructs. *Cancer Gene Ther*. 14(11):875-84. PMID: 17704753.
36. Balaji SA, Udupa N, Chamallamudi MR, Gupta V, Rangarajan A. (2016) Role of the Drug Transporter ABCC3 in Breast Cancer Chemoresistance. *PLoS One*. 11(5):e0155013. PMID: 27171227; PMCID: PMC4865144.
37. De Angelis PM, Fjell B, Kravik KL, et al. (2004) Molecular characterizations of derivatives of HCT116 colorectal cancer cells that are resistant to the chemotherapeutic agent 5-fluorouracil. *Int J Oncol.* 24(5):1279-88. PMID: 15067352.
38. De Angelis PM, Svendsrud DH, Kravik KL, Stokke T. (2006) Cellular response to 5-fluorouracil (5-FU) in 5-FU-resistant colon cancer cell lines during treatment and recovery. Mol Cancer. 5:20. doi: 10.1186/1476-4598-5-20. PMID: 16709241; PMCID: PMC1524802.
39. Fan S, el-Deiry WS, Bae I, et al. (1994) p53 gene mutations are associated with decreased sensitivity of human lymphoma cells to DNA damaging agents. Cancer Res. 54(22):5824-30. PMID: 7954409.
40. Sartorius UA, Krammer PH. (2002) Upregulation of Bcl-2 is involved in the mediation of chemotherapy resistance in human small cell lung cancer cell lines. *Int J Cancer*. 97(5):584-92. PMID: 11807782.
41. Sasaki K, Tsuno NH, Sunami E, et al. (2010) Chloroquine potentiates the anti-cancer effect of 5-fluorouracil on colon cancer cells. *BMC Cancer*. 10:370. PMID: 20630104; PMCID: PMC2914703.
42. Ohata Y, Shimada S, Akiyama Y, et al. (2017) Acquired Resistance with Epigenetic Alterations Under Long-Term Antiangiogenic Therapy for Hepatocellular Carcinoma. Mol Cancer Ther. 16(6):1155-1165. PMID: 28246302.
43. Parkin B, Ouillette P, Li Y, et al. (2013) Clonal evolution and devolution after chemotherapy in adult acute myelogenous leukemia. Blood. 121(2):369-77. PMID: 23175688; PMCID: PMC3653567.
44. Garnett MJ, Edelman EJ, Heidorn SJ, (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 483(7391):570-5. PMID: 22460902; PMCID: PMC3349233.
45. Iorio F, Knijnenburg TA, Vis DJ, et al. (2016) A Landscape of Pharmacogenomic Interactions in Cancer. *Cell*. 166(3):740-754. PMID: 27397505; PMCID: PMC4967469.
46. O'Driscoll L, Clynes M. (2006) Molecular markers of multiple drug resistance in breast cancer. *Chemotherapy*. 52(3):125-9. PMID: 16612055.
47. Houben R, Becker JC, Kappel A, et al. (2004) Constitutive activation of the Ras-Raf signaling pathway in metastatic melanoma is associated with poor prognosis. *J Carcinog.* 3:6. PMID: 15046639.
48. Irahara N, Baba Y, Nosho K, et al. (2010) NRAS mutations are rare in colorectal cancer. *Diagn Mol Pathol.* 19(3):157-63. PMID: 20736745; PMCID: PMC2929976.
49. Downward J. (2003) Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer*. 3(1):11-22. PMID: 12509763.

50. Schubbert S, Shannon K, Bollag G. (2007) Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer*. 7(4):295-308. Erratum in: Nat Rev Cancer. 7(7):563. PMID: 17384584.
51. Lau KS, Haigis KM. (2009) Non-redundancy within the RAS oncogene family: insights into mutational disparities in cancer. Mol Cells. 28(4):315-20. PMID: 19812895.
52. Bertoli E, Giavarra M, Vitale MG, Minisini AM. (2019) Neuroblastoma rat sarcoma mutated melanoma: That's what we got so far. *Pigment Cell Melanoma Res*. 32(6):744-752. PMID: 31403745.
53. Le K, Blomain ES, Rodeck U, Aplin AE. (2013) Selective RAF inhibitor impairs ERK1/2 phosphorylation and growth in mutant NRAS, vemurafenib-resistant melanoma cells. *Pigment Cell Melanoma Res*. 26(4):509-17. PMID: 23490205.
54. Wang Y, Velho S, Vakiani E, et al. (2013) Mutant N-RAS protects colorectal cancer cells from stress-induced apoptosis and contributes to cancer development and progression. *Cancer Discov*. 3(3):294-307. PMID: 23274911.
55. Haigis KM, Kendall KR, Wang Y, et al. (2008) Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentia-tion and tumor progression in the colon. *Nat Genet*. 40(5):600-8. PMID: 18372904.
56. Healy FM, Prior IA, MacEwan DJ. (2021) The importance of Ras in drug resistance in cancer. *Br J Pharmacol*. PMID: 33634485.
57. Johnson DB, Puzanov I. (2015) Treatment of NRAS-mutant melanoma. *Curr Treat Options Oncol.* 16(4):15. PMID: 25796376
58. Dummer R, Schadendorf D, Ascierto PA, et al. (2017) Binimetinib versus dacarbazine in patients with advanced NRAS-mutant melanoma (NEMO): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol*. 18(4):435-445. PMID: 28284557.
59. Garcia-Alvarez A, Ortiz C, Muñoz-Couselo E. (2021) Current Perspectives and Novel Strategies of *NRAS*-Mutant Melanoma. *Onco Targets Ther*. 14:3709-3719. PMID: 34135599.
60. Boespflug A, Caramel J, Dalle S, Thomas L. (2017) Treatment of *NRAS*-mutated ad-vanced or metastatic melanoma: rationale, current trials and evidence to date. *Ther Adv Med Oncol.* 9(7):481-492. PMID: 28717400.
61. Rajalingam K, Schreck R, Rapp UR, Albert S. (2007) Ras oncogenes and their downstream targets. *Biochim Biophys Acta.* 1773(8):1177-95. doi: 10.1016/j.bbamcr.2007.01.012. PMID: 17428555.
62. Mishra S. (2014) CSNK1A1 and Gli2 as Novel Targets Identified Through an Integrative Analysis of Gene Expression Data, Protein-Protein Interaction and Pathways Networks in Glioblastoma Tumors: Can These Two Be Antagonistic Proteins? *Cancer Inform*. 13;13:93-108. PMID: 25374452.

Chapter 2

Databases and Tools

2.1 Databases

A. Genomics of drug sensitivity in cancer

The GDSC is a wellcome funded joint collaborative project of The Cancer Genome Project at the Wellcome Sanger Institute and the Center for Molecular Therapeutics, Massachusetts General Hospital Cancer Center. The expertise from both places in this collaboration has focused toward the aim of identifying cancer biomarkers which can be used to identify genetically elucidated groups of patients in response to cancer treatment.

The GDSC database (www.cancerRxgene.org) is established to provide an information on the molecular properties of cancer cells that control drug response. GDSC carry and annotates massive amount of datasets related to drug sensitivity in cancer cell lines, and especially these data are linked with genomic information in detail to facilitate the molecular biomarkers discovery of drug response. The GDSC database basically describes three types of datasets are following;

i. **Drug sensitivity data in cancer cell lines:** The drug sensitivity (IC_{50}) data of cancer cell lines are generated from an ongoing high-throughput drug screened against a collection of >1000 cell lines at the Wellcome Trust Sanger Institute (WTSI) by the Cancer Genome Project and at Massachusetts General Hospital by the Center for Molecular Therapeutics. Anticancer therapeutics compounds that are selected for screening include both cytotoxic chemotherapeutics and targeted agents. These compounds are either approved for clinical use, under clinical development and investigation, or experimental drugs in the early phase of development (***Yang et al. 2013)***.

ii. **Genomics datasets for cancer cell lines:** The total collection available for drug screening includes around 1000 cancer cell lines from different tissue types. These cell lines have been selected to constitute the frequent and rare types of cancers in adult and childhood derived from the haematopoietic, epithelial and mesenchymal cells. These cell lines in

GDSC have been widely characterized genomically and are part of a project on the cancer cell line from the Cancer Genome Project. The GDSC contains genomic datasets for many different types of cancer cell lines. The datasets include huge amounts of information on somatic mutations for cancer-related genes, genome-wide copy numbers for amplification and deletion of the gene, markers of microsatellite instability, and pan-tissue type, along with transcriptional data. All these information regarding genomic alteration and others directly available in the Catalogue of Somatic Mutations in Cancer (COSMIC) database, a publically available open-source for the annotation and presentation of somatic gene mutations in cancer (***Yang et al. 2013)***.

iii. **Analysis of genomic features of drug sensitivity:** The systematic incorporation of a wide range of genomic and drug sensitivity data is a crucial element of the GDSC database. There are two complementary analytical approaches have been used to spot genomic markers of drug sensitivity in cancer. An analysis of variance (ANOVA) is used to correlate genomic alterations with drug sensitivity (IC50 values) in cancer, such as somatic mutations, gene deletions and amplifications of common cancer-related genes, rearrangements of genes and microsatellite instability. The ANOVA analysis point outs particular genomic alteration linked with drug sensitivity and also describe a size effect and calculate statistical significance for each drug-gene association. Other hand, elastic net regression has been used, a penalized linear regression modeling approach, to identify a variety of relevant genomic features which influences drug effectiveness. Elastic net regression analysis includes all of those genomic data used in the ANOVA analysis and also integrates the transcriptional profiles of the genome and pan-cancer tissue type (***Yang et al. 2013)***.

B. TCGA and CancerRx Tissue

TCGA is a landmark cancer genomics collaborative program between the National Cancer Institute (NCI), Therapeutically Applicable Research to Generate Effective Treatments (TARGET) and the National Human Genome Research Institute (NHGRI). This database provided molecularly characterized high-quality 20,000 primary cancer tissue samples and matched normal samples derived from 33 cancer types. Over the years, TCGA has produced more than 2.5 petabytes of genomic, transcriptomic, and epigenetic data along with proteomic data. To study the genomic and proteomic profiles of tissue samples from patients, TCGA has used various genomic approaches. It integrates clinical information about patients, metadata about sample information, and molecular information about coding and non-coding gene sequence, DNA methylation, somatic mutation and copy number variation.

The CancerRx tissue database was created for public users to visualize and download the predicted drug sensitivity (IC_{50}) data of 272 drugs in cancer tissue. Predictive models were built for 272 drugs using the gene expression data from the CCLE database in cancer cell lines and drug sensitivity (IC_{50}) data for cancer cell lines from GDSC by applying the genetic algorithm (GA) and *k*-nearest neighbours (KNN) algorithm. Subsequently, the same predictive models were applied to predict drug response (IC_{50} values) for ~17,000 samples, including both normal and tumor tissues, using RNA-seq gene expression profile data for the tissues from TCGA and GTEx. Predicted cancer tissue drug sensitivity (IC_{50}) data is available at the following link: https://manticore.niehs.nih.gov/cancerRxTissue (***Li et al. 2021***).

C. GeneCodis4:

GeneCodis is an online web server for functional enrichment analysis. Researchers from worldwide uses this tools to combine various sources of annotations. It retrieves sets of meaningful simultaneous annotations and allocates a valid statistical score to asses those outcomes that are remarkably enriched from the set of input genes/proteins list. In order to

elucidate the fundamental cellular and biological mechanisms, GeneCodis4 has been extensively used to investigate sets of genes/proteins. This web server supports functional annotations for genes, proteins, miRNA, CpG sites and TFs identifiers extracted from 16 different organisms, including *Homo sapiens*. However, GeneCodis4 categorizes annotations into three leading groups by integrating 19 various collections: functional, regulatory and perturbation annotations. The first functional group overspreads the following databases: Gene Ontology (GO) and its three subgroups (Biological Process, Molecular Function and Cellular Component) and pathway include; KEGG Pathways, Reactome, WikiPathways, and Mouse Genome Informatics database Panther Pathways. The second regulatory group holds two curated associations; TF-gene and miRNA-gene interactions. And finally, the perturbation group collects two types of associations, which encompasses gene-chemicals and gene-phenotype associations (***García-Moreno et al. 2021***). GeneCodis 4 can accepts various kinds of input ID/name lists: genes/proteins, TFs, CpG sites and miRNAs. GeneCodis4 web server is publically available at https://genecodis.genyo.es.

D. GeneMANIA

It is a user-friendly, freely available web interface, generally used for gene function predictions, comprising a widely adaptive algorithm. It is a simple interactive, intuitive interface and extendable database and also a Cytoscape plugin application. GeneMANIA collects network data from databases resource which are publicly available for users, such as Gene Expression Omnibus (GEO) database provides gene co-expression network data, BioGRID database provides physical and genetic interaction data and predicted protein-protein interaction data based on orthology from I2D, etc. These network data obtain from different sources like BioGRID, Human Protein Reference Database, IntAct, MINT and Reactome, etc., across the eight organisms like; *Homo sapiens*, *Arabidopsis thaliana*, *Mus musculus*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Danio rerio*, *Rattus norvegicus* and *Saccharomyces*

cerevisiae. GeneMANIA also collects individual data from organism-specific genomic datasets (***Warde-Farley et al. 2010***). It generates networks from a set of gene lists and categorizes them as gene co-expression, gene fusion, shared domain proteins, physically interacted genes, and predicted interactions and pathway genes. Users can download gene network data available at the following link *http://www.genemania.org.*

E. STRING

It is a database that provides experimentally validated and predicted protein-protein interactions. This database integrates direct (physical) and indirect (functional) interactions of protein obtained from *in silico* prediction, from knowledge conveyed between organisms, and interactions collected from other (primary) databases. Protein-protein interactions data in the STRING database are extracted from various sources, including genomic context predictions, high-throughput experimentally determined, co-expression, automated text-mining, computational prediction and curated database. STRING database version 11.5 currently contains data for approximately 24,584,628 proteins from 14,000 organisms. The database user can query the PPI network from STRING directly within Cytoscape apart from the online website. All protein-protein interaction evidence in the database that incorporates a given network is benchmarked and scored and these scores are included in a final 'consolidated score.' These scores mapped between zero to one and approximate STRING's confidence in whether a presented association is biologically significant, given all the contributing corroboration. Protein-protein interaction networks data available in the STRING database can be exported from the following link: https://string-db.org/.

F. Open-access Repository Transcription factors interactions (ORTI)

It is a huge and freely accessible database for transcriptional; transcription factor-target gene (TF-TG) interactions in the human and mouse, experimentally validated using high-throughput methods. It is followed by tools that can identify and anticipate transcriptional (TF-TG)

interactions. The ORTI database was created by combining several open-access databases and, from extensive literature searches to bring about a cluster of TF-TG interactions. Transcription factors (TFs) are known to have key roles in biological and cellular pathways. These TFs are the end target activated by various external stimuli through the cascade of intermediate molecules. These transcription factors, along with complex proteins, activate or suppress the transcription of specific target genes, which shows an impact on biological and cellular functions in the cells. The data for TF-TG interaction was retrieved from various database sources, including HTRI, TRED, TFactS, TRRD, PAZAR, and NFI-Regulome, and also from a literature search to construct an ORTI database. This database includes 20146 genes, 660 TFs, and 72,817 TF-TG interaction data. ORTI serves as an asset for unrevealing the context-specific topology of interaction networks. The interaction data flat file for TF-TG is available at the following link: https://orti.sydney.edu.au/index.html. This web portal also allows public users to search for TF or TG names, and the database provides suggestions when it comes to queries based content.

G. Harmonizome

It is a publically accessible web portal that lays out a pictorial user interface, a web service for browsing and downloading all of the accumulated data. The Harmonizome database was built by using a collection of various processed data assembled to provide and extract information about humans and mice, genes and proteins from 125 unique datasets hosted by 72 major open-access web resources. This database extracted and abstracted around 72 million functional consortiums between genes/proteins and arranged these data systematically. These gathered datasets cover information about mammalian genes or proteins, mainly divided into six broad categories, which include; transcriptomic profiles, genomic profiles, proteomic profiles, structural or functional annotations, disease and phenotype associations, and physical interactions. The Harmonizome home page features a search bar that can be used to enter any

key search term and system autocomplete search for users by autocomplete capabilities. The system searches for matching datasets, genes and attributes that may contain metadata and deliver various views (***Rouillard et al. 2016***). The data is available on the database at the following link: *http://amp.pharm.mssm.edu/Harmonizome*.

H. RCSB PDB

The Protein Data Bank (PDB) is a publicly available open-access database for the three-dimensional crystal structure data of macro biomolecules, which include primarily proteins, nucleic acids (DNA & RNA), and associated small molecules such as drugs, cofactors and inhibitors. PDB database was created in the year 1971 at Brookhaven National Laboratories (BNL) as an archive for macromolecular 3D structures, typically determined by X-ray crystallography and nuclear magnetic resonance (NMR) spectrometry and submitted by biologists and biochemists from different parts of the globe. The PDB database is supervised by an international consortium called 'Worldwide Protein Data Bank (wwPDB),' a collaboration among three countries; United States, Europe, and Japan. This database provides a tool for searching and exploring the data from PDB, including an interactive interface that lets users explore how chemical interactions affect the stability of macromolecules and leads to play key roles in their interactions and functions (***Berman et al. 2000; Christine et al. 2016***).

I. ZINC database

ZINC is an open-access database and tool set which is basically developed to enable ready access to compounds for virtual screening. It has become widely used for ligand discovery, pharmacophore screens and other aspects of drug discovery. ZINC is used by investigators in pharmaceutical companies, biotechnology companies and research universities. ZINC15 (current version) currently holds more than 120 million purchasable "drug-like" compounds effectively all of which are organic molecules. ZINC15 retrieves drug data from various other database sources such as ChEMBL, DrugBank, HMDB and https://ClinicalTrials.gov to

annotate the compounds with detailed information which are active in, or naturally originated, including FDA-approved drugs, pre-clinical drugs, experimental or investigational compounds, natural products, and metabolites, and others (***Irwin & Shoichet 2005; Sterling & Irwin 2015***). This database provides drug-like compounds in the form of several common file formats SMILES, mol2, 3D SDF, and DOCK Flexi base file format. These all are freely available at the following link: *http://zinc15.docking.org*.

2.2 Tools

A. MultiExperiment Viewer (MeV)

MeV is a cloud-based, freely available multifaceted software application. It supports advanced bioinformatics tools for combined data analysis, visualization and stratification of massive genomic data in the hands of bench biologists. MeV is a tool primarily used to analyze microarray and RAN-Seq data consolidating advanced algorithms for statistical analysis for differential expression, visualization, classification, clustering, and functional representation through a graphical approach.

B. RStudio and R programming

RStudio is an integrated development environment and open tools for R. R is a programming language used for data miners and statistical analysis, generating graphical representations and developing statistical tools. R is publically available on the GNU General Public License, and binary versions work for various operating systems like Linux, Windows and Mac. Various packages and code used for plot generation in R is freely available.

i. We used R code to generate the volcano plot given below:

```
library ("ggplot2")
>mydata<-read.csv("filename.farmet", header=T, sep=",")
>mydata$threshold = as.factor(mydata$Adj.p.value < 0.05)
>mydata$threshold = as.factor(abs(mydata$logFC) > 2 & mydata$Adj.p.value < 0.05)
>g <- ggplot(data=mydata, aes(x=logFC, y =-log10(Adj.p.value), colour=threshold))
+ geom_point(alpha=0.4, size=1.75)
```

+ ggtitle("plot_title")

+ xlim(c(-6, 6))

+ xlab("log2 fold change") + ylab("-log10 Adj p-value") +

theme_bw() + theme(legend.position="none", plot.title = element_text(size = rel(1.5), hjust = 0.5), axis.title = element_text(size = rel(1.25)))
>g

ii. R code used to generate bubble plot:

>mydata<-read.csv("filename.farmet", header=T, sep=",")

>p = ggplot(mydata,aes(Rich.factor,Biological.process))

>p=p + geom_point()

>p=p + geom_point(aes(size=Gene.number))

*>pbubble = p+ geom_point(aes(size=Gene.number,color=-1*log10(pval_adj)))*

>pr = pbubble+scale_color_gradient(low="green",high = "red")

>pr = pr+labs(color=expression(-log[10](p-value)),size="Gene number",x="Rich factor",y="Biological process",title="Top GO or KEGG terms enrichment")+theme(plot.title=element_text(hjust=0.5))

>pr=pr + theme_bw()

>pr

iii. R code used to generate balloon plot:

library ("ggplot2")

library (ggpubr)

>mydata<-read.csv("filename.farmet", header = T, sep = ",", row.names = "Gene")

>ggballoonplot(mydata)

>ggballoonplot(mydata, fill = "value",color = "lightgray",size = 5, show.label = F)+

gradient_fill(c("white", "white", "red")) + theme(axis.text.x = element_text(face="bold", color="black", size=10, angle=45), axis.text.y = element_text(face="bold", color="black", size=7, angle=360))

C. GenePattern and Comparative marker selection

GenePattern is a powerful, user-friendly, and freely available web interface that provides access to a variety of computational tools used to analyze genomic data, such as gene expression profile data (RNA-seq and microarray), sequence variation, and copy number, and network data analysis. These tools are all available (on the following link:

https://www.genepattern.org/) through a software package with no programming knowledge required. The comparative marker selection (CMS), a module of the GenePattern web interface, is used to analyze data derived from high-throughput experiments such as microarray or RNA-seq data. CMS uses a test statistic to calculate the relative changes in the gene expression that can discriminate between the two classes of the samples (such as drug-sensitive *vs* resistance) and assess the significance (*p*-value) of the test statistic score. CMS identifies marker genes by calculating the expression value for each profiled gene which assesses the correlation of the gene's expression profile in distinct classes. The test statistics values are calculated by CMS for each gene, which determines the differentially expressed genes between classes that are expected to be marker genes. The CMS takes two files as input: one for gene expression data from a different sample belonging to two classes and another file that specifies the class of each sample. CMS produces a structured output file with significance values that include several test statistic scores, such as p-value, logFC, FDR (BH), Q-value, maxT, and FWER for each gene. The results generated from the CMS algorithm are visualized as a heatmap using a comparative marker selection viewer, which accepts the output file and represents the results collectively (***Kuehn et al. 2008).***

D. Cytoscape

It (https://cytoscape.org) is one of the most frequently used open-source computational tools for visualizing and studying the molecular interaction networks as well as integrating with high throughput gene expression and metabolic data. Although the study of molecular elements and interactions is applicable to any model system, Cytoscape is extensively used in combination with many large databases that contain data of protein-protein, protein-DNA, and genetic interactions and are broadly available for humans as well as for other model organisms. It provides basic functionality that helps to import and query the networks and allows to visualize networks which come integrated with expression data from different phenotypes. In Cytoscape,

nodes represent biological molecules (genes or proteins) and edges are the connection between two nodes, which depict the kind of relationship or interaction between the nodes, such as inhibition or activation (***Shannon et al. 2003; Kohl et al. 2011***). Cytoscape provides a network Analyzer tool to compute comprehensive topological parameters for directed or undirected networks.

E. Open Babel

Open Babel is a designed toolbox for users which can read and speak the many different languages of chemical data. It is an open and publicly available platform used to search, convert, and analyze data from various areas, such as molecular modeling, chemistry, solid-state materials, and biochemistry. Open babel can read and interconvert more than 110 molecular file formats and also generates 2D and 3D coordinates for various file formats such as SDF, mol files (***O'Boyle et al. 2011***). It is freely available for users under a free license from the following web link: *http://openbabel.org*.

F. Swiss-PDB Viewer

Swiss-PdbViewer (DeepView) is a software platform that provides an easy-to-use interface so that users can analyze multiple proteins simultaneously. The proteins can be superimposed so that the positions of their active sites can be compared and deduce structural alignments. Mutation of amino acid residues in a protein, hydrogen bond interactions, angles and distances between atoms may also be observed. Swiss-PdbViewer was developed by Nicolas Guex in 1994. It was initially associated with SWISS-MODEL, an automated homology modeling server designed by the Structural Bioinformatics Group at the Swiss Institute of Bioinformatics (SIB), Biozentrum in Basel ***(Guex & Peitsch 1997***). In our study, SPDBV was used for energy minimization. SPDBV is available on the web link: *http://www.expasy.org/spdbv***.**

G. AutoDock Tools

AutoDock Tools (ADT) is a graphical user interface for setting up and running a computerized docking application program which used to predict the binding of small molecules, such as drug candidates, to a receptor of a known 3D structure. AutoDock Tools simplifies the process of configuring the input molecule files used for molecular docking. It provides with an array of methods that leads the user through molecule protonation, calculating various charges, and specifying rotatable bonds in the protein and ligand. ADT allows the user to identify the active site and find out visually the volume of space occupied by molecules in the docking simulation. It also allows users to view results from molecular docking experiments using a variety of methods, such as clustering and analyzing data ***(Morris et al. 2009***). For our study, we used the AutoDock tool to prepare protein, ligand molecules and grid box generation for molecular docking.

H. AutoDock Vina

AutoDock Vina is a user-friendly program for molecular docking as well as virtual screening (***Trott & Olson 2010***). It was originally developed and launched at The Scripps Research Institute by Dr. Oleg Trott in the Molecular Graphics Lab. AutoDock Vina is one of the docking engines of the AutoDock Suite. Vina significantly predicts the binding mode with average accuracy, efficient optimization and multithreading. It calculates the grid maps and clusters the docking results automatically, which is transparent to the researchers. Vina uses the same .pdbqt file format for its input and also gives the .pdbqt file format as the results output. Vina is freely available on the following link: https://vina.scripps.edu.

I. PyMOL

PyMOL is a cloud-based tool for molecular visualization interface developed by Warren Lyford DeLano. Currently, it is maintained and distributed by Schrödinger **(16).** It is a globally used tool in the field of research in structural biology that became globally available to scientific and educational groups. PyMOL can generate high-quality three-dimensional images of small molecules and macromolecules, such as proteins. For our research, PyMol has been used to analyze the docking of protein-ligand complexes and to convert from pdbqt to PDB format.

J. Discovery Studio

This software was developed and distributed by Dassault Systems BIOVIA, with multiple applications. *Discovery Studio* Visualizer is one of the leading visualization tools for viewing and analyzing proteins and modeling molecular structures, simulations including molecular mechanics, molecular dynamics, and quantum mechanics, and other data of relevance to life science researchers in structural biology. This BIOVIA product provides features for viewing and editing data, as well as basic data analysis tools. It also provides a huge platform for displaying plots and representations of 3D graphics of data. For our research, Discovery studio has been used to study protein-ligand interactions analysis post-virtual screening through molecular docking.

K. PockDrug

PockDrug is an online pocket druggability prediction tool that predicts the possible druggability of the pockets present in a protein. This web-based tool uses geometry, hydrophobicity, and aromaticity of pocket residues or atoms of a protein to predict possible druggable pockets that can be targeted by drugs. PockDrug uses two different methods for the estimation of pocket druggability, first is prox4 and prox5.5 estimation methods predict druggability by using the information about ligand position to guide its extraction of protein

atoms located within two fixed distance point thresholds of 4 and 5.5 Å from the bound ligand, respectively. The second is the fpocket estimation method, which is an automated geometry-based method. It is not guided by the position of a ligand to predict pocket druggability. This method uses the Voronoi polyhedral decomposition of a 3D protein to extract all the pockets volume from the apo- or holo-protein using spheres of varying diameters (***Hussein et al. 2015***). PockDrug-Server is publicly available at: *http://pockdrug.rpbs.univ-paris-diderot.fr*.

2.3 References

1. Yang W, Soares J, Greninger P, et al. (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*. 41(Database issue):D955-61. PMID: 23180760.
2. Li Y, Umbach DM, Krahn JM, Shats I, Li X, Li L. (2021) Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. BMC Genomics. 22(1):272. PMID: 33858332; PMCID: PMC8048084.
3. A. García-Moreno, R. López-Domínguez, A. Ramirez-Mena, et al. (2021) GeneCodis4: Expanding the modular enrichment analysis to regulatory elements.
4. Warde-Farley D, Donaldson SL, Comes O, et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38(Web Server issue):W214-20. doi: 10.1093/nar/gkq537. PMID: 20576703; PMCID: PMC2896186.
5. Szklarczyk D, Gable AL, Lyon D, et al. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47(D1):D607-D613. doi: 10.1093/nar/gky1131. PMID: 30476243; PMCID: PMC6323986.
6. Vafaee F, Krycer JR, Ma X, et al. (2016) ORTI: An Open-Access Repository of Transcriptional Interactions for Interrogating Mammalian Gene Expression Data. *PLoS One.* 11(10):e0164535. PMID: 27723773; PMCID: PMC5056720.
7. Rouillard AD, Gundersen GW, Fernandez NF, et al. (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford).* 2016:baw100. PMID: 27374120; PMCID: PMC4930834.
8. Kuehn H, Liberzon A, Reich M, Mesirov JP. (2008) Using GenePattern for gene expression analysis. Curr Protoc Bioinformatics. Chapter 7: Unit 7.12. doi: 10.1002/0471250953.bi0712s22. PMID: 18551415; PMCID: PMC3893799.
9. Kohl M, Wiese S, Warscheid B. (2011) Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol.* 696:291-303. doi: 10.1007/978-1-60761-987-1_18. PMID: 21063955.
10. Shannon P, Markiel A, Ozier O, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13(11):2498-504. PMID: 14597658; PMCID: PMC403769.
11. Berman HM, Westbrook J, Feng Z, et al. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28(1):235-42. PMID: 10592235; PMCID: PMC102472.
12. Zardecki Christine, Dutta Shuchismita, Goodsell, David S. et al. (2016) RCSB Protein Data Bank: A Resource for Chemical, Biochemical, and Structural Explorations of Large and Small Biomolecules. *J of Chem Education* 3(93):569-575
13. Irwin JJ, Shoichet BK. (2005) ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model*. 45(1):177-82. PMID: 15667143; PMCID: PMC1360656.
14. Sterling T, Irwin JJ. (2015) ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model*. 55(11):2324-37. PMID: 26479676; PMCID: PMC4658288.
15. O'Boyle NM, Banck M, James CA, et al. (2011) Open Babel: An open chemical toolbox. J Cheminform. 3:33. PMID: 21982300; PMCID: PMC3198950.
16. Guex N, Peitsch MC. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* 18(15):2714-23. PMID: 9504803.

17. Morris GM, Huey R, Lindstrom W, et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*. 30(16):2785-91. PMID: 19399780; PMCID: PMC2760638.
18. Trott O, Olson AJ. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 31(2):455-61. PMID: 19499576; PMCID: PMC3041641.
19. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
20. Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016.
21. Hussein HA, Borrel A, Geneix C, et al. (2015) PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. *Nucleic Acids Res*. 43(W1):W436-42. PMID: 25956651; PMCID: PMC4489252.

Chapter-3

Objective-1: Computational analysis of drug-dose responses from a panel of mutant NRAS pan-cancer cell lines to identify drug-sensitive and -resistant cell lines from the GDSC database.

3.1 Introduction

Drug sensitivity is one of the main reasons for individualized cancer chemotherapy since past experiments have shown that certain drugs work better with some people. Oncologists made therapeutic conclusions based on their patients' experiences, based on the pathological features of the tumor, prior to the appearance of drug sensitivity testing, rather than relying on their assessment and understanding of tumor responses to therapeutic drugs in clinics. Since the arrival of drug-sensitivity testing in cancer, oncologists have played a crucial role in the cure of cancer. Prediction of drug sensitivity has become quite accessible due to the development of computational approaches that can promote precision anticancer therapeutics (***Tang et al. 2021***). However, due to the high prevalence of drug resistance in cancer demand further, more research and development of new therapeutic treatments as the potency of cancers to develop resistance to conventional therapies are now day's increased.

Precision or personalized medicine is intended to provide the most appropriate treatment for each individual patient of cancer. In the development of advanced oncology techniques such as next-generation sequencing (NGS), transcriptome (RNA-sequencing), ChIP-sequencing, and mass spectrometry are extensively used to perform full molecular profiling for each cancer patient. However, because of some of the high degrees of tumor heterogeneity, it is quite challenging to suggest an appropriate treatment for a cancer patient on the basis of high-throughput molecular profiling. To be able to define the drug-sensitivity and -resistance of each individual cancer, an *in vitro* test can be carried out on cancer cells or tissue samples derived from a patient with a panel of therapeutic drugs (***Popova et al. 2020***).

The origin of advanced technology like; proteomics, microarray and targeted therapies provide new insight into conquering drug resistance in cancer. Although new chemotherapeutic agents are being designed in large numbers, still an effective chemotherapy agent is yet to be

uncovered for cancer in the advanced stage. There may be several factors, including genetic differences of the individual, especially in somatic cells of the tumor, which might be driving cancer cell resistance to anticancer agents (***Mansoori et al. 2017***).

Currently, large-scale of biological data is being generated at an economical cost by using advanced high-throughput technologies to investigate drug sensitivity and resistance in cancer (***Pouryahya et al. 2022***). There are some databases available, such as the NCI-60, GDSC and CCLE, which are pioneers of such datasets (***Shoemaker 2006; Iorio et al. 2016; Barretina et al. 2012***). Overall, studies from these different databases have illustrated that pharmacogenomics profiling of cancer cell lines derived from clinical tumor tissue samples can be used as a platform for biomarker discovery that could lead to the development of a new method for cancer treatment ***(Yang et al. 2013; Garnett et al. 2012)***. The NCI-60 database is one of the earliest established studies to screen drugs *in vitro* among these drug sensitivity databases. It has remarkably improved the philosophy and research on anticancer drugs (***Shoemaker 2006; Chabner 2016).***

The NCI-60 cell line panel and screened drugs link cell lines' drug sensitivity with genotype data which has guided to several key findings, including a general understanding of the basic phenomenon of drug sensitivity or resistance in cancer (***Shoemaker 2006; Weinstein 2004***). However, the NCI-60 database, although a good starting point for developing predictive models, is limited in its use because the panel contains only 60 cell lines.

By contrast, our study focused on the GDSC database, which annotates a comprehensive landscape of drug response data of around thousand (~1000) of human cancer cell lines from different tissue types for 265 anticancer drugs. Above all, the cancer cell lines involved in GDSC for drug screening are genomically and transcriptomically well characterized as a part of the COSMIC cell line project (CCLP, https://cancer.sanger.ac.uk). These resources provided

a platform for the new development of significant molecular biomarkers when it is used in conjunction with powerful analytical tools to deal with the high-dimensional and complex nature of cancer data. And these tools have the potential to link the drug sensitivity of cancer cell lines to their genomic feature.

Furthermore, several computational regression approaches have been designed to predict the sensitivity (IC50) of cancer cell lines toward the screened anti-cancer drugs (***Ahmadi Moughari & Eslahchi 2021***).

GDSC holds a large amount of drug sensitivity data of human pan-cancer cell lines. These data from the GDSC database facilitate the identification of novel biomarkers of drug response by linking the detailed genomic information of cell lines. The GDSC database was built to help researchers for a better understanding of the molecular features which influences drug efficacy in cancer cells and can empower the plan of advanced strategy for cancer treatment. This website is created to give direct access to browse the database and to provide easily interpretable summaries of data and analyses by using interactive graphical interfaces (***Yang et al. 2013***).

3.2 Materials and Methods

3.2.1 Cancer cell lines and drug data acquisition from the GDSC database

GDSC database (***https://www.cancerrxgene.org***) is one of the biggest publicly available open sources for detailed information on thousands of cancer cell lines that are drug sensitive with their molecular markers of drug response from various tissues along with gene expression, and copy number variation. Currently, this database carries data of around 2,12,774 drug dose-response toward drug sensitivity, describing 265 drugs screened against almost thousands of cancer cell lines originating from primary cells of pan-cancer tissues, which includes; CNS (58 cell lines), lung (179 cell lines), skin (62 cell lines), breast (52 cell lines) and hematopoietic & lymphoid tissues (175 cell lines) (**Fig. 1A)** (***Yang et al. 2012; Iorio et al. 2016*****).** These screened

anticancer compounds include clinical used drugs (n=48), pre-clinical drugs (n=76), and experimental compounds (n=141). These 265 compounds are targeted agents (n = 242) and cytotoxic drugs (n = 19) targeting a wide range of biomarkers and 20 key biological and cellular pathways such as protein kinases, transcription regulation, apoptosis, DNA repair, and cellular processes in cancer biology (**Fig. 1B**).

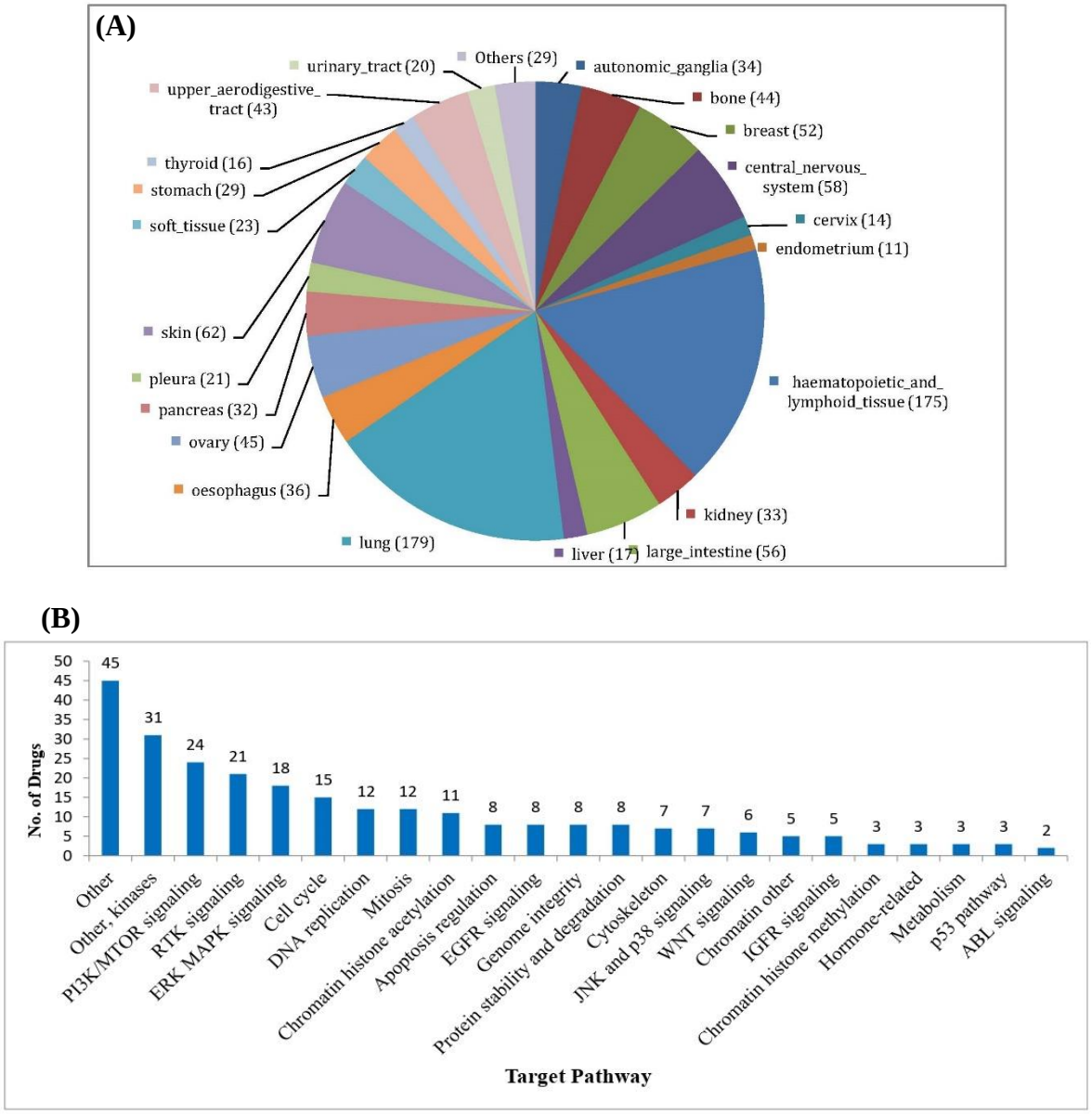


Figure 1: Cancer cell lines and drugs screened against cell lines. (A) Classification of cancer cell lines derived from different tissue types. (B) Anticancer drugs (265 drugs) are used in screening categories based on their therapeutic targets, and role in biological and cellular pathways/functions. A single drug may target multiple molecules.

3.2.2 NRAS mutant cancer drug sensitivity data acquisition from the GDSC

The GDSC database contains a huge amount of genomics and drug sensitivity datasets for NRAS mutant cancer cell lines. To reveal the drug-gene interactions for drug sensitivity and resistance in cancer cell lines harbouring *NRAS* gene mutation, an analysis of variance (ANOVA) test has been performed using drug IC_{50} value. The ANOVA analysis between *NRAS*-mutant vs. *NRAS*-wild type cancer cell lines revealed 12 drugs which were significantly associated (threshold p<0.001) with drug sensitivity or resistance **(Fig. 2)** and were enlisted (**Table 1**). In the cell lines with NRAS mutation, treatment with BRAF, MEK1/2, MAP4K2 and TAK inhibitors (*p values* = $3.38x10^{-4}$ for PLX4720, *p*=$3.04x10^{-10}$ for PD0325901, *p*=$1.55x10^{-5}$ for NG-25 and *p*=$1.05x10^{-5}$ for TL-1-85), respectively significantly attenuated cell viability. However, cancer cell lines harbouring NRAS mutation were significantly resistant to Foretinib, a MET inhibitor (*p-value* =$2.61x10^{-4}$) and also were resistant to Cabozantinib (*p-value* =$1.61x10^{-4}$) and Ponatinib (*p-value* =$2.99x10^{-5}$), where these two drugs are known to multiple target inhibitors.

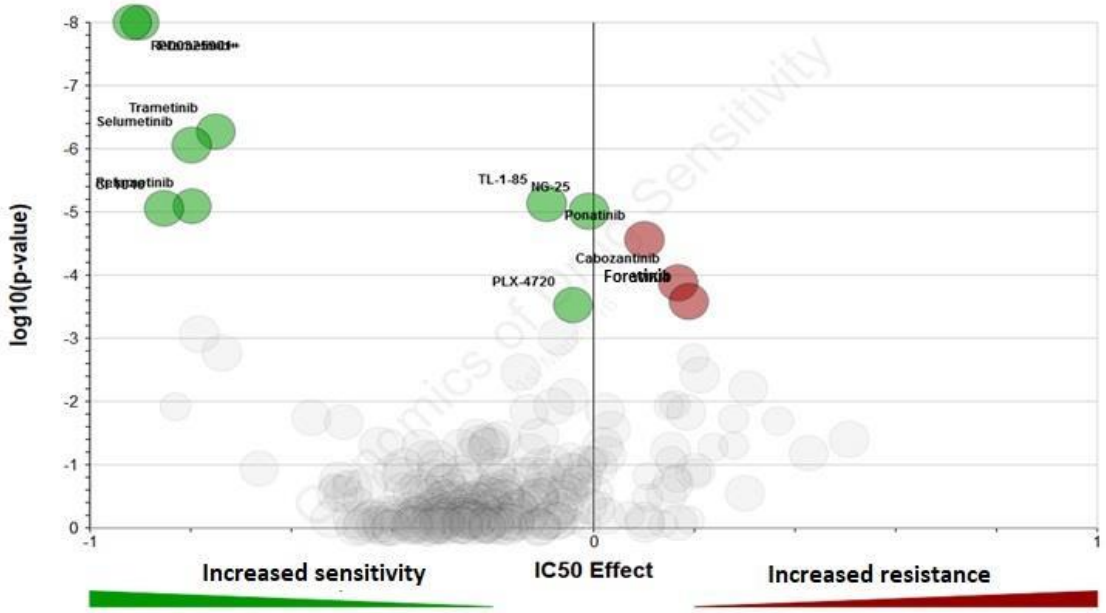


Figure 2: Volcano plot of ANOVA analysis result retrieved from GDSC database. Each circle in the volcano plot represents gene-drug interaction, where the green circle indicates drug sensitivity &

the red circle indicates drug resistance. The position of the circle shows how significant the interaction is, and the circle size is proportional to the number of cell lines altered. (https://www.cancerrxgene.org)

Table 1:- **ANOVA analysis result from GDSC.** Compounds with their targets showing effect size and number of altered cell lines against a target-specific drug. (https://www.cancerrxgene.org)

Sr. no	Drug	Drug Target	Effect size	P-value	No. of altered cell lines
1	PD0325901	MEK1, MEK2	-0.901	1.51E-10	54
2	RDEA119	MEK1, MEK2	-0.916	4.75E-10	54
3	Trametinib	MEK1, MEK2	-0.75	5.36E-07	55
4	Selumetinib	MEK1, MEK2	-0.798	8.74E-07	57
5	TL-1-85	TAK	-0.0932	7.33E-06	58
6	RDEA119	MEK1, MEK2	-0.797	8.08E-06	53
7	CI-1040	MEK1, MEK2	-0.853	8.77E-06	56
8	NG-25	TAK1, MAP4K2	-0.00937	9.73E-06	58
9	Ponatinib	ABL, PDGFRA, VEGFR2, FGFR1, SRC, TIE2, FLT3	0.101	2.74E-05	58
10	Cabozantinib	VEGFR, MET, RET, KIT, FLT1, FLT3, FLT4, TIE2,AXL	0.168	0.000132	58
11	Foretinib	MET	0.189	0.000261	57
12	PLX-4720	BRAF	-0.041	0.000297	55

3.2.3 Drug sensitivity (IC50) data analysis

Drug IC50 (Inhibitory concentration) values for 265 anticancer drugs that are frequently used to assess drug efficacy are available in the GDSC database. We chose only *NRAS*-mutant cancer cell lines that were present and responsive in the case of all 10 drugs. Drug-sensitivity (log normalized IC50) data was downloaded for these selected 10 drugs across the 41 pan-cancer cell lines harbouring *NRAS* mutation. Due to contradictory drug LN_IC_{50} values for the drug RDEA119 deposited twice in GDSC database, it was eliminated from our downstream analysis. From the GDSC database it was suggested that the cancer cell lines regarded as a drug-sensitive having LN_IC_{50} value smaller than the maximum concentration of a drug, and cell lines with value greater than the maximum concentration of a drug are regarded to be drug-resistant. We used a hierarchical clustering module in the online tool GenePattern v11 to generate clustered heatmap (unsupervised clustering) with distance measure uncentered

correlation and clustering method pair-wise average linkage using LN_IC_{50} (cancer cell lines) value. TreeView version 1.1 was used to visualize it.

Further, to correlate drug sensitivity and resistance of NRAS-mutant cancer cell line with that of cancer tissue for the same selected drugs, we were able to collect the predicted drug IC_{50} value for 8 drugs (Foretinib, Ponatinib, Selumetinib, Trametinib, PD-0325901, PLX4720, TL-1-85 and CI-1040) from the cancerRxTissue database, for NRAS mutation harbouring cancer tissue samples. The predicted drug sensitivity (IC_{50} value) data for two drugs (Cabozantinib and NG-25) were not available in the database, and we found the NRAS mutation information from TCGA database. Using the above method as used for cancer cell lines, we generated a clustered heatmap for cancer tissue also.

3.3 Results

3.3.1 Identification of pan-cancer drug-sensitive and -resistant NRAS mutant cell lines

To distinguish individual drug-sensitive and resistant cancer cell lines for each drug, we studied and analyzed the drug response of all present drugs retrieved from GDSC. We observed that 41 pan-cancer cell lines harbouring NRAS mutations were commonly responsive to 10 drugs [Selumetinib, PD-0325901, TL-1-85, Trametinib, NG-25, PLX4720, CI-1040, Foretinib, Xl-184 (Cabozantinib) and AP-24534 (Ponatinib)]. We performed uncentered hierarchical clustering by using the log normalized drug IC_{50} values of these 41 pan-cancer cell lines harbouring NRAS-mutation to distinguish the drug-sensitive and resistant cell lines and generated a clustered heatmap. The heatmap color represented dose response in terms of the drug-sensitivity and resistant (IC_{50} value) cell lines to a particular drug. The normalized IC_{50} value of cancer cell lines greater than zero were selected as drug-resistant cell lines, while LN_IC_{50} value of cell lines less than zero were attributed as sensitive cell lines to the drug **(*Jianting et al. 2015*) (Fig. 3A**). Each rows in the heatmap indicates the IC_{50} score for a screened compound and each column represents cancer cell lines in the generated heatmap.

From the heatmap, highly drug-sensitive (IC_{50}<-1) and resistant (IC_{50}>1) NRAS mutant cancer cell lines were selected based on IC_{50} value and color intensity (enlisted in **Table 2**). In case of four drugs (Cabozantinib, NG-25, TL-1-85, and PLX4720) there were no individual drug-sensitive cell lines observed, on the other hand, we found only one drug-resistant cell line for PD-0325901 as shown in the heatmap. All these cancer cell lines employed in GDSC were originated from different tissue types of primary tumor and classified at the TCGA matching label **(Table 3).**

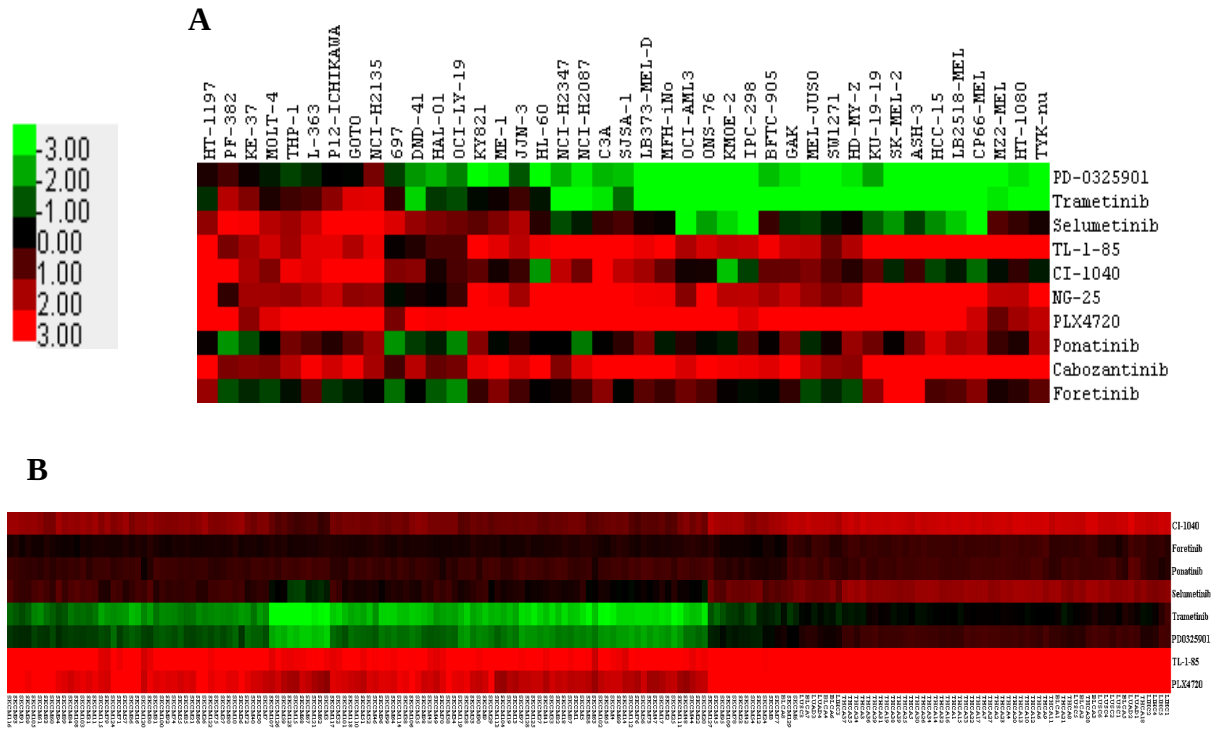


Figure 3: - **Clustered heatmap for drug dose-response in cell lines and cancer tissues**. Rows are drugs and columns are cell lines/cancer tissue sample. Drug-sensitive cell lines/cancer tissue sample are shown in green and drug-resistant cell lines/cancer tissue sample are shown in red color. (A) Heatmap for cancer cell lines, (B) heatmap for cancer tissue sample.

Table 2: Number of drug-sensitive and -resistant cancer cell lines identified by normalized IC_{50} score for 10 drugs.

Sr No.	Drug name	No. of drug-resistant cell lines	No. of drug-sensitive cell lines
1	Selumetinib	16	10
2	CI-1040	21	3
3	PD-0325901	1	31
4	Trametinib	5	24
5	TL-1-85	39	0
6	NG-25	36	0
7	Cabozantinib	40	0
8	PLX4720	41	0
9	Foretinib	12	3
10	Ponatinib	14	4

Table 3: Names of 41 cell lines studied similar to cancer types as identified from TCGA.

TCGA Classification	Cancer cell lines
ALL	P12-ICHIKAWA, DND-41, KE-37, MOLT-4, PF-382, HAL-01.
BLCA	HT-1197, KU-19-19, BFTC-905,
DLBC	OCI-LY-19
LIHC	C3A
LUAD	NCI-H2347, NCI-H2087.
LUSC	HCC-15
LAML	THP-1, ME-1, KY821, OCI-AML3, HL-60, KMOE-2
MB	ONS-76
MM	L-363, JJN-3
NB	GOTO
SCLC	SW1271
THCA	ASH-3
SKCM	IPC-298, LB373-MEL-D, GAK, MEL-JUSO, LB2518-MEL, CP66-MEL, SK-MEL-2, MZ2-MEL
Unclassified	MFH-ino, SJSA-1, HT-1080, TYK-nu, NCI-H2135, HD-MY-Z, 697

3.3.2 Correlation and validation of cell lines with cancer tissue drug-sensitivity status

Further, we wanted to investigate if there are any similarities in the pattern of drug-sensitivity and resistance in cell lines and cancer tissue samples. From the heatmap **(Fig. 3B)** of cancer tissue drug response, we have observed that almost all NRAS-mutant cancer tissue were showing resistance to drugs (CI-1040, Foretinib, Ponatinib, Selumetinib), while few of skin cutaneous melanoma (SKCM) cancer tissue were sensitive to drug Selumetinib. Most of the cancer tissue samples from SKCM were sensitive to the drug Trametinib and other cancer types were less responsive to the Trametinib. Correlating drug response of cancer cell lines with cancer tissue, we have observed that cancer cell lines from BLCA, LUSC, SKCM, and THCA cancer tissue were highly resistant and from LUAD cancer tissues were highly sensitive to Ponatinib, while all cancer tissues were resistant to Ponatinib. Cell lines from cancer BLCA, LIHC, SKCM, THCA and cancer tissue as well, were resistant to Foretinib. SKCM cancer cell lines and cancer tissue both were sensitive to Trametinib, while cell lines from BLCA, LIHC, LUAD and LUSC were sensitive to Trametinib, but cancer tissue samples were very less responsive. In the case of drug CI-1040 some cell lines from SKCM were resistant, as well as some were sensitive, while all SKCM cancer tissue samples were resistant. However, cell lines and cancer tissue samples from LIHC and LUAD were both resistant to CI-1040. Further, in the case of Selumetinib, some BLCA cell lines were resistant, while cell lines from BLCA, LUAD, LUSC, and SKCM were sensitive to Selumetinib. Whereas all cancer tissue samples from these cancers were resistant to Selumetinib except a few SKCM samples, which were observed as sensitive to Selumetinib. A similar correlation pattern between cell line and cancer tissue would depict a similarity in gene expression pattern that might be causing drug-sensitivity or drug-resistance. Cancer tissue samples for NRAS-mutant; LAML, ALL, MM, MB, NB, DLBC, and SCLC cancer types were not available in TCGA.

3.4 References

1. Tang YC, Gottlieb A. (2021) Explainable drug sensitivity prediction through cancer pathway enrichment. *Sci Rep.* 11(1):3128. doi: 10.1038/s41598-021-82612-7. PMID: 33542382; PMCID: PMC7862690.
2. Mansoori B, Mohammadi A, Davudian S, Shirjang S, Baradaran B. (2017) The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv Pharm Bull.* 7(3):339-348. doi: 10.15171/apb.2017.041. PMID: 29071215; PMCID: PMC5651054.
3. Tyner JW, Haderk F, Kumaraswamy A, (2022) Understanding Drug Sensitivity and Tackling Resistance in Cancer. *Cancer Res.* 82(8):1448-1460. doi: 10.1158/0008-5472.CAN-21-3695. PMID: 35195258; PMCID: PMC9018544.
4. Popova, Anna Alexandrovna; Levkin, Pavel Andreevich (2020). Precision Medicine in Oncology: In Vitro Drug Sensitivity and Resistance Test (DSRT) for Selection of Personalized Anticancer Therapy. *Advanced Therapeutics*, 1900100–. doi:10.1002/adtp.201900100.
5. Ahmadi Moughari, F., & Eslahchi, C. (2021). A computational method for drug sensitivity prediction of cancer cell lines based on various molecular information. *PloS one*, *16*(4), e0250620. https://doi.org/10.1371/journal.pone.0250620.
6. Wang L, Li X, Zhang L, Gao Q. (2017) Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer*. 17(1):513. doi: 10.1186/s12885-017-3500-5. PMID: 28768489; PMCID: PMC5541434.
7. Pouryahya M, Oh JH, Mathews JC, Belkhatir Z, Moosmüller C, Deasy JO, Tannenbaum AR. (2022) Pan-Cancer Prediction of Cell-Line Drug Sensitivity Using Network-Based Methods. Int J Mol Sci. 23(3):1074. doi: 10.3390/ijms23031074. PMID: 35163005; PMCID: PMC8835038.
8. Shoemaker RH. (2006) The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer. 6(10):813-23. doi: 10.1038/nrc1951. PMID: 16990858.
9. Iorio F, Knijnenburg TA, Vis DJ, et al. (2016) A Landscape of Pharmacogenomic Interactions in Cancer. Cell. 166(3):740-754. doi: 10.1016/j.cell.2016.06.017. PMID: 27397505; PMCID: PMC4967469.
10. Barretina J, Caponigro G, Stransky N, et al. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 483(7391):603-7. doi: 10.1038/nature11003. 492(7428):290. 565(7738):E5-E6. PMID: 22460905; PMCID: PMC3320027.
11. Yang W, Soares J, Greninger P, et al. (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* D955-61. doi: 10.1093/nar/gks1111. PMID: 23180760; PMCID: PMC3531057.
12. Garnett MJ, Edelman EJ, Heidorn SJ, et al. (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature.* 483(7391):570-5. doi: 10.1038/nature11005. PMID: 22460902; PMCID: PMC3349233.
13. Chabner BA. (2016) NCI-60 Cell Line Screening: A Radical Departure in its Time. *J Natl Cancer Inst.* 108(5):djv388. doi: 10.1093/jnci/djv388. PMID: 26755050.
14. Weinstein JN. (2004) Integromic analysis of the NCI-60 cancer cell lines. Breast Dis. 19:11-22. doi: 10.3233/bd-2004-19103. PMID: 15687693.

Chapter-4

Objective-2: Identification of differentially expressed genes (DEGs) between identified drug-sensitive and -resistant cancer cell lines to serve as possible biomarkers

4.1 Introduction

Protein-coding genes are defined as gene sequences which are transcribed into mRNA and later on translated into a protein. These sequences share a tiny fraction, close to 2%, of the entire human genome. The basic structure of protein-coding genes includes a promoter followed by a coding sequence that codes for mRNA, which is then translated into a protein and eventually, all of these are followed by a terminator which specifies the end of the mRNA transcript **(Fig. 1).**

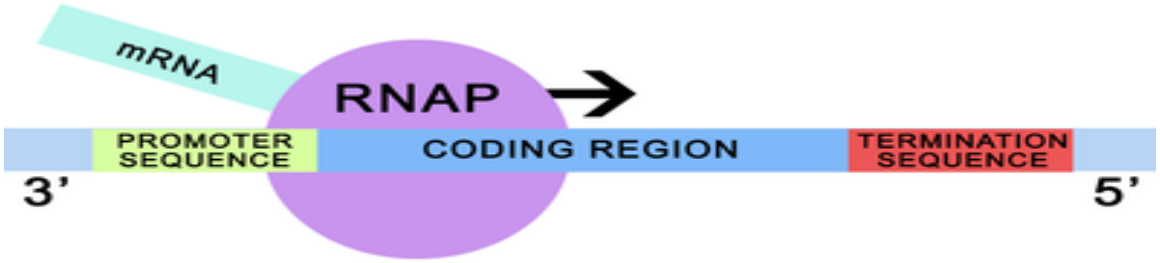


Figure 1: Schematic representation of promoter sequence, coding region and termination sequence on protein-coding gene sequence (*Modified from: https://en.wikipedia.org/wiki/Coding_region*).

Carcinogenesis, as well as chemoresistance, are driven by the accumulation of scores of alterations affecting the structure and function of the human genome where in this process, both genetic and epigenetic changes are equally important. Genomic defects play a critical role in cancer by influencing cell proliferation, growth and survival through the direct or indirect alterations of gene expression, a variety of protein activities, and molecular signaling pathways (***Hanahan & Weinberg 2011; Watson et al. 2013)***. There are several computational strategies have been developed to identify so-called driver mutations using a diverse range of somatic mutations characteristics, which includes generative conserved mutation sites in multiple species (***Reva et al. 2011***), and the influence of mutations on transcriptome ***(Hou & Ma 2014***), among others. Generally, insertion or deletion of DNA segments during biological processes

such as DNA replication, evolution, junctional diversity, and development of immune systems and particularly in cancer development, can lead to mutations. Among all, some mutations introduce premature stop codons, which can bring down the expression of a gene's corresponding mRNA transcript, and some affect protein activity by changing the sequence of the amino acid residues in encoded protein (***Jia & Zhao 2017***). Cancer-related gene mutations result in altered expression of particular genes/proteins and often produce a distinct phenotype in different cancers. Alteration in the genes' expression that encodes for a protein may encourage the initiation or progression of a tumor, as oncogenes do or may suppress its growth, as do tumor suppressor genes. Traditionally, only mutated genes/proteins are considered as a candidate for cancer-related genes/proteins. However, the relationship between mutated genes and cancer phenotypes is not always clear-cut since cancer phenotypes result from abnormal gene expression rather than direct mutations in DNA (***Sager 1997)***. In cancer, some genes are identified as driver genes. These genes are oncogenes, tumor suppressors, proto-oncogenes, and anti-apoptotic genes that may be involved in cancer development and chemoresistance.

For instance, cancer genes such as *myc* (oncogene) and *p53 (tumor suppressor)* encode transcription factors that transcriptionally regulate the expression of several downstream genes. (***Dang 2012; Sullivan et al. 2018)***. Since the function of Myc protein is crucial for the maintenance of tumors, it is possible that tumors might resist treatments by engaging a variety of resistance mechanisms (***Llombart & Mansour 2022***). Cancer cells harbouring p53 mutations are commonly characterized by a high rate of metastasis as well as genomic instability (***Liu et al. 2010***). This characteristic has important implications for the treatment of many cancers and also has been linked to drug resistance and mitogenic defects (***Hientz et al. 2017).***

Similarly, Mutations of the *RAS* proto-oncogenes are one of the widely known common genetic alterations observed in a variety of human cancers. They are encoded by three genes that are expressed ubiquitously: *HRAS, KRAS* and *NRAS* (***Prior et al. 2012***). These proteins are

GTPases that switch various pathways on and off, controlling proliferation and cell survival and also influencing drug resistance in cancer by altering gene expression. Among the RAS family members, NRAS is the second most mutated protein after KRAS mutation in human cancers (***Prior et al. 2012***). Most of these mutations involve codons 12, 13, and 61 and the mutation status is useful in guiding therapy for certain cancers (***Muñoz-Couselo et al. 2017***). Genetic mutation in the *NRAS* gene/protein is extensively associated with the biological or cellular mechanisms that involved in drug resistance (***Le et al. 2013; Nazarian et al. 2010***). Apart from these mutations, the overexpression of certain tyrosine kinase receptors, such as EGFR and hepatocyte growth factor receptor (HGFR/c-Met), also contributes to RAS hyperactivation (***Kawauchi et al. 2018***).

Genetic changes can occur not only in the genes but also in epigenetic regulators, which are involved in the regulation of histone modifications and DNA methylation to modulate the chromatinization of chromosomes (***Huether et al. 2014; Veitia et al. 2017***). This can affect gene expression by affecting metabolic factors, as well as genetic and epigenetic factors. Abnormalities or changes in these factors lead to genomic instability and abnormal gene expression in drug-resistant cancer. Biomarker genes are examples of possible drug targets, and they have been identified in individual cancer types. These genes have alterations in cancer at the genomic, transcript, and protein levels. They are also linked to drug resistance and may serve as potential drug targets for cancer treatment.

Microarray technology has been widely used to analyze gene expression and identify genetic variations such as mutation and single nucleotide polymorphism (SNP). Specifically, massive-amount microarray gene expression data analysis enables researchers to identify significant patterns in thousands of genes and analyze simultaneous changes in those genes. Because genome-wide expression profile data analysis in drug-resistant pan-cancer cell lines has not yet been done. We have used the latest Affymetrix human genome U219 array data for our analysis

of global gene expression into an analytical model to improve the potential to identify predictive biomarkers of drug response.

Integrative and comprehensive GDSC data analysis has identified and characterized (or profiled) molecular subtypes, possible driver biological processes, and pathways in mutant NRAS-harbouring drug-resistant pan-cancer systems.

4.2 Materials and Methods

4.2.1 Gene expression data collection from GDSC

In our study, we have used gene expression profile data generated using high-throughput technique microarray, downloaded from GDSC. Basal transcriptional profile raw data (E-MTAB-3610) deposited to GDSC for the 1000 cell lines generated using the latest mRNA expression array Affymetrix human genome U219 along with processed gene expression data. The gene expression data of around 17417 genes were normalized using a robust multi-array average (RMA) algorithm and deposited in GDSC. Gene expression data were taken and analyzed for drug-sensitive and resistant cell lines from GDSC for five drugs.

4.2.2 Significant differential gene expression analysis

To check the association between drug sensitivity and resistance of cell lines and gene expression, combined datasets were examined and statistical tests were performed for significant differential gene expression. The basal gene expression profile data were downloaded from the GDSC database. It was filtered to exclude the expression values that were missing gene names from the column. An unpaired t-test was performed using a cloud based tool Multi Experiment Viewer (MeV) version 4.9.0 with threshold cut-off p-value <0.05 with unequal group variance (Welch approximation) between two groups (drug-sensitive and -resistant cancer cell lines). We studied and analysed microarray gene expression profile data of 17417 genes in each cancer cell lines with NRAS mutation. A total of 68 drug-resistant and

44 drug-sensitive NRAS mutant cancer cell lines from our drug sensitivity heatmap, belonging to 5 out of 10 drugs were analysed. We were not able to classify cancer cell lines clearly into resistant/sensitive classes in the case of the rest 5 of these drugs. A volcano plot to visualize up-and down-regulated genes (identified differentially expressed genes, DEGs) between two groups was generated using the R package "ggplot2" with double filtration cut-off p-value <0.05 and logFC>2.

4.2.3 Heatmap to discriminate up- and down-regulated DEGs in drug-sensitive and resistant cell lines

To identify DEGs by their names and visualize their expression pattern in each drug-sensitive and -resistant cell line, we used the module (Comparative Marker Selection) in Gene Pattern version 11 (https://cloud.genepattern.org) to generate a heatmap with default parameters, at 10,000 permutations and to visualized them as a heatmap, we used "Comparative Marker Selection Viewer v9.1.

Further, to identify genes that were differentially expressed across multiple drugs, we imported DEG's list of all five drugs in an online web tool called BioInfoRX (http://apps.bioinforx.com/bxaf6/tools). Then we generated a bubble plot using the R package "ggplot2 & ggpubr" to visualize overlapping DEGs to multiple drugs (minimum for three drugs).

4.2.4 Functional gene enrichment annotation analysis

To investigate the functional implication of identified DEGs in drug-resistant/sensitive cancer cell lines, a gene enrichment analysis was carried out in the context of 5 drugs using a web-accessible bioinformatics tool GeneCodis version 4 to characterize their functions. In order to detect a significant functional enrichment of genes, the threshold hypergeometric p-value <0.05 (Benjamini-adjusted Fisher's exact test p-value) was used by default.

4.3 Results

4.3.1 Differentially expressed gene analysis between drug-sensitive and resistant cancer cell lines

Several studies have revealed that gene expression data is one of the important predictive biomarkers of molecular profile in drug-sensitivity and resistance studies (***Wildey et al., 2014***). We have analysed the basal gene expression data which are log normalized, retrieved from the GDSC database widely associated with pan-cancer drug sensitivity and resistance. We identified several hundreds of differentially expressed genes (DEGs) between drug-sensitive and resistant pan-cancer cell lines (harbouring NRAS-mutation) using Welch's t-test (unequal group variance) for 5 drugs and listed the number of DEGs (threshold $p<0.05$). We did not perform a statistical test for differential expression analysis in case of other 5 drugs because, as seen from **table 1** of chapter 3, all of the chosen cancer cell lines were either uniformly resistant (NG-25, TL-1-85, Cabozantinib, PLX4720) or uniformly sensitive (PD-0325901) to these mentioned drugs. Further, volcano plots were generated for each drug by applying (p value<0.05, log2FC >2) double filtration to statistically validate the results (**Fig. 2A-E).** The number of significantly DEGs varies from 38 DEGs for CI-1040 to 467 DEGs Foretinib, in case of each drug (**Table 1**). Significantly DEGs are shown in volcano plots as top blue dots and represented as down-regulated at the left side and up-regulated genes right side position, in resistant cancer cell lines ($p<0.05$, log2FC>2). As the volcano plot shows,

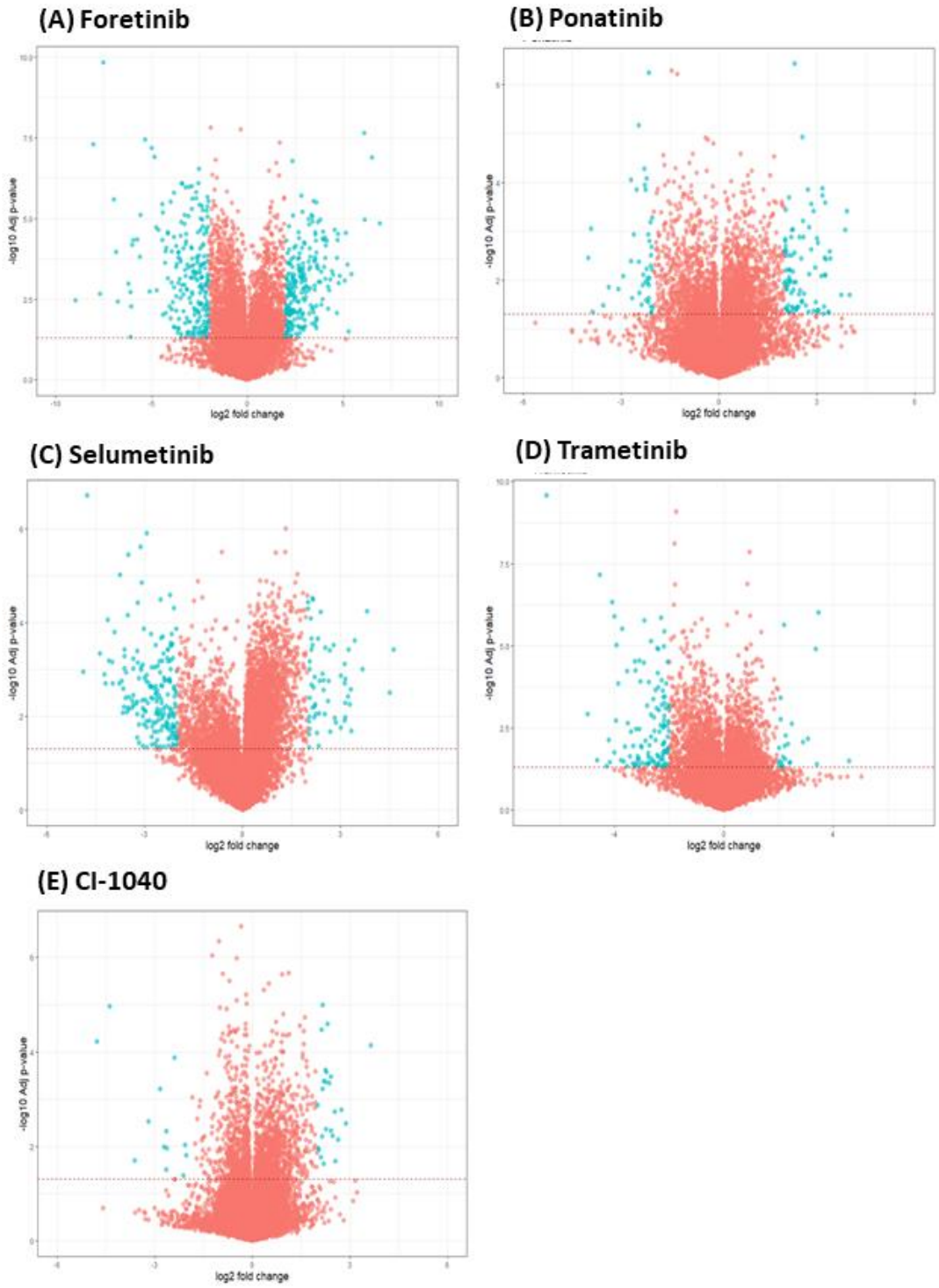


Figure 2: Volcano plot for significantly differentially expressed genes between drug-sensitive and –resistant cancer cell lines. (A to E) blue dots at the top represent significantly differentially expressed genes, bottom dots (red) represent non-significant differentially expressed genes. The x-axis shows fold change in gene expression (magnitude of change, logFC > 2), and the y-axis (p-value) shows statistically significant genes (threshold p-value <0.05) for each drug.

Table 1: Number of significantly DEGs (up- and down-regulated) in drug-sensitive and resistant pan-cancer cell lines.

Sr no.	Drugs	Up-regulated genes in drug-resistant cells	Down-regulated genes in drug-resistant cells	Total no. of DEGs genes
1	Selumetinib	60	189	249
2	CI-1040	25	13	38
3	Trametinib	23	122	145
4	Ponatinib	90	44	134
5	Foretinib	236	231	467

4.3.2 Heatmap of DEGs between drug-sensitive and resistant cancer cell lines

Further, to visualize the pattern of identified DEG expression in individual cancer cell lines and to discriminate between up- and down-regulated DEGs in resistant/sensitive cell lines, a heatmap was generated using a comparative marker selection module in the online web tool GenePattern **(Fig. 3A-E**). From the heatmap, we observed that the results were more or less coinciding with volcano plots. Further, these predictive microarray gene expression data analyses uncovered key DEGs as being strongly associated with either drug resistance or sensitivity.

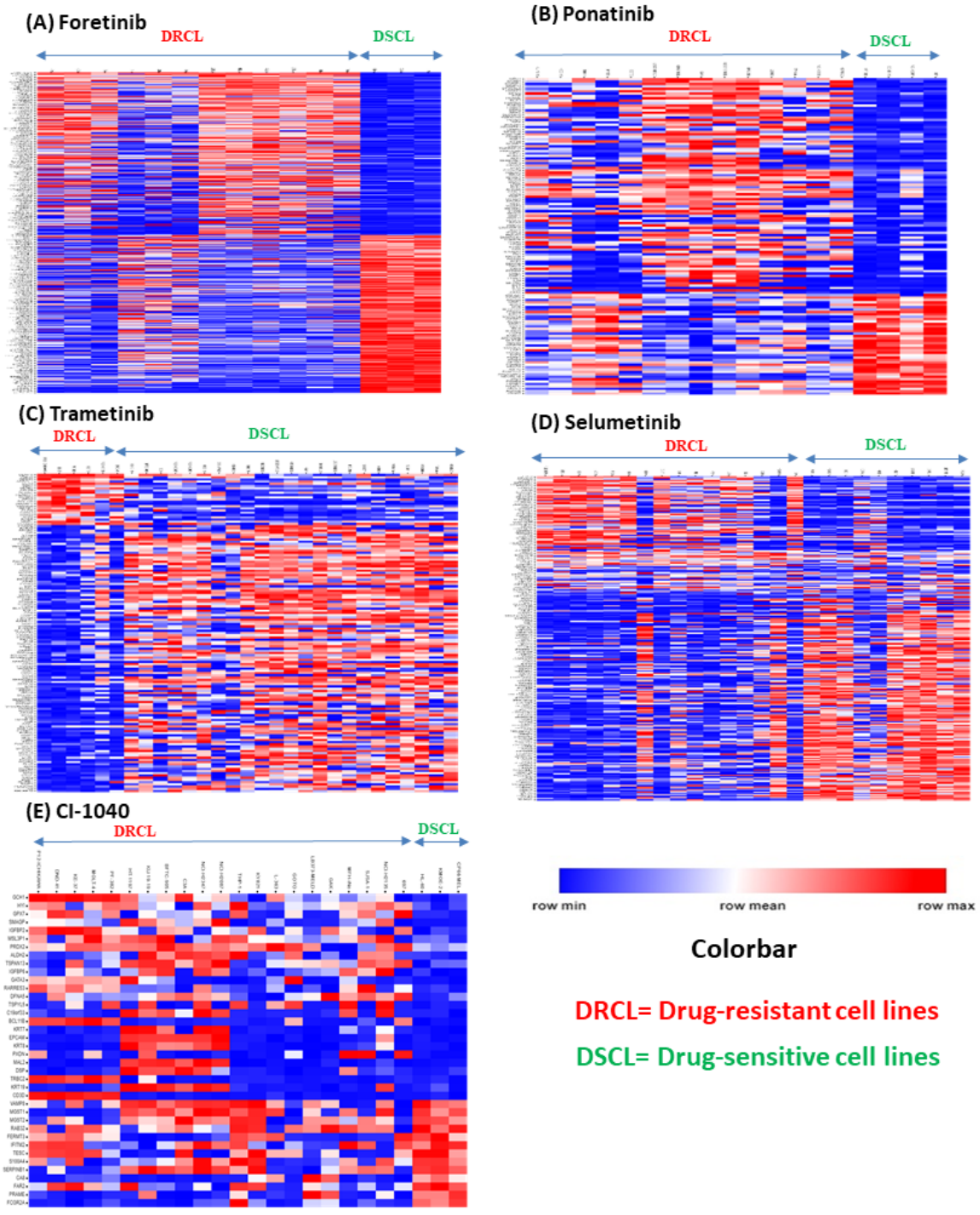


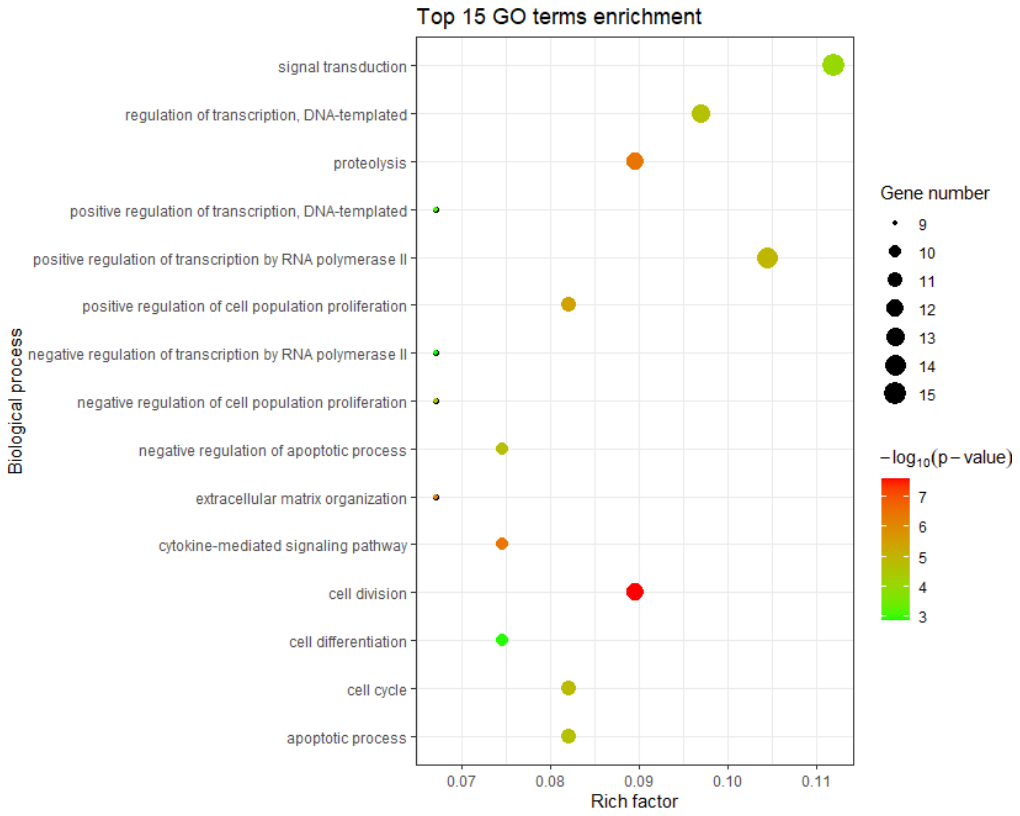
Figure 3: Heatmap of identified DEGs. (A-E) Identified DEGs expression pattern in drug- sensitive and resistant cancer cell lines for five drug. Up-regulated genes represented by Red color, down-regulated genes represented by blue color.

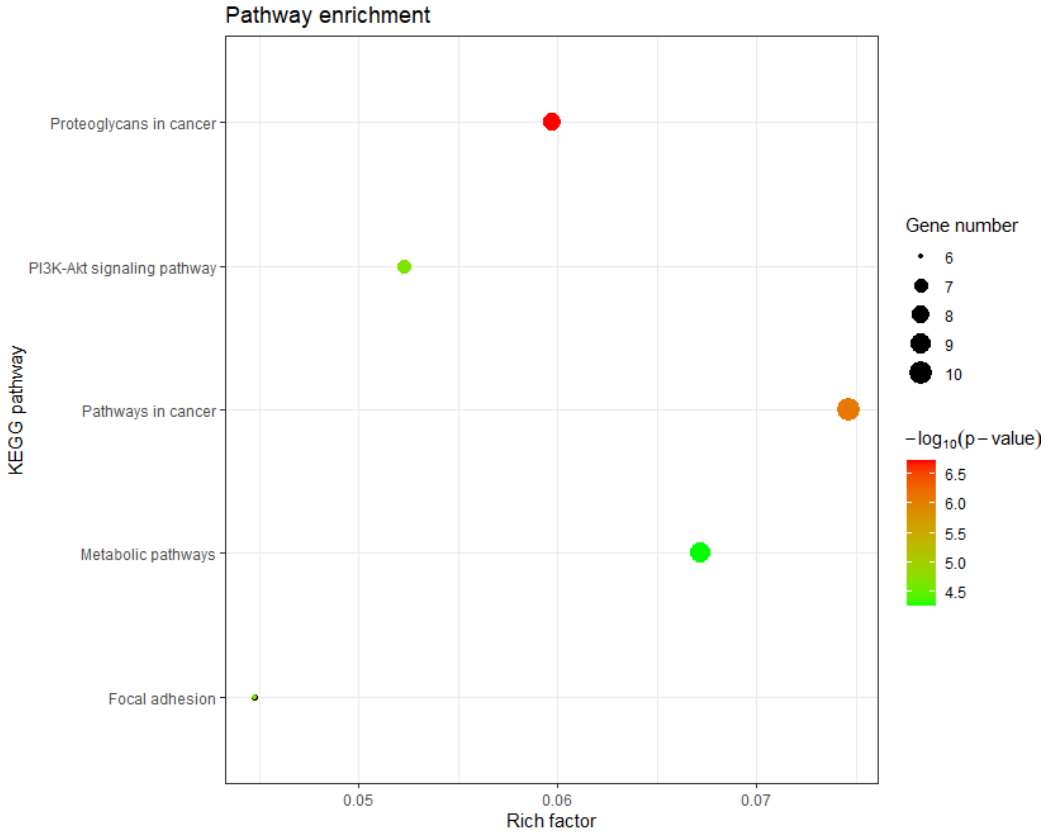
4.3.3 Functional enrichment analysis for gene ontology (GO) and KEGG pathway

The following step is to determine the functional enrichment analyses of the given set of DEGs across multiple drugs, for biological processes and KEGG pathways. To annotate the possible biological processes and the KEGG pathway we used GeneCodis4, which is a web-accessible tool with a default hypergeometric cut-off p-value<0.05. DEGs significantly enriched in GO terms for biological processes for the drug Ponatinib are; GO: 0006508 proteolysis, GO: 0007165- signal transduction, GO: 0008285-cell cycle, GO: 0006915-apoptotic process, and GO: 0006355-regulation of transcription, DNA-dependent and cell division (**Fig. 4A**). In addition, KEGG pathway analyses stipulate that DEGs significantly enriched in, hsa04151:PI3K-Akt signaling pathway, hsa01100: Metabolic pathway, hsa05200: Pathway in cancer and has04510: Focal adhesion (**Fig. 4A**).

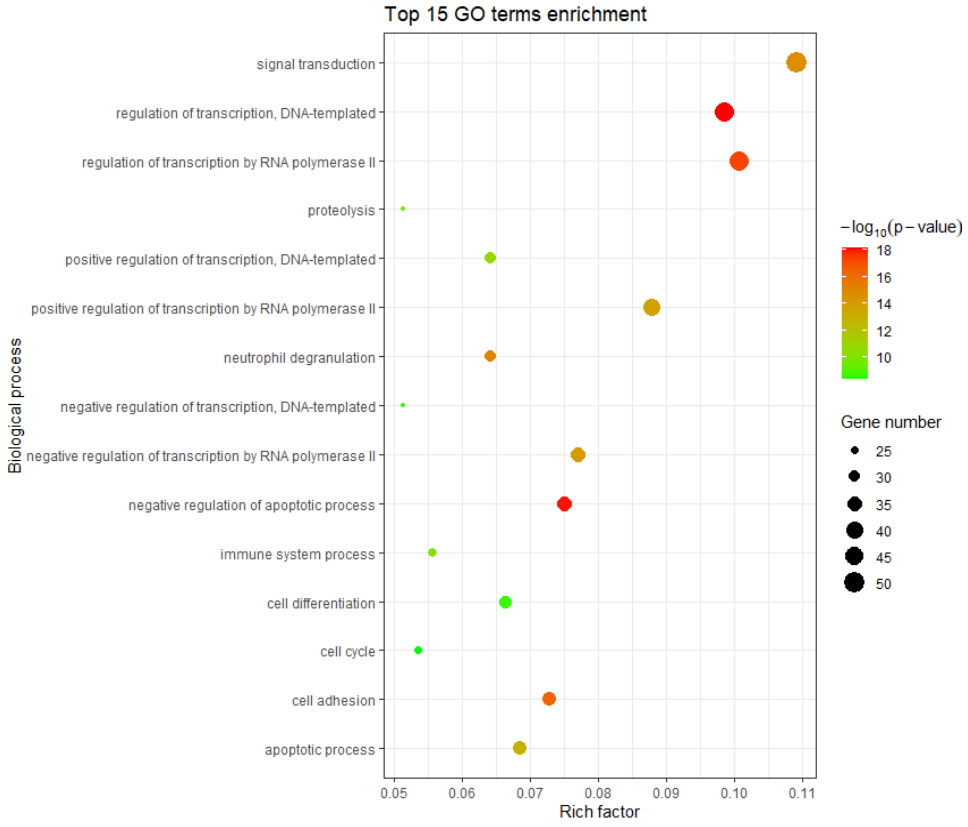
Interestingly, in the case of the other four drugs we also observed that the DEGs were significantly enriched in biological processes and KEGG pathways were more or less similar. GO terms such as GO: 0007165 signal transduction, GO: 0006508 proteolysis, GO: 0006915 apoptotic processes, GO: 0007155 cell adhesion, and KEGG pathways are hsa05205 proteoglycans in cancer, hsa05200: Pathway in cancer, hsa01100: Metabolic pathway and has04510: Focal adhesion, and (**Fig. 4A-E).**

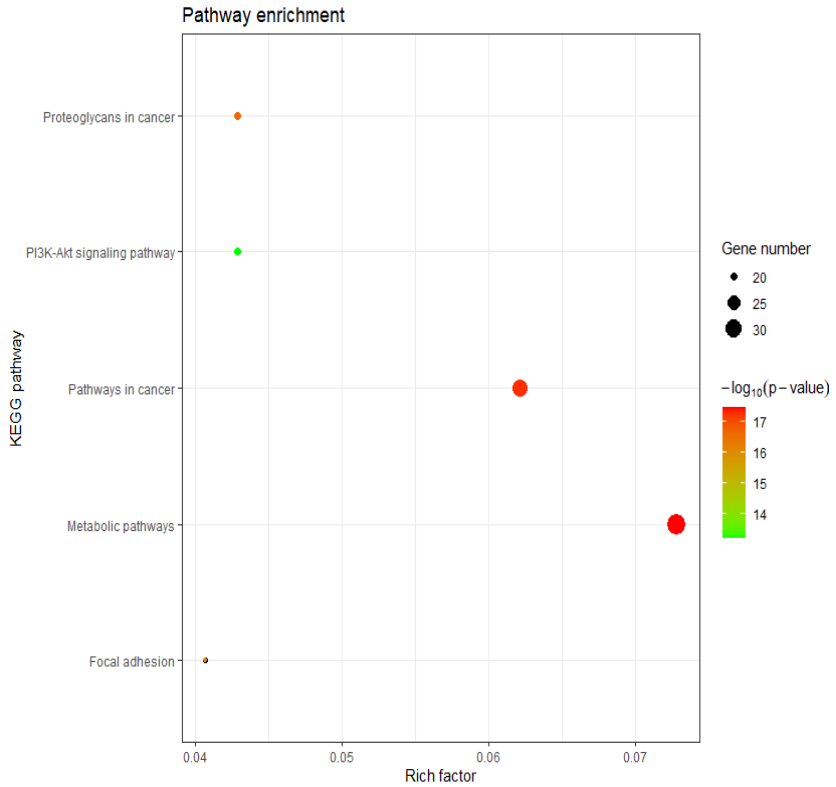
(A) Ponatinib


Top 15 GO terms enrichment
signal transduction
regulation of transcription, DNA-templated
proteolysis
positive regulation of transcription, DNA-templated
positive regulation of transcription by RNA polymerase II
positive regulation of cell population proliferation
negative regulation of transcription by RNA polymerase II
negative regulation of cell population proliferation
negative regulation of apoptotic process
extracellular matrix organization
cytokine-mediated signaling pathway
cell division
cell differentiation
cell cycle
apoptotic process
Biological process
0.07
0.08
0.09
0.10
0.11
Rich factor
Gene number
9
10
11
12
13
14
15
−log10(p−value)
7
6
5
4
3

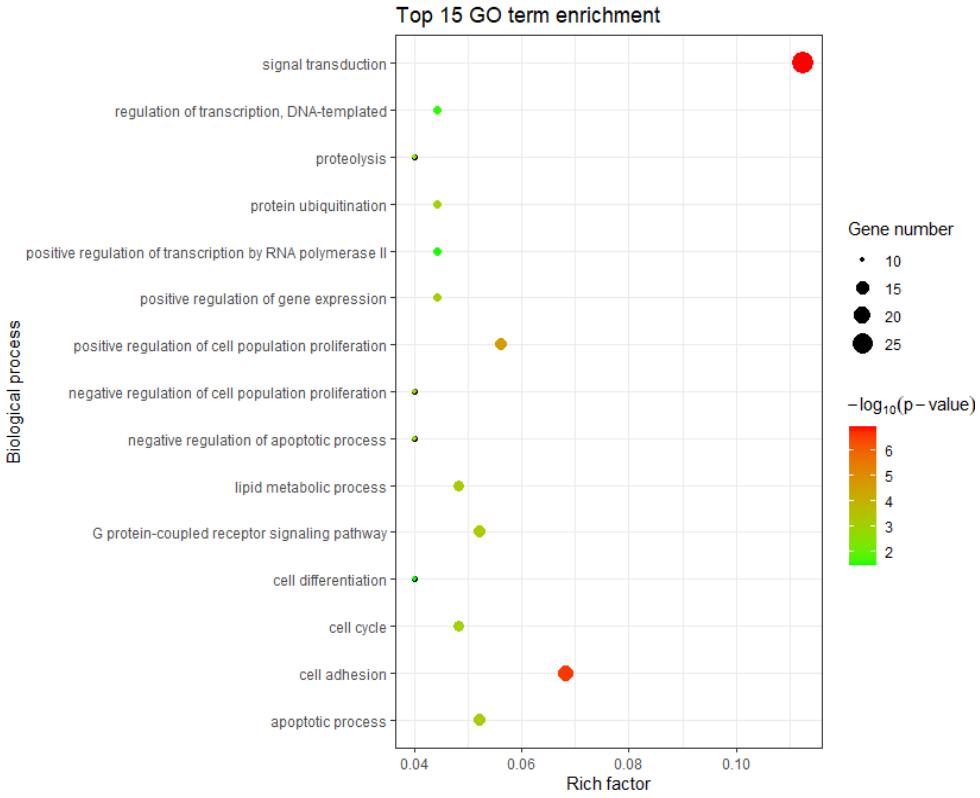

Pathway enrichment
Proteoglycans in cancer
PI3K-Akt signaling pathway
Pathways in cancer
Metabolic pathways
Focal adhesion
KEGG pathway
0.05
0.06
0.07
Rich factor
Gene number
6
7
8
9
10
−log10(p−value)
6.5
6.0
5.5
5.0
4.5

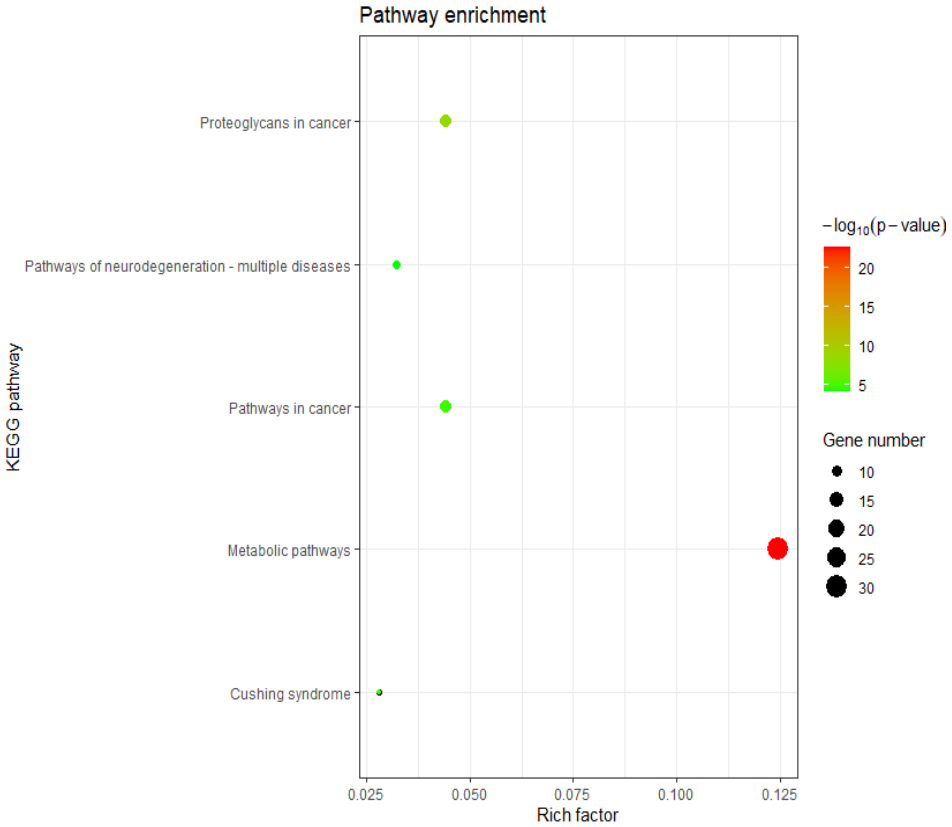
(B) Foretinib


Top 15 GO terms enrichment
signal transduction
regulation of transcription, DNA-templated
regulation of transcription by RNA polymerase II
proteolysis
positive regulation of transcription, DNA-templated
positive regulation of transcription by RNA polymerase II
neutrophil degranulation
negative regulation of transcription, DNA-templated
negative regulation of transcription by RNA polymerase II
negative regulation of apoptotic process
immune system process
cell differentiation
cell cycle
cell adhesion
apoptotic process
Biological process
0.05
0.06
0.07
0.08
0.09
0.10
0.11
Rich factor
−log10(p−value)
18
16
14
12
10
Gene number
25
30
35
40
45
50

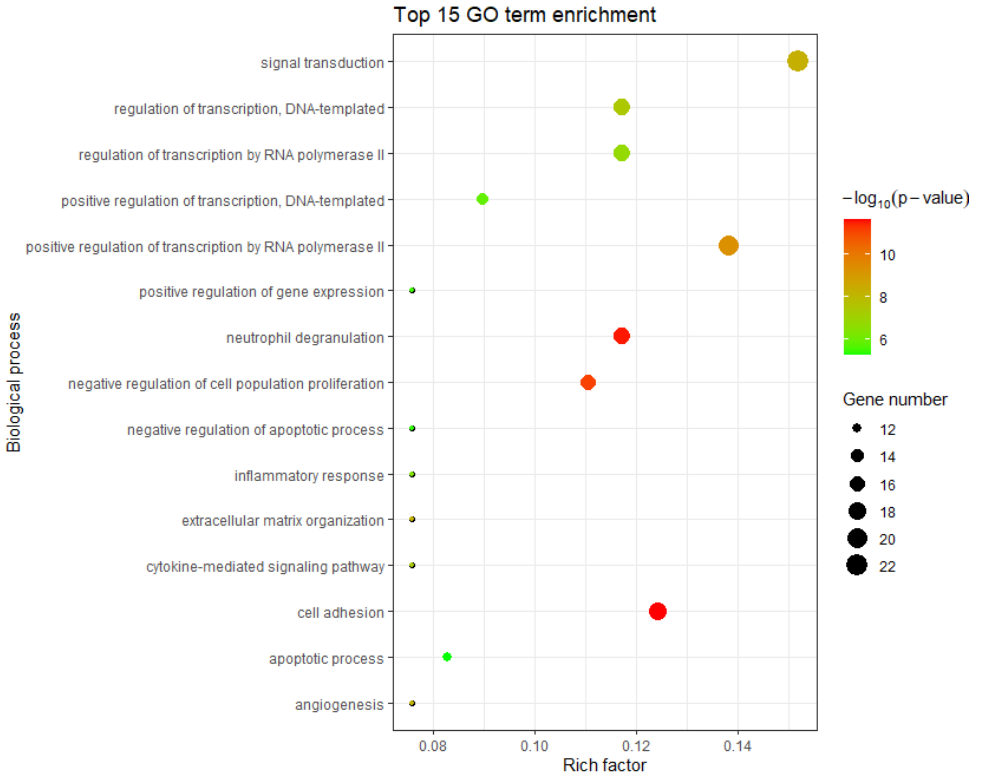

Pathway enrichment
Proteoglycans in cancer
PI3K-Akt signaling pathway
Pathways in cancer
Metabolic pathways
Focal adhesion
KEGG pathway
0.04
0.05
0.06
0.07
Rich factor
Gene number
20
25
30
−log10(p−value)
17
16
15
14

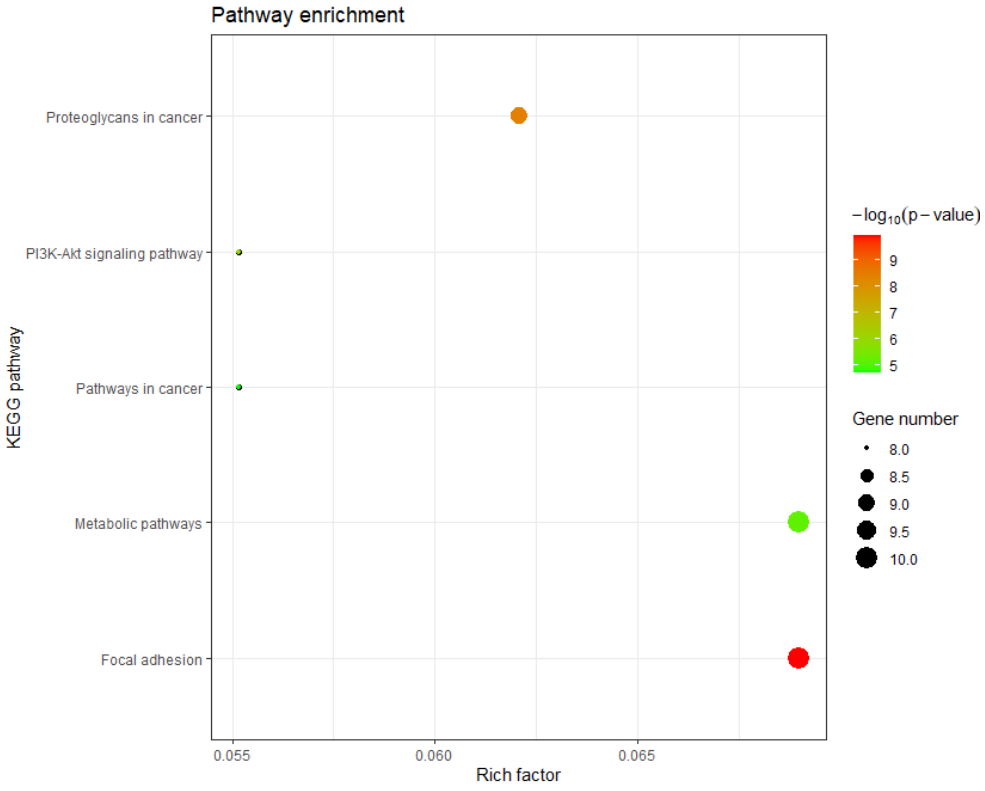
(C) Selumetinib





(D) Trametinib





(E) CI-1040

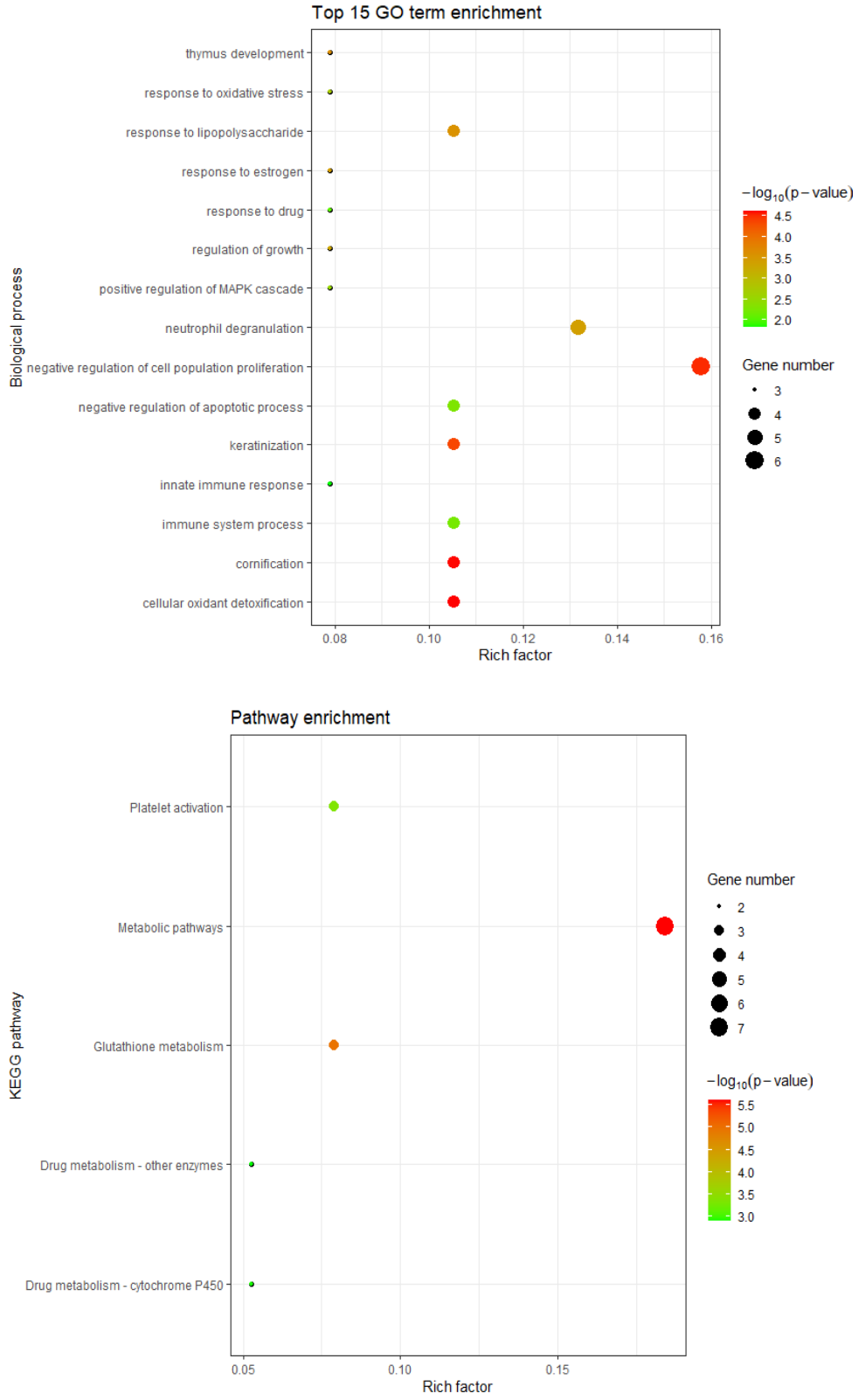


Figure 4: Functional enrichment analysis of DEGs. Balloon plot for Gene Ontology and KEGG pathway analysis of DEGs for five drugs, A) Ponatinib, B) Foretinib, C) Selumetinib, D) Trametinib,

E) CI-1040. Top 15 GO terms enriched for the biological process and top 5 KEGG pathway enrichment at a default hypergeometric p-value cut off < 0.05.

4.3.4 Common DEGs across multiple drugs

Further, we looked for the common DEGs across these selected 5 drugs. Interestingly, we found DEGs for multiple drugs but not for all drugs. Among these DEGs, nine coding genes, were found to be significantly differentially expressed across four drugs, including CD44, FN1, and TIMP3. Similarly, SPARC, SNAI2 and TIMP1, including other 34 genes are observed to be overlapped in the case of three drugs **(Fig. 5).** These genes might be associated with pan-cancer drug sensitivity and resistance.

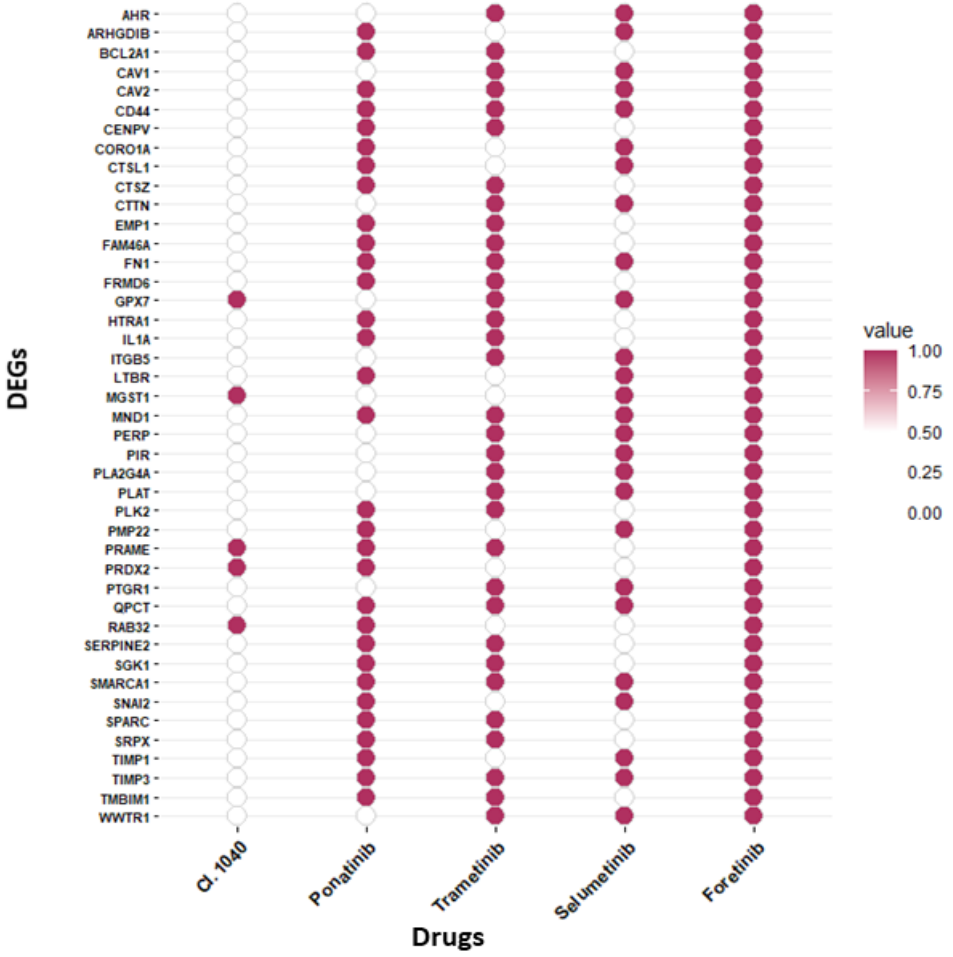


Figure 5: Bubble plot to identify common DEGs across five drugs. Genes are identified as differentially expressed gene (reddish-brown bubble) in the case of multiple drugs.

4.4 References

1. *https://en.wikipedia.org/wiki/Coding_region*. Accessed on august, 2022.
2. Hanahan D, Weinberg RA. (2011) Hallmarks of cancer: the next generation. *Cell.* 144(5):646-74. PMID: 21376230.
3. Watson IR, Takahashi K, Futreal PA, Chin L. (2013) Emerging patterns of somatic mutations in cancer. *Nat Rev Genet.* 14(10):703-18. PMID: 24022702; PMCID: PMC4014352.
4. Reva B, Antipin Y, Sander C. (2011) Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39(17):e118. PMID: 21727090; PMCID: PMC3177186.
5. Hou JP, Ma J. (2014) DawnRank: discovering personalized driver genes in cancer. *Genome Med.* 6(7):56. PMID: 25177370; PMCID: PMC4148527.
6. Jia P, Zhao Z. (2017) Impacts of somatic mutations on gene expression: an association perspective. *Brief Bioinform.* 18(3):413-425. PMID: 27127206; PMCID: PMC5862283.
7. Sager R. (1997) Expression genetics in cancer: shifting the focus from DNA to RNA. *Proc Natl Acad Sci U S A.* 94(3):952-5. doi: 10.1073/pnas.94.3.952. PMID: 9023363; PMCID: PMC19620.
8. Dang CV. (2012) MYC on the path to cancer. *Cell.* 149(1):22-35. PMID: 22464321; PMCID: PMC3345192.
9. Sullivan KD, Galbraith MD, Andrysik Z, Espinosa JM. (2018) Mechanisms of transcriptional regulation by p53. *Cell Death Differ.* 25(1):133-143. PMID: 29125602; PMCID: PMC5729533.
10. Llombart V, Mansour MR. (2022) Therapeutic targeting of "undruggable" MYC. *EBioMedicine.* 75:103756. PMID: 34942444; PMCID: PMC8713111.
11. Liu DP, Song H, Xu Y. (2010) A common gain of function of p53 cancer mutants in inducing genetic instability. *Oncogene.* 29(7):949-56. PMID: 19881536; PMCID: PMC2837937.
12. Hientz K, Mohr A, Bhakta-Guha D, Efferth T. (2017) The role of p53 in cancer drug resistance and targeted chemotherapy. *Oncotarget.* 8(5):8921-8946. PMID: 27888811; PMCID: PMC5352454.
13. Prior IA, Lewis PD, Mattos C. (2012)A comprehensive survey of Ras mutations in cancer. *Cancer Res.* 72(10):2457-67. PMID: 22589270; PMCID: PMC3354961.
14. Muñoz-Couselo E, Adelantado EZ, Ortiz C, García JS, Perez-Garcia J. (2017) *NRAS*-mutant melanoma: current challenges and future prospect. *Onco Targets Ther.* 10:3941-3947. PMID: 28860801; PMCID: PMC5558581.
15. Le K, Blomain ES, Rodeck U, Aplin AE. (2013) Selective RAF inhibitor impairs ERK1/2 phosphorylation and growth in mutant NRAS, vemurafenib-resistant melanoma cells. *Pigment Cell Melanoma Res.* 26(4):509-17. PMID: 23490205; PMCID: PMC3695051.
16. Nazarian R, Shi H, Wang Q, Kong X, Koya RC, Lee H, Chen Z, Lee MK, Attar N, Sazegar H, Chodon T, Nelson SF, McArthur G, Sosman JA, Ribas A, Lo RS. (2010) Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature.* 468(7326):973-7. PMID: 21107323; PMCID: PMC3143360.
17. Kawauchi K, Sugimoto W, Yasui T, et al. (2018) An anionic phthalocyanine decreases NRAS expression by breaking down its RNA G-quadruplex. *Nat Commun.* 9(1):2271. PMID: 29891945; PMCID: PMC5995912.

18. Huether R, Dong L, Chen X, et al. (2014) The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun*. 5:3630. PMID: 24710217; PMCID: PMC4119022.
19. Veitia RA, Govindaraju DR, Bottani S, Birchler JA. (2017) Aging: Somatic Mutations, Epigenetic Drift and Gene Dosage Imbalance. *Trends Cell Biol*. 27(4):299-310. PMID: 27939088.
20. Wildey G, Chen Y, Lent I, et al. (2014) Pharmacogenomic approach to identify drug sensitivity in small-cell lung cancer. *PLoS One.* 9(9):e106784. PMID: 25198282; PMCID: PMC4157793.

Chapter-5

Objective-3: Network analysis of the differentially expressed genes between drug-sensitive and -resistant cancer cell lines in order to identify key hub biomarkers.

I. *Gene co-expression network analysis*

II. *Protein-protein interaction analysis*

5.1 Introduction

Networks are a common part of the actual world and the usual approach of representing biological systems as they represent a different combination of binary interactions or relations between heterogeneous or homogeneous elements. With the interactive visualization of data and integration of multiple datasets for analysis, it enables a more comprehensive study of different systems in nature. For example, **biological networks**, food webs, or hierarchies in a systematic organization. A network or graph is a collection of nodes connected by edges which represent a relationship between the nodes (***Toor & Chana 2021***).

5.1.1 Biological Network

Basically, nearly all biological entities interact with one another, from the molecular level to the ecosystem level. Using a variety of networks, like those that are used to study ecological, metabolic, or molecular interaction and neurological networks, allows us to study biology. Even though heterogeneous properties of cancer present serious challenges for prevention, treatment, and a detailed understanding of the pathological mechanisms; thus it is important to discover an effective biomarker that is necessary and prime importance (***Yan et al. 2016***). Recent decades have seen a surge of research into identifying molecular biomarkers for pre-symptomatic diagnosis, stratification by cancer subtype, evaluation of cancer growth, prediction of cancer patient response to therapies and diagnosis of cancer relapses (***Sawyers 2008; Bolton et al. 2014***). However, biomarkers of the oncogenic process are not effective at predicting outcomes, so they are too unreliable to be used in clinical applications.

The biological network has been studied and used extensively to represent, quantify and design intracellular interactions in order to understand the cellular mechanisms in cancer (***Kreeger & Lauffenburger 2010***). These insights have led to the discovery of cancer-related biomarkers.

The network-based integrated analysis incorporates multifaceted high-throughput omics profiling data, including expression array, SNP array, CGH array, etc., from cancerous tissues, blood samples, and other samples have extensively increased the understanding of the molecular basis of carcinogenesis and identification of novel biomarkers (***Wang et al. 2015; Zhang et al. 2009***).

5.1.2 Types of biological networks

Different types of data create different general characteristics of a network, including connectivity, complexity and structure, where multiple layers of information can be conveyed through edges and nodes in the networks. Some most common types of biological networks are; the gene co-expression network (GCNs), protein-protein interaction (PPI) network, gene/transcriptional regulatory network, microRNA–mRNA network, metabolic network, and cell signalling network. We are providing a brief introduction about the co-expression and PPI networks as we have analyzed these two networks in our study.

i. **Gene co-expression network:** Gene co-expression networks (GCNs) are transcript–transcript expression-based association networks used for various purposes, such as annotation of genes with unknown biological functions or processes, categorizing candidate genes related to disease and determining transcriptional regulatory mechanisms. These networks are constructed using data from transcriptomics (microarray data) and next-generation sequencing. With the recent advances in these fields, it is now possible to infer functions and disease associations for non-coding genes and splice variants. GCNs are networks that connect genes with similar expression patterns across all over the samples. Various types of correlation measures have been used to construct and analyse GCNs, including Pearson and Spearman correlations (***Van Dam et al. 2018***).

ii. **Protein-protein interaction network (PPI):** PPI networks are physical and functional interactions between proteins and they carry information about how different proteins work together within a cell to enable a biological process. In the human interactome, there are approximately 40,000 to 200,000 protein-protein interactions available. These protein-protein interactions play an important role in most of the biological and cellular processes, such as signal transduction pathways, gene transcription, cell-to-cell communication, metabolism, and proliferation. Furthermore, these interactions are key in every step of the central dogma of molecular biology, thereby playing an important role in transmitting genetic information that is significantly extrapolated in cancer as drug targets and in immunotherapy (***Garner & Janda 2011***).

With advanced strategies for the construction, analysis, and interpretation of various biological networks, we can discover reliable and accurate molecular biomarkers that can be used to monitor cancer progression, treatment, and diagnosis. These biomarkers may lead toward the development of personalized/precision medicine against cancer (***Yan et al. 2016***). The surge of omics data has led to the creation of a variety of freely available databases that provides an extensive amount of gene, protein interaction, biological pathway and network information, which are being established so that biologists can analyze these data from the complex system using valuable tools. These databases include interactions from BioGRID, STRING, IntAct, PID, MINT, KEGG, GeneMANIA, and REACTOME provide extremely useful qualitative data on the physical and functional relation between important elements in canonical cellular pathways (***Wang et al. 2012***).

A biological network consists of two essential elements, which are nodes and edges. In a network, nodes represent genes or proteins. Edges, on the other hand, represent the type of interaction or relationship that exists between individual nodes. These relationships may

represent either protein-protein interaction or promoter interaction or gene expression regulation, or metabolic responses and can also validate genetic evidence.

5.1.3 Topological parameters of network: To measure the locations of nodes in a network, it has been a set of defined topological parameters to describe their properties and centrality or functionality (**Fig. 1**). There are some most commonly used topological parameters are the following-

a) **Node degree:** The degree of nodes is the sum of all it has. If a node that has a degree of n refers to the number of other nodes that have a connection with it.
b) **Betweenness centrality (BC):** It is a measurement of centrality to assess the significance of independent nodes in a network. BC is a value that represents the number of all the shortest paths between nodes divided by the total number of the shortest paths between all nodes.
c) **Closeness centrality** is the reciprocal of its average shortest path length in the network and measures of how fast information passes from one node to the other reachable nodes in the network.

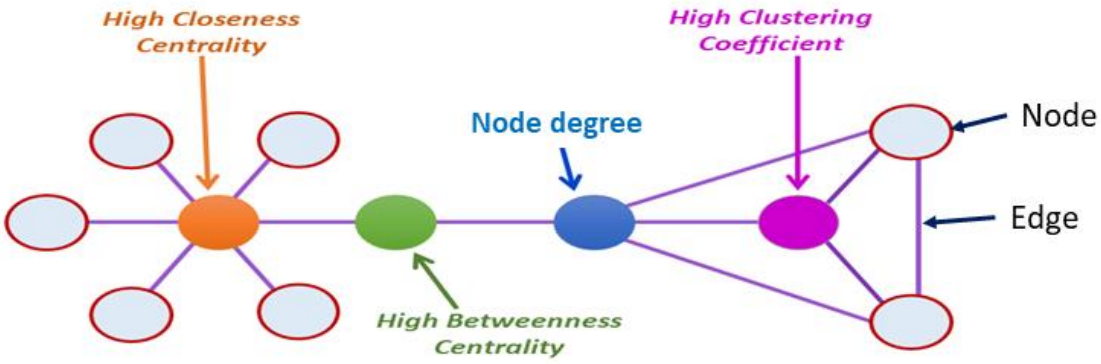


Figure 1: Diagrammatic representation shows the node degree, betweenness centrality closeness centrality, and clustering coefficient in the hypothetical network. *(Modified from Peng Zhang et al. 2016)*

Systems biology aims to understand complex biological entities at a systems level, looking not only at the individual components but also at how they interact and affect one another and using tools derived from graph theory to represent and analysis of biological systems. We have studied and analyzed gene co-expression and PPI network to identify probable biomarkers related to drug-resistant cancer using both qualitative and quantitative approaches by integrating data at genomic, transcriptomic and proteomic levels.

5.2 Materials and Methods

5.2.1 Generation and acquisition of gene co-expression network

We constructed a co-expression network to study the interaction between the DEGs in finer details. We subjected the set of DEGs list to an online web interface GeneMANIA program to query and construct the gene co-expression network of DEGs identified. It provides a meaningful gene-gene interaction network of DEGs, and as well as some predicted genes from GeneMANIA are automatically included to the network by default if it is found to be interacting with the presented DEGs list. We downloaded the co-expression network data from GeneMANIA.

5.2.2 Visualization and analysis of gene co-expression network

To visualize and analyze the gene co-expression network, we imported network data into Cytoscape 3.8.2. The co-expression network was analyzed by using a plugin of Cytoscape network analyzer to identify hub genes. Based on their node degree distribution top hub nodes were selected. A gene node considered as key/hub node that has higher number of edges with interacting gene nodes. To cluster the networks, we used a Cytoscape plugin app Glay (community cluster) from the clusterMaker module for network clustering (undirected edges), based on densely interacting nodes and functional relevance.

5.2.3 Generation and acquisition of PPI network

The PPI network analysis allows us to assess the corresponding protein interactions while gene-gene interaction network is only useful for identifying key hub genes. The PPI network was generated using a well validated online STRING v11.0 database, which provides physical (direct) and functional (indirect) protein association network data determined by experimental and computational methods. This database provides information about functional associations derived from various sources, including experimental, database, co-expression, co-occurrence text mining, neighborhood, etc., with a significant confidence score. We subjected the list of DEGs from co-expression network clusters into the STRING database to discern the physical and functional interaction among them. PPI network was generated with the cut-off interaction score for the network set to >0.400 (medium confidence), which implies that only interactions with a medium confidence score in the network were considered as reliable interactions.

5.2.4 Visualization and analysis of PPI network

We imported retrieved PPI network data into Cytoscape software (version 3.8.2) to visualize and analyze. Cytoscape Plugin Network Analyzer was used to analyze the PPI network. From the PPI network of each cluster we identified hub proteins based on highest node degree, which

is a measure of the protein's centrality in terms of its connection to other proteins nodes with key biological functions. To carry out GO and KEGG pathways analysis of the proteins in the PPI network, an online web server GeneCodis4 was used.

5.3 Results

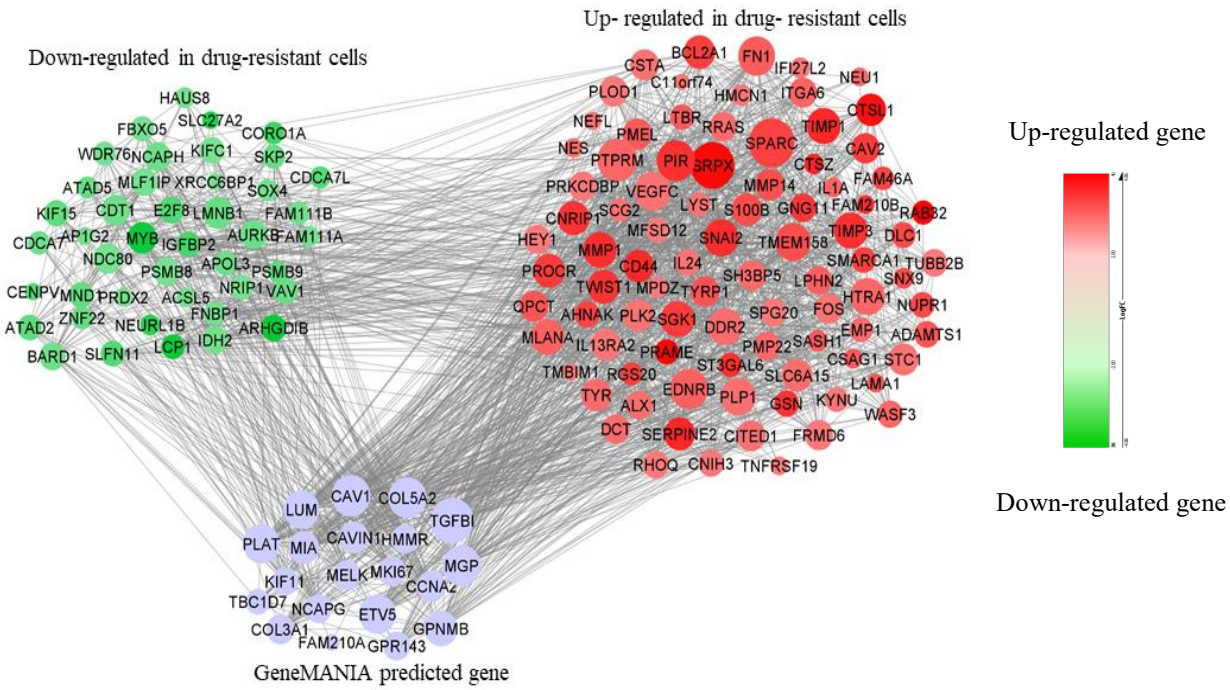
5.3.1 Gene co-expression network analysis and identification of hub genes

In order to examine the potential molecular interactions at the gene level and to undertake a deeper functional analysis of the identified DEGs in drug-resistant cancer cell lines, for the five drugs co-expression network was constructed and analyzed using DEGs enlisted from the above analysis. The generated co-expression networks of DEGs are shown in **Fig 2A-E**. The number of nodes (genes) and edges for the each networks was also enlisted (**Table 2)**. From the network analyses using quantitative method, we were able to identify the top 34 hub genes in the case of Ponatinib, having highest node connectivity (node degree) (**Table 1A**). Similarly, top 35, 48, 52, and 13 hub nodes were identified for Trametinib, Selumetinib, Foretinib, and CI-1040, respectively (**Table 1B-E**).

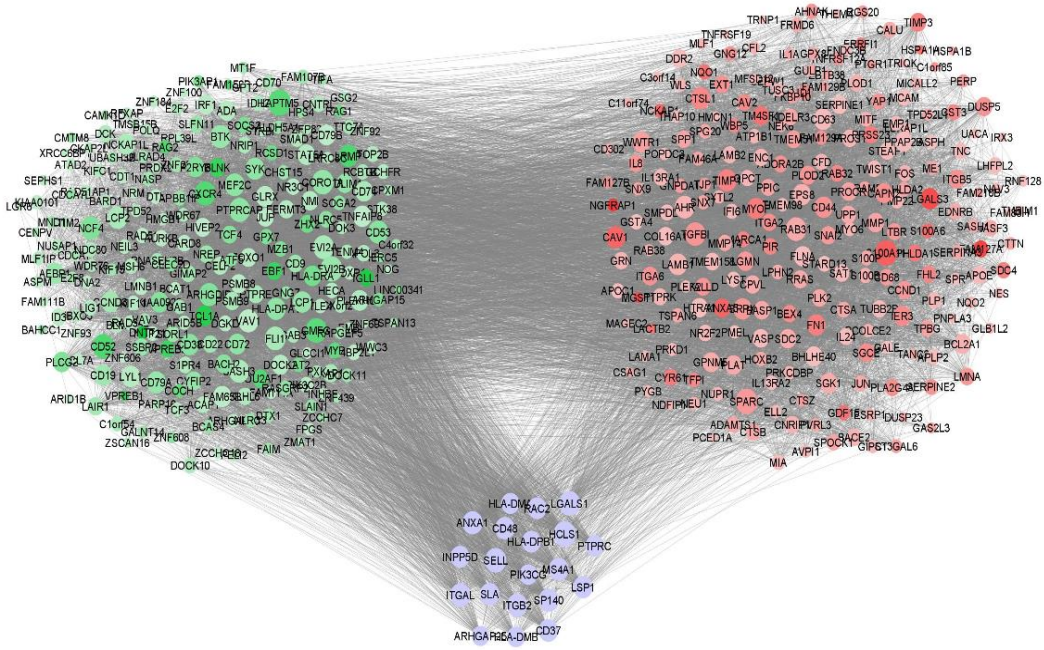
5.3.2 Clustering analysis of co-expression network

Glay, a Cytoscape plugin was used to cluster the gene co-expression network into modules. In case of Ponatinib, the generated clusters 1, 2, and 3 contained 35 nodes, 40 nodes, and 77 nodes, respectively (**Fig. 3A**). Then, different number of clusters were generated through the network clustering, in the case of the other four drugs: 3 clusters for CI-1040, 4 clusters for each Trametinib and Selumetinib, and 6 clusters for Foretinib (**Fig. 3B-E)**. After our co-expression networks analysis, further, we proceeded to generate PPI networks from clustered network genes in order to assess whether the same hub genes can be found as a hub node in the PPI network at the protein level.

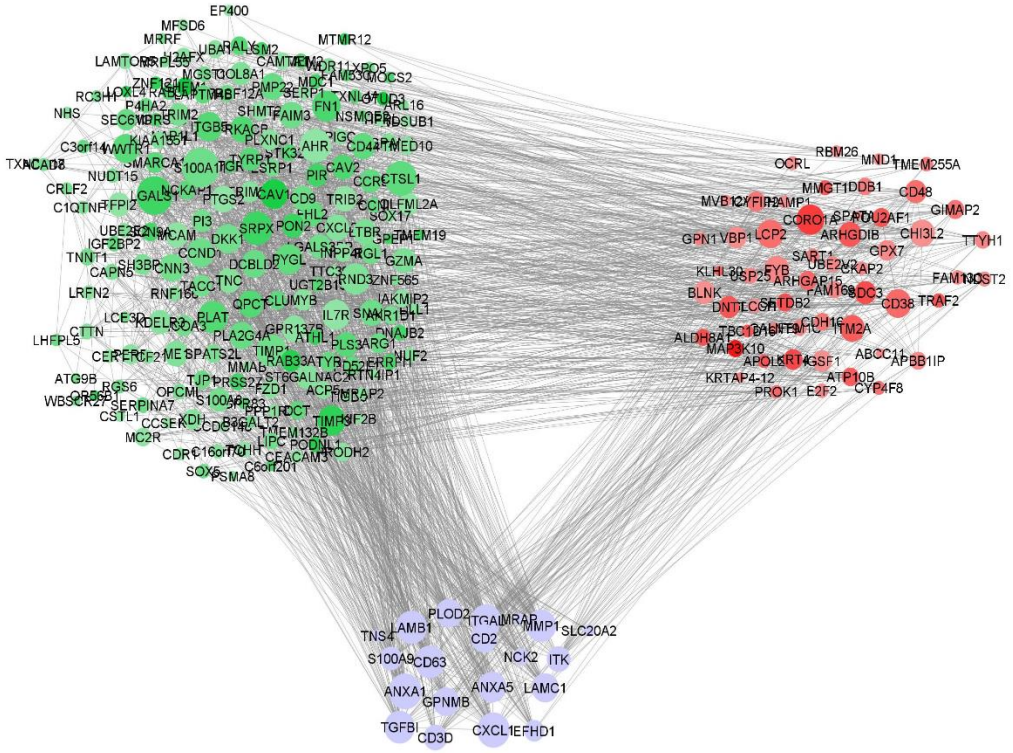
(A) Ponatinib



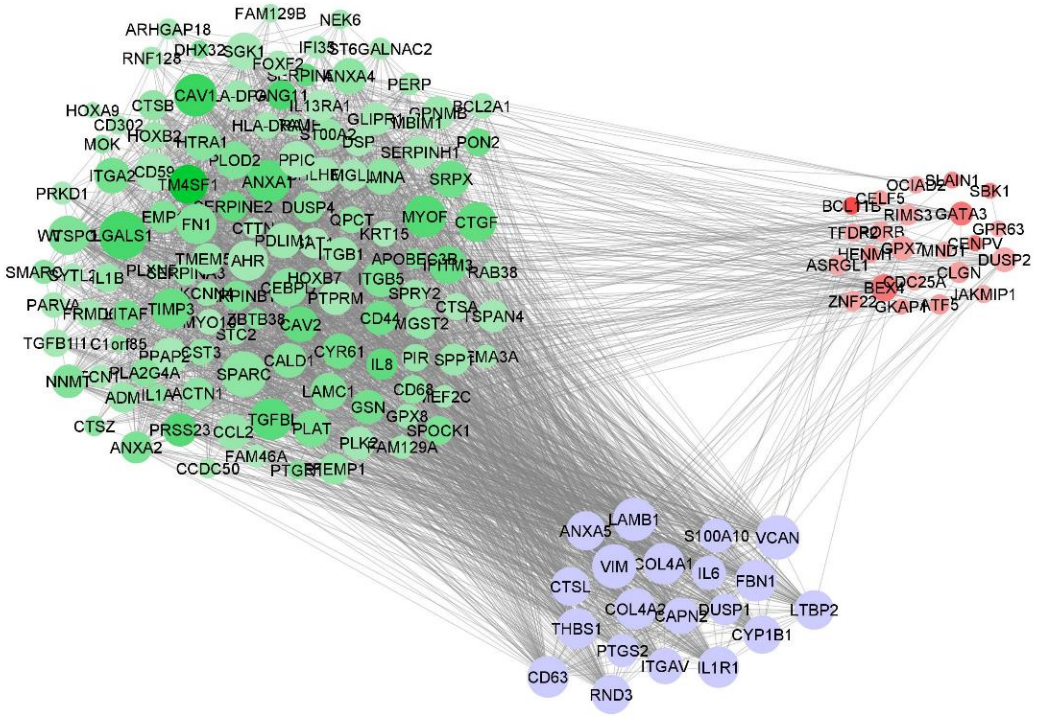
(B) Foretinib



(C) Selumetinib



(D) Trametinib



(E) CI-1040

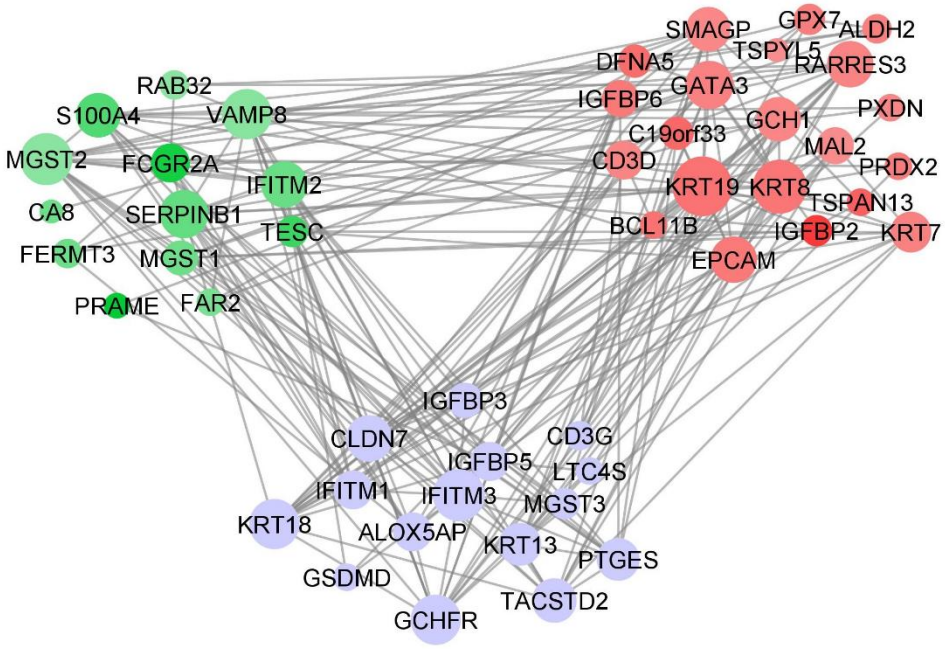


Figure 2: Co-expression network of DEGs. This network consists up- (red nodes) and down-regulated genes (green nodes) and genes predicted from GeneMANIA (blue nodes), hub genes were selected baesed on node degree (A) Ponatinib (B) Foretinib, (C) Selumetinib, (D) Trametinib, (E) CI-1040.

Table 1: List of identified hub genes from the co-expression network. (A) Ponatinib, (B) Foretinib, (C) Selumetinib, (D) Trametinib, (E) CI-1040.

(A) Ponatinib

S No.	Name	Degree	logFC	S No.	Name	Degree	logFC
1	*SPARC*	63	2.942045	18	*TWIST1*	35	3.171716
2	*SRPX*	59	3.861324	19	*PROCR*	35	3.053009
3	*PTPRM*	51	2.360718	20	*CNRIP1*	35	3.17705
4	*PIR*	48	3.213295	21	*BCL2A1*	34	2.995919
5	*EDNRB*	47	2.455901	22	*SERPINE2*	33	3.398535
6	*FN1*	45	2.48735	23	*TYR*	32	2.257251
7	*VEGFC*	44	2.159631	24	*S100B*	32	2.763242
8	*TMEM158*	42	2.762151	25	*CTSL1*	32	3.748661
9	*PLP1*	42	2.206968	26	*CD44*	32	3.355082
10	*HTRA1*	41	2.316941	27	*SLC6A15*	32	2.225731

11	*SNAI2*	41	3.26492	28	*IL13RA2*	31	2.086026
12	*DDR2*	40	2.261463	29	*LMNB1*	31	-2.29481
13	*TIMP3*	39	3.235401	30	*PLK2*	31	2.1479
14	*SGK1*	39	3.038964	31	*STC1*	31	2.149369
15	*TIMP1*	39	3.35654	32	*PMEL*	30	2.537572
16	*MLANA*	38	2.430249	33	*MYB*	30	-3.87657
17	*MMP1*	38	3.262725	34	*MMP14*	30	2.718661

(B) Foretinib

S No.	Name	Degree	LogFC	S No.	Name	Degree	LogFC
1	*LAPTM5*	145	-5.75001	27	*CAV1*	111	5.248772
2	*SPARC*	139	3.390596	28	*SYK*	111	-3.71093
3	*S100A11*	139	5.453968	29	*LRMP*	111	-6.19717
4	*EVI2B*	138	-3.34125	30	*LAMB1*	108	2.433101
5	*CORO1A*	132	-3.54753	31	*TNFAIP8*	106	-2.02497
6	*GMFG*	132	-5.60287	32	*AEBP1*	106	-3.25478
7	*CD52*	129	-6.07568	33	*LYL1*	106	-2.32261
8	*PTPRCAP*	128	-3.81623	34	*EPS8*	105	2.493185
9	*CXCR4*	128	-5.94196	35	*CD63*	105	2.180637
10	*EVI2A*	126	-2.75744	36	*CD38*	105	-3.94811
11	*FLI1*	125	-2.78249	37	*SRPX*	105	3.465235
12	*TGFBI*	124	3.527154	38	*CD44*	105	3.602315
13	*RAB31*	124	2.781301	39	*LGALS3*	105	6.130637
14	*BTK*	120	-3.43982	40	*LCP1*	105	-3.46115
15	*CD53*	120	-3.95217	41	*MYOF*	105	4.486707
16	*VAV1*	120	-2.10008	42	*FN1*	104	4.286871
17	*NCF4*	120	-4.28818	43	*FAM65B*	104	-2.17044
18	*SASH3*	120	-2.69993	44	*GLRX*	103	-2.23624
19	*ARHGDIB*	119	-4.13565	45	*CD19*	103	-3.10301
20	*AHR*	117	2.449523	46	*CCND3*	103	-3.1296
21	*MEF2C*	117	-3.81352	47	*NCKAP1L*	103	-2.59252
22	*LCP2*	115	-2.74972	48	*WWTR1*	102	3.119242
23	*TIMP1*	113	4.218703	49	*CD79B*	100	-2.09659
24	*CTSL1*	113	4.007319	50	*KIAA0922*	100	-4.13033
25	*HLA-DPA1*	112	-3.7146	51	*CD72*	100	-3.79728
26	*CD79A*	112	-3.436	52	*PLEKHO1*	100	-2.59832

(C) Selumetinib

S No.	Name	Degree	logFC	S No.	Name	Degree	logFC
1	*S100A11*	60	-2.66948	25	*FYB*	35	2.21787
2	*LGALS1*	58	-3.61629	26	*CHI3L2*	35	2.042024
3	*CTSL1*	50	-3.01495	27	*FAIM3*	34	-2.67386
4	*SRPX*	48	-3.6649	28	*GPR137B*	34	-2.10542
5	*AHR*	48	-2.01913	29	*PMP22*	34	-2.85202
6	*TIMP1*	46	-2.65478	30	*ITM2A*	34	3.022547
7	*PLAT*	45	-3.21702	31	*PTGS2*	34	-2.0082
8	*RND3*	43	-2.33281	32	*PRKACB*	33	-3.26241
9	*FN1*	42	-3.3788	33	*SNAI2*	33	-2.70462
10	*PYGL*	41	-2.95357	34	*CD44*	33	-3.02248
11	*IL7R*	41	-2.00598	35	*CXCL2*	33	-2.27643
12	*CORO1A*	41	3.689428	36	*ME1*	32	-2.32413
13	*TIMP3*	41	-3.97979	37	*TRIB2*	32	-2.00256
14	*CAV1*	40	-4.37341	38	*FHL2*	31	-2.98132
15	*DKK1*	39	-2.62356	39	*MCAM*	31	-2.06399
16	*CCND1*	39	-2.67084	40	*TNC*	31	-2.25594
17	*ITGB5*	39	-3.20911	41	*ARHGDIB*	31	3.13521
18	*WWTR1*	39	-3.10919	42	*GZMA*	31	-2.40187
19	*CD38*	38	2.785901	43	*LTBR*	31	-2.41978
20	*LCP2*	38	2.685341	44	*PLS3*	30	-3.18165
21	*CNN3*	36	-2.60362	45	*LGALS3BP*	30	-2.65608
22	*DCBLD2*	36	-2.7906	46	*CAV2*	30	-3.34135
23	*PLA2G4A*	35	-2.80208	47	*ST6GALNAC2*	30	-2.22252
24	*QPCT*	35	-3.21769	48	*CD9*	30	-2.7278

(D) Trametinib

S No.	Name	Degree	LogFC	S No.	Name	Degree	LogFC
1	*LGALS1*	87	-4.62733	19	*ITGB5*	58	-3.19813
2	*SPARC*	80	-2.73628	20	*SRPX*	57	-3.53466
3	*ANXA1*	74	-4.16594	21	*LAMC1*	57	-3.09847
4	*CAV1*	73	-4.98208	22	*PPAP2B*	56	-2.06307
5	*MYOF*	71	-4.2814	23	*PPIC*	56	-2.17838
6	*TGFBI*	70	-4.0868	24	*HTRA1*	56	-2.91928
7	*TIMP3*	67	-3.70926	25	*LMNA*	55	-2.75874
8	*AHR*	66	-2.20746	26	*ANXA4*	54	-2.73862
9	*FN1*	66	-3.07716	27	*ITGA2*	54	-3.38568
10	*CD59*	64	-2.16905	28	*CAV2*	54	-3.86202

11	*CTGF*	64	-3.92276	29	*CCL2*	53	-2.45051
12	*WWTR1*	63	-3.14903	30	*PDLIM1*	53	-2.31412
13	*TM4SF1*	62	-6.48639	31	*CYR61*	52	-3.52746
14	*CEBPD*	60	-2.86032	32	*BHLHE40*	51	-2.29155
15	*PLOD2*	59	-3.14062	33	*CALD1*	51	-3.0044
16	*IFITM3*	59	-3.75116	34	*CD44*	50	-3.60438
17	*PLAT*	58	-3.24524	35	*PTPRM*	50	-2.16834
18	*SGK1*	58	-2.06759				

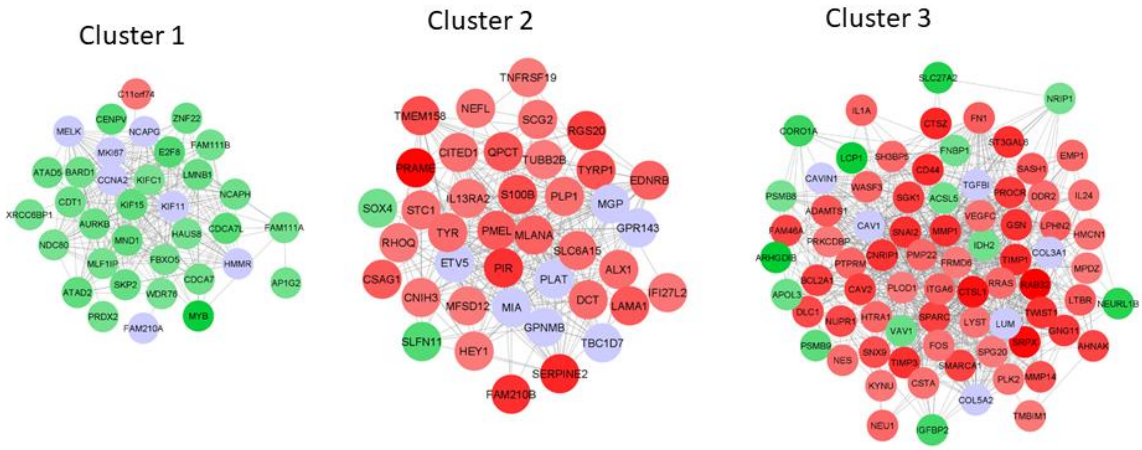
(E) CI-1040

S No.	Name	Degree	LogFC
1	*KRT19*	20	2.553488
2	*KRT8*	17	2.47259
3	*MGST2*	16	-2.06881
4	*VAMP8*	15	-2.12426
5	*GATA3*	14	2.268906
6	*SERPINB1*	14	-2.82802
7	*RARRES3*	13	2.195455
8	*IFITM2*	13	-2.71822
9	*EPCAM*	13	2.417885
10	*S100A4*	12	-3.19585
11	*GCH1*	12	2.156811
12	*SMAGP*	12	2.135293
13	*KRT7*	10	2.262705

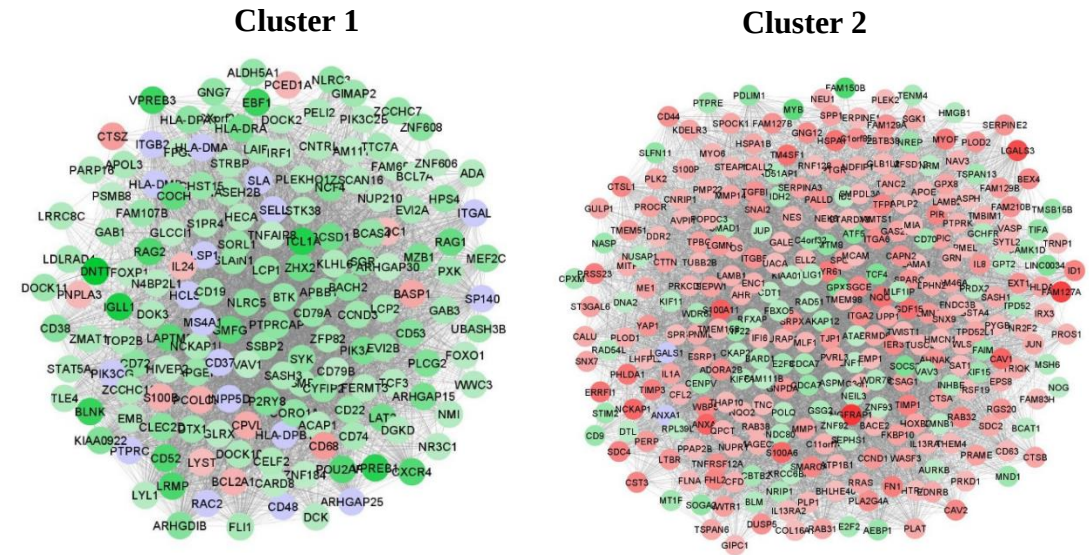
Table 2: List of number of nodes and hub nodes in gene co-expression network for all five drugs.

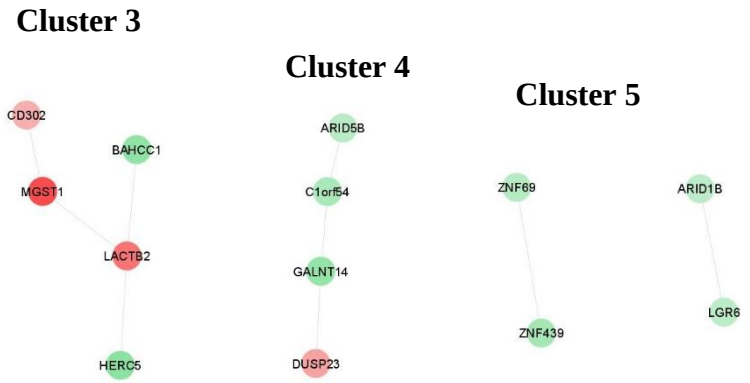
Drug name	Number of nodes in network	Number of edges in network	Number of hub nodes	Lowest node degree for hub node selection
Ponatinib	152	1826	34	30
Foretinib	483	14156	52	100
Selumetinib	256	2410	48	30
Trametinib	165	3019	35	50
CI-1040	49	211	13	10

(A) Ponatinib



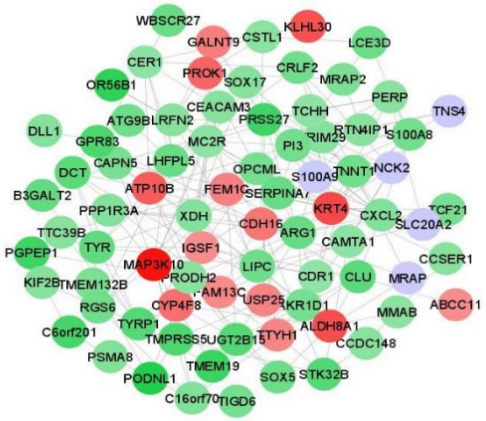
(B) Foretinib



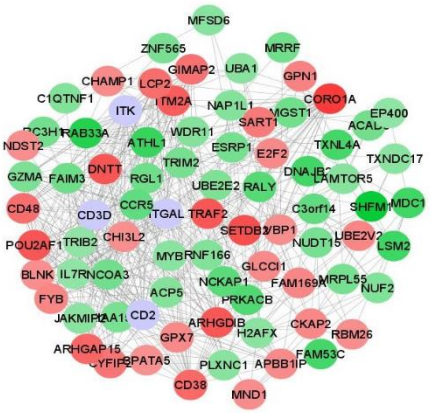


(C) Selumetinib

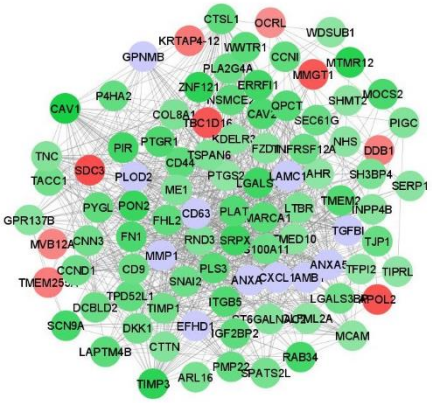
Cluster 1



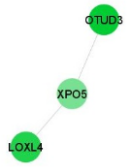
Cluster 2



Cluster 3

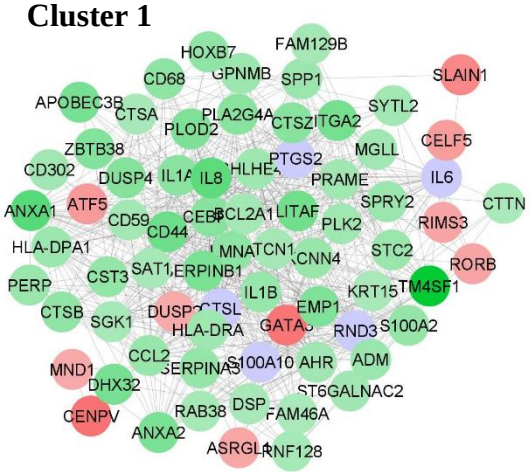


Cluster 4

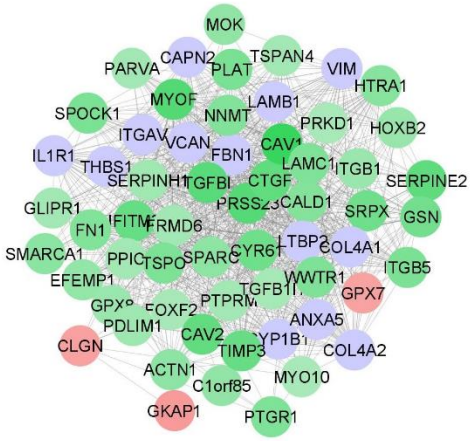


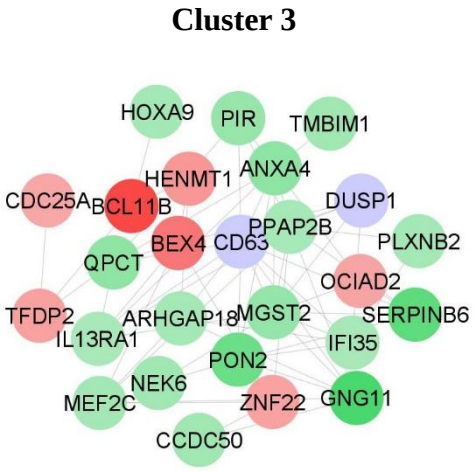
(D) Trametinib

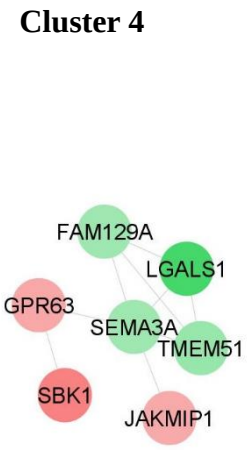
Cluster 1



Cluster 2







(E) CI-1040

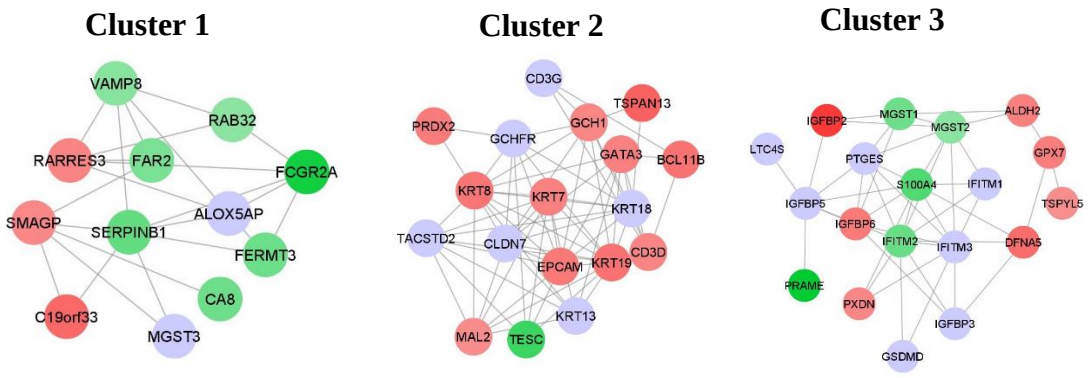


Figure 3: Co-expression network cluster generated by fast greedy (Glay) Cytoscape plugin clustering algorithm. Clusters for five drugs (A) Ponatinib, (B) Foretinib, (C) Selumetinib, (D) Trametinib, and (E) CI-1040. Red nodes: upregulated protein-coding genes, Green nodes: down-regulated protein-coding genes, Blue nodes: GeneMANIA predicted protein-coding genes.

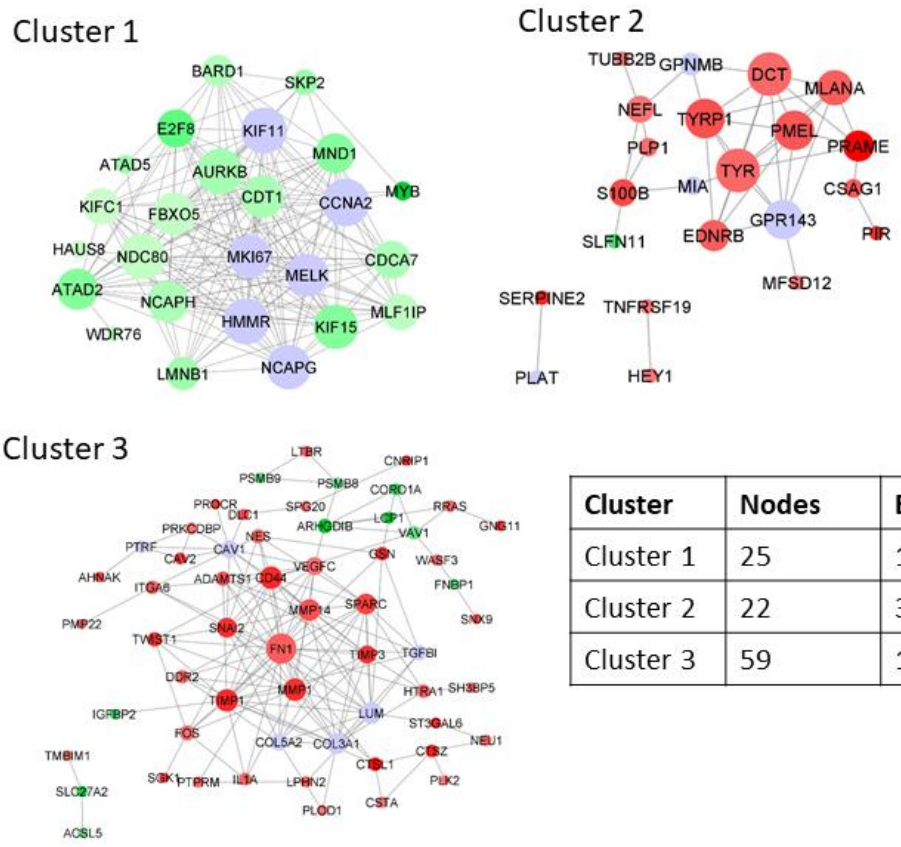
5.3.3 PPI network analysis and identification of hub proteins

The identification of key proteins in the network of proteins encoded by DEGs in drug-resistant cancer furnishes an important insight to understand the regulatory mechanisms that may cause cancer drug resistance. Genes/proteins significantly linked with drug resistance in cancer may act as a hub gene/protein. To study the same interactions at protein level, the PPI network was constructed by taking the genes from each cluster from the gene co-expression network in case of the selected drugs. We used a cut-off score 0.400, which is a median confidence score, quantifies the reliability of generated PPI network with corroborative evidence for the reported interactions (between two proteins) (***Bozhilova et al. 2019***). The PPI network for drug Ponatinib is shown in **Fig. 4A,** and for other four drugs are shown in **Fig. 4B-E**. Top hub proteins identified from all generated PPI networks are mentioned in **table 3A-E.** Only four hub protein nodes were identified for drug CI-1040 from PPI networks, whereas top 10 hub proteins node were identified for other four drugs (**Table 4)**. We were not able to acquire a PPI network for some of the clusters from the STRING database because of a lack of interaction data in STRING database.

5.3.4 Functional analysis of proteins from the PPI network

The functional enrichment analysis of these proteins are enriched in various biological processes term and KEGG pathways. Notably, the protein nodes from the PPI networks of the studied drugs were enriched in biological processes, such as signal transduction, proteolysis, melanogenesis, cell adhesion, cytokine-mediated signaling, cell population proliferation, and cell migration. From KEGG pathways analysis, we observed proteoglycans, and PI3K-Akt signaling pathways in cancer, etc. (**Fig. 5A-E**).

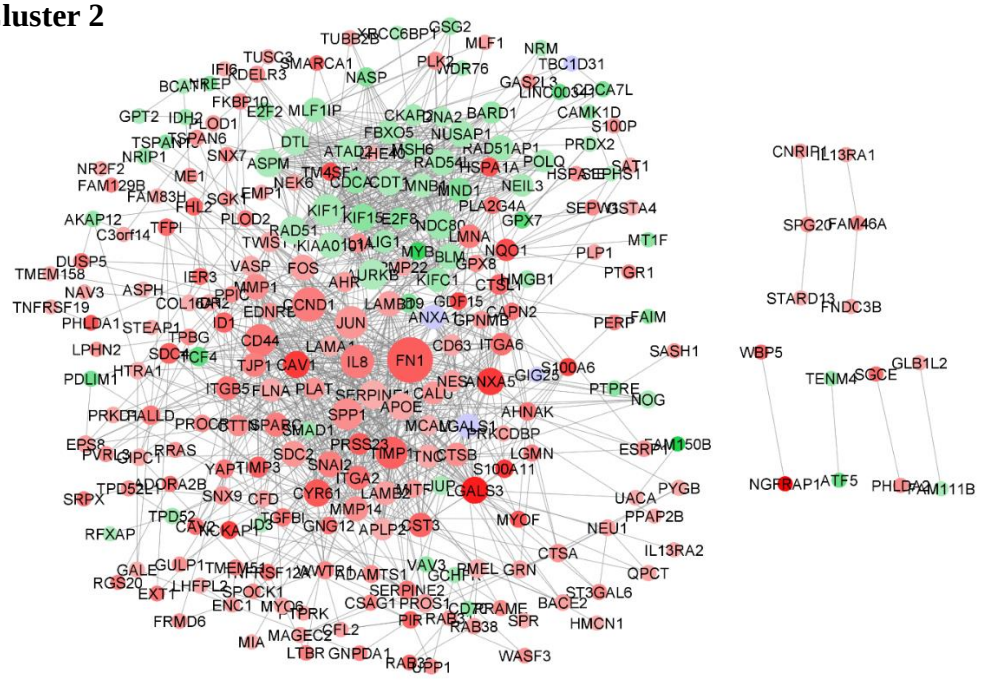
(A) Ponatinib



Cluster	Nodes	Edges
Cluster 1	25	179
Cluster 2	22	39
Cluster 3	59	155

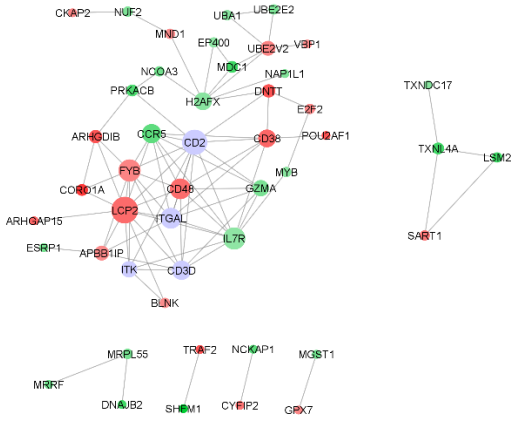
(B) Foretinib

Cluster 2

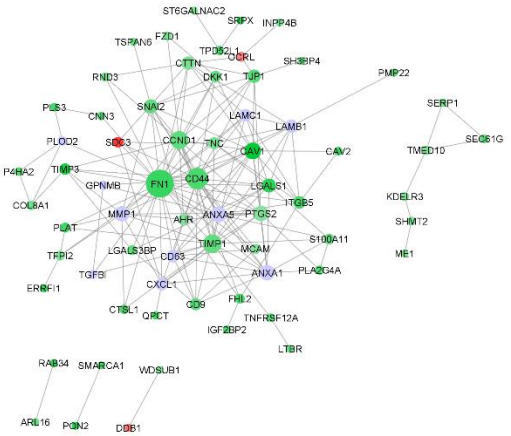


(C) Selumetinib

Cluster 2

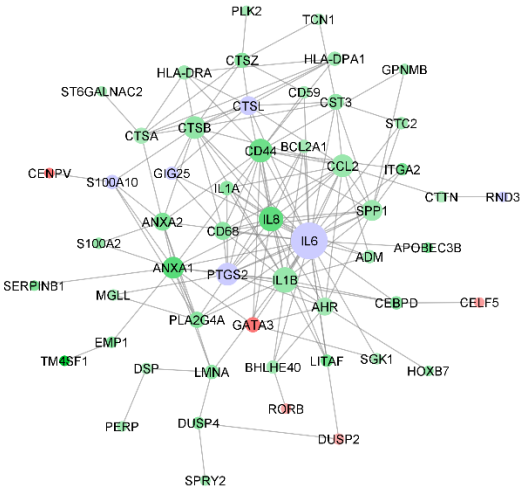


Cluster 3

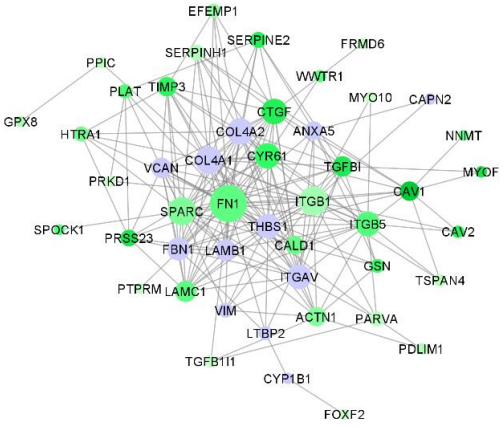


(D) Trametinib

Cluster 1



Cluster 2



(E) CI-1040

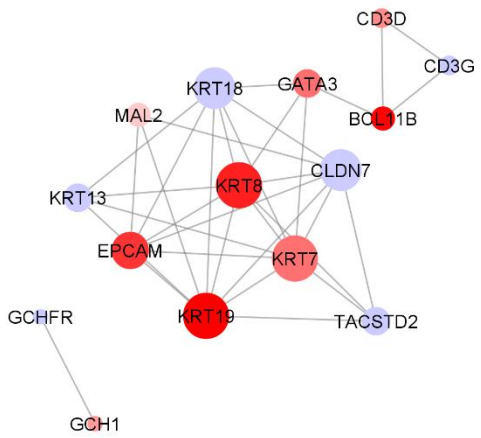


Figure 4: Protein-protein interaction network of cluster. PPI network of DEGs from the co-expression network clusters for four drugs and selection of hub proteins by analyzing node degree. (A) Ponatinib, (B) Foretinib, (C) Selumetinib, (D) Trametinib, (E) CI-1040. Red nodes: upregulated protein-coding genes, Green colored nodes: down-regulated protein-coding genes, Blue colored nodes: GeneMANIA predicted protein-coding genes. Node degree represented by circle size.

Table 3: List of identified hub proteins from the PPI network. (A) Ponatinib, (B) Foretinib, (C) Selumetinib, (D) Trametinib, (E) CI-1040.

(A) Ponatinib

Cluster 1

Hub protein	Degree	logFC
NDC80	20	-2.06481
AURKB	20	-2.26034
KIF15	19	-2.45912
FBXO5	18	-2.07972
CDT1	18	-2.22891
ATAD2	17	-2.45525
NCAPH	17	-2.2146
MND1	16	-2.43927
CDCA7	15	-2.21104
E2F8	15	-2.69675

Cluster 2

Hub protein	Degree	logFC
DCT	8	2.248641
TYR	8	2.257251
TYRP1	7	2.658401

PMEL	7	2.537572
MLANA	6	2.430249
PRAME	5	3.920346
EDNRB	5	2.455901

Cluster 3

Hub protein	Degree	logFC
FN1	24	2.48735
CD44	15	3.355082
MMP1	15	3.262725
TIMP1	15	3.35654
MMP14	14	2.718661
SPARC	13	2.942045
SNAI2	12	3.26492
VEGFC	10	2.159631
TIMP3	10	3.235401
FOS	7	2.148925

(B) Foretinib

Cluster 2

Hub protein	Degree	LogFC
FN1	66	4.286871
CCND1	42	3.398118
IL8	40	3.455896
JUN	38	2.687192
TIMP1	37	4.218703
CD44	37	3.602315
SPP1	36	2.925569
KIF11	36	-2.28524
AURKB	33	-2.06155
SERPINE1	30	2.088228

(C) Selumetinib

Cluster 2

Hub protein	Degree	logFC
LCP2	13	2.685341
FYB	10	2.21787
IL7R	10	-2.00598
CD48	9	2.679745
CCR5	8	-2.93984
CD38	7	2.785901
H2AFX	7	-2.10786
GZMA	6	-2.40187
APBB1IP	5	2.105772

UBE2V2	5	2.212951

Cluster 3

Hub protein	Degree	logFC
LCP2	13	2.685341
FYB	10	2.21787
IL7R	10	-2.00598
CD48	9	2.679745
CCR5	8	-2.93984
CD38	7	2.785901
H2AFX	7	-2.10786
GZMA	6	-2.40187
APBB1IP	5	2.105772
UBE2V2	5	2.212951

(D) Trametinib

Cluster 1

Hub protein	**Degree**	**LogFC**
IL1B	17	-2.48654
IL8	16	-3.99662
CD44	15	-3.60438
CCL2	14	-2.45051
CTSB	14	-2.65337
ANXA1	13	-4.16594
SPP1	12	-2.39734
AHR	9	-2.20746
ANXA2	9	-3.23904
CD68	9	-2.74891

Cluster 2

Hub protein	**Degree**	**LogFC**
FN1	31	-3.07716
ITGB1	22	-2.38848
SPARC	21	-2.73628
CYR61	18	-3.52746
CTGF	18	-3.92276
ITGB5	17	-3.19813
LAMC1	13	-3.09847
TGFBI	12	-4.0868
ACTN1	11	-2.7235
CAV1	11	-4.98208

(E) CI-1040
Cluster 2

Hub protein	Degree	LogFC
KRT7	8	2.262705
KRT8	8	2.47259
KRT19	8	2.553488
EPCAM	6	2.417885

Table 4: Number of top hub protein nodes identified from each PPI network cluster in the case of all five drugs.

Drug name	Clusters	Number of hub protein nodes from the PPI network	Number of nodes in the PPI network	Number of edges in the PPI network
Ponatinib	Cluster 1	10	25	179
	Cluster 2	7	22	39
	Cluster 3	10	59	155
Foretinib	Cluster 2	10	254	1174
Selumetinib	Cluster 2	10	47	87
	Cluster 3	10	67	179
Trametinib	Cluster 1	10	54	152
	Cluster 2	10	47	205
CI-1040	Cluster 2	4	15	34

(A) Ponatinib
Cluster 1

GO ID	Biological processes	genes_found	P-value
GO:0007049	cell cycle	13	1.29E-17
GO:0051301	cell division	10	3.33E-14
GO:0000278	mitotic cell cycle	5	2.45E-07
GO:0000086	G2/M transition of mitotic cell cycle	4	1.19E-05
GO:0008283	cell population proliferation	4	1.25E-05
GO:0006355	regulation of transcription, DNA-templated	4	0.00333

KEGG ID	KEGG Pathways	genes_found	P-value
hsa05169	Epstein-Barr virus infection	2	0.009102
hsa04110	Cell cycle	2	0.009265
hsa05203	Viral carcinogenesis	2	0.011292
hsa05200	Pathways in cancer	2	0.040845

Cluster 2

GO ID	Biological processes	genes_found	P-value
GO:0042438	melanin biosynthetic process	5	1.13E-12
GO:0043473	pigmentation	4	8.79E-08
GO:0007165	signal transduction	4	0.009021
GO:0032438	melanosome organization	3	5.74E-06
GO:0008284	positive regulation of cell population proliferation	3	0.006307
GO:0007399	nervous system development	3	0.006463
GO:0006583	melanin biosynthetic process from tyrosine	2	7.53E-06
GO:0006726	eye pigment biosynthetic process	2	1.81E-05
GO:0009637	response to blue light	2	5.02E-05
GO:0048066	developmental pigmentation	2	0.000236

KEGG ID	KEGG Pathways	genes_found	P-value
hsa04916	Melanogenesis	4	1.02E-07
hsa00350	Tyrosine metabolism	3	6.03E-07
hsa01100	Metabolic pathways	3	0.006029
hsa05014	Amyotrophic lateral sclerosis	2	0.007357
hsa05022	Pathways of neurodegeneration - multiple diseases	2	0.012748

Cluster 3

GO ID	Biological processes	genes_found	P-value
GO:0030198	extracellular matrix organization	12	8.64E-15
GO:0007165	signal transduction	12	2.75E-06
GO:0022617	extracellular matrix disassembly	8	9.83E-13
GO:0019221	cytokine-mediated signaling pathway	8	1.05E-07
GO:0006508	proteolysis	8	3.10E-06
GO:0007155	cell adhesion	7	5.26E-05
GO:0034097	response to cytokine	6	6.97E-09
GO:0030199	collagen fibril organization	6	8.84E-08
GO:0030335	positive regulation of cell migration	6	9.52E-06
GO:0008285	negative regulation of cell population proliferation	6	0.000155

KEGG ID	KEGG Pathways	genes_found	pval_adj
hsa05205	Proteoglycans in cancer	10	2.84E-14
hsa04510	Focal adhesion	6	2.13E-07
hsa05200	Pathways in cancer	6	2.24E-05
hsa04151	PI3K-Akt signaling pathway	5	3.35E-05
hsa04926	Relaxin signaling pathway	5	6.79E-07

(B) Foretinib

Cluster 2

GO ID	Biological processes	genes_found	P-value
GO:0007165	signal transduction	31	1.69E-10
GO:0045944	positive regulation of transcription by RNA polymerase II	27	3.83E-11
GO:0007155	cell adhesion	26	3.12E-16
GO:0043066	negative regulation of apoptotic process	25	2.95E-16
GO:0000122	negative regulation of transcription by RNA polymerase II	21	3.44E-09
GO:0030154	cell differentiation	21	5.00E-08
GO:0030198	extracellular matrix organization	20	1.00E-16
GO:0006915	apoptotic process	20	1.80E-09
GO:0043312	neutrophil degranulation	19	5.06E-11
GO:0007049	cell cycle	19	4.20E-09

KEGG ID	KEGG Pathways	genes_found	P-value
hsa01100	Metabolic pathways	23	1.17E-14
hsa05200	Pathways in cancer	21	8.42E-16
hsa04510	Focal adhesion	17	3.50E-18
hsa05205	Proteoglycans in cancer	17	7.89E-18
hsa04151	PI3K-Akt signaling pathway	13	4.99E-10

(C) Selumetinib

Cluster 2

GO ID	Biological processes	genes_found	P-value
GO:0007165	signal transduction	9	0.000456
GO:0007049	cell cycle	6	0.000472
GO:0006915	apoptotic process	6	0.000513
GO:0006955	immune response	5	0.000501
GO:0045944	positive regulation of transcription by RNA polymerase II	5	0.012815
GO:0006303	double-strand break repair via nonhomologous end joining	4	5.16E-05
GO:0050852	T cell receptor signaling pathway	4	0.000594
GO:0007166	cell surface receptor signaling pathway	4	0.00126
GO:0008380	RNA splicing	4	0.001276
GO:0006397	mRNA processing	4	0.002943

KEGG ID	KEGG Pathways	genes_found	P-value
hsa05200	Pathways in cancer	6	2.08E-05
hsa04640	Hematopoietic cell lineage	5	2.63E-07
hsa05169	Epstein-Barr virus infection	5	4.75E-06
hsa04015	Rap1 signaling pathway	4	0.00014
hsa05163	Human cytomegalovirus infection	4	0.000146

Cluster 3

GO ID	Biological processes	genes_found	P-value
GO:0007155	cell adhesion	15	5.35E-14
GO:0007165	signal transduction	11	4.47E-05
GO:0030198	extracellular matrix organization	9	3.04E-09
GO:0008285	negative regulation of cell population proliferation	9	2.41E-07
GO:0043066	negative regulation of apoptotic process	9	6.06E-07
GO:0019221	cytokine-mediated signaling pathway	8	2.66E-07
GO:0002576	platelet degranulation	7	2.31E-08
GO:0044267	cellular protein metabolic process	7	6.32E-08
GO:0016477	cell migration	7	1.13E-06
GO:0001525	angiogenesis	7	1.14E-06

KEGG ID	KEGG Pathways	genes_found	P-value
hsa05205	Proteoglycans in cancer	11	1.26E-15
hsa04510	Focal adhesion	8	1.32E-10
hsa05165	Human papillomavirus infection	8	9.11E-10
hsa01100	Metabolic pathways	8	3.35E-06
hsa05200	Pathways in cancer	7	3.40E-06

(D) Trametinib

Cluster 1

GO ID	Biological processes	genes_found	P-value
GO:0006954	inflammatory response	12	1.21E-12
GO:0043312	neutrophil degranulation	12	3.93E-12
GO:0019221	cytokine-mediated signaling pathway	10	5.47E-11
GO:0008285	negative regulation of cell population proliferation	10	1.42E-09
GO:0007165	signal transduction	10	3.84E-05
GO:0006955	immune response	8	2.27E-07
GO:0045944	positive regulation of transcription by RNA polymerase II	8	0.000133
GO:0071222	cellular response to lipopolysaccharide	7	3.40E-08
GO:0010628	positive regulation of gene expression	7	2.22E-05
GO:0043066	negative regulation of apoptotic process	7	2.24E-05

KEGG ID	KEGG pathways	genes_found	P-value
hsa05323	Rheumatoid arthritis	8	4.83E-14
hsa04640	Hematopoietic cell lineage	8	6.32E-14
hsa05164	Influenza A	7	1.92E-10
hsa05321	Inflammatory bowel disease	6	1.73E-10
hsa04142	Lysosome	6	1.75E-10

Cluster 2

GO ID	Biological processes	genes_found	P-value
GO:0007155	cell adhesion	16	1.07E-18
GO:0030198	extracellular matrix organization	15	2.97E-22
GO:0001525	angiogenesis	11	1.87E-14
GO:0044267	cellular protein metabolic process	7	4.02E-09
GO:0016477	cell migration	7	7.38E-08
GO:0043687	post-translational protein modification	7	1.79E-07
GO:0007165	signal transduction	7	0.001849
GO:0007229	integrin-mediated signaling pathway	6	2.15E-08
GO:0002576	platelet degranulation	6	4.34E-08
GO:0030335	positive regulation of cell migration	6	1.99E-06

KEGG ID	KEGG pathways	genes_found	P-value
hsa04510	Focal adhesion	14	5.36E-25
hsa04512	ECM-receptor interaction	9	7.62E-19
hsa05165	Human papillomavirus infection	9	3.73E-13
hsa04151	PI3K-Akt signaling pathway	9	1.63E-12
hsa05205	Proteoglycans in cancer	8	2.26E-12

(E) CI-1040

Cluster 2

GO ID	Biological process	genes_found	P-value
GO:0070268	cornification	5	1.14E-09
GO:0031424	keratinization	5	1.16E-09
GO:0065003	protein-containing complex assembly	3	0.000124
GO:0050852	T cell receptor signaling pathway	3	0.000185
GO:0043066	negative regulation of apoptotic process	3	0.001989
GO:0002376	immune system process	3	0.002215
GO:0045944	positive regulation of transcription by RNA polymerase II	3	0.005357

KEGG ID	KEGG pathways	genes_found	P-value
hsa05150	Staphylococcus aureus infection	3	2.54E-06
hsa04659	Th17 cell differentiation	3	5.89E-06
hsa04658	Th1 and Th2 cell differentiation	3	6.04E-06
hsa04915	Estrogen signaling pathway	3	6.73E-06

Figure 5: GO and KEGG pathways analysis of proteins from the PPI network. Gene ontology and KEGG pathway analysis using GeneCodis4 for five drugs. (A) Ponatinib, (B) Foretinib, (C) Selumetinib, (D) Trametinib, (E) CI-1040.

5.3.5 Selection of common hub nodes between co-expression and PPI network

Upon further analysis, combining hub node list from both PPI and gene co-expression networks, some common hub protein-coding genes were chosen. In case of Ponatinib there were nine hub proteins from clusters 3 and four hub proteins from cluster2, the hub genes

shared in common with the hub gene list from the gene co-expression network as well as with STRING-generated top hub proteins. Similarly, for the other four drugs also, common hub nodes were selected (**Table 5**). Through chronological analyses, the identified key hub proteins, common between co-expression and PPI network hub nodes, are enlisted as KRT7, KRT8, KRT19 and EPCAM for the drug CI-1040, TYR, MLANA, PMEL, EDNRB, FN1, CD44, MMP1, MMP14, TIMP1, SPARC, TIMP3, SNAI2, and VEGFC for Ponatinib, CD44, AHR, CCL2, ANXA1, FN1, CYR61, CTGF, SPARC, ITGB5, TGFB1 and LAMC1 for trametinib, Similarly, for other drugs, LCP2, FYB, IL7R, FN1, CD38, CD44, CCND1, TIMP1, PTGS2, CAV1, SNAI2 and LGALS1 for selumetinib, FN1, TIMP1 and CD44 for foretinib, were identified as driver proteins common between both co-expression and PPI network hub node list.

5.3.6 Common hub protein-coding genes across multiple drugs

Further, we intended to find out common hub proteins coding genes across all the drugs studied. Among the hub protein coding nodes from co-expression networks and PPI, as mentioned above, SPARC was common for two drugs (Trametinib and Ponatinib), TIMP1 was common across the three drugs (Ponatinib, Selumetinib, Foretinib), while CD44 and FN1 were found to be common for 4 drugs; Foretinib, Trametinib, Ponatinib, and Selumetinib (**Fig. 6**). These hub proteins might induce drug-resistant in cancer through various divergent pathways, including PI3K-Akt signaling pathways, proteoglycans pathway in cancer, metabolic, and focal adhesion pathway as these pathways are widely investigated to play a role in cancer. Therefore, targeting these hub proteins could be one possible approach for targeting the up/downstream pathways and biological processes and overcome pan-cancer drug resistance. The two important key hub proteins, FN1 and CD44, have interconnectivity with RAS and PI3K/Akt signaling pathways,

taken from KEGG Pathways database (*https://www.genome.jp/kegg/pathway.html*), shown in (**Fig. 7)** which might induce drug resistance.

Our big data analyses corroborate the same, albeit with a different drug and functioning in a pan-cancer context. This study of gene co-expression and PPI network might provide key driver protein-coding genes, which may be useful in further studies to improve drug sensitivity in pan-cancer therapy.

Table 5: List of hub nodes common between gene co-expression and PPI network for selected five drugs.

Drug name	Cluster	Common hub nodes between gene co-expression and cluster PPI network
Ponatinib	Cluster 1	-NA-
	Cluster 2	TYR, PMEL, MLANA, EDNRB
	Cluster 3	**FN1, CD44**, MMP1, **TIMP1**, MMP14, **SPARC**, **SNAI2**, VEGFC, TIMP3
Foretinib	Cluster 2	**FN1, TIMP1, CD44**
Selumetinib	Cluster 2	LCP2, FYB, IL7R, CD38
	Cluster 3	**FN1**, **CD44**, **TIMP1**, CCND1, CAV1, PTGS2, **SNAI2**, LGALS1
Trametinib	Cluster 1	**CD44**, CCL2, ANXA1, AHR
	Cluster 2	**FN1**, **SPARC**, CYR61, CTGF, ITGB5, LAMC1, TGFB1
CI-1040	Cluster 2	KRT7, KRT8, KRT19, EPCAM

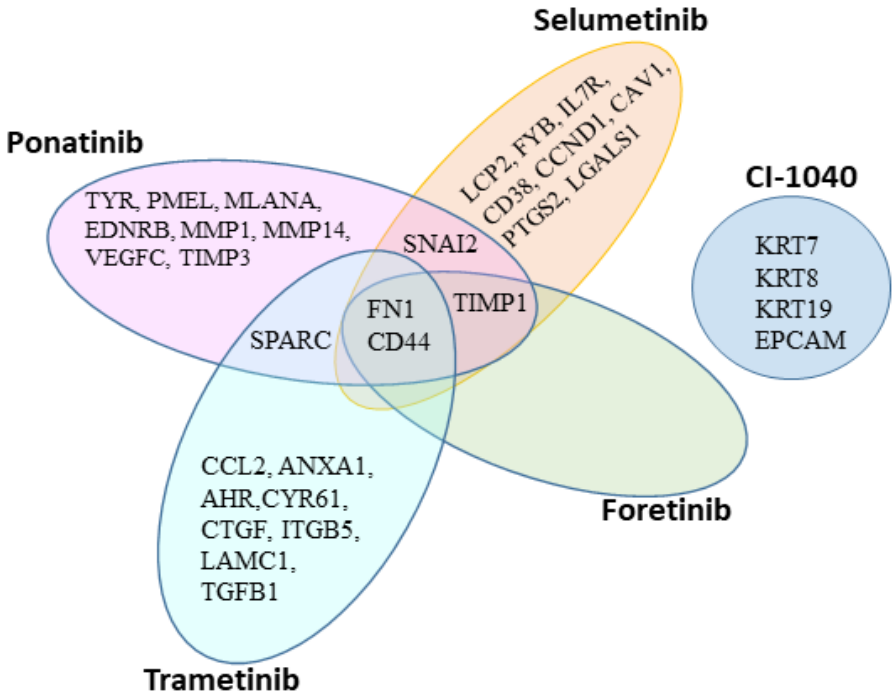


Figure 6: Venn diagram representing common hub protein-coding genes identified across the drugs.

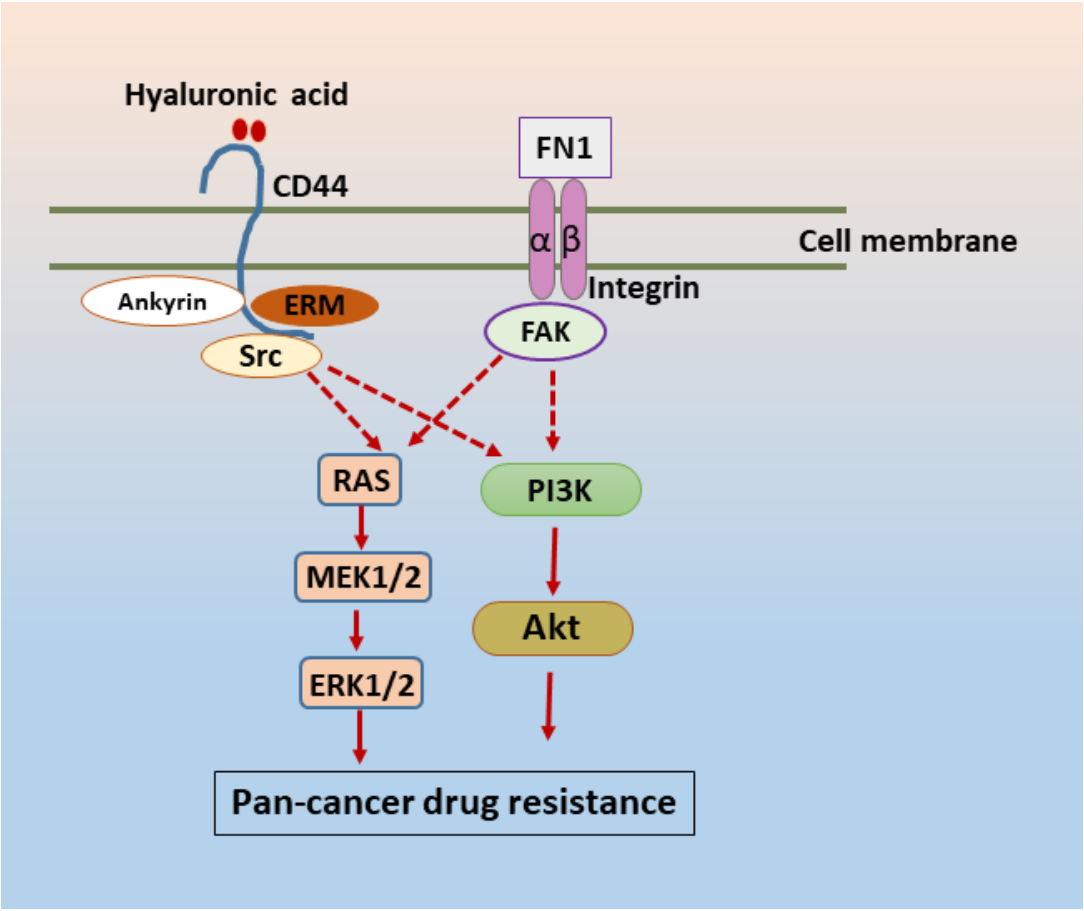


Figure 7: Schematic representation key genes CD44 and FN1 involved in RAS and PI3K/Akt signaling pathways to induce pan-cancer drug resistance. Several other intermediate proteins involved in the signaling cascade denoted by dashed arrows.

5.4 References

1. Toor, R., Chana, I. (2021) Network Analysis as a Computational Technique and Its Benefaction for Predictive Analysis of Healthcare Data: A Systematic Review. *Arch Computat Methods Eng* 28, 1689–1711. https://doi.org/10.1007/s11831-020-09435-z
2. Pablo Porras Millan, David Ochoa, Matt Z. Rogon, Luz Garcia Alonso, Denes Turei. (2020) Network analysis of protein interaction data-an introduction. https://www.ebi.ac.uk/training/online/courses/network-analysis-of-protein-interaction-data-an-introduction/#vf-tabs__section--contents.
3. Yan W, Xue W, Chen J, Hu G. (2016) Biological Networks for Cancer Candidate Biomarkers Discovery. Cancer Inform. 15(Suppl 3):1-7. doi: 10.4137/CIN.S39458. PMID: 27625573; PMCID: PMC5012434.
4. Sawyers CL. (2008) The cancer biomarker problem. *Nature.* 452(7187):548-52. doi: 10.1038/nature06913. PMID: 18385728.
5. Bolton EM, Tuzova AV, Walsh AL, Lynch T, Perry AS. (2014) Noncoding RNAs in prostate cancer: the long and the short of it. *Clin Cancer Res.* 20(1):35-43. doi: 10.1158/1078-0432.CCR-13-1989. PMID: 24146262.
6. Kreeger PK, Lauffenburger DA. (2010) Cancer systems biology: a network modeling perspective. *Carcinogenesis.* 31(1):2-8. doi: 10.1093/carcin/bgp261. PMID: 19861649; PMCID: PMC2802670.
7. Wang J, Zuo Y, Man YG, et al. (2015) Pathway and network approaches for identification of cancer signature markers from omics data. *J Cancer.* 6(1):54-65. doi: 10.7150/jca.10631. PMID: 25553089; PMCID: PMC4278915.
8. Zhang DY, Ye F, Gao L, et al. (2009) Proteomics, pathway array and signaling network-based medicine in cancer. *Cell Div.* 28;4:20. doi: 10.1186/1747-1028-4-20. PMID: 19863813; PMCID: PMC2780394.
9. Wang J, Zhang Y, Marian C, Ressom HW. (2012) Identification of aberrant pathways and network activities from high-throughput data. *Brief Bioinform.* 13(4):406-19. doi: 10.1093/bib/bbs001. PMID: 22287794; PMCID: PMC3404398.
10. van Dam S, Võsa U, van der Graaf A, et al. (2018) Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform*. 19(4):575-592. doi: 10.1093/bib/bbw139. PMID: 28077403; PMCID: PMC6054162.
11. Garner AL, Janda KD. (2011) Protein-protein interactions and cancer: targeting the central dogma. *Curr Top Med Chem*. 11(3):258-80. doi: 10.2174/156802611794072614. PMID: 21320057.

Chapter-6

Objective-4: LncRNAs-TFs-Hub genes (at mRNA level) interaction regulatory network analysis in order to identify likely master regulators of our identified biomarkers

6.1 Introduction

Genomes are transcribed extensively, which leads to the creation of more than thousands of non-coding RNAs, including lncRNAs. LncRNAs are described as RNAs extended with more than 200 nucleotides without having a significant open reading frame and therefore do not have the ability to translate (encode) into functional proteins. This broad definition includes many different types of transcripts, but each differs in their biogenesis and genomic origin (***Statello et al. 2021***). A Human Genome Project (GENCODE) database suggests that more than 16,000 lncRNA genes are present in the human genome; however, other database estimates indicate that there could be more than 100,000 human lncRNAs (***Uszczynska-Ratajczak et al. 2018; Statello et al. 2021***). Most of these lncRNAs mainly generated by RNA polymerase II (RNA Pol II), whereas some are by other RNA polymerases. These lncRNAs are transcribed from various region of genome, such as intergenic (lincRNAs) and intronic regions of genes. They can also be either sense or antisense transcripts that may coincide with other coding or non-coding genes. It is important to note that some of promoter and enhancer regions are also transcribed into promoter upstream transcripts and enhancer RNAs (eRNAs), respectively. The resulting lncRNAs are often capped at 5′ region by 7-methyl guanosine (m^7G), polyadenylated at 3′ region and spliced in similar manner as mRNAs **(Fig. 1**) (***Fang & Fullwood 2016*)**. The majority of lncRNAs are tended to be localized in the cytoplasm. However, some of the lncRNAs can be reside in both cytoplasm as well as nucleus to which they seem to be predominantly localized (***Bánfai et al. 2012; Derrien et al. 2012).***

6.1.1 Functional role of LncRNAs

The number of characterized lncRNAs is growing and they play a major role in negative or positive gene expression regulation in development and human disease, including cancer. In malignant tumors, lncRNAs mostly play a crucial role in regulating biological and cellular processes, such as cell proliferation, migration, invasion, metastasis, epithelial-mesenchymal

transition (EMT), cell apoptotic death, cell cycle, invasion and also in drug resistance (***Lecerf et al. 2019; Taniue & Akimitsu 2021***). LncRNAs are an emerging new molecular players in the cancer paradigm with potential roles in both tumor-suppressive and oncogenic pathways. These novel non-coding genes frequently show altered expression in human cancers, although the biological functions of the majority of lncRNAs are not fully understood (***Gibb et al. 2011***). LncRNAs are widely involved in nearly all the steps of a life cycle of genes and modulate through a variety of mechanisms that rely on interactions with multiple molecules. Several lncRNAs function to regulate gene expression through different molecular actions, including chromatin remodeling, transcriptional regulation, and posttranscriptional processing such as mRNA splicing, stability and translation or microRNA (miRNA) sponging (**Fig. 2**) ***(Hauptman & Glavač 2013).***

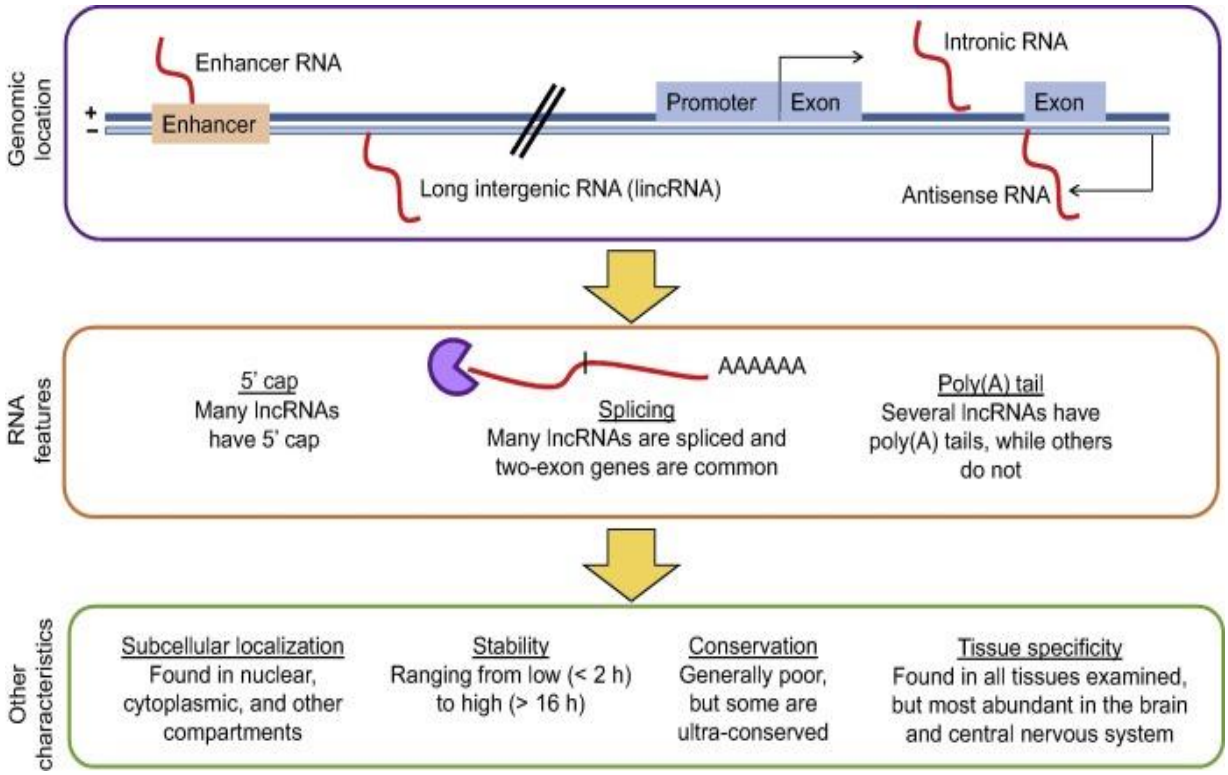


Figure 1: Diagrammatic representation of general characteristics of lncRNA (Modified from *Fang & Fullwood, 2016*)

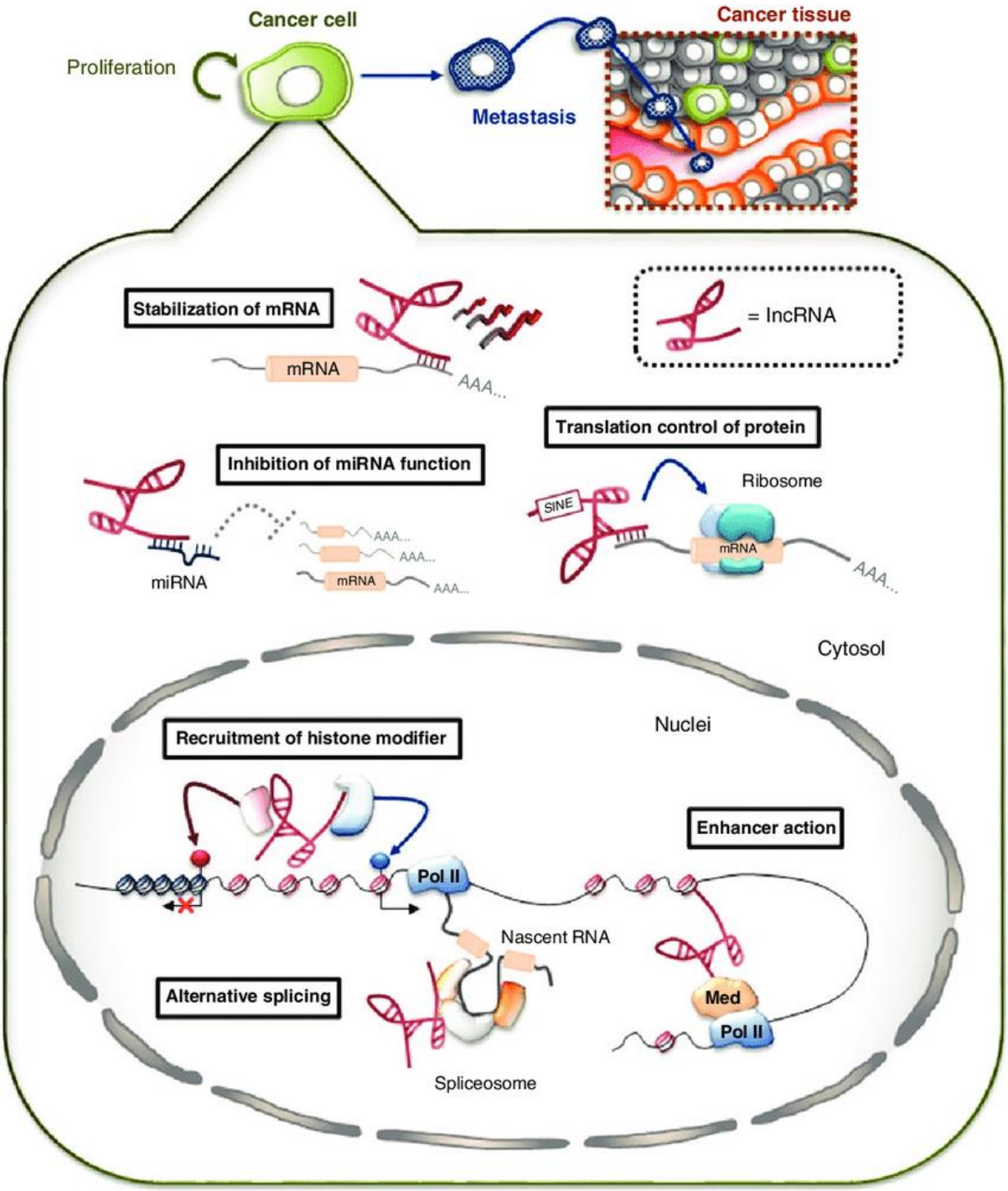


Figure 2: LncRNA molecular function in gene expression and the regulation mechanism. (Modified from *Joshi et al., 2019*)

6.1.2 LncRNA's role in drug-resistant cancer

LncRNAs widely alter gene expression in a wide range of human cancer types (***Bermudez et al. 2019, Clark & Mattick 2011; Saleembhasha & Mishra, 2018***). Many lncRNAs, including *PVT1, SNHG11* and *MIR22HG* are deduced to be vital regulatory molecules, have been implicated to function as master regulators of overexpressed several common coding genes, and are widely involved in primary pan-cancer development (***Saleembhasha & Mishra 2019***). Further, it was recently reported that many lncRNAs have a significant impact on cancer drug resistance in many cancer types like liver, breast, bladder, gastric, prostate, lung, and colorectal cancer (***Bermudez et al. 2019; Zinovieva et al. 2018***). There are some reported LncRNAs, such as *TP73-AS1,* that induce Temazolamide (TZM) resistance in glioblastoma cancer stem cells by altering ALDH1A1 expression (***Mazor et al. 2019***), while *HOTAIR1* promotes tamoxifen resistance in breast cancer through the activation of estrogen receptor (ER) signaling (***Xue et al. 2015***). However, the exact molecular mechanism of lncRNAs on cancer drug resistance has not been fully characterized.

In order to understand the molecular mechanism of lncRNA function in the drug-resistant pan-cancer system, we constructed and analyzed a comprehensive lncRNA-TFs-hub gene interaction regulatory network to deduce key master regulators (lncRNAs) of our identified coding hub genes (biomarkers) in mutant NRAS-harbouring drug-resistant pan-cancer systems.

6.1.3 Regulatory network Properties

Besides their high connectivity, hub genes/proteins are often described as being designated by other properties of the network, including degree and centrality being the most important, which refers to their central position in relationship to other proteins in the network (***Vandereyken et al. 2018).*** To understand the global gene regulatory interaction pattern in a cell, topological parameters such as betweenness centrality, clustering coefficient, neighborhood connectivity and node-degree help us to estimate a node's degree and also assess

the network dynamics by adding or subtracting nodes (genes) in a different biological context. Regulatory gene networks are used to understand how genes work as a network in biological pathways (**Fig. 3).**

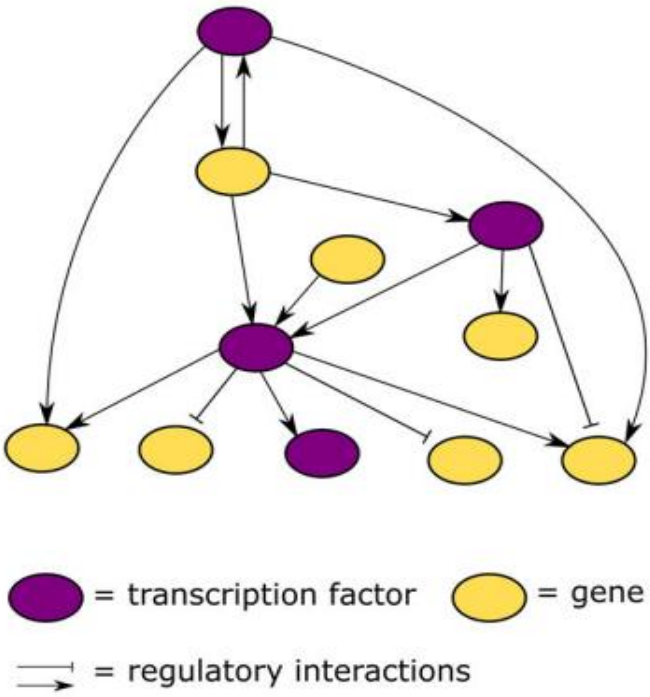


Figure 3: Schematic representation of gene regulatory network. Nodes represent genes or proteins (blue/yellow color circles) in the gene regulatory network and lines (edges) between them indicate regulatory interactions (*Modified from Vandereyken et al., 2018***).**

6.2 Materials and Methods

6.2.1 LncRNAs, TFs, Driver genes, interaction regulatory network data collection

We retrieved interaction data from the ORTI database to construct a TF-Driver genes interaction regulatory network, which consists transcriptional interactions data validated by experimental methods and text mining data from human and mouse source. ORTI database derived data from the high throughput ChIP-sec data and other database sources as well as from the literature, which harbours TF-TG (driver genes) interactions with data. Another data of lncRNAs- driver genes and lncRNAs-TFs regulatory interaction data have been retrieved from the extensive literature search. These regulatory interactions data were widely included in the

regulatory network, and it has a collection of lncRNA-target regulatory interaction validated from high throughput and low throughput experimental methods.

6.2.2 Analysis of regulatory network

To construct and analyze the regulatory interaction network, data from these two types of interaction were imported into Cytoscape 3.8.2 and then merged into a comprehensive/master network. Network analyses was done using quantitative directed method on the lncRNAs-TFs-mRNA regulatory interaction network, and lncRNAs/TFs are used as a source to target identified hub genes. We identified critical non-coding regulators (bottleneck hub node) of our identified coding biomarkers using topological network parameters; betweenness centrality and node degree (outdegree).

6.2.3 Sub-network analysis

From the master regulatory network we further generated a regulatory subnetwork to assess direct or indirect regulation of coding biomarker genes through the identified key regulator lncRNA. We also predicted lncRNA-driver genes' mRNA interaction by using a RNA-RNA interaction database, which contains huge data on lncRNA-lncRNA and lncRNA-mRNA interactions. This database provides information about the lncRNA interaction site on coding gene mRNA along with their binding energy.

6.3 Results

6.3.1 Gene-regulatory modules: LncRNA, Transcription factor (TF), protein-coding gene (hub genes) interaction regulatory network

During the identification of druggable targets (genes/proteins) in drug resistance, understanding the regulatory mechanisms occurring to regulate these druggable genes/proteins is important for further development as a drug target. Besides proteins, lncRNAs are imminent regulatory mechanisms in cancer drug resistance. During the investigation of key lncRNAs modulating our driver genes/proteins found across the multiple drugs that are typical for co-expression and PPI network, an interaction network between lncRNA-TF-hub genes was designed and analyzed. The interaction data for TFs and identified key driver genes were retrieved from the ORTI database. An extensive literature search was conducted for lncRNAs interacting with hub genes (*CD44, FN1, TIMP1, SPARC* and *SNAI2*) and TFs. A complex regulatory interaction network was generated which has a total of 91 nodes (genes/proteins) and 125 edges, together with 5 hub protein-coding genes, 38 TFs and 48 lncRNAs (**Fig. 4)**. Two topological parameters, betweenness centrality and out-degree, were set as criteria for the selection of hub node from the regulatory network. By examining the regulatory network, we established that *MALAT1* possessed the highest node outdegree and betweenness centrality among the lncRNAs. YBX1, EGR1 and AR were the TFs with the highest out-degree and among these top three EGR1 possessed the highest betweenness centrality .The other lncRNAs such as *lincRNA-p21* and *HOTAIR* were found to have the highest betweenness centrality and out-degree, respectively (**Table 1**). From our regulatory network analysis, we also observed that AR, YBX1, and EGR1 may regulate *FN1* and *CD44*; AR and YBX1 may regulate *SPARC;* YBX1 and EGR1 may regulate *TIMP1*.

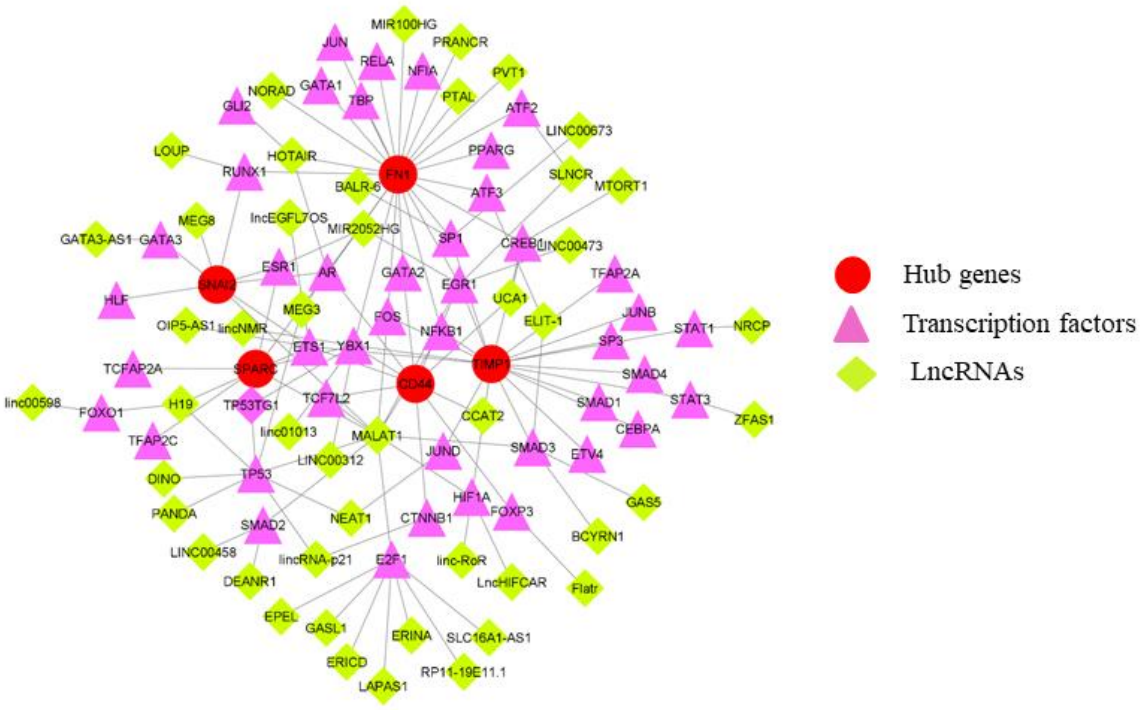


Figure 4: A master regulatory network of LncRNA-TF-Driver genes. Different types of regulatory interaction between lncRNAs, TFs and driver genes are depicted in this integrated network.

Table 1: LncRNAs-TFs-Genes (hub genes) regulatory network directed quantitative analyses result based on outdegree and betweenness centrality**.**

Sr No.	Gene name	Outdegree	Sr No.	Gene name	BetweennessCentrality
	MALAT1	8	1	*MALAT1*	0.336730123
2	EGR1	4	2	HIF1A	0.274132139
3	AR	4	3	TP53	0.205823068
4	YBX1	4	4	EGR1	0.174412094
5	NFKB1	3	5	YBX1	0.168868981
6	FOS	3	6	*lincRNA-p21*	0.138969765
7	ETS1	3	7	ETS1	0.063493841
8	*HOTAIR*	3	8	CTNNB1	0.051175812
9	*H19*	3	9	*SLNCR*	0.044456887
10	HIF1A	2	10	SP1	0.039193729

6.3.2 EGR1 and *MALAT1* sub-network analysis

Further, from the regulatory sub-network, we observed that two of the driver genes, including *SPARC* and *SNAI2*, and six transcription factors, including EGR1 might be directly regulated by lncRNA *MALAT1*. It was also observed that three driver genes, including *FN1*, *CD44*, and *TIMP1* were indirectly regulated by *MALAT1* through EGR1 (**Fig. 5A)**. *MALAT1* and EGR1 regulate each other through a two-way (mutual) interaction as depicted from the ENCODE dataset retrieved from the harmonozome database. It was shown that *MALAT1* is a transcriptional target of EGR1. And the interaction between EGR1 and *MALAT1* was determined by ChIP-Seq data.

We were interested to check if *MALAT1* directly interacted with the above transcripts. Therefore, we used RNA-RNA interaction database and predicted these interactions (**Fig. 5B**). We found that *MALAT1* and hub genes' interactions occur at different sites in mRNAs, such as 5´UTR, 3´UTR and CDS region. For example, *MALAT1* was observed to interact with *TIMP1* and *SNAI2* at the CDS region, *FN1* at the 5´UTR, and at the 3´UTR regions of *CD44* and *SPARC* (**Table 2**).

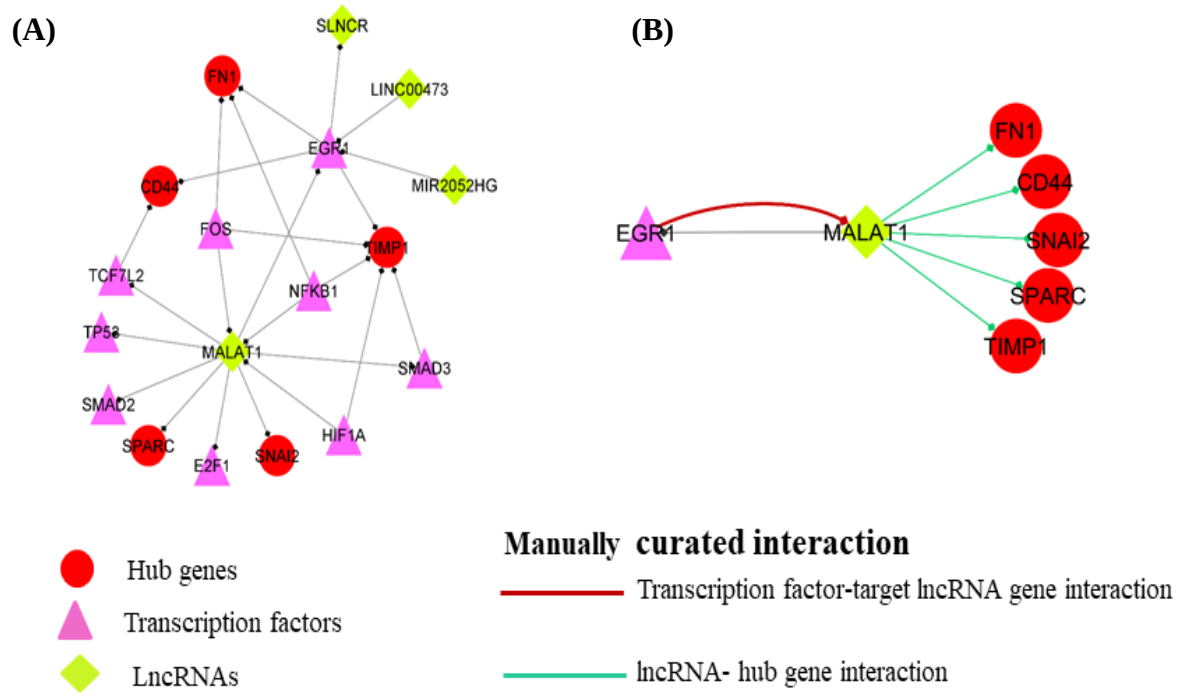


Figure 5: EGR1 and *MALAT1* subnetwork from the master regulatory network. **(A)** EGR1, which regulates *FN1*, *CD44* and *TIMP1*, being controlled by MALAT1. **(B)** *MALAT1*-mRNA interaction (hub genes, green-colored edge) and EGR1-*MALAT1* interaction (red-colored edge) were obtained from the lncRNA-mRNA interaction database and harmonizome database, respectively.

Table 2: Predicted lncRNA interaction site on mRNA of coding hub genes.

LncRNA	**Hub genes**	**Interaction site**
MALAT1	*FN1*	5'UTR
	CD44	3'UTR
	TIMP1	CDS
	SPARC	3'UTR
	SNAI2	CDS

6.3.3 *Cis* and *trans*-regulatory action of *MALAT1* on key driver genes

In order to illustrate the *cis* and *trans*-regulatory action of *MALAT1* on the target hub genes, we tried to investigate the genomic coordinates of the coding hub genes and *MALAT1* from the NCBI Gene database (*https://www.ncbi.nlm.nih.gov/gene/*). It was observed that, due to their

same chromosomal location (**Table 3**), *MALAT1* might regulate *CD44* in the *cis*-regulatory mode of action. While due to their different genomic location, *MALAT1* may regulate *FN1, TIMP1, SNAI2* and *SPARC* in the *trans*-regulatory mode of action.

Table 3: Genes with their chromosomal location.

Genes	**Chromosomal location**	**Gene ID**
MALAT1	11q13.1	378938
FN1	2q35	2335
CD44	11p13	960
TIMP1	Xp11.3	7076
SNAI2	8q11.2	6591
SPARC	5q33.1	6678

6.4 References

1. Statello L, Guo CJ, Chen LL, et al. (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol*. 22(2):96-118. doi: 10.1038/s41580-020-00315-9. PMID: 33353982.
2. Fang Y, Fullwood MJ. (2016) Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics Proteomics Bioinformatics*. 14(1):42-54. doi: 10.1016/j.gpb.2015.09.006. PMID: 26883671.
3. Hauptman N, Glavač D. (2013) Long non-coding RNA in cancer. *Int J Mol Sci*. 26;14(3):4655-69. doi: 10.3390/ijms14034655. PMID: 23443164.
4. Joshi P, Katsushima K, Zhou R, et al. (2019) The therapeutic and diagnostic potential of regulatory noncoding RNAs in medulloblastoma. *Neurooncol Adv*. 1(1):vdz023. doi: 10.1093/noajnl/vdz023. PMID: 31763623.
5. Uszczynska-Ratajczak B, Lagarde J, Frankish A, et al. (2018) Towards a complete map of the human long non-coding RNA transcriptome. *Nat Rev Genet.* 19(9):535-548. doi: 10.1038/s41576-018-0017-y. PMID: 29795125; PMCID: PMC6451964.
6. Bánfai B, Jia H, Khatun J, Wood E, et al. (2012) Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* 22(9):1646-57. doi: 10.1101/gr.134767.111. PMID: 22955977.
7. Derrien T, Johnson R, Bussotti G, et al. (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 22(9):1775-89. doi: 10.1101/gr.132159.111. PMID: 22955988.
8. Gibb EA, Brown CJ, Lam WL. (2011) The functional role of long non-coding RNA in human carcinomas. *Mol Cancer*. 10:38. doi: 10.1186/1476-4598-10-38. PMID: 21489289.
9. Taniue K, Akimitsu N. (2021) The Functions and Unique Features of LncRNAs in Cancer Development and Tumorigenesis. *Int J Mol Sci*. 22(2):632. doi: 10.3390/ijms22020632. PMID: 33435206.
10. Lecerf C, Le Bourhis X, Adriaenssens E. (2019) The long non-coding RNA H19: an active player with multiple facets to sustain the hallmarks of cancer. *Cell Mol Life Sci*. 76(23):4673-4687. doi: 10.1007/s00018-019-03240-z. PMID: 31338555.
11. Vandereyken K, Van Leene J, De Coninck B, et al. (2018) Hub Protein Controversy: Taking a Closer Look at Plant Stress Response Hubs. *Front Plant Sci*. 9:694. doi: 10.3389/fpls.2018.00694. PMID: 29922309; PMCID: PMC5996676.

Chapter-7

Objective-5: Database search of FDA-approved drugs targeting the identified hub gene/s for drug repurposing studies and in silico virtual screening of drugs against target protein

7.1 Introduction

Despite the many improvements in cancer treatment that have been made over the years, cancer still remains one of the leading causes of death in the world. It is one of the most common and severe health issues worldwide due to its high mortality and incidence rates (***Sung et al. 2021***). The development of drug resistance increases the mortality rate among cancer patients, which is one of the biggest challenges to getting better cancer treatment. Although there are many therapeutic strategies are available to treat cancer, the currently used therapeutic schemes are sometimes accompanied by drug-resistance development in the malignant tumor cells, resulting in a decline in the effectiveness of the therapeutic agents (***Falvo et al. 2021; Bukowski et al. 2020***). In order to overcome this phenomenon, it is necessary to develop new therapies or new anti-cancer drugs to overcome it. Developing new drugs is a lengthy and costly process involving clinical trials that often fail in the early phases of development. Developing new drugs is a long, expensive process, and clinical trials are often rejected in the early stages of development. An approach to encounter these disadvantages is known as drug repurposing, which involves finding a drug that has been approved for another purpose but still meets its original criteria (***Rodrigues et al. 2022***).

7.1.1 Drug repurposing strategies

Due to the potential for discovering new uses for existing drugs, the concept of drug repurposing has attracted considerable attention, it primarily includes approved, pre-clinical, discontinued, abandoned and experimental or investigational drugs. In the pharmaceutical research and industry for developing new drugs uses the repurposing method due to its high efficiency in saving time and economic over the conventional de novo approaches ***(Rudrapal et al. 2020)***. Drug repurposing is also known as drug repositioning, drug reprofiling and therapeutic switching ***(Jarada, et al. 2020)***. In recent years, several successes of repurposing drugs have brought worldwide attention to the old drug space for their potential off-target

effects that may be advantageous to certain kinds of diseases, such as cancer. Since existing drugs have well-established dose regimens and have already been used in humans with favourable pharmacodynamics (PD) and pharmacokinetics (PK) properties along with tolerable side effects, making old drugs are valuable sources of new therapeutic drug discovery ***(Shim & Liu 2014)***. There are two approaches of drug repurposing; activity based (experimental) and *in silico* (computational) approaches ***(Oprea & Overington 2015)***. And the computational approach has been categorized into ligand-based, target-based, and machine learning-based approaches. However, *in silico* based methods identify potential bioactive molecules based on the molecular interaction of drug and protein molecules.

7.1.2 Repurposed drug for cancer

There are some previously repurposed drugs for cancer, such as; Metformin was approved for type 2 diabetes which is currently in trial phase III/IV for cancer ***(Zhe Zhang et al. 2020)***. Rapamycin is an inhibitor of mTORC1 used as an immunosuppressant, but due to ineffectiveness, it was repurposed and approved for renal cell carcinoma treatment in 2007 ***(Malizzia & Hsu 2008)***. Itraconazole, an antifungal drug also repurposed as an anticancer agent by using *in silico* approach ***(Dhorje et al. 2020; Rudrapal et al. 2020)***. In recent years, during the coronavirus disease 2019 (COVID-19) pandemic, several drugs have been repurposed against SARS-CoV-2 due to which many drugs were approved for the COVID-19 patient's treatment ***(Chakraborty et al. 2021; Elmezayen et al. 2021)***.

Therefore, we were interested and aimed to target these key biomarker (CD44) through *in silico* drug repurposing approach to improve drug sensitivity in pan-cancer and to identify important residues of protein interacting with drug molecules.

From our previous objectives (chapter 5), we identified CD44, FN1, TIMP1, SNAI2, and SPARC **(Fig. 6)** as biomarkers in mutant NRAS-harbouring pan-cancer drug resistance across

multiple drugs from gene co-expression and protein-protein interaction network study and we detected CD44 and FN1 as our major key biomarkers across four drugs. Expression of both of these key biomarkers results in a multitude of cellular functions such as migration, proliferation, tumor microenvironment, adhesion, and also induced drug-resistant in cancers. Identified biomarkers have been reported to be highly expressed in most of the tumors and also expressed in drug-resistant cancer to support various biological processes and signaling pathways involved signal transduction, proteolysis, cell adhesion, proteoglycans in cancer and PI3K/Akt-signaling pathway. In our study, we also have observed that these biomarkers genes were significantly up-regulated in drug-resistant pan-cancer.

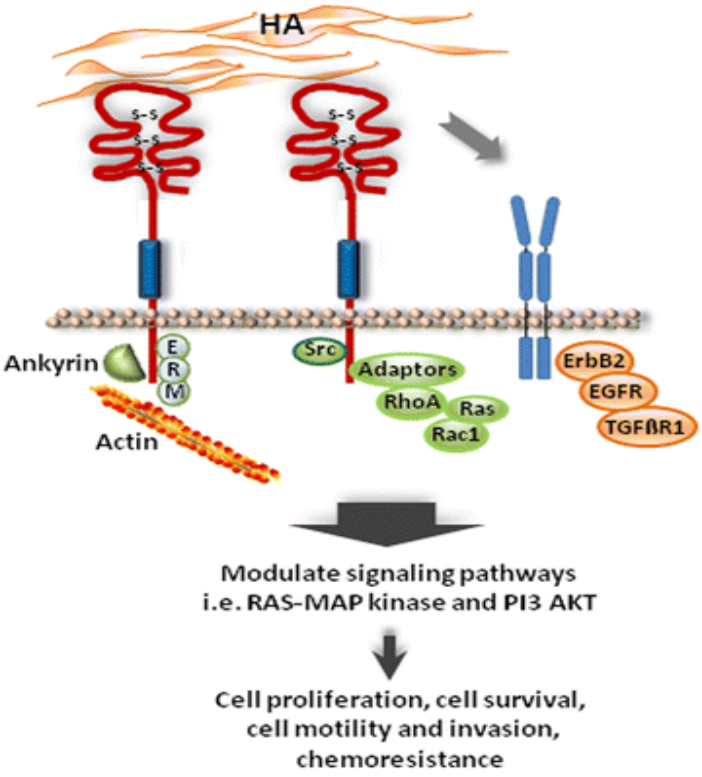
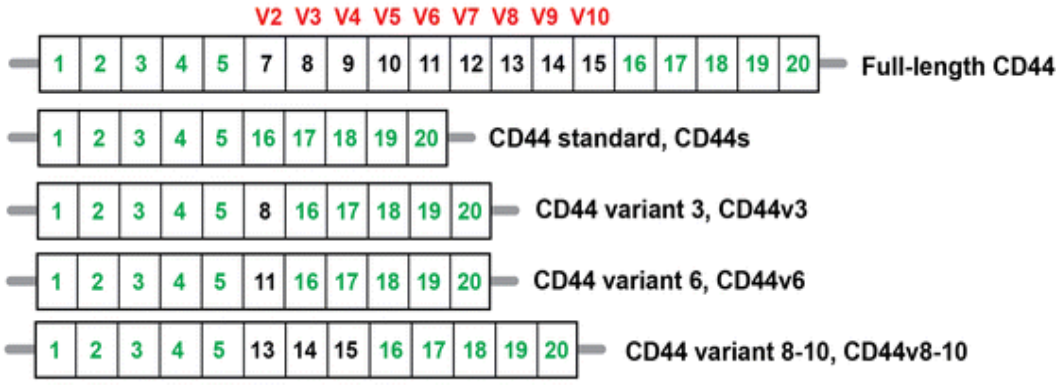


Figure 1: CD44 role in signaling pathways. HA binding to the extracellular domain of CD44 activates various downstream signaling pathways, including MAPK and PI3K/Akt pathways through the cytoplasmic domain to regulate several biological and cellular processes. (*Modified from Cortes-Dericks et al., 2017*).

7.1.3 Cluster of differentiation 44

Cluster of differentiation 44 **(**CD44; 90 kDa) is a widely distributed cell surface non-kinase transmembrane protein, an integral part of the extracellular matrix that is involved mostly in cell adhesion, migration, metastasis and also activates various signaling pathways, including RAS-MAPK and PI3k/Akt signaling and also induces chemoresistance in cancer **(Fig. 1)** (***Jamison et al. 2010; Herishanu et al. 2011; Cortes-Dericks et al. 2017***). It is activated by the binding of hyaluronic acid (HA) at the N-terminal region of the extracellular domain and HA is the most common activating endogenous (linear polysaccharide) ligand molecule of CD44 ***(Cortes-Dericks et al. 2017).*** CD44, a proteoglycan, is also functionally involved in the binding and presentation of growth factor and chemokine. The extracellular region of *CD44* gene contains 20 exons, 10 types of alternative splicing variants give rise to multiple *CD44* isoforms such as *CD44*s and *CD44*v (**Fig. 2A&B**) (***Bajorath et al., 1998***). Exon 1-17 form the extracellular domain (exons 1-5 are HA binding domain conserved across CD44s and their variants), exon 18 give rise transmembrane domain, and exons 19-20 are responsible for forming the cytoplasmic domain ***(Xu et al. 2015).***

(A)



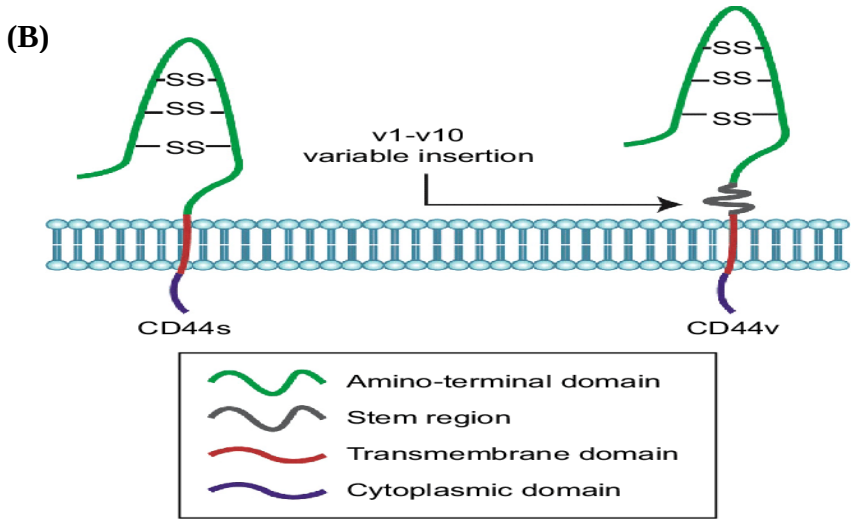


Figure 2: *CD44* gene illustration and alternatively spliced variants isoforms and key protein domain structure. (A) *CD44* gene contains 20 exons, out of which some exons form a constant region in every *CD44* variants exons (red bars) and protein (green bars) and are selected by alternative splicing. CD44v protein isoforms are the result of alternative splicing. **(B)** CD44 glycoprotein is composed of extracellular domains; HA binding domain (green), a variable domain (grey), a transmembrane domain (red), and a cytoplasmic tail (blue). (*Modified from Chen et al., 2018; Xu et al., 2015)*

7.2 Materials and methods

7.2.1 hCD44 and mCD44 protein sequence and structure alignment

As the 3D crystal complex structure of hCD44 with HA is not available in the PDB database, we retrieved the human CD44 protein sequence from UniprotKB (code: P16070) in FASTA format. We performed a pairwise protein sequence similarity search using the NCBI BLASTp server to determine the HA binding domain similarity with CD44 protein sequence available in the PDB database. We also assessed the structural similarity between human and mouse CD44 HABD (hCD44 PDB ID- 4PZ4 & mCD44 PDB ID- 2JCQ) using PyMol.

7.2.2 Protein preparation

Human CD44 protein HABD (PDB ID- 4PZ4) with resolution 1.60 Å was retrieved from the RCSB-PDB database and all heteroatoms and water molecules in the PDB file were removed manually. Energy-minimization of the protein structure was done using the Swiss-PDB viewer

tool to get the stable and low-energy conformation state of the protein. Protein was prepared using AutoDock Tools v1.5.6, in which water molecules were removed, hydrogens (only polar) were added and Kollman charges were assigned to the protein. The prepared protein was saved in pdbqt file format. The Druggability of ligand binding pocket was predicted using the online server PockDrug by estimation methods Prox 5.5, where it used holo-protein for pocket druggability prediction.

7.2.3 Ligand preparation

A library of 1615 chemical structures (only FDA-approved drugs) was retrieved from the ZINC15 database in 3D SDF file format and converted into PDB format and then split into individual drug PDB files using an open Babel suite. All the ligands were prepared using Auto-DockTools v1.5.6. Hydrogens and Gasteiger charges were added to the ligands and the nonpolar hydrogens were merged. The prepared ligands were saved in pdbqt file format.

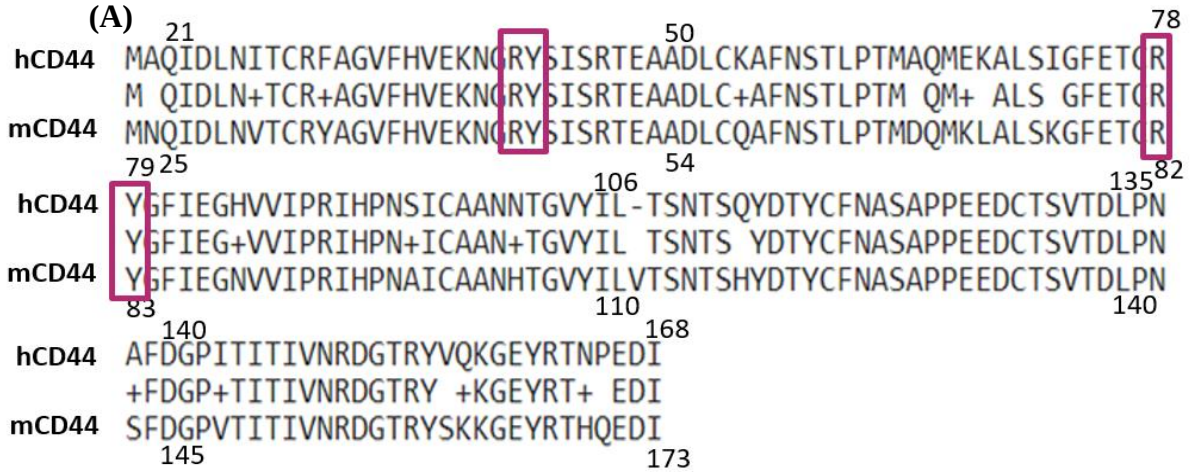
7.2.4 Virtual screening through molecular docking

Grid box was set at 60, 60, and 60 and center with x= 12.143, y=-6.510, and z=6.243, with default spacing 0.375Å to include all the present amino acid residues of ligand-binding pockets of the receptor. Virtual screening through molecular docking was performed in the Autodock vina program using Perl script. Final docked conformations were obtained using the AutoDock vina. The obtained lowest energy docking conformations and orientations were subjected to energy minimization. The resultant protein-ligand complexes were visualized and analyzed by using PyMol and Discovery studio4 (BIOVIA).

7.3 Results

7.3.1 hCD44 and mCD44 similarity assessment to identify HA binding cavity residues

With no hCD44 protein (in complex with) bound hyaluronic acid, we performed HABD protein sequence and structural similarity analysis of the mCD44 protein. Pairwise sequence alignment showed that hCD44 protein HABD shares around **87%** sequence identity with mCD44 HABD protein (PDB ID- 2JCQ) **(Fig. 3A)** and structural alignment showed significant similarity between human and mouse CD44 protein with RMSD value **0.311** within an acceptable value for protein similarity due to the presence of conserved amino acids residues in human and mouse HABD. From this similarity assessment, we observed that hCD44 and mCD44 have the same binding site for hyaluronic acid at the N-terminal domain **(Fig. 3B)**. Some previous mutagenesis studies have identified Arg41, Tyr42, Arg78 and Tyr79 are, as crucial residues in hCD44 for HA interaction stabilization (***Peach et al. 1993; Bajorath et al. 1998***). We further used HA binding pocket residues coordinates for grid box generation.



(B)

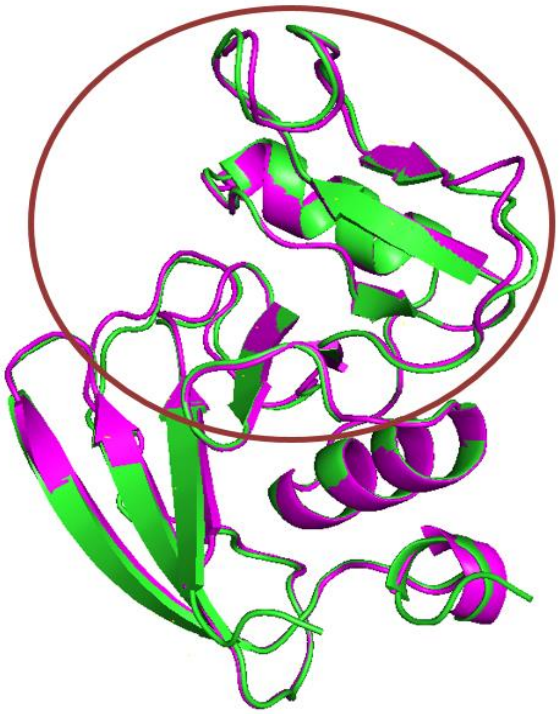


Figure 3: CD44 HABD protein sequence and structure alignment. **(A)** hCD44 HABD protein sequence similarity with mCD44 HABD. Highlighted (red) amino acid residues are key for HA binding to CD44. **(B)** hCD44 and mCD44 protein HABD 3D structure alignment, hCD44 (green) and mCD44 (magenta) and red circled portion is HA binding pocket.

7.3.2 HA binding pocket druggability prediction

We further assessed the pocket druggability of the HA binding pocket of hCD44 using the Prox5.5 method in the online PockDrug server and the HA binding pocket showed **0.89** druggability, suggesting that HA binding pocket is highly druggable with high affinity. The volume hull of the pocket was 5488.22 (**Fig. 4).**

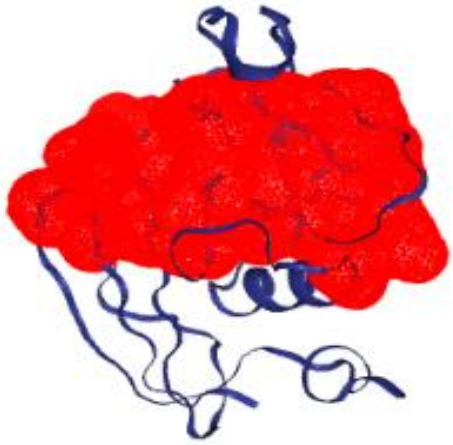


Figure 4: Schematic illustration of pocket druggability. Red covered volume is HA binding pocket in CD44.

7.3.3 Ligand binding site analysis through molecular docking

For the *in silico* virtual screening of 1615 FDA-approved drug molecules against our selected target protein CD44 through molecular docking, we utilized the AutoDockVina, which is one of the most commonly used docking software with an effective scoring function. After molecular docking, we selected top 16 protein-drug complexes with high binding affinity for further analysis (**Table 1).** Among these 16, only seven drugs bind to HA binding pocket of hCD44 (**Table 2**). As there is no known allosteric binding site at CD44 protein and other 9 drugs were binding at different sites than HA binding cavity, they were excluded from further analysis.

Further, we analyzed seven protein-drug complexes. **Fig 5A** shows the docked drug molecule binding at HA binding cavity (e.g., Glecaprevir) and HA docked with hCD44 protein and binding in the cavity with binding affinity -7.6 kcal/mol (**Fig. 5B**).

Table 1: Top 16 drugs selected from *in silico* virtual screening of 1615 drugs with their binding affinity.

Sr. No	ZINC ID	Name	Binding energy (kcal/mol)
1	ZINC000003813047	Oxandrolone	-9.3
2	ZINC000164528615	Glecaprevir	-9.3
3	ZINC000003860453	Ak-Fluor	-9.1
4	ZINC000052955754	Ergotamine	-9.0
5	ZINC000100378061	Naldemedine	-9.0
6	ZINC000252286875		-9.0
7	ZINC000203757351	Paritaprevir	-8.9
8	ZINC000004212851	Lokara	-8.7
9	ZINC000003978005	Dihydroergotamine	-8.6
10	ZINC000000968264	Cyproheptadine	-8.6
11	ZINC000100013130	Midostaurin	-8.5
12	ZINC000004097308	Cordran	-8.5
13	ZINC000005764759	Methylnaltrexone	-8.5
14	ZINC000169289767	Trypan Blue	-8.5
15	ZINC000003874185	Mefloquine	-8.5
16	ZINC000252286876		-8.5

Table 2: List of top 16 selected drug molecules binding at HA binding cavity or alternate cavity of CD44.

Drug molecules bind at HA binding pocket	Drug molecules bind at the alternate binding pocket
Glecaprevir, Ergotamine, Naldemedine, Midostaurin, Trypan Blue, ZINC000252286875, ZINC000252286876	Oxandrolone, Ak-Fluor, Paritaprevir, Lokara, Dihydroergotamine, Cyproheptadine, Cordran, Methylnaltrexone, Mefloquine

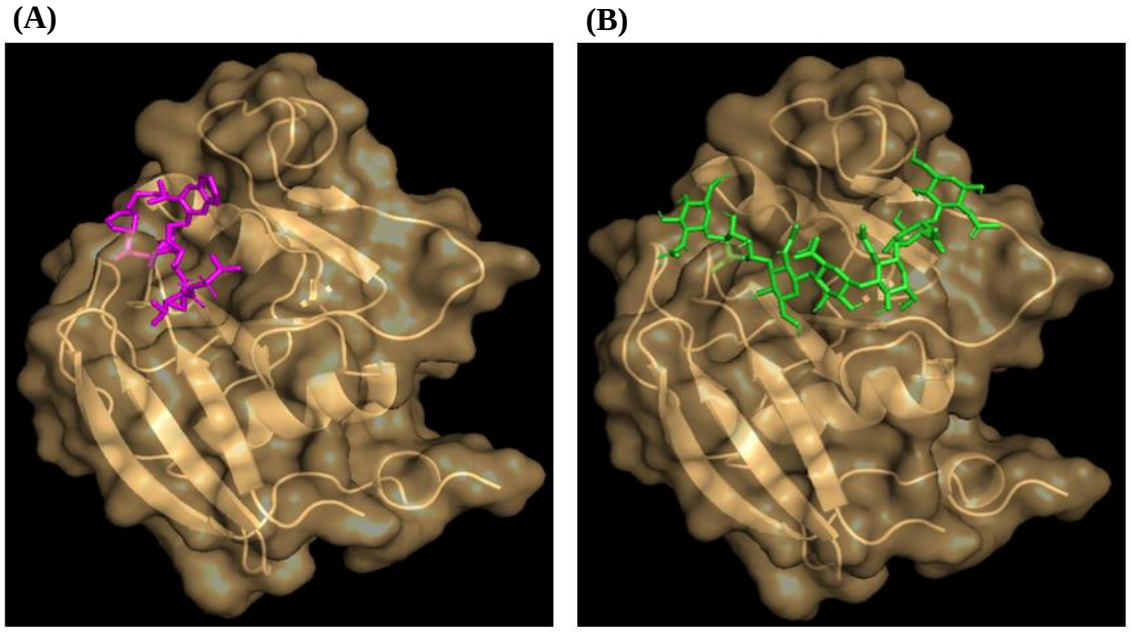


Figure 5: Docked drug molecule and HA in the cavity of CD44. (A) 3D structure of binding site of protein (CD44) showing the orientation of Glecaprevir (magenta) in HA binding groove of CD44 (B) Hyaluronic acid (green).

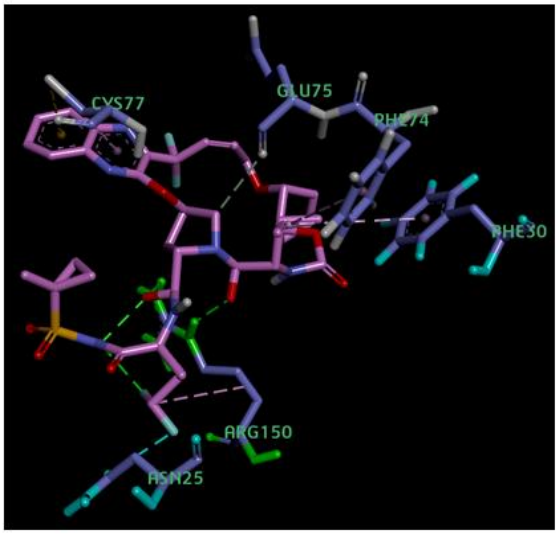
7.3.4 Protein-ligand interaction analysis

The docking results showed the different binding poses of virtually screened drugs at HA binding pocket (**Fig. 6**) showed interaction with various residues from CD44. Among these seven, four drugs, Glecaprevir and Ergotamine with high binding affinity -9.3 kcal/mol, -9.0 kcal/mol, respectively and both Midostaurin and Trypan Blue with binding affinity -8.5 kcal/mol formed no unfavourable interaction, and three drugs (Naldemedine with Arg150, ZINC000252286875 with Cys77 and ZINC000252286876 with Arg90) formed unfavorable interaction with CD44 residues (**Table 3**). HA binding pocket residue Arg150 from CD44 showed frequent H-bond as well as hydrophobic interactions with three drugs (Glecaprevir, Midostaurin and Naldemedine) and only H-bond interaction with Trypan Blue and ZINC000252286876, while with Ergotamine form only hydrophobic interaction. Other pocket residues Asn25, Thr27, Phe30, His35, Phe74, Thr76, Cys77, and Arg78, were commonly involved in non-bonded contacts with most of these 7 drugs *via* van der Waals interactions.

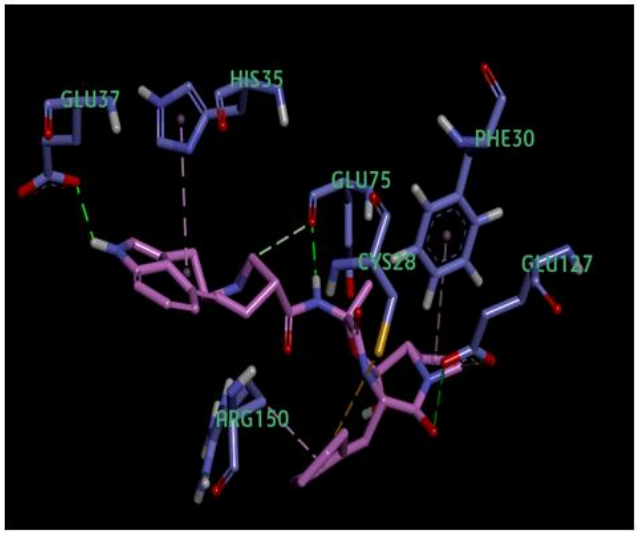
Ans25 was also observed to be interacting with Glecaprevir through halogen bond interaction. Cys77, previously observed to be essential for stabilizing the HA binding groove ***(Kellett-Clark et al. 2015),*** also shows hydrophobic interaction with two drugs (Glecaprevir and Midostaurin) and H-bond interaction with the drug ZINC000252286875 and van der Waals interaction with multiple drugs.

It was observed that among these small molecule drugs, Trypan blue formed the highest number of H-bond as well as van der Waals interactions (8 and 14, respectively) and Midostaurin formed more number of hydrophobic (6) interactions with CD44. At the same time, Ergotamine showed the second-highest number of H-bond (5) interactions with CD44 (**Table 4).**

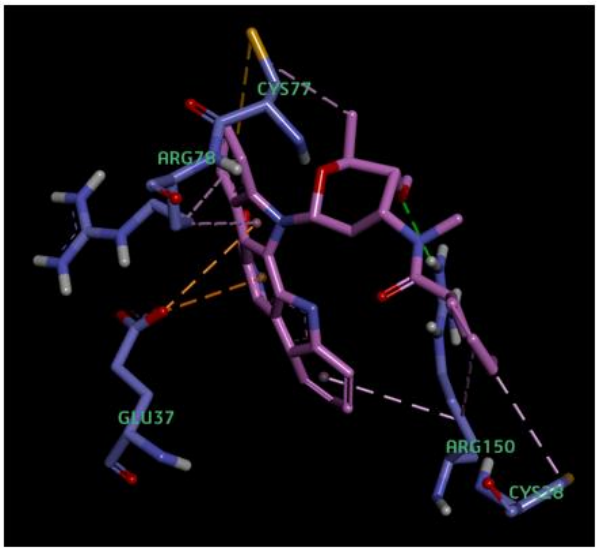
(A) Glecaprevir



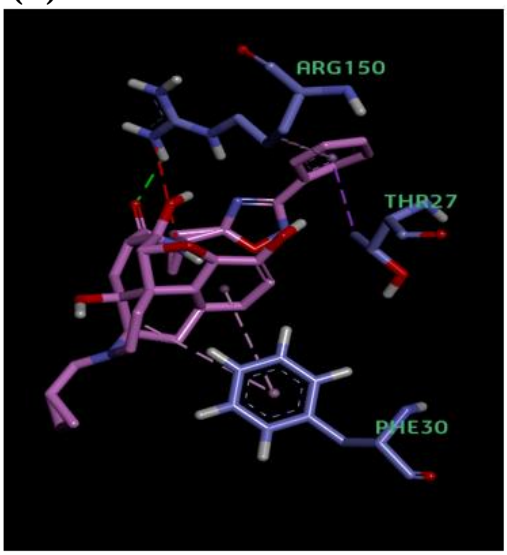
(B) Ergotamine



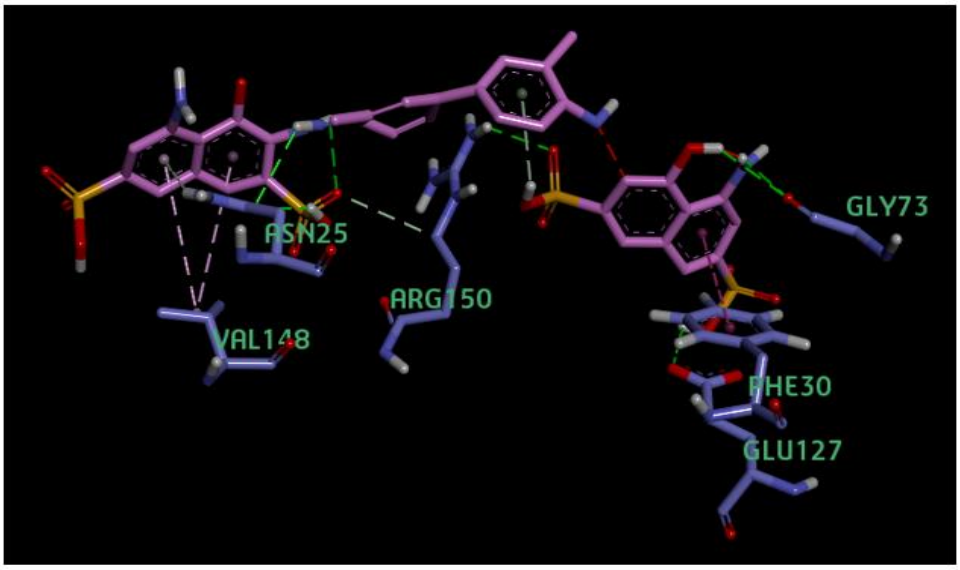
(C) Midostaurin



(D) Naldemedine



(E) Trypan blue



(F) ZINC000252286875 **(G) ZINC000252286876**

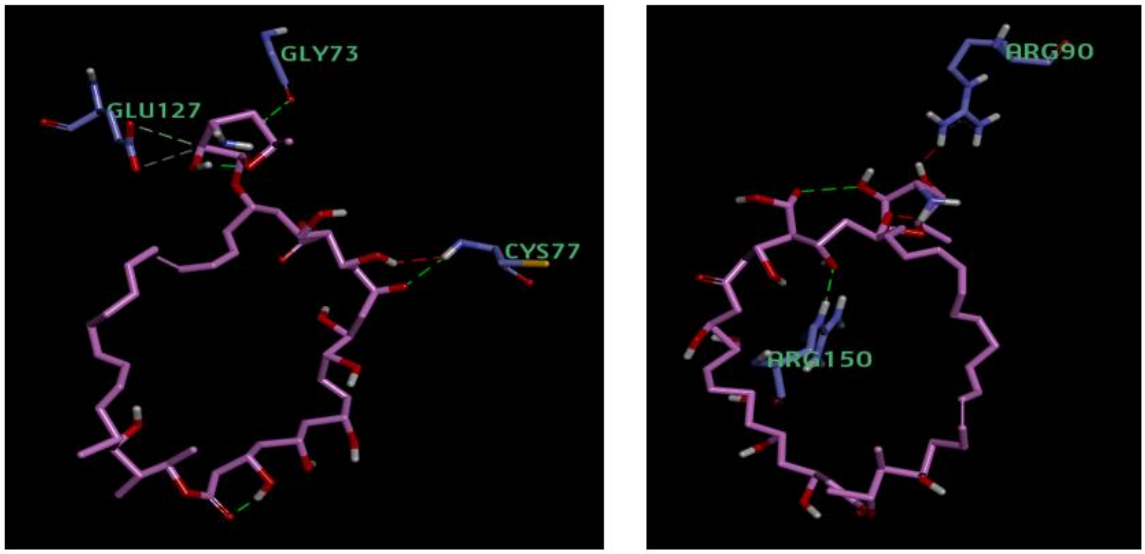


Figure 6: 3D representation of protein-ligand interactions of CD44 for seven drugs. (**A-G)** Ligand molecules are shown in the stick model in magenta color, and CD44 protein amino acid residues are in blue color.

Table 3: List of protein residues involved in various types of interaction with the top seven drugs.

Drug ID or name	Unfavorable interaction	H-bond interaction	Hydrophobic interaction	Van der Waals interaction
Glecaprevir		Arg150	Phe30, Phe74, Cys77, Arg150	His35, Glu37, Thr76, Arg78,
Ergotamine		Glu37, Glu75, Glu127	Phe30, His35, Arg150	Asn25, Thr27, Phe74, Thr76, Cys77, Arg78
Midostaurin		Arg150	Cys28, Cys77, Arg78, Arg150	Asn25, Thr27, Phe30, His35, Phe74, Thr76
Naldemedine	Arg150	Arg150	Thr27, Phe30, Arg150	His35, Phe74, Thr76, Cys77
Trypan Blue		Asn25, lu127, Arg150, Gly73	Phe30, Val148	Thr27, Phe30, His35, Phe74, Thr76, Cys77,
ZINC000252286875	Cys77	Cys77, Gly73		Asn25, Thr27, Phe30, His35, Phe74, Thr76, Arg78,
ZINC000252286876	Arg90	Arg150		Asn25, Thr27, Phe30, His35, Thr76, Cys77, Arg78,

Table 4: Number of different types of interaction between protein and ligands.

Ligand-protein complex	No. of H-bond interaction	No. of hydrophobic bond interaction	No. of van der Waals interaction
Glecaprevir-CD44	3	4	11
Ergotamine-CD44	5	3	7
Midostaurin-CD44	1	6	8
Naldemedine-CD44	1	4	14
Trypan Blue-CD44	8	3	15
ZINC000252286875-CD44	4	-	14
ZINC000252286876-CD44	1	-	14

From a previous *in vitro* study, it was observed that small fragment molecules show inhibitory action on CD44 by binding at HA binding pocket (***Liu & Finzel 2014***). The comparison of fragment similarity with our screened drug shows some similarity in a part of the drugs. So our top selected drugs binding at HA binding cavity may show a similar mode of inhibitory action; however, this needs to be assessed further and validated experimentally.

7.4 References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 71(3):209-249. PMID: 33538338.
2. Falvo P, Orecchioni S, Roma S, et al. (2021) Drug Repurposing in Oncology, an Attractive Opportunity for Novel Combinatorial Regimens. *Curr Med Chem*. 28(11):2114-2136. PMID: 33109033.
3. Bukowski K, Kciuk M, Kontek R. (2020) Mechanisms of Multidrug Resistance in Cancer Chemotherapy. *Int J Mol Sci.* 21(9):3233. PMID: 32370233; PMCID: PMC7247559.
4. Rodrigues R, Duarte D, Vale N. (2022) Drug Repurposing in Cancer Therapy: Influence of Patient's Genetic Background in Breast Cancer Treatment. *Int J Mol Sci*. 23(8):4280. PMID: 35457144; PMCID: PMC9028365.
5. Rudrapal, M., Khairnar, S. J., & Jadhav, A. G. (2020). Drug Repurposing (DR): An Emerging Approach in Drug Discovery. In (Ed.), Drug Repurposing - Hypothesis, Molecular Aspects and Therapeutic Applications. IntechOpen. https://doi.org/10.5772/intechopen.93193
6. Jarada TN, Rokne JG, Alhajj R. (2020) A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J Cheminform.* 12(1):46. PMID: 33431024; PMCID: PMC7374666.
7. Shim JS, Liu JO. (2014) Recent advances in drug repositioning for the discovery of new anticancer drugs. *Int J Biol Sci.* 10(7):654-63. PMID: 25013375; PMCID: PMC4081601.
8. Oprea TI, Overington JP. (2015) Computational and Practical Aspects of Drug Repositioning. *Assay Drug Dev Technol.* 13(6):299-306. PMID: 26241209; PMCID: PMC4533090.
9. Zhang Z, Zhou L, Xie N, (2020) Overcoming cancer therapeutic bottleneck by drug repurposing. *Signal Transduct Target Ther.* 5(1):113. PMID: 32616710; PMCID: PMC7331117.
10. Malizzia LJ, Hsu A. (2008) Temsirolimus, an mTOR inhibitor for treatment of patients with advanced renal cell carcinoma. *Clin J Oncol Nurs.* 12(4):639-46. PMID: 18676330.
11. Dhorje S, Lavhate P, Srivastav A (2020) In silico drug repurposing: An antifungal drug, itraconozole, repurposed as an anticancer agent using molecular docking. *MGM J Med Sci*, 7:110-8.
12. Chakraborty C, Sharma AR, Bhattacharya M, et al. (2021) The Drug Repurposing for COVID-19 Clinical Trials Provide Very Effective Therapeutic Combinations: Lessons Learned From Major Clinical Studies. *Front Pharmacol.* 12:704205. PMID: 34867318; PMCID: PMC8636940.
13. Elmezayen AD, Al-Obaidi A, Şahin AT, Yelekçi K. (2021) Drug repurposing for coronavirus (COVID-19): *in silico* screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *J Biomol Struct Dyn.* 39(8):2980-2992. PMID: 32306862; PMCID: PMC7189413.
14. Herishanu Y, Gibellini F, Njuguna N, (2011) Activation of CD44, a receptor for extracellular matrix components, protects chronic lymphocytic leukemia cells from spontaneous and drug induced apoptosis through MCL-1. *Leuk Lymphoma.* 52(9):1758-69. PMID: 21649540; PMCID: PMC3403533.

15. Peach RJ, Hollenbaugh D, Stamenkovic I, Aruffo A. (1993) Identification of hyaluronic acid binding sites in the extracellular domain of CD44. *J Cell Biol.* 122(1):257-64. PMID: 8314845; PMCID: PMC2119597.
16. Cortes-Dericks L, Schmid RA. (2017) CD44 and its ligand hyaluronan as potential biomarkers in malignant pleural mesothelioma: evidence and perspectives. *Respir Res.* 18(1):58. PMID: 28403901; PMCID: PMC5389171.
17. Jamison FW 2nd, Foster TJ, Barker JA, et al. (2011) Mechanism of binding site conformational switching in the CD44-hyaluronan protein-carbohydrate binding interaction. *J Mol Biol.* 406(4):631-47. PMID: 21216252.
18. Bajorath J, Greenfield B, Munro SB,et al. (1998) Identification of CD44 residues important for hyaluronan binding and delineation of the binding site. *J Biol Chem.* 273(1):338-43. PMID: 9417085.
19. Xu H, Tian Y, Yuan X, et al. (2015) The role of CD44 in epithelial-mesenchymal transition and cancer development. Onco Targets Ther. PMID: 26719706.
20. Chen C, Zhao S, Karnad A, Freeman JW. (2018) The biology and role of CD44 in cancer progression: therapeutic implications. *J Hematol Oncol*. PMID: 29747682.
21. Liu LK, Finzel BC. (2014) Fragment-based identification of an inducible binding site on cell surface receptor CD44 for the design of protein-carbohydrate interaction inhibitors. *J Med Chem.* 57(6):2714-25. PMID: 24606063.

Chapter-8

Discussion and Conclusion

8.1 Discussion

Mutant NRAS protein can be very difficult to target directly but it has frequently been found to be closely involved in drug resistance in a variety of cancer types. Utilizing NRAS-mutant pan-cancer cells lines, we performed an extensive data analysis of coding genes to enlist signaling molecules directly or indirectly connected to NRAS signaling pathway and possibly involved in pan-cancer drug sensitivity or resistance using drug dose-response of five select drugs. We also analyzed regulatory network to understand their regulation by long non-coding RNAs apart from proteins.

Crucial DEGs identified between drug-sensitive and resistant cancer cell lines, were observed to be significantly enriched in signal transduction, cell adhesion, apoptotic process, proteolysis and cell cycle biological processes observed using GO; and in proteoglycans pathway in cancer, focal adhesion pathway, PI3K/Akt signaling pathway, and metabolic pathway observed using KEGG Pathways. Since these pathways are found widely involved in cancers, these enriched DEGs likely play a more significant role in drug resistance development in cancer. ***Lee et al. 2015*** found signaling pathways involved in drug resistance, while our studies pinpointed these key DEGs, which are also involved in some of these pathways. Further analyses utilizing gene co-expression and PPI network of the clusters confirmed that similarity in functional modules of biological processes, as well as the KEGG pathway.

In order to identify an effective therapeutic biomarker, it is important that the mRNA concentration and protein abundance profiles should be correlated. In addition to a gene co-expression network analyses, the construction and analyses of a PPI network allows us to assess the functional roles. For common drugs, hub (driver) proteins were identified from the PPI network similar to the co-expression network hub gene list. Our study found that FN1, CD44, TIMP1, SPARC and SNAI2 are common protein-coding hub genes, which are frequently

associated with most of the drug-resistant cancer conditions. Further, it was seen that all of these protein-coding hub genes were up-regulated in the case of Ponatinib-resistant cancer cell lines and *FN1, CD44* and *TIMP1* were up-regulated in Foretinib-resistant cancer cell lines. *FN1* and *CD44* were down-regulated in Selumetinib- and Trametinib-resistant cancer cell lines, TIMP1 and SNAI2 were down-regulated only in Selumetinib-, while SPARC was down-regulated in the case of Trametinib-resistant cancer cell lines. Some of these identified key hub genes function as biomarkers in several cancer types (***Amundson et al. 2010***; ***Cheon et al. 2014***). It has been shown that overexpression of FN1 induces drug-resistance in breast cancer (***Saatci et al. 2020***) and activates Akt signaling pathway (***Yoshihara et al. 2020***). CD44 is a non-kinase transmembrane proteoglycan (***Jalkanen et al. 1992***). Higher expression of CD44s isoform, is known to induce acquired drug-resistance in cancer through multiple signaling pathways (***Chen et al. 2018).*** TIMP1 is a secretory protein that plays a crucial role in cancer progression and invasion in MMPs independent manner ***(Park et al. 2015***) and is reported to mediate chemoresistance in NSCLC ***(Xiao et al. 2019).*** SNAI2 is observed to be highly expressed in fulvestrant-resistant and tamoxifen-resistant breast cancer and also known to have an involvment in human malignancies (***Cobaleda et al. 2007; Alves et al. 2018***). Similarly, SPARC is a cysteine-rich secreted protein known to be associated with highly aggressive cancer; however, in less aggressive cancer, it is reported to act as a tumor suppressor (***Tai &Tang 2008***).

Further, our studies also focused on identifying non-coding RNA (e.g., lncRNAs) as master regulators of these hub biomarker genes involved in drug resistance, apart from proteins regulators. LncRNAs have been associated with drug resistance (***Corrà et al. 2018; Barth et al. 2020; Pandya et al. 2020; Liu et al. 2020***). Therefore, we wanted to identify key lncRNAs that could regulate our key driver genes identified from both co-expression and PPI network studies, to alter their expression in drug-resistant cancer.

From the directed regulatory interaction network analyses of lncRNA, TFs and mRNA (biomarker genes), we have identified *MALAT1* among the lncRNAs, to be the major interacting component (node) based on two important topological network parameters (outdegree and betweenness centrality). *MALAT1* regulates driver genes by interacting with mRNAs at 5' UTR of *FN1*; CDS of *TIMP1, SNAI2*; and 3' UTR of *CD44* and *SPARC*. In drug-resistant cancer, *MALAT1* could regulate hub genes expressions through the processes such as mRNA splicing, stability and degradation (***Amodio et al. 2018; Bhat et al. 2016***). Moreover, *MALAT1* is a widely studied lncRNA in a variety of cancers and was originally reported to be associated with metastasis in the early stage of non-small cell lung cancer (***Yoshimoto et al. 2016; Amodio et al. 2018).*** *MALAT1* transcripts are localized to the nuclear speckles, which is a site for the pre-mRNA splicing process, after being transcribed from human chromosome 11q13.1 (***Arun et al. 2020; Yoshimoto et al., 2016; Jadaliha et al. 2016; Gordon et al. 2019***). The initial studies of *MALAT1* overexpression were shown to be associated with tumor growth, metastasis, cell adhesion, migration, and poor prognosis in cancer (***Yoshimoto et al. 2016)***. From our gene regulatory interaction network study, driver genes *SPARC* and *SNAI2* interacting with *MALAT1* were corroboratively found to be down-regulated in *MALAT1*-depleted breast cancer reported by ***Jadaliha et al. 2016***. Further, studies suggest that *MALAT1* also positively modulates the expression of EGR1 (***Spreafico et al. 2018***) while results from our studies using harmonizome ChIP-Seq data from ENCODE dataset shows that EGR1 could transcriptionally regulate *MALAT1*. EGR1 and *MALAT1* might be possibly regulating each other through a positive feedback loop regulatory system. Many studies have reported that the mechanism of *MALAT1* action on mRNA splicing could serve as decoy processing (***Bhat et al. 2016; Nguyen et al. 2020),*** and apart from interacting with mRNAs of driver genes at different mRNA regions, *MALAT1* could also be interacting with their respective proteins. *MALAT1* might be regulating *CD44* in a *cis*-regulatory manner because both *MALAT1* and protein-

coding gene *CD44* are located on the same chromosome 11, while with respect to other driver coding genes (*FN1, TIMP1, SNAI2, SPARC*), *MALAT1* could be regulating, in a *trans*-regulatory manner as other genes are located on different chromosomes than *MALAT1*. This study further confirms that two possible scenarios might exist. First, that several lncRNAs may interact with one coding gene at a time, and second, that one or multiple lncRNAs may interact/regulate many coding genes simultaneously at the transcriptional level.

From our above studies, we came up with a working model of the mechanisms of driver genes regulation, specifically, *FN1, CD44, TIMP1, SPARC,* and *SNAI2,* by EGR1-*MALAT1* regulatory axis, which has been identified from our network analysis study using genes involved in NRAS-mutant pan-cancer drug resistance (**Fig. 1**). A few coding as well as non-coding genes (lncRNAs) can function as key targets replacing recalcitrant NRAS as a drug target.

Taken together, our data suggest that these identified driver genes' expression may be induced or suppressed via direct interaction with *MALAT1,* which leads to context-dependent drug resistance/sensitivity, corroborated by literature studies.

Key insights gained from these findings may improve our understanding of drug resistance development in pan-cancer systems. Further studies are needed to assess the clinical relevance of these findings as therapeutic targets in the cancer types harbouring NRAS mutation as we have used experiments conducted on cell lines, whereas tumor microenvironment is found to play a crucial role in regulating the overall cancer phenotypes, so some deviations from our study may be observed. Enlisted few driver genes/lncRNAs can be further studied for their specific expression in drug-resistant cancer cell lines, alongwith transcriptional dysregulation and its implications on regulatory activity.

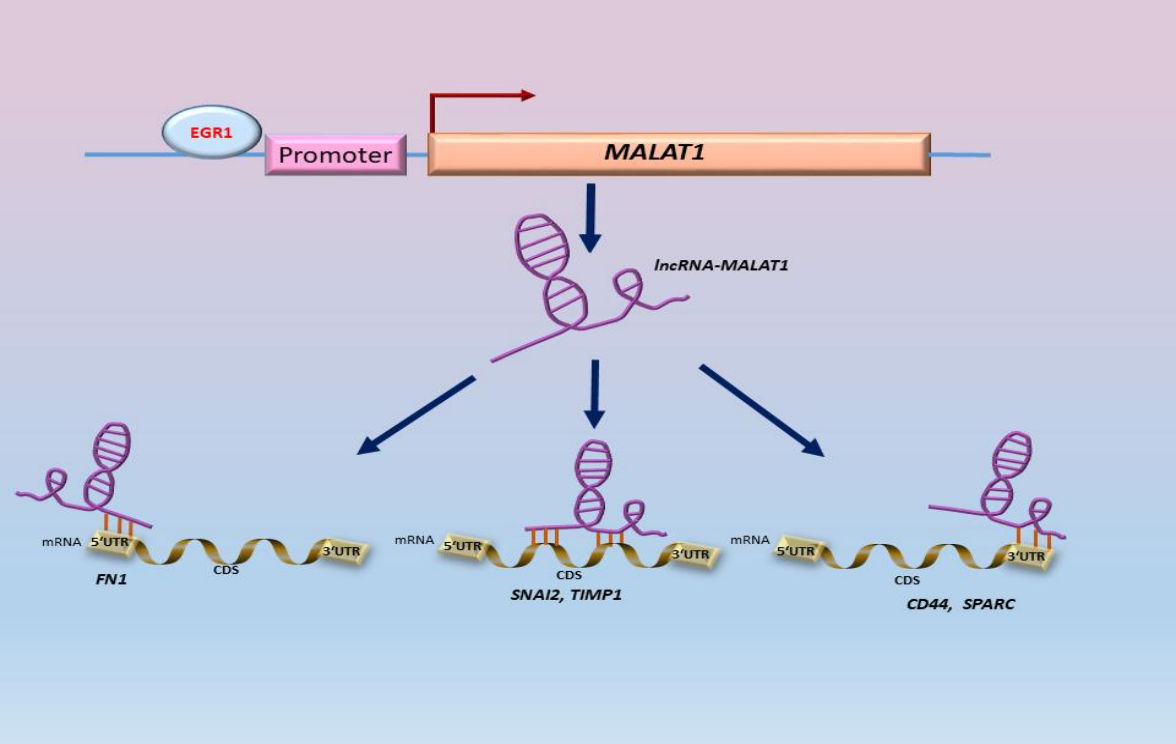


Figure 1: *MALAT1* may regulate driver genes associated with drug-resistance in cancer. EGR1 may bind to the promoter region of *MALAT1* to transcriptionally regulate it. After transcription, *MALAT1* may regulate driver genes by binding at 5'UTR, CDS, and 3'UTR regions of key genes.

Drug repurposing

A structure-based *in silico* virtual screening was done to discover novel inhibitary candidates of CD44, by using a drug repurposing approach. In summary, 1615 FDA-approved drugs from the ZINC15 database were screened against CD44 to discover potent inhibitors with a high binding affinity toward the target protein. Sixteen ligands showed a high binding affinity with CD44, and 7 of them were found to bind at the HA binding cavity of the target protein (CD44) with high affinity. Among these seven drugs, three drugs (Naldemedine, ZINC000252286875, and ZINC000252286876) showed unfavorable interaction with CD44 due to steric clashes and unfavorable donor–donor interaction between atoms. Other four drugs (Glecaprevir, Ergotamine, Midostaurin, and Trypan blue) displayed strong molecular interaction with residues of CD44, involving in the HA binding (Arg150 and Arg78) and also including some essential residues, without any unfavorable interaction. Protein residues Asn25, Glu37, and

Arg78 are interacting with drug molecules and were previously reported to induce conformational changes that are in direct contact with the loop of Arg41, which create a high-affinity HA-bound form of the HABD (***Liu & Finzel 2014***). Among the identified four residues (Arg41, Tyr42, Arg78 and Tyr79) from CD44 crucial HA interaction ***(Peach et al. 1993; Bajorath et al. 1998)***, Arg78 is the only residue observed to show interaction with multiple drugs in our drug repurposing study. Arg150 is important for HABD affinity to HA, and a previous study suggests that mutation at Arg150 reduced HABD affinity toward HA binding (***Banerji et al. 2007)***.

Unlike the oligomeric carbohydrate, which extended across a large and exposed binding cavity, the drug molecules induce conformational changes that allow them to bind with high ligand potency, which is much higher than HA. So, the interaction of CD44 protein residues with virtually screened drug molecules might abolish the binding of HA at the HABD of the CD44. Based on our extensive observations, Glecaprevir, Ergotamine, Midostaurin, and Trypan blue could be potential therapeutic inhibitors of CD44 with high binding affinity and without any unfavorable interactions with CD44. These four drugs may elicit a blocking effect on HA-binding to CD44 by competitive inhibition.

8.2 Conclusions

In our study, we analyzed basal gene expression using microarray data set from pan-cancer drug-sensitive and resistant cell lines from GDSC. From the significant differential gene expression analyses, gene co-expression and PPI networks; *FN1, CD44, TIMP1, SPARC* and *SNAI2,* were identified as common driver genes in drug-resistant cancer that might provide new biomarkers in NRAS-mutant pan-cancer drug resistance. *MALAT1,* as a key regulator of these coding biomarker genes in drug resistance, could be a master biomarker to regulate these driver genes' expression and provide key insights to improve drug sensitivity in a pan-cancer context.

In silico approach of drug repurposing can be used to discover new drug molecules that might be able to improve drug-sensitivity in the mutant *NRAS* pan-cancer system by inhibiting CD44. FDA-approved drugs Glecaprevir, Ergotamine, Midostaurin, and Trypan Blue, may be potential therapeutic inhibitors of identified hub node CD44 with high binding affinity with a view of repurposing these drugs. Drug molecules that have the potential to inhibit CD44 may serve as a lead molecules to improve drug-sensitivity and fight against mutant NRAS-harbouring pan-cancer.

References

1. Lee YS, Hwang SG, Kim JK, et al. (2015) Topological network analysis of differentially expressed genes in cancer cells with acquired gefitinib resistance. *Cancer Genomics Proteomics*. 12(3):153-66. PMID: 25977174.
2. Amundson SA, Smilenov LB. (2010) Integration of biological knowledge and gene expression data for biomarker selection: FN1 as a potential predictor of radiation resistance in head and neck cancer. *Cancer Biol Ther.* 10(12):1252-5. PMID: 20948301; PMCID: PMC3230290.
3. Cheon DJ, Tong Y, Sim MS, Dering J, et al. (2014) A collagen-remodeling gene signature regulated by TGF-β signaling is associated with metastasis and poor survival in serous ovarian cancer. *Clin Cancer Res.* 20(3):711-23. PMID: 24218511; PMCID: PMC3946428.
4. Saatci O, Kaymak A, Raza U, et al. (2020) Targeting lysyl oxidase (LOX) overcomes chemotherapy resistance in triple negative breast cancer. *Nat Commun*. 11(1):2416. PMID: 32415208; PMCID: PMC7229173.
5. Yoshihara M, Kajiyama H, Yokoi A, et al. (2020) Ovarian cancer-associated mesothelial cells induce acquired platinum-resistance in peritoneal metastasis via the FN1/Akt signaling pathway. *Int J Cancer*. 146(8):2268-2280. PMID: 31904865; PMCID: PMC7065188.
6. Jalkanen S, Jalkanen M. (1992) Lymphocyte CD44 binds the COOH-terminal heparin-binding domain of fibronectin. *J Cell Biol.* 116(3):817-25. PMID: 1730778; PMCID: PMC2289325.
7. Chen C, Zhao S, Karnad A, Freeman JW. (2018) The biology and role of CD44 in cancer progression: therapeutic implications. *J Hematol Oncol*. 11(1):64. PMID: 29747682; PMCID: PMC5946470.
8. Park SA, Kim MJ, Park SY, et al. (2015) TIMP-1 mediates TGF-β-dependent crosstalk between hepatic stellate and cancer cells via FAK signaling. *Sci Rep*. 5:16492. PMID: 26549110; PMCID: PMC4637930.
9. Xiao W, Wang L, Howard J, Kolhe R, Rojiani AM, Rojiani MV. (2019) TIMP-1-Mediated Chemoresistance via Induction of IL-6 in NSCLC. *Cancers (Basel).* 11(8):1184. PMID: 31443242; PMCID: PMC6721590.
10. Cobaleda C, Pérez-Caro M, Vicente-Dueñas C, et al. (2007) Function of the zinc-finger transcription factor SNAI2 in cancer and development. *Annu Rev Genet*. 41:41-61. PMID: 17550342.
11. Alves CL, Elias D, Lyng MB, et al. (2018) SNAI2 upregulation is associated with an aggressive phenotype in fulvestrant-resistant breast cancer cells and is an indicator of poor response to endocrine therapy in estrogen receptor-positive metastatic breast cancer. *Breast Cancer Res.* 20(1):60. PMID: 29921289; PMCID: PMC6009053.
12. Tai IT, Tang MJ. (2008) SPARC in cancer biology: its role in cancer progression and potential for therapy. *Drug Resist Updat*. 11(6):231-46. PMID: 18849185.
13. Corrà F, Agnoletto C, Minotti L, et al. (2018) The Network of Non-coding RNAs in Cancer Drug Resistance. *Front Oncol.* 8:327. PMID: 30211115; PMCID: PMC6123370.
14. Barth DA, Juracek J, Slaby O, et al. (2020) lncRNA and Mechanisms of Drug Resistance in Cancers of the Genitourinary System. *Cancers (Basel).* 12(8):2148. PMID: 32756406; PMCID: PMC7463785.

15. Pandya G, Kirtonia A, Sethi G, Pandey AK, Garg M. (2020) The implication of long non-coding RNAs in the diagnosis, pathogenesis and drug resistance of pancreatic ductal adenocarcinoma and their possible therapeutic potential. *Biochim Biophys Acta Rev Cancer*. 1874(2):188423. PMID: 32871244.
16. Liu K, Gao L, Ma X, Huang JJ, Chen J, Zeng L, Ashby CR Jr, Zou C, Chen ZS. (2020) long non-coding RNAs regulate drug resistance in cancer. *Mol Cancer*. 9(1):54. PMID: 32164712; PMCID: PMC7066752.
17. Amodio N, Raimondi L, Juli G, et al. (2018) MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. *J Hematol Oncol.* 11(1):63. PMID: 29739426; PMCID: PMC5941496.
18. Bhat SA, Ahmad SM, Mumtaz PT, et al. (2016) long non-coding RNAs: Mechanism of action and functional utility. *Noncoding RNA Res*. 1(1):43-50. PMID: 30159410; PMCID: PMC6096411.
19. Yoshimoto R, Mayeda A, Yoshida M, Nakagawa S. (2016) MALAT1 long non-coding RNA in cancer. *Biochim Biophys Acta.* 1859(1):192-9. PMID: 26434412.
20. Arun G, Aggarwal D, Spector DL. (2020) *MALAT1* Long Non-Coding RNA: Functional Implications. *Noncoding RNA.* 6(2):22. doi: 10.3390/ncrna6020022. PMID: 32503170; PMCID: PMC7344863.
21. Jadaliha M, Zong X, Malakar P, Ray T, Singh DK, Freier SM, Jensen T, Prasanth SG, Karni R, Ray PS, Prasanth KV. (2016) Functional and prognostic significance of long non-coding RNA MALAT1 as a metastasis driver in ER negative lymph node negative breast cancer. *Oncotarget.* 7(26):40418-40436. doi: 10.18632/oncotarget.9622. PMID: 27250026; PMCID: PMC5130017.
22. Gordon MA, Babbs B, Cochrane DR, Bitler BG, Richer JK. (2019) The long non-coding RNA MALAT1 promotes ovarian cancer progression by regulating RBFOX2-mediated alternative splicing. *Mol Carcinog.* 58(2):196-205. doi: 10.1002/mc.22919. PMID: 30294913.
23. Spreafico M, Grillo B, Rusconi F, Battaglioli E, Venturin M. (2018) Multiple Layers of *CDK5R1* Regulation in Alzheimer's disease Implicate Long Non-Coding RNAs. *Int J Mol Sci.* 19(7):2022. doi: 10.3390/ijms19072022. PMID: 29997370; PMCID: PMC6073344.
24. Nguyen TM, Alchalabi S, Oluwatoyosi A, Ropri AS, Herschkowitz JI, Rosen JM. (2020) New twists on long noncoding RNAs: from mobile elements to motile cancer cells. *RNA Biol.* 17(11):1535-1549. doi: 10.1080/15476286.2020.1760535. PMID: 32522127; PMCID: PMC7567495.
25. Liu LK, Finzel BC. (2014) Fragment-based identification of an inducible binding site on cell surface receptor CD44 for the design of protein-carbohydrate interaction inhibitors. *J Med Chem.* 57(6):2714-25. PMID: 24606063.
26. Banerji S, Wright AJ, Noble M, et al. (2007) Structures of the Cd44-hyaluronan complex provide insight into a fundamental carbohydrate-protein interaction. *Nat Struct Mol Biol.* 14(3):234-9. PMID: 17293874.

27. Peach RJ, Hollenbaugh D, Stamenkovic I, et al. (1993) Identification of hyaluronic acid binding sites in the extracellular domain of CD44. *J Cell Biol.* 122(1):257-64. PMID: 8314845; PMCID: PMC2119597.
28. Bajorath J, Greenfield B, Munro SB, et al. (1998) Identification of CD44 residues important for hyaluronan binding and delineation of the binding site. *J Biol Chem.* 273(1):338-43. PMID: 9417085.

Publications & Posters

scientific reports

OPEN

MALAT1 as master regulator of biomarkers predictive of pan-cancer multi-drug resistance in the context of recalcitrant NRAS signaling pathway identified using systems-oriented approach

Santosh Kumar & Seema Mishra✉

NRAS, a protein mutated in several cancer types, is involved in key drug resistance mechanisms and is an intractable target. The development of drug resistance is one of the major impediments in targeted therapy. Currently, gene expression data is used as the most predictive molecular profile in pan-cancer drug sensitivity and resistance studies. However, the common regulatory mechanisms that drive drug sensitivity/resistance across cancer types are as yet, not fully understood. We focused on GDSC data on NRAS-mutant pan-cancer cell lines, to pinpoint key signaling targets in direct or indirect associations with NRAS, in order to identify other druggable targets involved in drug resistance. Large-scale gene expression, comparative gene co-expression and protein–protein interaction network analyses were performed on selected drugs inducing drug sensitivity/resistance. We validated our data from cell lines with those obtained from primary tissues from TCGA. From our big data studies validated with independent datasets, protein-coding hub genes *FN1*, *CD44*, *TIMP1*, *SNAI2*, and *SPARC* were found significantly enriched in signal transduction, proteolysis, cell adhesion and proteoglycans pathways in cancer as well as the PI3K/Akt-signaling pathway. Further studies of the regulation of these hub/driver genes by lncRNAs revealed several lncRNAs as prominent regulators, with *MALAT1* as a possible master regulator. Transcription factor EGR1 may control the transcription rate of *MALAT1* transcript. Synergizing these studies, we zeroed in on a pan-cancer regulatory axis comprising EGR1-*MALAT1*-driver coding genes playing a role. These identified gene regulators are bound to provide new paradigms in pan-cancer targeted therapy, a foundation for precision medicine, through the targeting of these key driver genes in the improvement of multi-drug sensitivity or resistance.

Cancer is a serious health issue and the second leading cause of death worldwide as estimated by World Health Organization[1]. Drug resistance which can be acquired or intrinsic, develops due to the failure of chemotherapeutic drugs to treat cancer cells because of limited effectiveness[2–4]. While intrinsic antibiotic/drug resistance is a naturally occurring phenomenon primarily present before chemotherapy[5,6], acquired drug resistance arises after the chemotherapeutic treatment of cancer[2].

Intrinsic drug resistance may arise due to existential mutations in crucial genes, intrinsic heterogeneity of tumors, and/or activation of certain molecular pathways against anti-cancer drugs[6]. In one study, transcriptional repressors Snail and Slug were observed to induce radioresistance and chemoresistance in ovarian cancer through the antagonism of p53-mediated apoptosis[7]. Acquired drug resistance may be the result of activation of secondary proto-oncogenes, mutations or altered expression of drug targets and post-treatment changes in the tumor microenvironment. Reiterating, there are several possible mechanisms involved in cancer drug resistance, including altered expression and mutation in target oncogenes, compensatory activation of the downstream signaling pathways, epigenetic abnormalities and histological transformations[8,9]. Drug resistance can also occur due to

Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad, Telangana 500046, India. ✉email: seema_uoh@yahoo.com

scientific reports

Check for updates

OPEN

Structural exploration with AlphaFold2-generated STAT3α structure reveals selective elements in STAT3α-GRIM-19 interactions involved in negative regulation

Seema Mishra✉, Santosh Kumar, Kesaban Sankar Roy Choudhuri, Imliyangla Longkumer, Praveena Koyyada & Euphinia Tiberius Kharsyiemiong

STAT3, an important transcription factor constitutively activated in cancers, is bound specifically by GRIM-19 and this interaction inhibits STAT3-dependent gene expression. GRIM-19 is therefore, considered as an inhibitor of STAT3 and may be an effective anti-cancer therapeutic target. While STAT3 exists in a dimeric form in the cytoplasm and nucleus, it is mostly present in a monomeric form in the mitochondria. Although GRIM-19-binding domains of STAT3 have been identified in independent experiments, yet the identified domains are not the same, and hence, discrepancies exist. Human STAT3-GRIM-19 complex has not been crystallised yet. Dictated by fundamental biophysical principles, the binding region, interactions and effects of hotspot mutations can provide us a clue to the negative regulatory mechanisms of GRIM-19. Prompted by the very nature of STAT3 being a challenging molecule, and to understand the structural basis of binding and interactions in STAT3α-GRIM-19 complex, we performed homology modelling and *ab-initio* modelling with evolutionary information using I-TASSER and *avant-garde* AlphaFold2, respectively, to generate monomeric, and subsequently, dimeric STAT3α structures. The dimeric form of STAT3α structure was observed to potentially exist in an anti-parallel orientation of monomers. We demonstrate that during the interactions with both unphosphorylated and phosphorylated STAT3α, the NTD of GRIM-19 binds most strongly to the NTD of STAT3α, in direct contrast to the earlier works. Key arginine residues at positions 57, 58 and 68 of GRIM-19 are mainly involved in the hydrogen-bonded interactions. An intriguing feature of these arginine residues is that these display a consistent interaction pattern across unphosphorylated and phosphorylated monomers as well as unphosphorylated dimers in STAT3α-GRIM-19 complexes. MD studies verified the stability of these complexes. Analysing the binding affinity and stability through free energy changes upon mutation, we found GRIM-19 mutations Y33P and Q61L and among GRIM-19 arginines, R68P and R57M, to be one of the top-most major and minor disruptors of binding, respectively. The proportionate increase in average change in binding affinity upon mutation was inclined more towards GRIM-19 mutants, leading to the surmise that GRIM-19 may play a greater role in the complex formation. These studies propound a novel structural perspective of STAT3α-GRIM-19 binding and inhibitory mechanisms in both the monomeric and dimeric forms of STAT3α as compared to that observed from the earlier experiments, these experimental observations being inconsistent among each other.

... the structure of every organic being is related, in the most essential yet often hidden manner, to that of all the other organic beings ...
—Charles Darwin, The Origin of Species.

Department of Biochemistry, School of Life Sciences, University of Hyderabad, Hyderabad 500046, India. ✉email: seema_uoh@yahoo.com

सी डैक CDAC | प्रगत संगणन विकास केंद्र CENTRE FOR DEVELOPMENT OF ADVANCED COMPUTING

BRAF

NSM NATIONAL SUPERCOMPUTING MISSION

bio INFORMATICS

प्रगत संगणन विकास केंद्र

Centre for Development of Advanced Computing

CERTIFICATE OF PARTICIPATION AND POSTER PRESENTATION

This is to certify that

Santosh Kumar

has participated and presented poster in

the Symposium on "Accelerating **Biology** 2018"

...Digitizing Life

held during January 09 -11, 2018 at C-DAC, Pune

Dr. Rajendra Joshi
Convener

www.incob2018.org

17th International Conference on Bioinformatics

INCOB 2018
New Delhi

September 26-28, 2018
Jawaharlal Nehru University, New Delhi, India

Certificate of Poster Presentation

It is certified that **Santosh Kumar** has presented a poster in the 17th International Conference on Bioinformatics held at Jawaharlal Nehru University, New Delhi, India on September 26-28, 2018.

Shandar Ahmad
(JNU, General Chair)

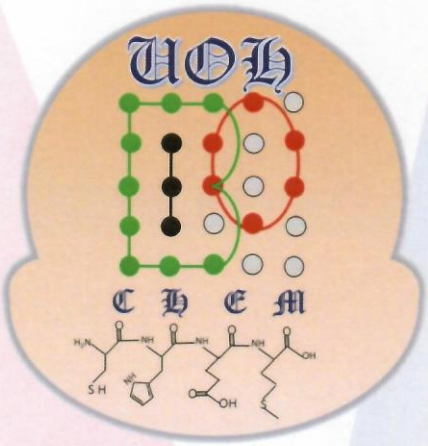
Michael Gromiha
(IIT Madras, Co-Chair)

GPS Raghava
(IIIT Delhi, Co-Chair)

Christian Schönbach
(President APBioNet)

UGC- SAP funded

National Seminar on "Biomolecular Interactions in Development and Diseases"



September 26 - 28, 2019

Organized by
Department of Biochemistry
School of Life Sciences
University of Hyderabad

Probing of biomarkers predictive of pan-cancer drug sensitivity and resistance

Santosh Kumar*[#], Seema Mishra
Department of Biochemistry, School of Life Sciences, University of Hyderabad, India
[#]*skp2259@gmail.com*

In the recent years, gene expression data has been extensively used to identify biomarkers as the most predictive molecular profile in pan-cancer drug sensitivity and resistance studies. We have studied and analysed basal gene expression data in context at systems scale using data from the GDSC database. Since NRAS is difficult to target in cancers, using NRAS-mutant cancer types and samples, we aimed to identify other suitable biomarkers that can serve as a potential drug targets which are involved directly or indirectly in the NRAS signaling pathway. At cut off $p < 0.01$ and $logFC > 2$, we have identified significantly differentially expressed genes (DEGs) that vary between drug-sensitive and drug-resistant cell lines for each drug (7drugs) studied in GDSC. Functional annotation of identified DEGs revealed that these genes plays major role in signal transduction, apoptotic process and cell adhesion processes among others. We have identified four hub genes common in both gene co-expression network and PPI network: CXCR4, *CAV1*, *TIMP1*, and *CD48*. These genes might play an important roles in drug-resistant NRAS-mutant cancers and provide a new molecular markers as therapeutic target in cancer therapy to improve drug sensitivity.

University of Hyderabad
Hyderabad-500046, India

PLAGIARISM FREE CERTIFICATE

This is to certify that the similarity index of this thesis as checked by the Library of the University of Hyderabad, India, is 20%. Out of this, 13% similarity has been found from the candidate's (**Santosh Kumar**) own publication which forms the adequate part of the thesis. The details of this student's publication are as follows:

	Paper	% Similarity
1.	Kumar S., Mishra S. MALAT1 as master regulator of biomarkers predictive of pan-cancer multi-drug resistance in the context of recalcitrant NRAS signaling pathway identified using systems-oriented approach. *Sci Rep* 12, 7540 (2022). DOI: 10.1038/s41598-022-11214-8	12% (Source no. 1) 1% (Source no. 9)

About 7% similarity was identified from external sources in the present thesis which is in accordance with the regulations of the University. All the publications related to this thesis have been appended at the end of the thesis. Hence the present thesis may be considered to be plagiarism-free.

Dr. Seema Mishra

Supervisor

Dr. SEEMA MISHRA
Assistant Professor
Deptt. of Biochemistry
School of Life Sciences
University of Hyderabad
HYDERABAD-500 046 INDIA

Probing of biomarkers predictive of pan-cancer drug sensitivity and resistance and drug repurposing

by Santosh Kumar

Submission date: 04-Jan-2023 04:50PM (UTC+0530)
Submission ID: 1988487239
File name: Santosh_Kumar.pdf (5.9M)
Word count: 24890
Character count: 136711

Probing of biomarkers predictive of pan-cancer drug sensitivity and resistance and drug repurposing

ORIGINALITY REPORT

20%	16%	17%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet Source	12%
2	Submitted to University of Hyderabad, Hyderabad Student Paper	2%
3	www.nature.com Internet Source	<1%
4	"The Chemical Biology of Long Noncoding RNAs", Springer Science and Business Media LLC, 2020 Publication	<1%
5	www.mdpi.com Internet Source	<1%
6	A. García-Moreno, R. López-Domínguez, A. Ramirez-Mena, A. Pascual-Montano et al. "GeneCodis 4: Expanding the modular enrichment analysis to regulatory elements", Cold Spring Harbor Laboratory, 2021 Publication	<1%

7	helda.helsinki.fi Internet Source	<1%
8	publikationen.bibliothek.kit.edu Internet Source	<1%
9	Santosh Kumar, Seema Mishra. "MALAT1 as master regulator of biomarkers predictive of pan-cancer multi-drug resistance in the context of recalcitrant NRAS signaling pathway identified using systems-oriented approach", Scientific Reports, 2022 Publication	<1%
10	Luisa Statello, Chun-Jie Guo, Ling-Ling Chen, Maite Huarte. "Gene regulation by long non-coding RNAs and its biological functions", Nature Reviews Molecular Cell Biology, 2020 Publication	<1%
11	Kanisha Shah, Rakesh M. Rawal. "Genetic and Epigenetic Modulation of Drug Resistance in Cancer: Challenges and Opportunities", Current Drug Metabolism, 2020 Publication	<1%
12	"Molecular Pathology of Breast Cancer", Springer Science and Business Media LLC, 2016 Publication	<1%
13	Garrett M. Morris, Ruth Huey, William Lindstrom, Michel F. Sanner, Richard K. Belew,	<1%

David S. Goodsell, Arthur J. Olson. "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility", Journal of Computational Chemistry, 2009
Publication

14	oaepublishstorage.blob.core.windows.net Internet Source	<1 %
15	Francesco Iorio, Theo A. Knijnenburg, Daniel J. Vis, Graham R. Bignell et al. "A Landscape of Pharmacogenomic Interactions in Cancer", Cell, 2016 Publication	<1 %
16	Teague Sterling, John J. Irwin. "ZINC 15 – Ligand Discovery for Everyone", Journal of Chemical Information and Modeling, 2015 Publication	<1 %
17	ANDREW LIMAN, Qian Zhu. "TMEM88 is a protective gene in HCC and has the function of promotesimmune escape", Research Square Platform LLC, 2022 Publication	<1 %
18	Sowmiya J, S. Thilagamani. "Malnutrition: an Unrecognized and Untreated Complication in Cancer", Research Square Platform LLC, 2022 Publication	<1 %
19	rforbiochemists.blogspot.co.uk Internet Source	<1 %

20	Natsumi Irahara, Yoshifumi Baba, Katsuhiko Nosho, Kaori Shima et al. "NRAS Mutations Are Rare in Colorectal Cancer", Diagnostic Molecular Pathology, 2010 Publication	<1 %
21	docshare.tips Internet Source	<1 %
22	search.yahoo.com Internet Source	<1 %
23	Maryam Pouryahya, Jung Hun Oh, James C. Mathews, Zehor Belkhatir, Caroline Moosmüller, Joseph O. Deasy, Allen R. Tannenbaum. "Network-based clustering for drug sensitivity prediction in cancer cell lines", Cold Spring Harbor Laboratory, 2019 Publication	<1 %
24	academic.oup.com Internet Source	<1 %
25	Advances in Experimental Medicine and Biology, 2005. Publication	<1 %
26	Submitted to RMIT University Student Paper	<1 %
27	Submitted to University of East London Student Paper	<1 %

28	Asish Mohapatra. "Software tools for toxicology and risk assessment", Elsevier BV, 2020 Publication	<1 %
29	Molecular Pathogenesis of Colorectal Cancer, 2013. Publication	<1 %
30	Submitted to Celebration HIgh School Student Paper	<1 %
31	Submitted to University of Macau Student Paper	<1 %
32	researchspace.ukzn.ac.za Internet Source	<1 %
33	dash.harvard.edu Internet Source	<1 %
34	digitalcommons.odu.edu Internet Source	<1 %
35	assets.researchsquare.com Internet Source	<1 %
36	core.ac.uk Internet Source	<1 %
37	hdl.handle.net Internet Source	<1 %
38	lib.smu.edu.cn Internet Source	<1 %

39	Sayed Haidar Abbas Raza, Rajwali Khan, Gong Cheng, Feng Long et al. "RNA-Seq reveals the potential molecular mechanisms of bovine KLF6 gene in the regulation of adipogenesis", International Journal of Biological Macromolecules, 2022 Publication	<1 %
40	Sike, Ádám, Enikő Nagy, Balázs Vedelek, Dávid Pusztai, Péter Szerémy, Anikó Venetianer, and Imre M. Boros. "mRNA Levels of Related Abcb Genes Change Opposite to Each Other upon Histone Deacetylase Inhibition in Drug-Resistant Rat Hepatoma Cells", PLoS ONE, 2014. Publication	<1 %
41	zenodo.org Internet Source	<1 %
42	hbctraining.github.io Internet Source	<1 %
43	Bimal Krishna Banik, Biswa Mohan Sahoo. "Green synthesis and biological evaluation of anticancer drugs", Elsevier BV, 2020 Publication	<1 %
44	Submitted to Zhejiang University Center of Modern Educational Technology Student Paper	<1 %
45	dpcpsi.nih.gov	

	Internet Source	<1%
46	www.spandidos-publications.com Internet Source	<1%
47	Érica Aparecida de Oliveira, Colin R. Goding, Silvya Stuchi Maria-Engler. "Chapter 369 Organotypic Models in Drug Development “Tumor Models and Cancer Systems Biology for the Investigation of Anticancer Drugs and Resistance Development”", Springer Science and Business Media LLC, 2020 Publication	<1%
48	dr.ntu.edu.sg Internet Source	<1%
49	Shamik Mitra, Martin Lauss, Rita Cabrita, Jiyeon Choi et al. "Analysis of DNA methylation patterns in the tumor immune microenvironment of metastatic melanoma", Molecular Oncology, 2020 Publication	<1%
50	Submitted to University of Southampton Student Paper	<1%
51	www.frontiersin.org Internet Source	<1%
52	Emanuel Gonçalves, Rebecca C. Poulos, Zhaoxiang Cai, Syd Barthorpe et al. "Pan-	<1%

cancer proteomic map of 949 human cell lines", Cancer Cell, 2022
Publication

53	Submitted to Higher Education Commission Pakistan Student Paper	<1%
54	Submitted to University College London Student Paper	<1%
55	link.springer.com Internet Source	<1%
56	res.mdpi.com Internet Source	<1%
57	Submitted to Ain Shams University Student Paper	<1%
58	Submitted to La Serna High School Student Paper	<1%
59	Warde-Farley, D., S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G. D. Bader, and Q. Morris. "The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function", Nucleic Acids Research, 2010. Publication	<1%

60	pubs.acs.org Internet Source	<1%

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		

Date of Registration: 06/08/2016
Date of Expiry : 31/12/2022

UNIVERSITY OF HYDERABAD
SCHOOL OF LIFE SCIENCES
DEPARTMENT OF BIOCHEMISTRY

PROFORMA FOR SUBMISSION OF THESIS/DISSERTATION

1. Name of the Candidate : SANTOSH KUMAR
2. Roll No. : 16LBPH04
3. Year of Registration : 2016
4. Topic : Probing of biomarkers predictive of pan-cancer drug-sensitivity & resistance & drug-repurposing.
5. Supervisor(s) and affiliations : Dr. Selma Mishra, Dept. of Biochemistry, UoH
6. Whether extension/re-registration granted: As per UoH rules (Specify below)
7. Date of submission : 06/01/2023
8. Last class attended on (M.Phil/Pre-Ph.D): 06/09/2022
9. Whether tuition fee for the current period Paid : Receipt No.& Dt :
10. Whether thesis submission fee paid : Receipt No & Dt : 22121413882951 6, 14th Dec, 2022
11. Whether hostel dues cleared : Receipt No & Dt : Rs-1200/—
12. Whether a summary of the dissertation enclosed:

05.01.23
Supervisor(s) Signature:
SELMA MISHRA
Assistant Professor
Dept. of Biochemistry
School of Life Sciences
University of Hyderabad
HYDERABAD-500 046. INDIA

Signature of the Head of the Department.
HEAD
Dept. of Biochemistry
SCHOOL OF LIFE SCIENCES
UNIVERSITY OF HYDERABAD
HYDERABAD-500 046.

Signature of the Dean, of the School :
6/1/2023
DEAN
School of Life Sciences
University of Hyderabad
Hyderabad-500 046.

Forwarded on :

BC/03 5/1/2023

Dean/SLS/38 6/1