Shailza Singh *Editor*

Systems Biomedicine Approaches in Cancer Research

Springer

Systems Biomedicine Approaches in Cancer Research

Shailza Singh
Editor

Systems Biomedicine Approaches in Cancer Research

Springer

Editor
Shailza Singh
Pathogenesis and Cellular Response
National Centre for Cell Science
Ganeshkhind, Pune, Maharashtra, India

ISBN 978-981-19-1952-7 ISBN 978-981-19-1953-4 (eBook)
https://doi.org/10.1007/978-981-19-1953-4

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Dear Readers,

Cancer is a deadly disease that has claimed the lives of millions of people. Research on cancer has been going on for well over a century. There have been a huge number of books and research papers on cancer. For cancer diagnosis, therapy, and prevention, this integrated book brings out the numerous facets which are timely and invaluable. Cancer systems biomedicine, with its unique synthesis of experimental biology, computational and mathematical analysis, is particularly positioned to address the complexity associated with cancer. The purpose of the book is to provide a bird's eye view of the evolving cancer ecosystem, allowing cancer biologists and oncologists to understand and forecast how one modification affects an entire tumor system, rather than viewing cancer through the lens of a single mutation or alteration. The collection of chapters represents the first systematic efforts to demonstrate all the different facets of Systems Biomedicine Approach in Cancer Research. Chapter 1 talks about the role of medicinal plants and its impact on Systems Biomedicine. The authors evaluated the recent initiatives aimed at improving natural product biosynthesis pathway discovery, activation, and modulation. Drug discovery could benefit from the use of these new biosynthetic pathway models and the integration of "omics" data. Chapter 2 summarizes lucidly artificial intelligence (AI) and machine learning (ML) as players in cancer research. This chapter looks at how ML can be used to improve diagnosis and therapy. The authors lay out a vision for how ML might revolutionize three areas of biomedicine, viz. clinical diagnostics, precision therapeutics, and health monitoring, with the key focus on health. Due to the proliferation of data, cancer systems biology is an ever-growing field of study; the difficulty is figuring out how to mine that data and extract relevant information. To gain a better understanding of carcinogenesis, researchers must methodically exploit a variety of resources, including databases, microarrays, and next-generation sequencing. The management and analysis of cancer data, the creation and deposition of databases, whole transcriptome and genome comparison, analyzing results from high-throughput experiments to uncover cellular pathways and molecular interactions, and the design of effective algorithms to identify potential biomarkers are all topics in the field of Cancer Biomarkers. Chapter 3 encompasses Cancer Biomarkers in the era of Systems Biology. Chapter 4 highlights mPGES as inhibitors in Cancer Biomedicine. Chapter 5 talks about the

Emerging Role of Structural and Systems Biology in Anticancer Therapeutics. The goal of systems biology is to anticipate the behavior of biological systems based on the molecules that are involved. As a result, understanding the interactions between these molecules is critical to these endeavors. Only three-dimensional structures provide a complete understanding of how molecules interact, yet structural biology is still difficult for complexes of two or more macromolecules. As a result, the methodologies utilized to anticipate structural features for interactions are extremely important. Protein docking, homology modeling, or detecting repeated interaction-sequence signatures, either a pair of domains or a domain, can all be used to anticipate the chemical interactions. It is conceivable to anticipate the architecture of massive molecular assemblies or the details of how biological pathways function using these methods. Complementing the interactome with structural data will result in a more complete whole-cell framework at the atomic level, which will have a significant impact on cancer system biomedicine research. In recent times, gene fusions, which comprise genomic rearrangements that fuse regulatory or coding regions from two separate genes, are a common form of mutation in numerous cancer types. The genetics of tumors including fusion oncogenes and the proteins they encode has improved cancer diagnosis and treatment in some circumstances. However, little is known about the impact of fusion genes' complicated structure on the biogenesis of the chimeric transcripts they produce. An improved understanding of fusion transcript synthesis and the diversity of chimeric RNAs found in fusion-driven cancers would improve the chances of RNA-based therapies being successful in cancer malignancy. Chapter 6 envisions the computational tools and methods for fusion transcripts used to study cancer biology. Chapter 7 is about understanding molecular kinetics in non-small cell lung cancer which talks about the research focused on a small number of genes and proteins that have yielded crucial insights into the complex interactions which occur within and between cells. To turn complex datasets that span diverse length and time dimensions into usable knowledge, systems analysis and predictive modeling are required. Systems biomedicine will face hurdles in terms of technology, experimental design, data processing, and data integration. Given the system medicine promise and interest in the topic, as evidenced by the rise in the number of literature, researchers worldwide appear to believe that systems biomedicine has significant potential to aid in the study of cancer biology.

Pune, India Shailza Singh

Acknowledgments

Systems Medicine interdisciplinary and collaborative character gives a rich environment for sharing the most innovative and cutting-edge findings. Today's generation whether be faculties, scientists, engineers, or researchers should be able to use computational and experimental tools to do analysis, create, and invent, as well as tinker further in a lab on their own with the various elements of science involved in day-to-day study. This is critical to innovation because it unlocks science's hidden treasures, since a biological model enables interdisciplinary approaches to real-world challenges. Collaboration across disciplines gives rise to new paths for solutions, with the goal of a more sustainable environment. This book is a culmination of efforts bringing together the support and love of our own near and dear ones during these unprecedented times.

I'd like to express my gratitude to Bhavik Sawhney, Ashok Kumar, Lenold Esithor, and the Springer Nature Publishing Group for their help and dedication in preparing this book for publication. My special mention to all my outstanding graduate students, Ritika Kabra, Anurag, Prajakta Nimsarkar, Nikhil, Shweta, Vrushali, Pooja, Prajakta Ingale, and Komal Kharat, for their unwavering support throughout the process. Finally, I'd like to express my gratitude to my family, especially Isha and Akshaya, my parents, for their patience with me during the process. This book would not have been possible without their love and support.

Shailza Singh, PhD

Contents

About the Editor

Shailza Singh is serving as Scientist E and in charge of the bioinformatics and high performance computing facility. Her lab focuses on systems and synthetic biology of infectious disease and cancer model systems, wherein she is trying to integrate the action of regulatory circuits, cross-talk between pathways, and nonlinear kinetics of biochemical processes through mathematical modeling. She is the recipient of several awards such as RGYI, DST-Young Scientist, INSA Bilateral Exchange, and SAKURA Exchange Programme. Dr. Singh is also serving as a reviewer and academic editor of various international journals of repute.

Medicinal Plants for Indigenous Cancer Drug Discovery: Current to Future

1

Pragya Misra, Prajakta Nimsarkar, and Shailza Singh

Abstract

Medicinal plants serve as a rich source of therapeutic modalities against various diseases especially cancer, which gained attention by researchers, due to the problems associated with available treatments. Indian traditional medicinal systems have a rich repository of medicinal plants and could serve as a source of safe and cost-effective alternative therapy. A major limitation with natural products is meager information about their mode of action, hindering their wider clinical use. Herein, we are focusing on anticancer potential of few medicinal plants based on common chemical constituents, viz. *Peganum harmala*, *Quercus infectoria*, *Melissa officinalis,* and *Plumbago zeylanica,* with well-cited use in folk medicines. Anticancer mechanism of these plants/active components has been well deciphered with targeted pathways involved in their anticancer potential. *Peganum harmala* and its active constituents showed involvement of various signaling pathways for anticancer effect. Cancer cell death by *Quercus infectoria* and its constituents also involved pathways such as AKT, NF-κB, and JAK/STAT. *Melissa officinalis* showed anticancer potential by affecting various transcription factors such as NF-κB, TNF-α, and COX-2. Anticancer activity mediated by *plumbagin* isolated from *Plumbago* also showed effect on various transcription factors/signaling pathways as pTEN, mTOR, and Akt pathways, thereby inhibiting survival signaling. However, overall data suggested the need for more extensive studies focused on clinical investigation for pharmacokinetics, bioavailability, and toxicity of these plants/plant products. We propose in this chapter that studies based on computational biology using identified pure compounds and their target to develop pharmacophore model for drug discovery would be a fast and feasible way to design new potential derivatives.

P. Misra · P. Nimsarkar · S. Singh (✉)
National Centre for Cell Science, Pune, India

S. Singh (ed.), *Systems Biomedicine Approaches in Cancer Research*,
https://doi.org/10.1007/978-981-19-1953-4_1

Abbreviations

COX-2	Cyclooxygenase-2
DYRK	Dual-specificity tyrosine phosphorylation-regulated kinase
GQI	*Galls of Quercus infectoria*
MA	*Melissa officinalis*
MRP1	Multidrug resistance-associated protein 1
PG	*Peganum harmala*
PU	*Plumbagin*

1.1 Introduction

Cancer, a metabolic syndrome, is associated with second highest mortality and morbidity worldwide with increasing cases. Due to the associated complexities, it has been a major area of research for new therapeutic/preventive strategies (Bakitas 2007). Current treatment regimen includes chemotherapy, radiotherapy, and novel strategies like monoclonal antibodies against immune checkpoint inhibitors (Farkona et al. 2016; Falzone et al. 2018). However, all these treatment strategies are associated with severe side effects (Schirrmacher 2019; Kennedy and Salama 2020). Therefore, there is an urgent need for developing alternative treatments.

Medicinal plants have been the source of treatment for various ailments worldwide through ages. With associated advantage of safety and tolerance, drug development program based on medicinal plants has gained a lot of attention in last decades. Various plant extracts of traditional/folk medicinal system and their phytochemicals have been studied for treatment/prevention of cancer. Indian Ayurveda system is a rich repository of plants with medicinal value, and these have been extensively studied for various ailments (Behere et al. 2013). Herein, we have selected four such plants based on their known anticancer/medicinal potential since ancient times and have explored steps ahead, in terms of identifying active phyto-constituents and mode of action, which could help in providing an interface for traditional medicinal to be used in clinics (Vaidya and Devasagayam 2007; Pandey et al. 2016; Roy and Bharadvaja 2017). Another major criteria for selection of these plants were their common phyto-constituents including flavonoids, alkaloids, phenolics, and essential oils, which have known pharmacological functions (Yin et al. 2013). We would discuss these medicinal plants, viz. *Peganum harmala*, *Quercus infectoria*, *Melissa officinalis,* and *Plumbago zeylanica,* with focus on their active constituents and mode of action.

***Peganum harmala* (PG)**
Kingdom: Plantae
Division: Magnoliophyta
Class: Magnoliopsida—Dicotyledons
Subclass: Rosidae

Order: Sapindales
Family: Zygophyllaceae
Genus: *Peganum L.—Peganum P*
Species: *Peganum harmala L.*

Peganum harmala (PG) also known harmala, Syrian rue, wild rue, harmal, and Africa rue is native of Africa, Middle-East, and Mediterranean region along with historical reports of its presence in Turkey, Iran, Iraq, Uzbekistan, Tajikistan, Russia, China, Mongolia, Afghanistan, India, and Pakistan (Moloudizargari et al. 2013). Seeds, fruits, root, and bark of *Peganum harmala* have been used as folk medicine. It has been used for various ailments such as antispasmodic, antipyretic, and diseases of digestive system (Mamedov et al. 2017).

Studies conducted to screen the phytochemical composition of PG showed the presence of flavonoids, alkaloids, tannins, triterpenes, sterol, anthraquinones, coumarins and volatile oils in seeds, roots, and aerial parts of plants. B-Carboline alkaloids such as harmaline and harmine have been found in seeds and roots along with harmalol, harman, and harmol, and quinazoline alkaloids such as peganine, deoxypeganine, deoxyvasicinone, isopeganine, pegaminem, peganol, dipegene, and peganones were obtained from seeds and whole plant (Moloudizargari et al. 2013). Most of the medicinal properties of PG have been attributed to presence of these alkaloids.

1.2 Anticancer Potential of *Peganum harmala* (PG)

Traditional medicinal uses of PG (Pandey et al. 2016) had drawn the attention of researchers worldwide to carry out various pharmacological studies evaluating its anticancer role and identifying active components for anticancer efficacy. Seed extract of PG induced apoptosis in breast cancer cells and decreased expression of Bcl-1 along with increased expression of Bax, Puma, TRAIL, and caspase-8. This concluded that PG mediated breast cancer cell death via apoptosis involved both intrinsic and extrinsic pathways (Hashemi Sheikh Shabani et al. 2015). Mode of action of *Peganum harmala*/its derivatives against various forms of cancer is represented in Fig. 1.1.

Alkaloids present in various plant parts of PG are major contributors for its anticancer potential. Studies conducted in breast cancer showed that β-carboline alkaloid from PG, harmine inhibited the overexpression of ABC transporter protein BCRP, which has a role in multidrug resistance. Harmine also showed its anticancer potential by inhibiting resistance of these cells, for anticancer drugs mitoxantrone and camptothecin, found to be mediated by overexpression of BCRP (Ma and Wink 2010).

Harmine was found effective against childhood malignancy neuroblastoma (NB). Poor prognosis of NB is associated with overexpression of MYCN gene. Therefore, to assess the role of harmine against NB, four cells lines, two having amplification of MYCN and two without it, were selected. It was found that harmine induced apoptotic cell death by cleavage of PARP, mediated by caspase, and it was more

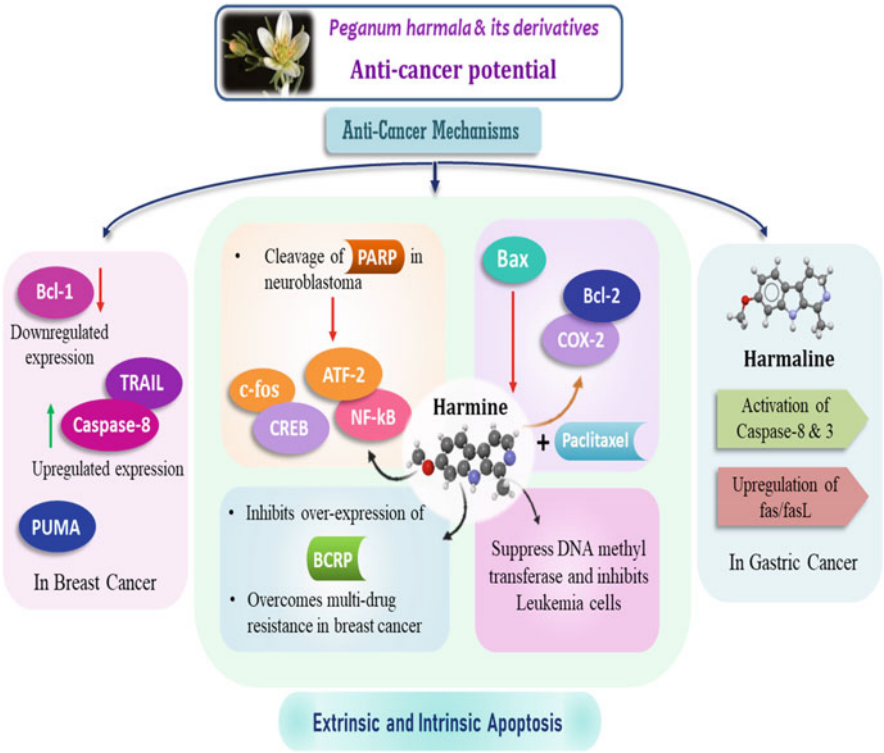


Fig. 1.1 Mode of action of *Peganum harmala* and its derivatives against various forms of cancer

efficient in cells with amplified MYCN gene. Patients having high expression of dual-specificity tyrosine phosphorylation-regulated kinase (DYRK), DYRK2 and DYRK3, show poor prognosis and are also co-related to expression of MYCN. Based on this clinical data, it was believed harmine might have DYRK inhibitory potential (Uhl et al. 2018).

Effect of harmine was also assessed against melanoma cells and found to be mediated by both extrinsic and intrinsic apoptotic pathways, as evident by activation of caspase-3, caspase-8, and caspase-9 along with overexpression of Bax and Bid. Various transcription factors such as NF-κB subunits, c-Fos, ATF-2, and CREB were found to be downregulated in harmine-treated cells. Most of these transcription factors are negative regulators of apoptosis (Hamsa and Kuttan 2011).

Synergistic anticancer effect of harmine was tested with paclitaxel in gastric cancer cells. These cells expressed cyclooxygenase-2 (COX-2) which helps in progression of tumor. This combination therapy was found to be significantly efficient in inhibiting cancer cell growth by inducing apoptosis and downregulating COX-2, Bcl-2, and overexpression of Bax (Yu et al. 2016). Suppression of DNA methyltransferases was explored as mechanism for anticancer potential of harmine against leukemia. This strategy has been used to develop therapeutic targets against

various malignancies (Blum and Marcucci 2005; Brueckner et al. 2005). It was observed that treatment with harmine reduced the proliferation of cells along with suppressing DNMT1 gene (DNA methyltransferase 1). Due to suppression of DNMT1, tumor suppressor promoter p53 was hypomethylated leading to its activation, thereby inducing cell death (Oodi et al. 2017). Harmaline, isolated from PG, was found to show efficacy against gastric tumor mediated through apoptosis involving activation of caspase-8 and caspase-3 along with upregulation of Fas/FasL and cell cycle arrest (Wang et al. 2015).

Various human and animal studies have reported intoxication induced by PG and extract of PG is found to be toxic at higher dose causing complications such as paralysis, liver degeneration, and digestive issues (Lamchouri et al. 2002; Herraiz et al. 2010). However, a clinical trial study wherein PG oral capsules were used for improving urinary symptoms of benign prostate enlargement showed that in regulated doses, this extract did not affect vital parameters of body (Shirani-Boroujeni et al. 2017). Most of these studies suggested that special attention should be paid by researchers and clinicians for therapeutic usage of this plant by conducting detailed studies for its safer dose.

***Melissa officinalis* (MA)**
Kingdom: Plantae
Division: Magnoliophyta
Class: Magnoliopsida—dicotyledons
Subclass: Asteridae
Order: Lamiales
Family: Lamiaceae—mint family
Genus: *Melissa L.*—balm P
Species: *Melissa officinalis L.*

It is commonly known as lemon balm, honey balm, garden balm, etc., due to its lemon-like fragrance (Singh Verma et al. 2015). Plant is native of Eastern Mediterranean Region and Western Asia, Germany, France, Italy Romania, Bulgaria, and North America (Meftahizade et al. 2010). Medicinal importance of this plant has been documented back to 50–80 BC (Kennedy et al. 2003). The plant has been used in Europe, Austria (Vogl et al. 2013), and Iranian medicinal system for various disorders (Shakeri et al. 2016). Indian Ayurvedic system has also mentioned the importance of *Melissa officinalis* (MA) in improving memory (Singhal et al. 2012). This plant contains terpenes, phenols, flavonoids, and essential oils as major chemical constituents (Cohen et al. 1964; Herrmann and Kucera 1967; Mulkens and Kapetanidis 1987; Žiaková et al. 2003; Allahverdiyev et al. 2004; Awad et al. 2009; Meftahizade et al. 2010; Moradkhani et al. 2010). Important active constituents include volatile oils such as geranial, neral, citronellal and geraniol, triterpenes such as ursolic acid, and phenolics such as luteolin, naringin, and hesperidin.

Melissa officinalis (MA) and its active components have been explored for anticancer potential and pharmacological studies which have revealed a lot about its mode of action as discussed below and depicted in Fig. 1.2.

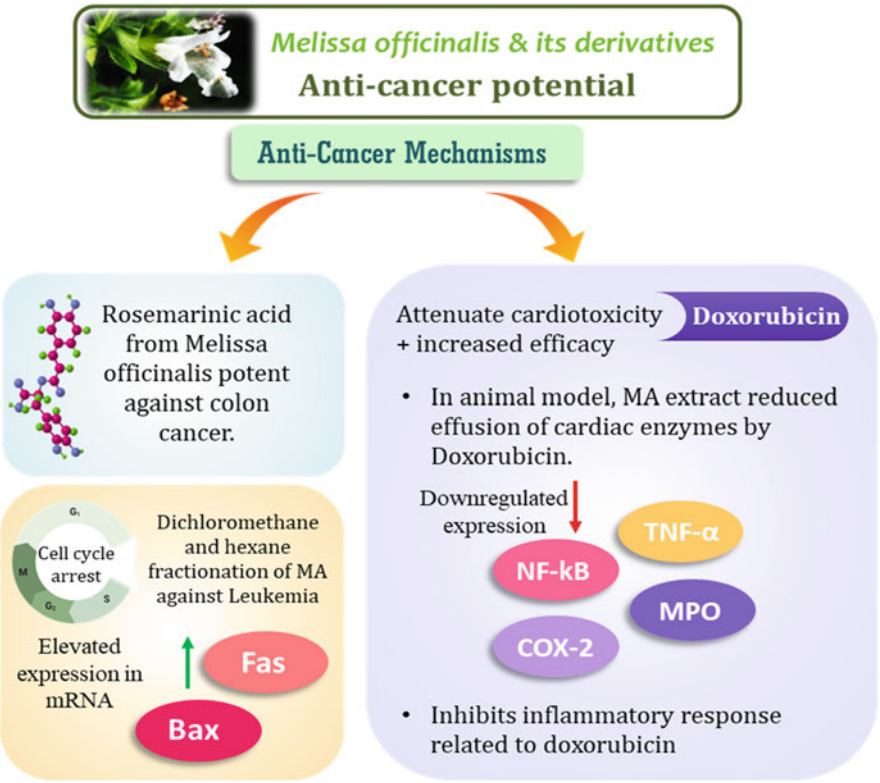


Fig. 1.2 Mode of action of *Melissa officinalis* and its derivatives against various forms of cancer

1.3 Anticancer Potential of *Melissa officinalis* (MA)

Hydroalcoholic extract of MA showed anticancer potential against various cancer cell lines including lung cancer, breast adenocarcinoma, ovarian cancer, and prostate cancer (Jahanban-Esfahlan et al. 2015). MA was also found to attenuate the cytotoxicity induced by doxorubicin (DOX), a known anticancer drug against breast cancer. It was observed that when rats were treated with MA, it reduced the leakage of cardiac enzymes by doxorubicin along with decreasing the oxidative stress. Treatment with MA also reduced the expression of NF-κB, TNF-α, and COX-2. It also reduced the activity of myeloperoxidase, thereby suppressing the inflammatory response induced by DOX. This data suggested the synergistic potential of MA with DOX against human breast cancer cells (MCF-7) (Saraydin et al. 2012).

Anticancer effect of ethanolic/aqueous extract of MA was evaluated against human colon cancer cells, and it was found that MA showed potent antiproliferative effect mediated by rosmarinic acid present in it (Encalada et al. 2011). Dichloromethane and *n*-hexane fractions of MA showed significant inhibitory

potential against leukemia cell lines. Effect of dichloromethane fraction was found to be mediated by cell cycle arrest and apoptosis. Both intrinsic and extrinsic apoptotic pathways were involved as evident by upregulation of Fas and Bax mRNA expression as well as the Bax/Bcl-2 ratio posttreatment with MA fraction (Ebrahimnezhad Darzi and Amirghofran 2013).

Glioblastoma multiforme develops from glial cells, and no treatment regimen has been fully effective against it. Effect of essential oils of MA and citral was evaluated against GBM cells and was found to show anticancer potential mediated by apoptosis. Citral downregulated the expression of multidrug resistance-associated protein 1 (MRP1). It can be inferred that essential oils of MA could be an effective therapeutic strategy against glioblastoma (De Queiroz et al. 2014).

Various studies have been conducted to assess the toxicity and tolerability of MA in humans and showed varied data. A clinical trial in healthy individuals and the other one in stressed volunteers showed no cytotoxicity by MA containing rosmarinic acid. A double-blind, randomized trial to treat heart palpitation also did not indicate any serious side effects. Some side effects like vomiting, dizziness, wheezing, agitation, abdominal pain, and nausea were observed in a double-blind, randomized, placebo-controlled trial of Alzheimer's disease patients (Cases et al. 2011; Alijaniha et al. 2015).

***Galls of Quercus infectoria* (GQI)**

Kingdom: Plantae

Division: Magnoliophyta—flowering plants

Class: Magnoliopsida—dicotyledons

Order: Fagales

Family: Fagaceae

Genus: *Quercus L.*

Species: *Quercus infectoria Olivier*

Quercus infectoria Olivier habitat includes Turkey, Greece, along with Asia Minor, Europe, and North Africa (Harris 2003). Galls are irregular growth on Quercus and which develop due to interaction between plant hormones and chemicals produced by insects (Bartlett and Connor 2014). Galls also known as *Galla Turcica* have been used for wide range of medicinal potential in folk medicines (The Wealth of India 1952; The Indian Pharmacopoeia Commission 2007; Bhati et al. 2012). The main chemical constituent present in these galls is tannins along with syringic acid, ellagic acid, sitosterol, amentoflavone, hexamethyl ether, isocryptomerin, methyl betulate, methyloleanate, and hexagalloylglucose; 50–70% constituent of galls is gallotannic acid (Harris 2003). Its pharmacological efficacy has further been explored and found to include anticancer, anti-inflammatory, antidiabetic, antifungal, and anti-MRSA (methicillin-resistant *Staphylococcus aureus*)/antibacterial (Vermani et al. 2009; Basri and Khairon 2012; Basri et al. 2013; Sithisarn et al. 2015). Herein, this chapter we are focusing specifically on its anticancer potential.

1.4 Anticancer Effect of *Galls of Quercus infectoria* (GQI)

Anticancer potential of *galls of Quercus infectoria* (GQI) and its active constituents has been well studied. It was observed that ethyl acetate extract of GQI inhibited EGFR (epidermal growth factor receptor), a potent anticancer drug target (Wang et al. 2014).

Ellagic acid from GQI showed potent anticancer effect mediated via apoptosis along with inhibiting the migration and invasion of cancer cells. This effect was mediated via impacting many signaling pathways such as PKC pathway, TGF-β/Smad3 pathway (Vanella et al. 2013; Zhang et al. 2014; Chen et al. 2015; Mishra and Vinayak 2015; Salimi et al. 2015). Other major chemical constituent of GQI is methyl gallate, which blocks AKT, NF-κB and JAK/STAT pathways leading to apoptosis-mediated cell death (Chaudhuri et al. 2015; Afsar et al. 2016). It also helps in overcoming the immune suppression state in tumor by inhibiting infiltration of regulatory T cells (Lee et al. 2010).

Gallic acid isolated from GQI showed anticancer effect by cell cycle arrest leading to apoptosis. It also affects angiogenesis and metagenesis of cancer cells. Other important targets for anticancer potential of gallic acid include activation of ATM kinases, inhibition of COX, depletion of GSH, and inhibition of VEGF along with inhibition of NF-κB (Madlener et al. 2007; Lu et al. 2010, 2016; You and Park 2010; Sun and McKallip 2011; Chandramohan Reddy et al. 2012; Ho et al. 2013; He et al. 2016; Kennedy and Salama 2020).

1,2,3,4,6-*Penta-O*-galloylglucose isolated from GQI was found to be effective against breast cancer cells via affecting metabolic genes, for example, targeting overexpression of lactic acid dehydrogenase-A (Deiab et al. 2015).

Toxicity of GQI was evaluated in mice model, and it was found that even at 300 times higher dose mice survived along with few biochemical and hematological parameters changed initially which further resolved with time. In spite of promising data, further studies need to be done before clinical trials in human (Iminjan et al. 2014) (Fig. 1.3).

Plumbago zeylanica along with many other species of Plumbago, viz. *Plumbago rosea, Plumbago capensis, Plumbago europaea, and Plumbago scandens,* has been reported for its medicinal use in history since second century (Bhati et al. 2012). Root and extract of other part of this plant were used for treating dyspepsia, piles, diarrhea, skin diseases, tuberculosis, and leprosy (Jain et al. 2014). Most of the medicinal properties of *Plumbago zeylanica* are attributed to one of the naphthoquinone metabolite (5-hydroxy-2-methyl-1,4-naphthoquinone) isolated from genus Plumbago and have been found as a miracle molecule with activity against various forms of cancer. Herein, this section we will focus on its anticancer potential.

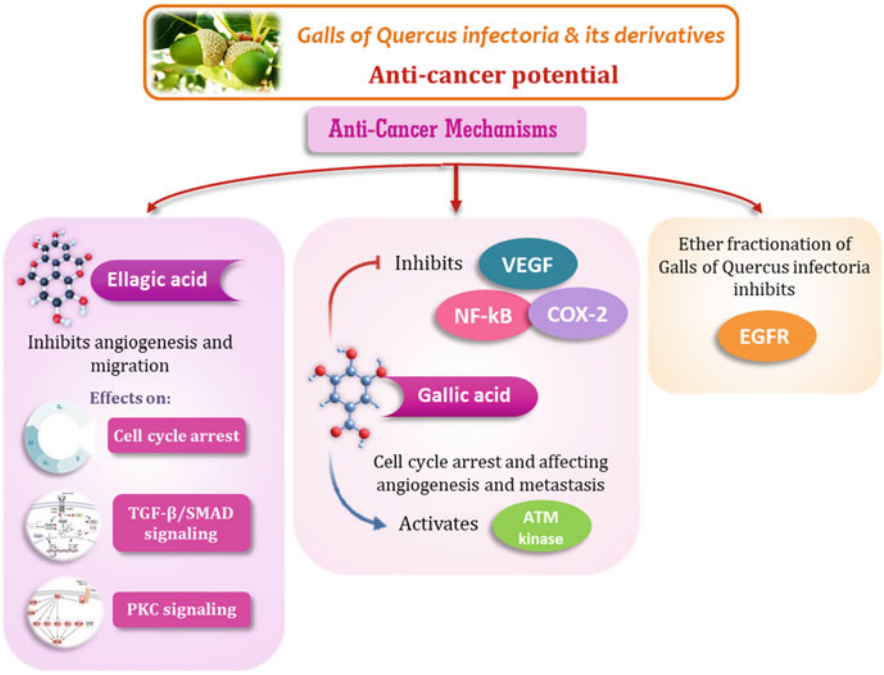


Fig. 1.3 Depiction of anticancerous properties of *Quercus infectoria* galls and its mechanisms

1.5 Anticancer Potential of *Plumbago* with Focus on *Plumbagin* (PU)

Plumbago zeylanica along with other species has shown anticancer potential against various forms of cancer and studies suggested that most of the anticancer activity is due to presence of *Plumbagin* (Hiradeve et al. 2011; Sundari et al. 2017). *Plumbagin* (PU) has been found to be effective against hormone-refractory phase, i.e., androgen-independent (AI) phase of prostate cancer which is the last stage leading to death (Edwards and Bartlett 2005; Quinn et al. 2005). It was found that PU inhibited the progression of AI prostate cancer cells both in vitro and in vivo. Expression of multiple targets such as PKCε, PI3K, pAKT, pJAK-2, pStat3, and NF-κB was inhibited both in prostate cancer cell line and DU145 xenografts (Aziz et al. 2008). PU generated ROS and depleted GSH levels in androgen-independent prostate cancer cells which lacks p53 and induced apoptosis. It also altered the expression of superoxide dismutase 2 (Powolny and Singh 2008). The anticancerous properties of Plumbago are depicted in Fig. 1.4 with the modes of action.

PU was found to induce cell cycle arrest and autophagy-mediated cell death in human breast cancer cells. It blocked the activation of Akt pathway and other downstream molecules such as mammalian target of rapamycin (mTOR), forkhead

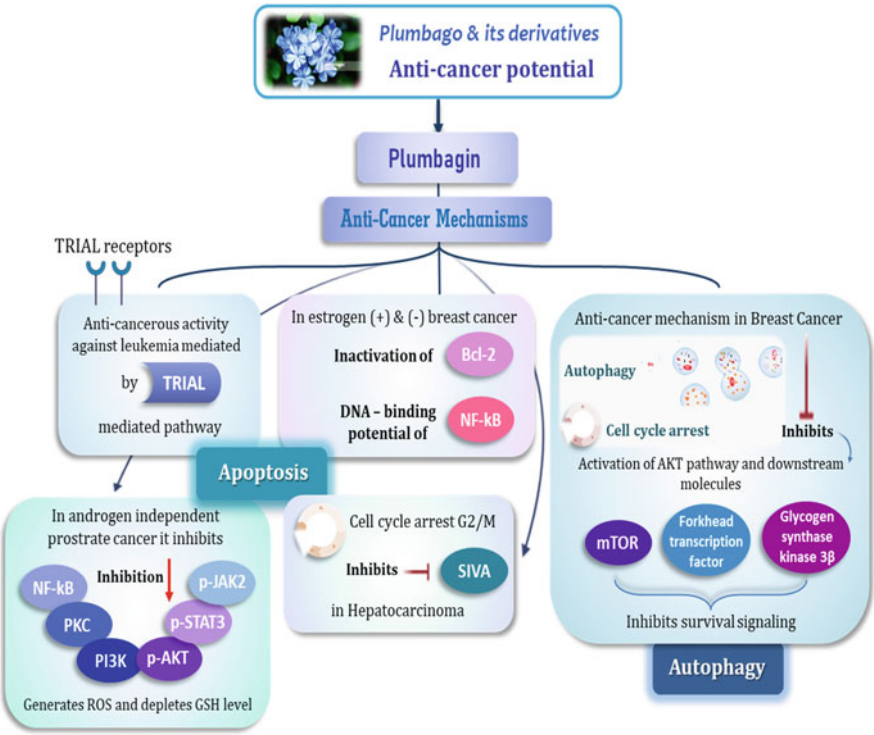


Fig. 1.4 Anticancerous potential of Plumbagin and its mode of action

transcription factors, and glycogen synthase kinase 3β, thereby inhibiting survival signaling (Kuo et al. 2006).

Hepatocarcinoma was also found to be inhibited both in vitro and in vivo by PU, and this anticancer effect was found to be mediated by cell cycle arrest at G2/M phase along with inhibition of expression of p53-targeted gene SIVA and mTOR. SIVA activates mTOR signaling and therefore contributes to tumorigenesis (Prasad et al. 1997; Van Nostrand et al. 2016). Another study using Wistar male rats wherein hepatoma was induced by 3-methyl-4-dimethyl aminoazobenzene (3Me-DAB) showed tumor regression on treatment with PU along with increased levels of glycolic enzymes and decreased levels of enzymes of gluconeogenesis pathway (Parimala and Sachdanandam 1993).

It was found that PU was cytotoxic to leukemia cells and role of ROS was established in the cell death. It was also observed that treatment with PU significantly enhanced the expression of TRAIL-R1 (DR4) and TRAIL-R2 (DR5) which mediate the TRIAL-dependent apoptosis of tumor cells. Data suggested that PU shows its anticancer activity by TRIAL-mediated pathway (Sun and McKallip 2011).

Anticancer potential of PU against non-small cell lung cancer was mediated by G2/M phase arrest and apoptosis. Accumulation of p53 and phosphor-p53 was also

observed along with increased levels of p21 and decreased levels of cyclin B1, Cdc2, and Cdc25C. It was also found that c-Jun Nh2-terminal kinase (JNK), which has proven role in apoptosis mediated by various agents, was an important mediator for PU-induced inhibition of cell growth (Liu and Lin 2005). Downregulation of oncogenic growth factors EGFR/Neu along with downstream signaling targets was observed in H460 lung cancer cells treated with PU. It was also found that PU activated JNK/p38 signaling pathway, promoting apoptosis in cancer cells (Gomathinayagam et al. 2008).

Ovarian cancer cells PEO-1 and PEO-4 were found to be sensitive to treatment with PU. Inhibition of VEGF-A and Glut-1 was also observed on treatment with PU. OVCAR-5 tumor-bearing mice when treated with PU showed tumor regression (Sinha et al. 2013).

BG1 ovarian cancer cells were treated with few standard anticancer drugs and PU. It was found that PU induced cell death most efficiently. These cells were positive for estrogen receptor, and it was found that PU can bind with active site of ER-α and induce its truncated form, thereby inhibiting the classical ER-α signaling pathway (Srinivas et al. 2004; Thasni et al. 2008). Pure compounds from plants showing anticancerous properties have been tabulated in Table 1.1.

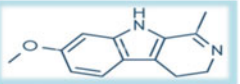
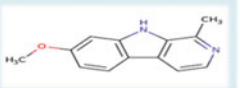
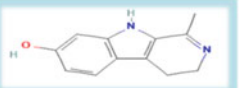
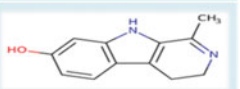
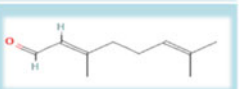
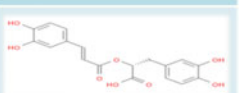
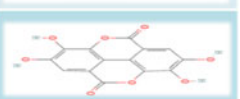
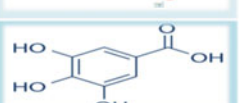
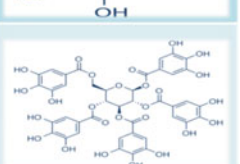
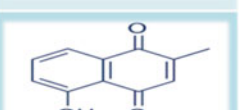
1.6 Computational Approach Toward Natural Products and Drug Discovery

The process of developing new pharmaceutical drugs is known as drug discovery. It is one of the most critical aspects of pharmaceutical research and development (validating, testing, and launching a new drug). Since the advent of modern medicine, the preponderance of systematic drug discovery has focused on small-molecule prospects. For example, small molecules make up over 86% of the medications (both authorized and investigational) in the DrugBank database (Wishart et al. 2017). Lipinski's "rule of five" is a set of established best practice rules for identifying potential orally active pharmacological candidates that reflects small molecules' pervasive nature in drug development (Lipinski 2004).

Sequence-based characteristics, interactions with body structures (proteins, metabolites, tissues, cells, and so on), pathway disruptions, and toxicity are only a few examples of bioinformatics methodologies used in drug discovery. High-throughput sequencing with multiomics is two examples of bioinformatics sciences. Bioinformatics may be utilized in the drug development process in a variety of ways (Wishart et al. 2017; Thomford et al. 2018). In the case of natural products, scientists can employ a wide range of methodologies relating to the organisms that manufacture the substances. Evolution and phylogenetics, for instance, provide a variety of avenues for drug discovery.

In recent decades, computational strategies have become mainstream in research, extending to medication development (Sliwoski et al. 2014). For example, cheminformatics is the application of computer science to understand and characterize the molecular properties and chemical behavior of particular compounds. These

Table 1.1 2D structure of few pure compounds from plants showing anticancer potential

S.no.	Plant	Pure Compound	Structure
1	Peganum harmala	Harmaline	
		Harmine	
		Harmol	
		Harmalol	
2	Melissa officinalis	Citral	
		Rosmarinic acid	
3	Quercus infectoria	Ellagic acid	
		Methyl gallate	
		Gallic acid	
		1,2,3,4,6-Penta-O-galloyl glucose	
4	Plumbago zeylanica	Plumbagin	

techniques have resulted in enormous libraries of tiny compounds that may be used to search for specific therapeutic activities (Blaney and Martin 1997). Several cheminformatics methods may be used to generate libraries of chemically and structurally related compounds after identified candidates have been identified to enhance stability, toxicity, and kinetics. Bioinformatics approaches may also be used to figure out how potential medications produce therapeutic activity in the human body, such as predicting drug–protein interactions, determining the impact on biological pathways and functions, as well as genetic differences that determine drug response (Drews 2000). Regardless of these advancements in drug discovery techniques, new therapeutic pharmaceutical approvals have stalled in recent years. The number of distinct molecular entities authorized by the US FDA decreased from 53 to 17 each year between 1996 and 2007, the same rate as it has been for more than 50 years (FitzGerald 2008; Munos 2009). These studies might be supplemented with a special debate confined to in silico techniques for natural product drug development, considering the developments as mentioned earlier in emerging computational tools and advances in traditional informatics for translational applications.

A rising move toward data-driven drug discovery is another trend in drug development that has been supported by informatics and computational approaches (Tatonetti et al. 2012; Lusher et al. 2014). The traditional drug development process was as follows: researchers would first find a target structure in the human body associated with a disease or condition and then screen for "lead" chemicals that are associated with the target. The most promising ideas are then narrowed down and placed through the development process, where they are tested for safety and efficacy in model organisms before being tested on humans. An overview of drug discovery by systems biology means has been highlighted in Fig. 1.5.

1.7 Therapeutic Natural Products Divided into Several Categories

Natural Products: "Natural products" are seemed to be limited to small-molecule secondary metabolites (Nature Publishing Group 2007) for some authors, while others define them more generally as any chemical component created by a live organism (Nature Publishing Group 2007; National Center for Complementary and Integrative Health 2017).

Fungal Metabolites: Plant and fungal metabolites have several similarities and are classified in the same way (the flavonoid compounds to be most notable one). Like plant metabolites, fungi can treat a wide range of ailments and problems, although they are best recognized for producing numerous powerful antibiotics. Statins (mevastatin, lovastatin), immunosuppressants (ciclosporin), antimalarials (antiamoebin), and other drugs have previously been utilized effectively (Thomford et al. 2018).

Phytochemicals: Plant-produced substances, or phytochemicals, are found in many natural goods. Phytochemicals can be cytotoxic, provide essential nutrients (such as amino acids, antioxidants, and dietary fiber), or be inert in humans.

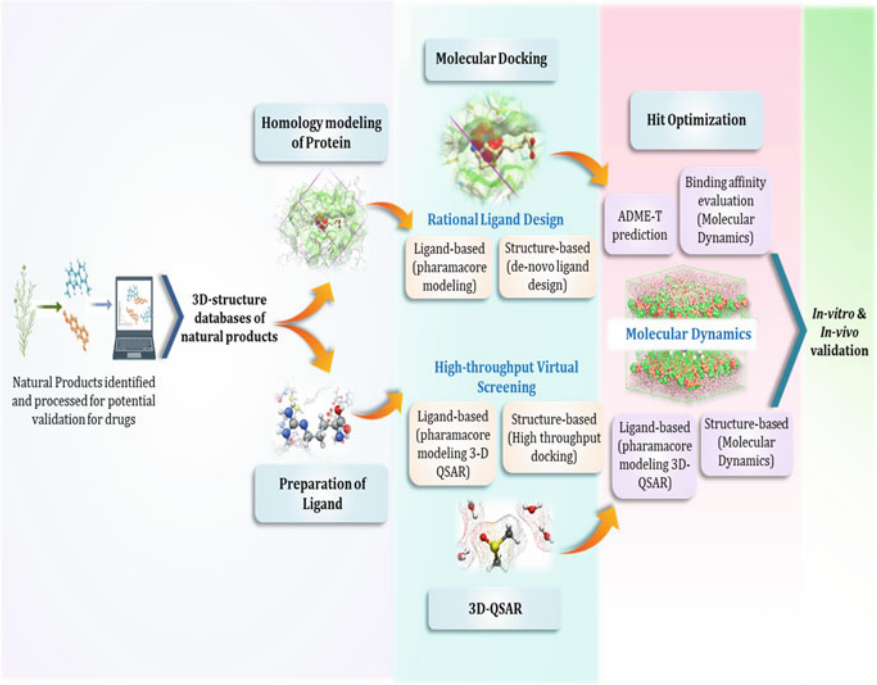


Fig. 1.5 Highlighting the process of drug discovery from natural products using system biology approaches

Phytochemicals are small molecules (rather than macromolecules, prevalent in many of the other classes) found in primary and secondary metabolites in plants. They are divided into phenolic acids, stilbenes, and flavonoids (further subdivided into more specific subclasses) (Harborne 1999).

1.8 Natural Products and Cheminformatics

Many typical cheminformatics methodologies are difficult to apply to some types of natural products, especially those with extensive chemical structures. We classify cheminformatics into three primary kinds of methodologies that have proven successful with natural products, discussing the limitations that must be addressed for natural products specifically.

1.8.1 QSAR Analysis of Natural Products

Quantitative structure–activity relationship (QSAR) analysis is a common approach in cheminformatics for predicting a response variable from a set of structural,

chemical, and maybe physical input elements (known as molecular descriptors). Different classes of natural products have been subjected to QSAR, with specific classes dictating the molecular descriptors employed. Symbolic (1D or 2D) descriptors, 3D spatial organization, higher-order conformational characteristics (e.g., time-dependent or ligand-bound) (Polanski 2009), and nonnatural product applications frequently use experimental data such as partition coefficient, polarizability, and refractivity. In terms of predicting antibody binding affinity to proteins, QSAR has worked well. For each amino acid location in a library of single-chain monoclonal antibodies, a framework consisting of 26 physicochemical parameters (including hydrophobicity, polarity, and electronegativity) was presented (Mandrika et al. 2007). While this technique has yet to be used in the identification of NP drugs, it appears to be a viable option.

1.8.2 Molecular Docking and Dynamics

Molecular docking is a technique for determining if and how two molecules (usually a target and a ligand) will interact physically. Typically, this is accomplished in two steps: (1) determining novel conformational fits and (2) scoring those that are discovered. Molecular dynamics is a prominent simulation tool for docking. The role (target vs. ligand) that a natural product compound plays in docking simulations is usually dictated by the compound's class. Small molecules and short polypeptides are frequently utilized as ligands, whereas targets are more significant proteins and protein complexes (although exceptions are typical). This distinction is crucial, particularly when screening a high number of possible compounds: The target is customarily established, but the ligand can be chosen from a vast number of molecules. Docking many small-molecule compounds is computationally conceivable when a specific molecular target has been determined (Khan et al. 2009; Lee et al. 2011; Ma et al. 2011). Docking simulations may be used to determine which metabolites might bind to a natural macromolecular product if endogenous small-molecule metabolites are suspected of interacting with it (Pithayanukul et al. 2009). When both a target and a ligand have been predicted using other approaches, docking is commonly used as an additional validation step (e.g., QSAR or other methods). Albrand et al. employed molecular dynamics and nuclear magnetic resonance in 1995 to explain how the toxin FS2 (from Black Mamba venom) blocks L-type calcium channels, causing significant cardiotoxicity (Albrand et al. 1995).

1.8.3 Library Construction

The generation of enormous libraries of compounds that may be screened in parallel is one of the most frequent strategies for generating drug candidates, understanding that just a small percentage will result in "hits" (potential therapeutic activity). Combinatorial chemistry (i.e., using combinatorics to enumerate chemical structures) comprises many approaches to creating such libraries (Terrett et al.

1995). ChEMBL and PubChem databases consist of numerous natural products annotated by compound classes (Li et al. 2010; Gaulton et al. 2016). For better annotations, databases such as Dictionary of Marine Natural Products and ArachnoServer are helpful as natural product libraries accumulate characteristic and appeal features (Pineda et al. 2017; Romano et al. 2018).

1.9 Conclusion and Perspectives

This chapter suggests that medicinal plants used in folk medicine and their active components show remarkable anticancer potential which involved various signaling/metabolic pathways. These extended studies not only deciphered the molecular mechanism behind anticancer potential of these plants which would help in bringing these molecules/extracts from laboratory to clinics, but also helped in identification of various biological targets for novel computational-based drug designing strategies. Such extended efforts would not only introduce novel anticancer molecules with safe and potent activity but also reduce their cost due to origin from natural products and also the computational pipeline adopted. However, more detailed in vivo and clinical trial studies are needed to bring these molecules/extracts for clinical use along with toxicity evaluation.

Most of the literature available for cancer treatment indicates various comorbidities along with lower average survival. Medicinal plants have proven strong anticancer potential directly or by immunomodulatory function along with being safe. Bioassay-guided fractionation has identified various pure compounds with strong anticancer potential which could be further synthesized in laboratories. To increase the availability, shelf life and targeted delivery of these molecules, various new technologies such as delivery vehicles (nanoparticles, liposomes, and virosomes) could be used. This would not only allow sustained release of the drug for better activity but also help in implying tissue-specific strategy of drug designing to reduce side effects if any.

References

Afsar T, Trembley JH, Salomon CE, Razak S, Khan MR, Ahmed K (2016) Growth inhibition and apoptosis in cancer cells induced by polyphenolic compounds of *Acacia hydaspica*: involvement of multiple signal transduction pathways. Sci Rep 6:23077. https://doi.org/10.1038/srep23077

Albrand JP, Blackledge MJ, Pascaud F, Hollecker M, Marion D (1995) NMR and restrained molecular dynamics study of the three-dimensional solution structure of toxin fs2, a specific blocker of the l-type calcium channel, isolated from black mamba venom. Biochemistry 34: 5923–5937. https://doi.org/10.1021/bi00017a022

Alijaniha F, Naseri M, Afsharypuor S, Fallahi F, Noorbala A, Mosaddegh M et al (2015) Heart palpitation relief with *Melissa officinalis* leaf extract: double blind, randomized, placebo controlled trial of efficacy and safety. J Ethnopharmacol 164:378–384. https://doi.org/10.1016/j.jep.2015.02.007

Allahverdiyev A, Duran N, Ozguven M, Koltas S (2004) Antiviral activity of the volatile oils of *Melissa officinalis* L. against herpes simplex virus type-2. Phytomedicine 11:657–661. https://doi.org/10.1016/j.phymed.2003.07.014

Awad R, Muhammad A, Durst T, Trudeau VL, Arnason JT (2009) Bioassay-guided fractionation of lemon balm (*Melissa officinalis* L.) using an in vitro measure of GABA transaminase activity. Phyther Res 23:1075–1081. https://doi.org/10.1002/ptr.2712

Aziz MH, Dreckschmidt NE, Verma AK (2008) Plumbagin, a medicinal plant-derived naphthoquinone, is a novel inhibitor of the growth and invasion of hormone-refractory prostate cancer. Cancer Res 68:9024–9032. https://doi.org/10.1158/0008-5472.CAN-08-2494

Bakitas MA (2007) Background noise: the experience of chemotherapy-induced peripheral neuropathy. Nurs Res 56:323–331. https://doi.org/10.1097/01.NNR.0000289503.22414.79

Bartlett L, Connor EF (2014) Exogenous phytohormones and the induction of plant galls by insects. Arthropod Plant Interact 8:339–348. https://doi.org/10.1007/s11829-014-9309-0

Basri DF, Khairon R (2012) Pharmacodynamic interaction of *Quercus infectoria* galls extract in combination with vancomycin against MRSA using microdilution checkerboard and time-kill assay. Evid Based Complement Alternat Med 2012:1–6. https://doi.org/10.1155/2012/493156

Basri DF, Jaffar N, Zin NM, Raj LS (2013) Electron microscope study of gall extract from *Quercus infectoria* in combination with vancomycin against MRSA using post-antibiotic effect determination. Int J Pharm 9:150–156. https://doi.org/10.3923/ijp.2013.150.156

Behere PB, Das A, Yadav R, Behere AP (2013) Ayurvedic concepts related to psychotherapy. Indian J Psychiatry 55:310. https://doi.org/10.4103/0019-5545.105556

Bhati R, Singh A, Saharan V, Ram VA, Bhandari A (2012) *Strychnos nux-vomica* seeds: pharmacognostical standardization, extraction, and antidiabetic activity. J Ayur Integr Med 3(2):80

Blaney JM, Martin EJ (1997) Computational approaches for combinatorial library design and molecular diversity analysis. Curr Opin Chem Biol 1:54–59

Blum W, Marcucci G (2005) Targeting epigenetic changes in acute myeloid leukemia. Clin Adv Hematol Oncol 3(11):855–865

Brueckner B, Boy RG, Siedlecki P, Musch T, Kliem HC, Zielenkiewicz P et al (2005) Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. Cancer Res 65:6305–6311. https://doi.org/10.1158/0008-5472.CAN-04-2957

Cases J, Ibarra A, Feuillère N, Roller M, Sukkar SG (2011) Pilot trial of *Melissa officinalis* L. leaf extract in the treatment of volunteers suffering from mild-to-moderate anxiety disorders and sleep disturbances. Med J Nutr Metab 4:211–218. https://doi.org/10.1007/s12349-010-0045-4

Chandramohan Reddy T, Bharat Reddy D, Aparna A, Arunasree KM, Gupta G, Achari C et al (2012) Anti-leukemic effects of gallic acid on human leukemia K562 cells: downregulation of COX-2, inhibition of BCR/ABL kinase and NF-κB inactivation. Toxicol In Vitro 26:396–405. https://doi.org/10.1016/j.tiv.2011.12.018

Chaudhuri D, Ghate NB, Singh SS, Mandal N (2015) Methyl gallate isolated from *Spondias pinnata* exhibits anticancer activity against human glioblastoma by induction of apoptosis and sustained extracellular signal-regulated kinase 1/2 activation. Pharmacogn Mag 11:269–276. https://doi.org/10.4103/0973-1296.153078

Chen HS, Bai MH, Zhang T, Li GD, Liu M (2015) Ellagic acid induces cell cycle arrest and apoptosis through TGF-β/Smad3 signaling pathway in human breast cancer MCF-7 cells. Int J Oncol 46:1730–1738. https://doi.org/10.3892/ijo.2015.2870

Cohen RA, Kucera LS, Herrmann EC (1964) Antiviral activity of *Melissa officinalis* (lemon balm) extract. Proc Soc Exp Biol Med 117:431–434. https://doi.org/10.3181/00379727-117-29600

De Queiroz RM, Takiya CM, Guimarães LPTP, Rocha GDG, Alviano DS, Blank AF et al (2014) Apoptosis-inducing effects of *Melissa officinalis* L. essential oil in glioblastoma multiforme cells. Cancer Invest 32:226–235. https://doi.org/10.3109/07357907.2014.905587

Deiab S, Mazzio E, Eyunni S, McTier O, Mateeva N, Elshami F et al (2015) 1,2,3,4,6-Penta-*O*-galloylglucose within Galla Chinensis inhibits human LDH-A and attenuates cell proliferation

in MDA-MB-231 breast cancer cells. Evid Based Complement Alternat Med 2015:1–11. https://doi.org/10.1155/2015/276946

Drews J (2000) Drug discovery: a historical perspective. Science 287:1960–1964. https://doi.org/10.1126/science.287.5460.1960

Ebrahimnezhad Darzi S, Amirghofran Z (2013) Dichloromethane fraction of *Melissa officinalis* induces apoptosis by activation of intrinsic and extrinsic pathways in human leukemia cell lines. Immunopharmacol Immunotoxicol 35:313–320. https://doi.org/10.3109/08923973.2013.768268

Edwards J, Bartlett JMS (2005) The androgen receptor and signal-transduction pathways in hormone-refractory prostate cancer. Part 2: androgen-receptor cofactors and bypass pathways. BJU Int 95:1327–1335. https://doi.org/10.1111/j.1464-410X.2005.05527.x

Encalada MA, Hoyos KM, Rehecho S, Berasategi I, de Ciriano MGÍ, Ansorena D et al (2011) Anti-proliferative effect of *Melissa officinalis* on human colon cancer cell line. Plant Foods Hum Nutr 66:328–334. https://doi.org/10.1007/s11130-011-0256-y

Falzone L, Salomone S, Libra M (2018) Evolution of cancer pharmacological treatments at the turn of the third millennium. Front Pharmacol 9:1300. https://doi.org/10.3389/fphar.2018.01300

Farkona S, Diamandis EP, Blasutig IM (2016) Cancer immunotherapy: the beginning of the end of cancer? BMC Med 14:623. https://doi.org/10.1186/s12916-016-0623-5

FitzGerald GA (2008) Drugs, industry, and academia. Science 320:1563. https://doi.org/10.1126/science.1161006

Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D et al (2016) The ChEMBL database in 2017. Nucleic Acids Res 45:D945–D954. https://doi.org/10.1093/nar/gkw1074

Gomathinayagam R, Sowmyalakshmi S, Mardhatillah F, Kumar R, Akbarsha MA, Damodaran C (2008) Anticancer mechanism of plumbagin, a natural compound, on non-small cell lung cancer cells. Anticancer Res 28:785–792

Hamsa T, Kuttan G (2011) Harmine activates intrinsic and extrinsic pathways of apoptosis in B16F-10 melanoma. Chinas Med 6:11. https://doi.org/10.1186/1749-8546-6-11

Harborne JB (1999) Classes and functions of secondary products from plants. Chem Plants 26:1–25

Harris R (2003) Trease and Evans' pharmacognosy. Int J Aromather 13:55. https://doi.org/10.1016/s0962-4562(03)00051-1

He Z, Chen AY, Rojanasakul Y, Rankin GO, Chen YC (2016) Gallic acid, a phenolic compound, exerts anti-angiogenic effects via the PTEN/AKT/HIF-1α/VEGF signaling pathway in ovarian cancer cells. Oncol Rep 35:291–297. https://doi.org/10.3892/or.2015.4354

Herraiz T, González D, Ancín-Azpilicueta C, Arán VJ, Guillén H (2010) β-Carboline alkaloids in *Peganum harmala* and inhibition of human monoamine oxidase (MAO). Food Chem Toxicol 48:839–845. https://doi.org/10.1016/j.fct.2009.12.019

Herrmann EC, Kucera LS (1967) Antiviral substances in plants of the mint family (Labiatae). II. Nontannin polyphenol of *Melissa officinalis*. Proc Soc Exp Biol Med 124:869–874. https://doi.org/10.3181/00379727-124-31873

Hiradeve S, Danao K, Kharabe V, Mendhe B (2011) Evaluation of anticancer activity of *Plumbago zeylanica Linn* leaf extract. Int J Biomed Res 1:52. https://doi.org/10.7439/ijbr.v1i2.52

Ho HH, Chang CS, Ho WC, Liao SY, Lin WL, Wang CJ (2013) Gallic acid inhibits gastric cancer cells metastasis and invasive growth via increased expression of RhoB, downregulation of AKT/small GTPase signals and inhibition of NF-κB activity. Toxicol Appl Pharmacol 266:76–85. https://doi.org/10.1016/j.taap.2012.10.019

Iminjan M, Amat N, Li XH, Upur H, Ahmat D, He B (2014) Investigation into the toxicity of traditional uyghur medicine *Quercus infectoria* galls water extract. PLoS One 9:e90756. https://doi.org/10.1371/journal.pone.0090756

Jahanban-Esfahlan A, Modaeinama S, Abasi M, Abbasi MM, Jahanban-Esfahlan R (2015) Anti proliferative properties of *Melissa officinalis* in different human cancer cells. Asian Pac J Cancer Prev 16:5703–5707. https://doi.org/10.7314/APJCP.2015.16.14.5703

Jain P, Sharma HP, Basri F, Baraik B, Kumari S, Pathak C (2014) Pharmacological profiles of ethno-medicinal plant: *Plumbago zeylanica* I. A review. Int J Pharm Sci Rev Res 24:157–163

Kennedy LB, Salama AKS (2020) A review of cancer immunotherapy toxicity. CA Cancer J Clin 70:86–104. https://doi.org/10.3322/caac.21596

Kennedy DO, Wake G, Savelev S, Tildesley NTJ, Perry EK, Wesnes KA et al (2003) Modulation of mood and cognitive performance following acute administration of single doses of *Melissa officinalis* (lemon balm) with human CNS nicotinic and muscarinic receptor-binding properties. Neuropsychopharmacology 28:1871–1881. https://doi.org/10.1038/sj.npp.1300230

Khan MTH, Orhan I, Şenol F, Kartal M, Şener B, Dvorská M et al (2009) Cholinesterase inhibitory activities of some flavonoid derivatives and chosen xanthone and their molecular docking studies. Chem Biol Interact 181:383–389. https://doi.org/10.1016/j.cbi.2009.06.024

Kuo PL, Hsu YL, Cho CY (2006) Plumbagin induces G2-M arrest and autophagy by inhibiting the AKT/mammalian target of rapamycin pathway in breast cancer cells. Mol Cancer Ther 5:3209–3221. https://doi.org/10.1158/1535-7163.MCT-06-0478

Lamchouri F, Settaf A, Cherrah Y, El Hamidi M, Tligui NS, Lyoussi B et al (2002) Experimental toxicity of *Peganum harmala* seeds. Ann Pharm Fr 60:123–129

Lee H, Lee H, Kwon Y, Lee J-H, Kim J, Shin M-K et al (2010) Methyl gallate exhibits potent antitumor activities by inhibiting tumor infiltration of CD4 + CD25 + regulatory T cells. J Immunol 185:6698–6705. https://doi.org/10.4049/jimmunol.1001373

Lee KW, Bode AM, Dong Z (2011) Molecular targets of phytochemicals for cancer prevention. Nat Rev Cancer 11:211–218. https://doi.org/10.1038/nrc3017

Li Q, Cheng T, Wang Y, Bryant SH (2010) Pubchem as a public resource for drug discovery. Drug Discov Today 15:1052–1057. https://doi.org/10.1016/j.drudis.2010.10.003

Lipinski CA (2004) Lead-and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 1:337–341. https://doi.org/10.1016/j.ddtec.2004.11.007

Liu J, Lin A (2005) Role of JNK activation in apoptosis: a double-edged sword. Cell Res 15:36–42. https://doi.org/10.1038/sj.cr.7290262

Lu Y, Jiang F, Jiang H, Wu K, Zheng X, Cai Y et al (2010) Gallic acid suppresses cell viability, proliferation, invasion and angiogenesis in human glioma cells. Eur J Pharmacol 641:102–107. https://doi.org/10.1016/j.ejphar.2010.05.043

Lu YC, Lin ML, Su HL, Chen SS (2016) ER-dependent Ca^{2+}-mediated cytosolic ROS as an effector for induction of mitochondrial apoptotic and ATM–JNK signal pathways in gallic acid-treated human oral cancer cells. Anticancer Res 36:697–705

Lusher SJ, McGuire R, van Schaik RC, Nicholson CD, de Vlieg J (2014) Data-driven medicinal chemistry in the era of big data. Drug Discov Today 19:859–868. https://doi.org/10.1016/j.drudis.2013.12.004

Ma Y, Wink M (2010) The beta-carboline alkaloid harmine inhibits BCRP and can reverse resistance to the anticancer drugs mitoxantrone and camptothecin in breast cancer cells. Phyther Res 24:146–149. https://doi.org/10.1002/ptr.2860

Ma D-L, Chan DS-H, Leung C-H (2011) Molecular docking for virtual screening of natural product databases. Chem Sci 2:1656–1665. https://doi.org/10.1039/C1SC00152C

Madlener S, Illmer C, Horvath Z, Saiko P, Losert A, Herbacek I et al (2007) Gallic acid inhibits ribonucleotide reductase and cyclooxygenases in human HL-60 promyelocytic leukemia cells. Cancer Lett 245:156–162. https://doi.org/10.1016/j.canlet.2006.01.001

Mamedov NA, Pasdaran A, Mamadalieva NZM (2017) Pharmacological studies of Syrian rue (*Peganum harmala* L., Zygophyllaceae). Int J Second Metab 5:1–6. https://doi.org/10.21448/ijsm.335539

Mandrika I, Prusis P, Yahorava S, Tars K, Wikberg JE (2007) QSAR of multiple mutated antibodies. J Mol Recognit 20:97–102. https://doi.org/10.1002/jmr.817

Meftahizade H, Moradkhani H, Naseri B, Lofti M, Naseri A (2010) Improved in vitro culture and micropropagation of different *Melissa officinalis* L. genotypes. J Med Plant Res 4:240–246

Mishra S, Vinayak M (2015) Role of ellagic acid in regulation of apoptosis by modulating novel and atypical PKC in lymphoma bearing mice. BMC Complement Altern Med 15:281. https://doi.org/10.1186/s12906-015-0810-5

Moloudizargari M, Mikaili P, Aghajanshakeri S, Asghari M, Shayegh J (2013) Pharmacological and therapeutic effects of *Peganum harmala* and its main alkaloids. Pharmacogn Rev 7:199–212. https://doi.org/10.4103/0973-7847.120524

Moradkhani H, Sargsyan E, Bibak H, Naseri B, Sadat-Hosseini M, Fayazi-Barjin A et al (2010) *Melissa officinalis* L., a valuable medicine plant: a review. J Med Plant Res 4:2753–2759

Mulkens A, Kapetanidis I (1987) Flavonoids of the leaves of *Melissa officinalis* L. (Lamiaceae). Pharm Acta Helv 62:19–22

Munos B (2009) Lessons from 60 years of pharmaceutical innovation. Nat Rev Drug Discov 8:959. https://doi.org/10.1038/nrd2961

National Center for Complementary and Integrative Health (2017) Natural products research—information for researchers. NCCIH, Bethesda

Nature Publishing Group (2007) All natural. Nat Chem Biol 3:351. https://doi.org/10.1038/nchembio0707-351

Oodi A, Norouzi H, Amirizadeh N, Nikougoftar M, Vafaie Z (2017) Harmine, a novel DNA methyltransferase 1 inhibitor in the leukemia cell line. Indian J Hematol Blood Transfus 33:509–515. https://doi.org/10.1007/s12288-016-0770-z

Pandey R, Shukla S, Tiwari R, Chauhan N (2016) *Peganum harmala* Indian traditional plant: a scientific update. Spat. DD—Peer Rev. J Complement Med Drug Discov 6:103. https://doi.org/10.5455/spatula.20170407075541

Parimala R, Sachdanandam P (1993) Effect of plumbagin on some glucose metabolising enzymes studied in rats in experimental hepatoma. Mol Cell Biochem 125:59–63. https://doi.org/10.1007/BF00926835

Pineda SS, Chaumeil P-A, Kunert A, Kaas Q, Thang MW, Le L et al (2017) Arachnoserver 3.0: an online resource for automated discovery, analysis and annotation of spider toxins. Bioinformatics 34:1074–1076. https://doi.org/10.1093/bioinformatics/btx661

Pithayanukul P, Leanpolchareanchai J, Saparpakorn P (2009) Molecular docking studies and anti-snake venom metalloproteinase activity of Thai mango seed kernel extract. Molecules 14:3198–3213. https://doi.org/10.3390/molecules14093198

Polanski J (2009) Receptor dependent multidimensional QSAR for modeling drug–receptor interactions. Curr Med Chem 16:3243–3257. https://doi.org/10.2174/092986709788803286

Powolny AA, Singh SV (2008) Plumbagin-induced apoptosis in human prostate cancer cells is associated with modulation of cellular redox status and generation of reactive oxygen species. Pharm Res 25:2171–2180. https://doi.org/10.1007/s11095-008-9533-3

Prasad KVS, Ao Z, Yoon Y, Wu MX, Rizk M, Jacquot S et al (1997) CD27, a member of the tumor necrosis factor receptor family, induces apoptosis and binds to Siva, a proapoptotic protein. Proc Natl Acad Sci U S A 94:6346–6351. https://doi.org/10.1073/pnas.94.12.6346

Quinn DI, Henshall SM, Sutherland RL (2005) Molecular markers of prostate cancer outcome. Eur J Cancer 41:858–887. https://doi.org/10.1016/j.ejca.2004.12.035

Romano J, Nwankwo V, Tatonetti N (2018) VenomKB v20: a knowledge repository for computational toxicology. bioRxiv 2018:295204. https://doi.org/10.1101/295204

Roy A, Bharadvaja N (2017) A review on pharmaceutically important medical plant: *Plumbago zeylanica*. J Ayurvedic Herb Med 3:225–228. www.ayurvedjournal.com

Salimi A, Roudkenar MH, Sadeghi L, Mohseni A, Seydi E, Pirahmadi N et al (2015) Ellagic acid, a polyphenolic compound, selectively induces ROS-mediated apoptosis in cancerous B-lymphocytes of CLL patients by directly targeting mitochondria. Redox Biol 6:461–471. https://doi.org/10.1016/j.redox.2015.08.021

Saraydin SU, Tuncer E, Tepe B, Karadayi S, Özer H, Şen M et al (2012) Antitumoral effects of *Melissa officinalis* on breast cancer in vitro and in vivo. Asian Pac J Cancer Prev 13:2765–2770. https://doi.org/10.7314/APJCP.2012.13.6.2765

Schirrmacher V (2019) From chemotherapy to biological therapy: a review of novel concepts to reduce the side effects of systemic cancer treatment (review). Int J Oncol 54:407–419. https://doi.org/10.3892/ijo.2018.4661

Shabani SH, Tehrani SS, Rabiei Z, Enferadi ST, Vannozzi GP (2015) *Peganum harmala* L.'s anti-growth effect on a breast cancer cell line. Biotechnol Rep 8:138–143. https://doi.org/10.1016/j.btre.2015.08.007

Shakeri A, Sahebkar A, Javadi B (2016) *Melissa officinalis* L.—a review of its traditional uses, phytochemistry and pharmacology. J Ethnopharmacol 188:204–228. https://doi.org/10.1016/j.jep.2016.05.010

Shirani-Boroujeni M, Heidari-Soureshjani S, Hafshejani ZK (2017) Impact of oral capsule of *Peganum harmala* on alleviating urinary symptoms in men with benign prostatic hyperplasia; a randomized clinical trial. J Renal Inj Prev 6:127–131. https://doi.org/10.15171/jrip.2017.25

Singh Verma PP, Singh A, Rahaman L, Bahl JR (2015) Review article lemon balm (*Melissa officinalis* L.) an herbal medicinal plant with broad therapeutic uses and cultivation practices: a review. Int J Recent Adv Multidiscip Res 2(11):928–933

Singhal A, Bangar O, Naithani V (2012) Medicinal plants with a potential to treat Alzheimer and associated symptoms. Int J Nutr Pharmacol Neurol Dis 2:84. https://doi.org/10.4103/2231-0738.95927

Sinha S, Pal K, Elkhanany A, Dutta S, Cao Y, Mondal G et al (2013) Plumbagin inhibits tumorigenesis and angiogenesis of ovarian cancer cells in vivo. Int J Cancer 132:1201–1212. https://doi.org/10.1002/ijc.27724

Sithisarn P, Gritsanapan W, Jateleela S (2015) Formulation and quality control of readily dissolving drink powder from *Antidesma ghaesembilla* fruits. Chiang Mai Univ J Nat Sci 14:89–102. https://doi.org/10.12982/cmujns.2015.0073

Sliwoski G, Kothiwale S, Meiler J, Lowe EW (2014) Computational methods in drug discovery. Pharmacol Rev 66:334–395. https://doi.org/10.1124/pr.112.007336

Srinivas G, Annab LA, Gopinath G, Banerji A, Srinivas P (2004) Antisense blocking of BRCA1 enhances sensitivity to Plumbagin but not tamoxifen in BG-1 ovarian cancer cells. Mol Carcinog 39:15–25. https://doi.org/10.1002/mc.10164

Sun J, McKallip RJ (2011) Plumbagin treatment leads to apoptosis in human K562 leukemia cells through increased ROS and elevated TRAIL receptor expression. Leuk Res 35:1402–1408. https://doi.org/10.1016/j.leukres.2011.06.018

Sundari BKR, Telapolu S, Dwarakanath BS, Thyagarajan SP (2017) Cytotoxic and antioxidant effects in various tissue extracts of *Plumbago zeylanica*: implications for anticancer potential. Pharm J 9:706–712. https://doi.org/10.5530/pj.2017.5.111

Tatonetti NP, Patrick PY, Daneshjou R, Altman RB (2012) Data-driven prediction of drug effects and interactions. Sci Transl Med 4:125ra31. https://doi.org/10.1126/scitranslmed.3003377

Terrett NK, Gardner M, Gordon DW, Kobylecki RJ, Steele J (1995) Combinatorial synthesis—the design of compound libraries and their application to drug discovery. Tetrahedron 51:8135–8173

Thasni KA, Rakesh S, Rojini G, Ratheeshkumar T, Srinivas G, Priya S (2008) Estrogen-dependent cell signaling and apoptosis in BRCA1-blocked BG1 ovarian cancer cells in response to plumbagin and other chemotherapeutic agents. Ann Oncol 19:696–705. https://doi.org/10.1093/annonc/mdm557

The Indian Pharmacopoeia Commission (2007) The unani pharmacopoeia of India. The Indian Pharmacopoeia Commission, Ghaziabad

The Wealth of India (1952) A dictionary of Indian raw materials and industrial products. Raw Mater 3:53–54

Thomford NE, Senthebane DA, Rowe A, Munro D, Seele P, Maroyi A et al (2018) Natural products for drug discovery in the 21st century: innovations for novel drug discovery. Int J Mol Sci 19:1578. https://doi.org/10.3390/ijms19061578

Uhl KL, Schultz CR, Geerts D, Bachmann AS (2018) Harmine, a dual-specificity tyrosine phosphorylation-regulated kinase (DYRK) inhibitor induces caspase-mediated apoptosis in neuroblastoma. Cancer Cell Int 18:574. https://doi.org/10.1186/s12935-018-0574-3

Vaidya ADB, Devasagayam TPA (2007) Current status of herbal drugs in India: an overview. J Clin Biochem Nutr 41:1–11. https://doi.org/10.3164/jcbn.2007001

Van Nostrand JL, Brisac A, Mello SS, Jacobs SBR, Luong R, Attardi LD (2016) The p53 target gene SIVA enables non-small cell lung cancer development. Cancer Discov 5:622–635. https://doi.org/10.1158/2159-8290.CD-14-0921

Vanella L, Di Giacomo C, Acquaviva R, Barbagallo I, Cardile V, Kim DH et al (2013) Apoptotic markers in a prostate cancer cell line: effect of ellagic acid. Oncol Rep 30:2804–2810. https://doi.org/10.3892/or.2013.2757

Vermani A, Navneet, Prabhat (2009) Screening of *Quercus infectoria* gall extracts as anti-bacterial agents against dental pathogens. Indian J Dent Res 20(3):337. https://doi.org/10.4103/0970-9290.57380

Vogl S, Picker P, Mihaly-Bison J, Fakhrudin N, Atanasov AG, Heiss EH et al (2013) Ethnopharmacological in vitro studies on Austria's folk medicine—an unexplored lore in vitro anti-inflammatory activities of 71 Austrian traditional herbal drugs. J Ethnopharmacol 149:750–771. https://doi.org/10.1016/j.jep.2013.06.007

Wang Z, Lin S, Wang D, Huang L (2014) Anti-epidermal growth factor receptor tyrosine kinase activities of traditional Chinese medicine for cancer treatment. Eur J Integr Med 6:565–570. https://doi.org/10.1016/j.eujim.2014.05.006

Wang Y, Wang C, Jiang C, Zeng H, He X (2015) Novel mechanism of harmaline on inducing G2/M cell cycle arrest and apoptosis by up-regulating Fas/FasL in SGC-7901 cells. Sci Rep 5:18613. https://doi.org/10.1038/srep18613

Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR et al (2017) Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Res 46:D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Yin SY, Wei WC, Jian FY, Yang NS (2013) Therapeutic applications of herbal medicines for cancer patients. Evid Based Complement Alternat Med 2013:1–15. https://doi.org/10.1155/2013/302426

You BR, Park WH (2010) Gallic acid-induced lung cancer cell death is related to glutathione depletion as well as reactive oxygen species increase. Toxicol In Vitro 24:1356–1362. https://doi.org/10.1016/j.tiv.2010.04.009

Yu XJ, Sun K, Tang XH, Zhou CJ, Sun H, Yan Z et al (2016) Harmine combined with paclitaxel inhibits tumor proliferation and induces apoptosis through down-regulation of cyclooxygenase-2 expression in gastric cancer. Oncol Lett 12:983–988. https://doi.org/10.3892/ol.2016.4696

Zhang T, Chen H-S, Wang L-F, Bai M-H, Wang Y-C, Jiang X-F et al (2014) Ellagic acid exerts anti-proliferation effects via modulation of Tgf-B/Smad3 signaling in MCF-7 breast cancer cells. Asian Pac J Cancer Prev 15:273–276. https://doi.org/10.7314/APJCP.2014.15.1.273

Žiaková A, Brandšteterová E, Blahová E (2003) Matrix solid-phase dispersion for the liquid chromatographic determination of phenolic acids in *Melissa officinalis*. J Chromatogr A 983:271–275. https://doi.org/10.1016/S0021-9673(02)01712-0

Artificial Intelligence and Machine Learning Techniques Using Omics Data for Cancer Diagnosis and Treatment

2

Priyanka Gawade, Sutanu Nandi, Chandrakala Meena, and Ram Rup Sarkar

Abstract

Cancer is a heterogeneous disease concerning molecular, functional and clinical behaviour, and poses a challenge for timely detection and treatment. Early detection and prognosis of cancer type may facilitate refined clinical management of cancer treatment. Recent technological development, such as next-generation sequencing, generated a large number of omics datasets in cancer genomics. The genome-wide biological information, such as cancer driver mutations, aberrantly methylated regions, gene, and miRNA expression profiles, is helpful for predicting the cancer onset, subtypes, and treatment response and is valuable for improving diagnosis and therapeutic and clinical decisions. In this context, machine learning (ML) algorithms and artificial intelligence have been beneficial and essential for the better accuracy of cancer-related predictions. Here, we mainly focus on research based on these omics data, paying close attention to machine learning methods. We summarize various kinds of omics data and different ML algorithms effective in cancer prediction. We also highlighted the

P. Gawade
Chemical Engineering and Process Development, CSIR-National Chemical Laboratory, Pune, Maharashtra, India

Bioinformatics Centre, Savitribai Phule Pune University, Pune, Maharashtra, India

S. Nandi · R. R. Sarkar (✉)
Chemical Engineering and Process Development, CSIR-National Chemical Laboratory, Pune, Maharashtra, India

Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh, India
e-mail: rr.sarkar@ncl.res.in

C. Meena
Chemical Engineering and Process Development, CSIR-National Chemical Laboratory, Pune, Maharashtra, India

S. Singh (ed.), *Systems Biomedicine Approaches in Cancer Research*,
https://doi.org/10.1007/978-981-19-1953-4_2

applications of the ML algorithm on genomic information in cancer, including cancer classification, therapy response, survival, metastasis, and biomarker identification. Further we discussed the novel approaches in machine learning for improving cancer prediction. These data-driven approaches can potentially provide a new solution for enhancing the precise treatment of cancer.

2.1 Introduction

Cancer shows significant disease burden globally due to its high prevalence and death rate. It occurs due to the development of atypical cells that divide in an uncontrolled manner. A central feature of cancer malignancy is metastasis. In the metastasis stage, cancerous cells leave their pre-neoplastic lesions, enter the bloodstream, disseminate throughout the body, and acclimate to new cellular surroundings in a secondary site, ultimately destroying the normal body tissue (Kang and Pantel 2013; Welch and Hurst 2019). These abilities of cancer cells viz., dissemination and invasion, eventually prove fatal to the host.

Cancer is a multistep and progressive disease in which gene expression alters because of the accretion of numerous genetic and epigenetic aberrations within a genome. The genomic complexity of cancer cells arises due to intrinsic factors and/or extrinsic factors that cause gross-scale abnormalities, i.e., variation in chromosome numbers (including aneuploidy and whole-genome duplication) (Hasty and Montagna 2014). Also, small-scale/local changes, i.e., genome rearrangements (consist of gene amplification, deletions, and non-reciprocal translocations), occurs due to causative agents and are responsible for genomic complexity. In addition to this, aberrant alterations in genes encoding epigenetic players that control epigenetic mechanisms are also responsible for increasing the complexity of cancer by causing the inappropriate onset (initiation/inhibition) of genetic expressions and promoting tumorigenesis. The epigenetic changes modify DNA (via methylation), histones (by post-translational modifications PTMs, namely methylation, acetylation, and phosphorylation, etc.), and non-coding RNAs (small and long ncRNAs) regulations and nucleosome remodeling, to form a regulatory system that controls accessibility between DNA elements and histones/non-coding RNAs (Ilango et al. 2020; Lu et al. 2020). The epigenetic players that participate in these modifications are susceptible to extrinsic factors, and changes caused by these players are reversible. These genetic and epigenetic alterations are often found in two kinds of genes, namely, proto-oncogenes and tumor suppressor genes. The activation changes like gain-of-function mutations and hypomethylation converts the proto-oncogenes into oncogenes (OGs), which are overactive positive cell cycle regulators responsible for cell survival, growth, and division, ultimately leading to cancer progression. The changes like loss-of-function mutations, epigenetic silencing like hypermethylation, proteasomal degradation by ubiquitination, and abnormal cellular localization of tumor suppressor genes (TSGs) leads to their inactivation. As a consequence of this, tumor development occurs due to elimination of negative regulatory proteins that

usually restrict cell growth (by apoptosis or activating DNA repair and cell cycle checkpoint) (Wang et al. 2018; Kontomanolis et al. 2020). Numerous studies have been done for identifying the genetic and epigenetic changes in different cancer types. For example, glioblastoma is associated with genetic alterations in the number of tumor suppressors, viz., PTEN, TP53, PIK3R1, NF1, RB1, and oncogenes, i.e., EGFR, PIK3CA, and IDH1 (Zhang et al. 2019a). The other tumor suppressor genes such as BRCA1/2, P53, PTEN, ATM, Rb, LKB, Nm23, P16, and oncogenes like HER2, c-MYC, and ERBB2, MYC, PIK3CA are very frequently mutated in breast cancer (Oliveira et al. 2005; Perera and Bardeesy 2012).

The complexity of cancer is further enhanced due to *tumor heterogeneity* that can occur during cancer evolution. Tumor heterogeneity is of two types as follows: (a) *Intra-tumor heterogeneity*, in which subsets of cancer cells within a tumor of a single patient possess discrete phenotypic and molecular characteristics and (b) *Inter-tumor heterogeneity*, which comprises tumor genotype variations among tumors of the same histological type between different patients (Meacham and Morrison 2013). This heterogeneity can arise from genetic, epigenetic, transcriptomic, or phenotypic changes (McQuerry et al. 2017). Genomic-level studies of *tumor heterogeneity* showed that cells in a tumor are highly diverse, spatio-temporally by analyzing their genetic variations like single-nucleotide variants (SNV), insertion–deletion mutations (indels), and copy number variation (CNV) (Murtaza et al. 2015; Li et al. 2017). Several studies provided information on *epigenetic heterogeneity* by inspecting DNA methylome and micro-RNA (miRNA) pools (Liu et al. 2018; Dietz et al. 2019; Wang et al. 2019; Guo et al. 2019; Alfardus et al. 2021). Studies of *tumor heterogeneity* at the transcriptome level revealed variation in the gene expression pattern of particular pathways like cell cycle, MAPK signaling pathway, immune/complement system pathways, and biological programs, namely hypoxia and epithelial–mesenchymal transition (EMT) (Patel et al. 2014; Zhang et al. 2016). Some studies supported the proteomic heterogeneity of tumors, but it is less prominent than genomic and transcriptomic heterogeneity (Ahmed et al. 2016; Sood et al. 2016). *Tumor heterogeneity* also includes heterogeneity of the tumor microenvironment (consists of endothelial cells, fibroblasts, adipocytes, immune cells, mesenchymal stroma/stem-like cells, and extracellular matrix), that sends physical and chemical signals to tumor cells and influences epigenetic machinery (Hass et al. 2020). Such a dynamic and highly variable nature of cancer hinders diagnosis and prognosis and leads to treatment resistance, relapse, and eventually death (Dagogo-Jack and Shaw 2018; Marusyk et al. 2020). Hence, understanding the mechanism of cancer development at different biological levels and early prediction of cancer may help in designing better therapeutic strategies.

2.2 Omics Data in Cancer Research

Advancements of high-throughput sequencing (or next-generation sequencing, NGS) techniques and the availability of omics data provide genome-wide measurements of genomic features (including genetic variants, DNA methylation,

and transcripts etc.) at various levels and resulted in remarkable progress in cancer research. Several databases are available online which provide free access to genomic information related to cancer. Among these, a few popular and important resources are listed in Table 2.1.

In order to properly analyze various kinds of omics data and to perform exploratory analysis, several computational tools are freely available online (see Table 2.1). Integration of different omics data will decode interrelationships between these features and their functions. This holistic approach seems to be promising to understand cancer development, recurrence, therapy response, and patient survival. The subsequent sections will discuss different types of omics data produced by various high-throughput sequencing approaches.

2.2.1 Genomic Data

Genomic information helps to unravel functional information present in DNA sequences.

2.2.1.1 Genomic Variation Data

Genetic variation is an alteration in the nucleotide order of DNA sequences that occur either due to mutation or genetic recombination. It can be grouped into following classes on the basis of size: (1) Small-scale sequence variation (<1 kb) consists of single-nucleotide variants (SNV), single nucleotide insertions/deletions (indels), etc., (2) Large-scale structural variation includes copy number variations (CNV) (loss or gain) and chromosomal rearrangement (genomic inversions, translocations) (Cardoso et al. 2015). SNVs are the most prevalent variants and can be present in different genomic locations: (1) protein-coding sequences, (2) non-coding regions like splice sites, promoters, ribosome binding sites, etc. Indels cause frameshift mutations within a coding region, whereas chromosomal rearrangements affect the spatial organization of chromosomes and cause nuclear reorganization. This kind of genomic variation is a fundamental constituent of genomics data and provides an opportunity to explore associations between genes, tissues, individuals, and phenotypes. DNA-sequencing (DNA-seq) techniques have been used to study genomic alterations, which include whole-genome sequencing (WGS), whole-exome sequencing (WES), and targeted massively parallel sequencing (TS) (Lightbody et al. 2019). WGS technique analyzes entire genomes and allows investigation of changes within coding and regulatory sites (Meienberg et al. 2016). It offers identification of CNVs, chromosomal rearrangements, and other structural variations that may be missed by targeted sequencing. This technique provides global insight into novel genomic changes in cancer samples as it gives the base-pair resolution of complete cancer genome in a single run (Zhao et al. 2019). Yates et al. (2017) performed WGS of primary as well as metastatic tumor samples of breast cancer and observed that cell clones causing metastasis or relapse migrate late from the primary tumors; however, they constantly gain alterations, mostly in the same biological process as the primary tumor. Another WGS study of

Table 2.1 Resources (data repositories and analysis tools) for cancer genomics study

Database/tools	Features	Link	References
The Cancer Genome Atlas (TCGA)	Exhaustive data repository of genomic, epigenomic data of cancer and control samples	https://portal.gdc.cancer.gov/	Tomczak et al. (2015)
Gene Expression Omnibus (GEO)	Public repository of genomic and proteomic data from array- and sequencing-based techniques	https://www.ncbi.nlm.nih.gov/geo/	Barrett et al. (2012)
International Cancer Genome Consortium (ICGC)	Data portal consists of somatic mutations and molecular data of major tumor types for competent visualization and analysis	https://dcc.icgc.org/	Zhang et al. (2019b)
Database of DNA methylation and gene expression in human cancer (MethHC)	DNA methylomes and mRNA/microRNA expression database; provides clinical and genomic variation data; multiplicity of information present	https://awi.cuhk.edu.cn/~MethHC/methhc_2020/php/index.php	Huang et al. (2021)
The database of human DNA methylation and cancer (MethyCancer)	Database comprises of DNA methylation data, cancer-related gene and mutations; also provides an efficient visualization tool, MethyView	http://methycancer.psych.ac.cn/	He et al. (2007)
Chinese Glioma Genome Atlas (CGGA)	Database contains mRNA/miRNA expression profiles and DNA methylation data of brain tumors from Chinese cohorts	http://www.cgga.org.cn/	Zhao et al. (2021)
UCSC Xena	Graphical viewer for gene- and genomic-coordinate across multiple data types of tumors	http://xena.ucsc.edu/	Goldman et al. (2020)
cBioPortal	Data portals provide genetic alterations across samples, genes, and pathways by analyzing multi-omics cancer data	https://www.cbioportal.org/	Gao et al. (2013)
SomamiR	Comprehensive resource for somatic and germline alterations in miRNA and their target sites in cancer	https://compbio.uthsc.edu/SomamiR/	Bhattacharya et al. (2013)
Database of Epigenetic Modifiers (dbEM)	Data resource for genomic information of epigenetic modifiers in cancer and healthy samples	https://webs.iiitd.edu.in/raghava/dbem/	Nanda et al. (2016)

glioblastoma (GBM) tumors identified novel non-coding constraint mutations for GBM-associated genes (Sakthikumar et al. 2020). In contrast to WGS, targeted sequencing approaches examine specific genomic regions of interest for the detection of rare variants and include WES and TS. WES covers coding genomic portions (i.e., genes and their flanking regions) to find out disease-causing variants in these portions (Gupta et al. 2017; Mueller et al. 2018). Mainly, WES is useful for identifying indels and SNV/SNPs inside the genome's coding sites. TS technique is helpful when prior information of disease is available and performed on particular locations of the genome (Davis et al. 2021). Recently, Weigelt et al. (2018) performed WES of breast tumors and TS of 410 breast cancer genes to investigate the somatic changes and the phenotypic characteristics associated with breast cancer which is originated from ataxia–telangiectasia (ATM) germline mutation. Garrett et al. (2020) carried out WES study of GBM tumor samples to analyze their genetic profile and correlated this information with drug treatment response to develop personalized treatments against GBM. Targeted sequencing was used to identify somatic mutations and CNV alterations in 30 genes which are most frequently altered in gliomas in order to detect biomarkers associated with the long-term survival of GBM patients (Cantero et al. 2018).

2.2.2 Epigenomic Data

Epigenomic information is useful to map the dynamic state of the genome in order to elucidate phenotypic characteristics observed via gene expression studies.

2.2.2.1 DNA Methylation Data

DNA methylation process is an epigenetic mechanism which incorporates a methyl (CH_3) group into the cytosine residue of DNA via the action of DNA methyltransferase enzymes. It controls gene expression and chromatin remodeling by influencing the interactions of DNA with histone or specific transcription factors. Whole-genome bisulfite-sequencing (WGBS) is a high-throughput technique used to quantify genome-wide DNA methylation. It provides a higher resolution to allele-specific DNA methylation as compared to DNA methylation assays and DNA microarrays. This technique allows identification of differentially methylated positions (DMPs) and differentially methylated regions (DMRs) which are the genomic positions/regions having distinct of DNA methylation levels in various biological circumstances (Wu et al. 2015). These DMPs and DMRs in disease conditions are useful for the development of potential epigenetic biomarkers which may help in early detection and diagnosis. The methylation changes in circulating DNA of metastatic breast cancer were studied using WGBS and found 21 DNA hypermethylation hotspots that could be potential blood-based biomarkers (Legendre et al. 2015). Bam et al. (2021) analyzed the global methylation status of both tumor-infiltrating and blood CD4+ T-cell from glioblastoma patients. The study found that the epigenetic modifications in tumor-infiltrating helper T-cells are affected by tumor cells.

2.2.2.2 Histone Modification Data

Chromosomal DNA tightly wraps around histone proteins and forms a chromatin structure in the nucleus. The post-translational modifications (PTMs) of histone proteins are crucial in chromatin remodelling which influence transcription. There are two mechanisms by which histone modifications exert their effect: (1) by directly altering overall chromatin structure either over short or long distances and (2) regulating (either positively or negatively) the binding of histone modifiers (Bannister and Kouzarides 2011). The detection of various histone modifications enables a greater understanding of epigenetic regulation and leads to the development of therapeutic strategies against histone-modifying enzymes. Chromatin immunoprecipitation-sequencing (i.e., ChIP-seq), an effective method for detecting DNA, targets for histone modifications as well as for transcription factors (TFs) at genomic scale with base-pair resolution (O'Geen et al. 2011). It identifies differences in the histone modification patterns which help in understanding epigenetic mechanisms that regulate various biological processes in diseases and thus a powerful tool to analyze chromatin structure and gene expression. ChIP-seq data also reveals how the genome is organized and the functional domains across the entire genome which aid in predicting and validating a set of large, non-coding RNAs. Xi et al. (2018) used the ChIP-seq technique to profile the distributions of 8 key histone modifications (i.e., H3K4me1, H3K4me3, H3K9ac, H3K9me3, H3K27ac, H3K27me3, H3K36me3 and H3K79me2) across 13 breast cancer cell lines and from the epigenetic landscape of 5 molecular subtypes of breast cancer defined subtypes-specific key chromatin signatures to determine potential biomarkers. ChIP-seq analysis of histone H3 Lys27 acetylation (H3K27ac) revealed that alteration in the metabolite acetyl-CoA stimulates site-specific regulation of H3K27ac through which acetyl-CoA impacts the expression of distinct sets of genes associated with malignant phenotypes of glioblastoma, i.e., cell adhesion and migration (Lee et al. 2018).

2.2.3 Transcriptomics Data

Gene expression data are useful to obtain information on the abundance of complete sets of RNA transcripts that are produced by the genome within a biological sample simultaneously.

2.2.3.1 Transcript Profiling Data

The RNA molecules are used to form proteins that serve a crucial part of the cell. Thus, RNA expression reveals active transcription of cell and core activities in cells and tissue under specific conditions. Different types of RNA molecules present in eukaryotic cells play different biological functions like: mRNAs—carry the genetic blueprint from a cell's DNA to its ribosomes to make protein; microRNAs (miRNAs)—involve in gene silencing by repressing translation; long ncRNAs (lncRNAs)— involve in regulating chromatin function, modulating mRNA translation and also interfere with signalling pathways by acting as decoys, scaffolds or

enhancer RNAs. RNA sequencing (RNA-seq) technique is useful to study expression level of transcripts under particular conditions, namely, different environmental conditions, disease scenarios, and therapeutics exposure etc. Analysis of transcriptome data reveals which genes are activated or silent in cells/tissue (qualitative information) and to what extent genes are expressed (quantitative information) (Wang et al. 2009). RNA-seq methods provide information on differentially expressed genes to detect both known and novel transcripts. The profiling of mRNA molecules can be done using several RNA-seq assays viz., mRNA-seq, single-cell RNA-seq (scRNA-seq), strand-specific RNA-seq, ultra-low input RNA-seq and isoform sequencing (Iso-seq). However, the small RNA-seq technique is useful for expression profiling of small non-coding RNAs (like miRNA, siRNA, and piRNA). Total RNA-seq technique provides genome-wide expression data of both coding and non-coding RNAs. For example, using total RNA-seq technology, miRNA associated with metastatic breast cancer response to systemic treatment was identified based on miRNA count (Martinez-Gutierrez et al. 2019). Gao et al. (2021) showed that circular RNA (circRNA)-encoded unique E-cadherin variant *circ-E-Cad* (C-E-Cad) activates oncogenic EGFR signalling by directly binding to it and contributes to glioma stem cell tumorigenicity. Recently, Ren et al. (2021) discussed usage of scRNA-seq technology in breast cancer heterogeneity, metastasis, drug resistance, and prognosis and highlighted the importance of scRNA-seq for development of better treatment strategies.

High-throughput technologies generate massive amount of omics data which present a challenge due to its high dimension and redundancy. There is still a gap in understanding of these data that are often publicly and freely available. The traditional simplex classification algorithms are not suitable to handle large data sets as they contain a small sample size and large gene count. In this scenario, machine learning-based methods provide an excellent tool for analyzing such large and complex data, thus promoting clinical diagnosis and precision medicine against cancer.

2.3 Machine Learning Approaches

Nowadays, machine learning (ML), a subset of artificial intelligence (AI), is extensively applied in growing areas of healthcare, like medical imaging and gene expression pattern analysis, etc. and is extremely useful for high-dimensional data analysis and prediction. It is a data driven approach, which handles large datasets and automatically learns inherent patterns in the data that are useful to make decisions for new sets of data (Witten and Frank 2000). These characteristics make ML a suitable approach to design effective strategies for cancer diagnosis and treatment. Recent developments in ML models have indicated pronounced potential in preclinical conditions. The following sections give details of machine learning algorithms and their applications in cancer research.

The terminologies used in machine learning are mentioned below:

Dataset: It is a matrix containing features from which the machine learns and class label/target to predict. Each column in the matrix represents a feature or target, whereas each row represents an instance/observation. An initial dataset from which the model learns any relationships between features and targets during model training is called as training dataset. However, testing dataset is a subset of data which is not provided during model training but is useful for unbiased model evaluation by comparing predictions with the true value of the dataset.

Instance: An observation or data point is denoted as instance.

Feature/Attribute/Variable: This describes instances by measurable values and acts as input for prediction.

Target/Class Label: A value of an observation that a machine learns to predict is called as target or class label. For example, molecular subtype identification of breast cancer is a multi-classification task. Here, four class labels, i.e., luminal A, luminal B, HER2, and triple negative, are available.

Cross-Validation (*CV*): It is a technique that uses a subset of the original dataset for model training and utilizes other subset for model evaluation. This is generally useful to reduce model overfitting during training time. This method generates a fixed number of subset (fold) of data and performs the analysis for each subset. Further, it averages the final error estimate. Types of cross-validation methods are mentioned below.

a) k-Fold Cross-Validation: The *k*-fold cross-validation method performs random splitting on the original dataset to generate k equal size subsets and uses $(k - 1)$ subsets for training. For the testing purpose, it uses one subset.

b) Leave One Out Cross-Validation: The "leave one out cross-validation" method selects one instance from the original dataset for testing and the remaining instances for model training. The iteration is performed for each instance, and the final outcome is the average of results obtained from each iteration.

c) Bootstrap Cross-Validation: In this method, the complete original dataset is used for model training with sample replacement technique, and the remaining instances are used for model testing.

Machine learning algorithms are mainly grouped into following three categories on the basis of the availability of class labels/targets (Kotsiantis et al. 2007).

1. *Supervised Learning Algorithms*: It uses known targets during training and a model learn the relationship between features and targets. This information can be used for predicting unknown instances.
2. *Unsupervised Learning Algorithms*: Targets for unsupervised machine learning algorithms are unknown. It is used to find hidden structures/patterns or groups of similar samples during training the model. For clustering and pattern detection in biological research, these algorithms are mostly applied and also, useful for identification of gene signature in cancer and survival prediction.
3. *Semi-Supervised Learning Algorithms*: In this case, limited class labels are available, and thus, both labeled and unlabeled data are used during model building to improve accuracy. These algorithms are self-learning and show great potential in cancer prediction problems.

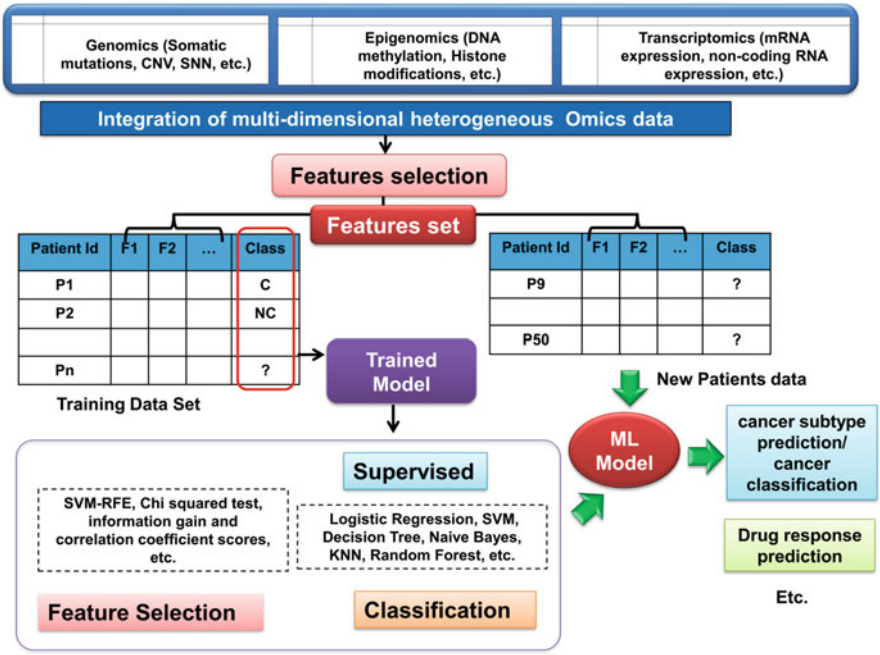


Fig. 2.1 Schematic overview of a machine learning workflow for cancer prediction using multi-omics data. Cancer patient's omics data, i.e., genomics, epigenomics, and transcriptomics, can act as input for machine learning models. Once a model is trained, it can be used to make predictions like cancer classification and drug response etc. for new patient's data

Depending on the nature of target values, above-discussed ML algorithms are further divided into either classification or regression type. The classification algorithms are used to predict categories of new instances by training the input dataset. However, regression algorithms learn from input datasets and then predict the outcome for continuous values. The schematic representation of machine learning workflow in the case of cancer prediction is given in Fig. 2.1.

The commonly used methods during model construction for improving model performance are discussed below.

2.3.1 Feature Selection Methods

Post-genomics era generated a large amount of transcriptomic, mutational, copy number variation (CNV), DNA methylation, histone modification, and miRNA expression data from various high-throughput techniques when applied on cancer cell lines or patients. These different types of data act as features in machine learning models and hold predictive power. To improve prediction accuracy, the feature selection method selects relevant features and removes irrelevant features present

in the original dataset without changing its original value. This method is very important when a dataset contains a large number of features. In such cases, there is no need to give every feature to the algorithm but only important ones for model prediction. This step will make the algorithm to perform fast, and will decrease the model complexity, increase model accuracy, reduce overfitting and simplify interpretation.

The feature selection is mainly grouped into three classes, namely, filter, wrapper, and embedded (Hira and Gillies 2015).

Filter Method: This feature selection algorithm employs some ranking over features that decide the importance of each feature for prediction. In this way, it selects best N features without depending on any ML algorithms. This method is used as pre-processing step. Few examples of this method are Pearson's correlation, *t* test, variance thresholds, information gain (IG), and Bayesian networks.

Wrapper Method: This method selects the best N features using machine learning classifiers. It uses forward selection or backward elimination or bi-directional elimination techniques to decide which features to retain or remove.

Embedded Method: It combines the filter and wrapper techniques to check feature importance. This method is useful for avoiding the overfitting. The gradient boosting machine (GBM), ridge regression, recursive feature elimination (RFE), and LASSO are few examples of embedded feature selection algorithms.

2.3.2 Dimension Reduction Methods

In model construction, the feature selection method selects a subset of relevant features, resulting in a reduction in the dataset's dimension. It retains a subset of original features. However, in the case of high-dimensional data (i.e., 100 or 1000 features), the dimension reduction approach is used to reduce the high number of features into low numbers by transforming the original values. The implementation of this method will reduce computational time and provide quick visualization. Few commonly used dimension reduction techniques are, principal component analysis (PCA) (Pearson 1901), metric dimensional scaling (MDS) (Torgerson 1952), and t-distributed stochastic neighbor embedding (t-SNE) (Hinton and Roweis 2002). The high-dimensional data in the biological area like high-throughput gene expression data, can be analyzed using the above techniques, and some of them are discussed below.

Principal Component Analysis (Pearson 1901): This method uses the orthogonal transformation process to convert instances of correlated features into a group of linearly uncorrelated features. In this way, it reduces the dimension of the dataset with the most negligible information loss, and newly formed features are known as principal components. If data are nonlinear, kernel PCA is beneficial with nonlinear kernel mapping. PCA works well on the dataset which shows the Gaussian distribution.

Metric Dimensional Scaling (Torgerson 1952): This statistical method uses data that contains dissimilarities among pairs of instances. MDS denotes these

dissimilarities as distances among instances and obtain low dimension data points from the high-dimensional dataset by keeping pairwise distances the same.

2.3.3 Overview of Machine Learning Algorithms

2.3.3.1 Supervised Machine Learning Algorithms

Supervised machine learning algorithms have been used for cancer diagnosis and prognosis. Different supervised ML algorithms are available to analyze multi-omics data with categorical and quantitative variables in cancer research and build prediction models. Omics data of individual cancer patient at variety of molecular levels can also be used with these classifiers to develop personalized predictions; they are also useful for personalized predictions models. A detailed description of some of the supervised ML algorithms useful in cancer prediction/prognosis is given below.

Support Vector Machine (SVM)

SVM is a commonly applied supervised machine learning algorithm that searches hyperplane with maximal separation from each data class. Vapnik first described such a kind of classifier to classify data classes using only a hyperplane (Cortes and Vapnik 1995). The general principle of SVM is presented in Fig. 2.2a. SVM uses a multidimensional function known as kernel to transform input data points from the feature space to target space so as to differentiate complex real-life datasets. The classification, as well as regression problems, can be solved using SVM. The proper selection of kernel functions and their parameters significantly helps to improve the model performance.

The following function describes SVM:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i^+ + \xi_i^-\right)$$

$$s.t.\begin{cases} y_i - f(x_i) \le \varepsilon + \xi_i^+ \\ y_i - f(x_i) \le -\varepsilon - \xi_i^- \\ \xi_i^+, \xi_i^- \ge 0 \end{cases},$$

where f, y, and ε represent prediction, actual class label, and free threshold parameter, respectively. The constant C is a coefficient of adjustment between the margin of separation and error on the hyper-plane. The ξ_i^+ and ξ_i^- parameters representing slack variables for error calculation.

Naive Bayes (NB)

Naive Bayes, another supervised ML algorithm (Rish 2001), is a probabilistic method based on Bayes' law. It assumes that a particular feature in a class is independent of another feature in the same class and each feature is equally

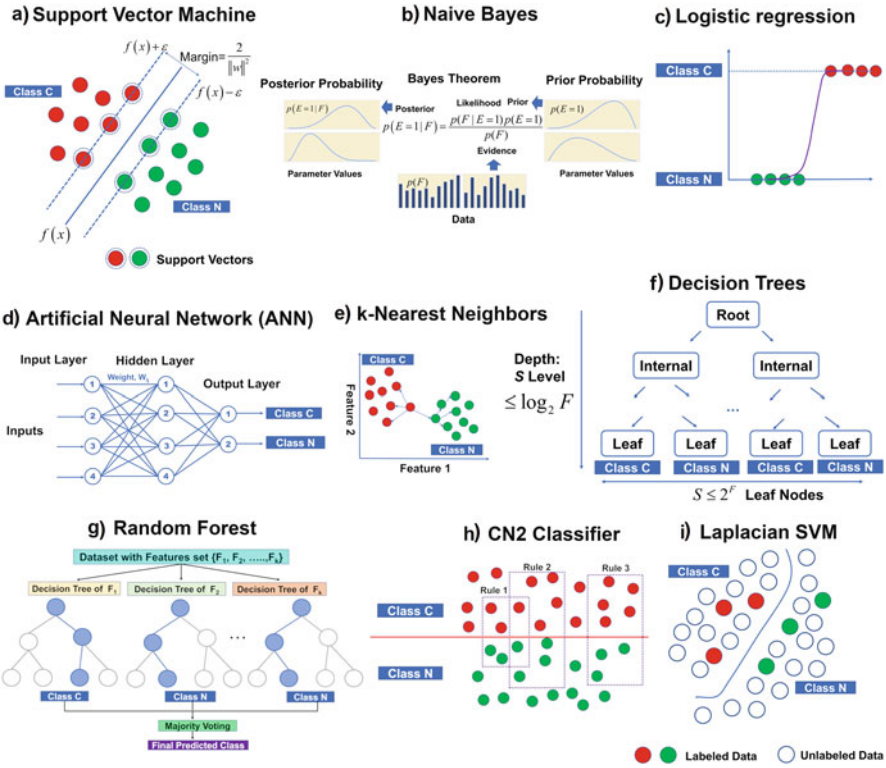


Fig. 2.2 The basic principles of different types of machine learning algorithms. (**a**) Support vector machine (SVM), (**b**) Naive Bayes, (**c**) logistic regression, (**d**) artificial neural network (ANN), (**e**) *k*-nearest neighbors (KNN), (**f**) decision tree, (**g**) random forest, (**h**) CN2, and (**i**) Laplacian SVM

contributing to target class. See Fig. 2.2b. This algorithm is used for classification purposes.

Naïve Bayes classifier can be defined as follows:

$$p(E=1|F) = \frac{p(E=1)\prod_{i=1}^{n} p(f_i|E=1)}{p(F)},$$

where, $F = (f_1, f_2, \ldots . f_n)$ denotes all the features, $p(E=1)$ is obtained from a training set and known as target class prior probability, $p(F)$ is the feature prior probability, $p(f_i|E=1)$ is likelihood that is probability of feature given target, and $p(E=1|F)$ is the posterior probability of target class given feature.

The below function finds class with maximum probability:

$$\text{classify}(f_1, f_2, \ldots . f_n) = \arg\max_{E=1,0} p(E) \prod_{i=1}^{n} p(f_i|E).$$

Logistic Regression Classifier

Logistic regression is a classification algorithm utilized for probability prediction of target class by logistic function (refer Fig. 2.2c). To use this algorithm, the target class must be categorical, and multi-collinearity should not be in features. This classifier helps to detect the best fitting model so as to represent the association between features and the target class.

The logistic function with the feature set $F = \{f_1, f_2, \ldots \ldots \ldots \ldots, f_n\}$ is

$$p(E = 1|F) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 f_1 + \ldots . \beta_n f_n)}}.$$

Artificial Neural Networks (ANNs)

ANN (Hagan et al. 1997), also known as a neural network or simulated neural network (SNN), simulates behaviour of the human nervous system. This computational network comprises numerous interconnected layers (i.e., multi-layer perceptron) that learn (without any programming), generalize training data, and give output from complex data. It mainly contains three layers, namely, (1) input layer, only one input layer in which input data are fed, (2) hidden layers, one or more hidden layers in which processing takes place to derive results based on the weighted sum of connections, and (3) output layer, demonstrating the results. Each layer consists of multiple processing units called nodes, which possess an "activation function" that converts input signal to output signal. Model performance gets affected by the number of nodes and hidden layers. Figure 2.2d shows the computational scheme of ANN.

ANN's objective function is as follows:

$$\arg\min_{w} E(w) = \frac{1}{2} \sum_{i=1}^{m} (N(w, x_i) - y_i)^2,$$

where x, w, and y represents input vector, weight between nodes, and target vector, respectively. ANN algorithms are useful for prediction, classification, regression, and pattern recognition.

k-Nearest Neighbors (KNNs)

The *k*-nearest neighbors are a distance-based algorithm as it first finds all the closest points around new unknown data point and calculates the distance between them to determine the class of new data points (Aha et al. 1991) (shown in Fig. 2.2e). The number of closest points near new unknown data points is denoted as "*k*" symbol, and fine-tuning of this value improves the model performance. This method helps to

solve the classification task by considering the majority of votes, while for the regression problem, KNN takes the mean for all the closest points.

Decision Trees (DTs)

Decision tree, a supervised machine learning algorithm, is a tree-structured classifier that continuously divides the data based on specific parameters. This classifier starts with the root node (i.e., entire dataset), which further expands based on features into a number of branches (represent decision rule) and finally forms leaf nodes (viz., final outcome) (Breiman et al. 2017). Figure 2.2f illustrates the decision tree classifier. Decision tree has two types as follows: (1) classification tree (for categorical class variable) and (2) regression tree (for continuous class variable).

The following measures in a decision tree are used to check the impurity of a node t:

$$\text{Entropy}(t) = -\sum_{i\in(0,\,1)} p(i|t)\log_2 p(i|t)$$

$$\text{Gini}(t) = 1 - \sum_{i\in(0,\,1)} [p(i|t)]^2$$

$$\text{Classification error} = 1 - \max_i [p(i|t)].$$

The gain ratio is as follows:

$$= \frac{I(\text{parent}) - \sum_{i=1}^{n} \frac{N(\text{child}_i)}{N} I(\text{child}_i)}{-\sum_{i=1}^{n} p(\text{child}_i)\log_2 p(\text{child}_i)}.$$

In the decision tree, the gain ratio is used to measure the goodness of a node's split. This measure decides which feature in a tree should be the parent node and which should be set down after being split as a child node.

Random Forest (RF)

As the name suggests, the random forest comprises multiple decision trees and can provide more accurate predictions by combining all of them (Fig. 2.2g). However, this algorithm solves the problem of overfitting associated with the decision tree. Each decision tree in a random forest makes prediction of a class, and the class with highest number of hits will be the prediction of model. If an optimal classifier is unfeasible, a random forest classifier is especially helpful (Ditterrich 1997; Breiman 2001). This classifier applies bagging and feature randomness to generate uncorrelated trees in a forest, ultimately giving a more accurate and stable result. It is used to solve classification and regression problems.

CN2 Classifier

The CN2 algorithm induces classification rules "if. . .then.." from data using entropy (Clark and Niblett 1989). This classifier is used only for classification purposes and works well with imperfect/noisy training data. Figure 2.2h gives the schematic representation of this classifier.

The advantages and disadvantages of above supervised ML algorithms are given in Table 2.2.

2.3.3.2 Semi-Supervised Classifier

Semi-supervised ML algorithms use a combination of supervised learning on a small amount of labeled data and unsupervised learning on large amount of unlabeled data (Chapelle et al. 2009). This approach is applicable when a large number of labeled data is not available and overcomes the drawbacks of supervised (i.e., require sufficient labels and costly process) and unsupervised (i.e., limited range of applications) algorithms. This algorithm works on the basis of any of these three assumptions, viz., (1) continuity assumption, data points around each other belong to the same class; (2) cluster assumption, data can be split into distinct clusters and data points in the same cluster tend to share class; and (3) the manifold assumption, assumes that data points are present on the manifold of lower dimensions than input space. The manifold assumption is useful in condition where data points may locate in high dimensions, and is very difficult to map data points in those dimensions. Semi-supervised classifier includes Laplacian SVM, generative models, and transductive SVM.

Laplacian SVM

Laplacian support vector machine (LapSVM) is based on a support vector machine algorithm and obeys manifold regularization (Belkin et al. 2006). This is a graph-based approach in which nodes are formed from labeled and unlabeled data. The KNN algorithm is employed to compute edge weight to define similarity between data points in a graph. Through this procedure unlabeled nodes can be labeled by transferring the information of labeled data points to other nodes. See Fig. 2.2i for pictorial representation of LapSVM.

LapSVM solves the following optimization problem.

$$\arg\min_{f \in H_k} \frac{1}{n_\mathrm{l}} \sum_{i=1}^{n_i} |1 - y_i f(x_i)|_+ + \lambda_\mathrm{a} \|f\|_K^2 + \frac{\lambda_\mathrm{b}}{(n_\mathrm{l} + n_\mathrm{u})^2} \times f^\mathrm{T} L f,$$

where $\|f\|_K^2$, n_l, n_u are a regularization function for smoothness, number of labeled data points, number of unlabeled data points, respectively, and λ_a, λ_b are hyperparameters.

$$\text{loss function} = |1 - y_i f(x_i)|_+ = \max(0, 1 - yf(x)),$$

Table 2.2 Advantages and disadvantages of different supervised ML algorithms

Algorithm name	Advantages	Disadvantages
Support Vector Machine (SVM)	• High prediction accuracy • Handle high-dimensional space • Generalized well with small amount of data • Less prone to condition of overfitting • Less influence of outliers	• Extensive memory required for optimization • Not appropriate for large datasets • High time complexity • Selection of proper kernel function is challenging • Difficult to fine-tune some hyper-parameters
Naive Bayes	• Its implementation is easy and simple • Computationally very fast • If conditional independence assumption holds, it quickly generates outcomes • Works well with categorical and continuous data	• The conditional independence assumption does not always hold in the complex biological problems • Not suitable for imbalanced data • Shows decrease in performance with increase in sample size of dataset
Logistic regression classifier	• Simplest algorithm to use • Very fast • Do not suffer from overfitting in case of low-dimensional dataset • Very efficient for linearly-separable dataset	• Causes model overfitting on high-dimensional dataset • Shows decrease in performance with increase in number of samples and features in dataset • Not suitable for non-linear data • Sensitive to outliers
Artificial Neural Network (ANN)	• Robust to noise • Shows good fault tolerance • Works well on complex nonlinear association among dependent and independent features • Able to perform parallel processing	• Training performance increases with increase in training dataset • Unexplained functioning • The algorithm may be stuck into local minima • Hardware dependent • Long training time is required • Difficult to determine network structure • Suffers from overfitting
k-Nearest neighbors	• Implementation is simple and easy • Fast, as no training time is require • Highly reserved for local information • Versatile as it performs classification, regression, and search tasks • Analytically tractable	• Huge storage space requires • Different values of *k* give different outcomes • Takes long computation time for large dataset • Larger *k* values increase the time complexity • Sensitive to noisy data and outliers • Difficult to work with high dimensional data • Standardization and normalization steps require
Decision Trees	• Simple, easy to understand and interpret • Can handle irrelevant features and nonlinear associations	• Small changes affect stability of decision tree structure • Suffer from overfitting without proper tree pruning

(continued)

Table 2.2 (continued)

Algorithm name	Advantages	Disadvantages
	• Not sensitive to missing values • Runs fast • No normalization and scaling require • Data preparation takes less efforts	• Stuck in local minima • Not suitable for regression problem and prediction of continuous values • Difficult to find optimal decision tree
Random Forest	• High predictive performance • Works well with both classification and regression problems • Efficiently handles large datasets • Reduces overfitting and variance • Easy to understand model predictions	• Complex thus requires more computational power and resources • Suffers from overfitting for noisy datasets • Takes more time than decision tree
CN2 classifier	• Implementation is simple and easy to understand • Can handle irrelevant features and nonlinear relationships • Works fast	• In case of large number of features, difficult to define rules for training datasets

$$\sum_{i,j=1}^{n} W_{ij}\left(f(x_i) - f\left(x_j\right)\right)^2 = f^T L f,$$

W_{ij}, is the edge weights in the graph,
Laplacian operator, $L = D - W$.

This classifier is less vulnerable to overfitting, robustness to noise and outliers, and has high prediction power and good generalization ability with small labeled data. However, it does not work well with large number of data points because it needs high memory to construct a graph and is time-consuming.

2.3.4 Model Performance Evaluation

The major part of building an effective ML model is evaluation of model's performance. ML requires evaluation metrics for selecting the best model. Following are the primary building blocks of several evaluation metrics, formed from confusion matrix, which is obtained from actual and predicted class labels:

True Positive (*TP*): It represents an outcome in which positive samples are accurately predicted as positive by the model.

True Negative (*TN*): It represents an outcome in which negative samples are accurately predicted as negative samples by the model.

False Positive (*FP*): It represents an outcome in which negative samples are incorrectly predicted as positive by the model.

False Negative (*FN*): It represents an outcome in which positive samples are wrongly predicted as negative samples by the model.

Based on these 4 outcomes, model performance metrics are given below.

True-Positive Rate (*TPR*) (*Also Known as Sensitivity*): The probability that positive samples will predict positive.

$$\text{True Positive Rate (TPR) or Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{TPR} \in [0, 1].$$

False-Positive Rate (*FPR*): The probability that negative samples will predict positive.

$$\text{False Positive Rate (FPR)} = \frac{\text{FP}}{\text{FP} + \text{TN}}, \quad \text{FPR} \in [0, 1].$$

Precision: It estimates positive sample predictions that are genuinely from the positive class label.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Precision} \in [0, 1].$$

Recall: It estimates positive sample predictions from all actual positives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{Recall} \in [0, 1].$$

F-Measure: It balances precision and recalls both together to provide a single score.

$$F\text{-measure} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{(\text{Precision} + \text{Recall})}, \quad F\text{-measure} \in [0, 1].$$

Accuracy: It gives the total correct predictions (TP + TN) made by the model.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad \text{Accuracy} \in [0, 1].$$

The Area Under the Receiver Operating Characteristic Curve (auROC): It shows whether the model is capable of correctly discriminating between class labels.

$$\text{auROC} \in [0, 1].$$

Matthews Correlation Coefficient (*MCC*): It computes the correlation between actual and predicted labels and is calculated by the following formula.

$$\text{MCC} = \frac{(\text{TP} \cdot \text{TN}) - (\text{FP} \cdot \text{FN})}{\sqrt{(\text{TP} + \text{FP}) \cdot (\text{TP} + \text{FN}) \cdot (\text{TN} + \text{FP}) \cdot (\text{TN} + \text{FN})}}, \quad \text{MCC} \in [-1, 1].$$

Sufficient labeled data should be available to get statistically significant measures. The above performance metrics are useful to evaluate various ML algorithms to check how good the model is in predicting the outcome.

To implement different machine learning algorithms, various functions are available in scikit-learn (Python) (Pedregosa et al. 2011; Kramer 2016), e1071 (R) (Dimitriadou et al. 2008; Meyer et al. 2019), and Weka (Java) (Witten et al. 1999; Dimov et al. 2007).

2.4 Application of AI and Machine Learning Techniques in Cancer

As mentioned previously, cancer is a heterogeneous disease and shows distinct molecular as well as phenotypic characteristics within a tumor. This heterogeneous nature of the tumor poses a challenge for successful treatment and recovery. Different machine learning and artificial intelligence techniques have been effectively applied to combat the disease.

2.4.1 Cancer Classification

In order to determine the proper treatment regime and to reduce cancer-related mortality, correct classification of cancer is needed. RNA-seq provides genome-wide gene expression data that can be useful to determine cancer types and unravel cancer subtypes, indicating a profound impact on cancer prediction/diagnosis.

However, gene expression data have several limitations like small sample sizes, large number of genes, and presence of some uninformative genes. All these factors decrease classification performance. This indicates the need for filtration and feature selection steps before model building. With the stringent threshold, these two steps ensure that only informative and sufficiently differentially expressed genes between the target classes can be used in building the classifiers. Various supervised and unsupervised algorithms are developed using gene expression data for cancer classification purposes. For instance, Flynn et al. (2018) identified primary site of 33 cancers and the molecular subtype of 11 cancers by applying several machine learning approaches like diagonal linear discriminant analysis (DLDA), KNN, RF, and SVM on gene expression profiles from the TCGA. The gene expression data of breast cancer provide an information based on which ML methods classified the disease into triple-negative breast cancer (TNBC) and non triple-negative breast cancer (non-TNBC) (Wu and Hicks 2021). The authors evaluated four different classification algorithms, namely SVM, KNN, NB and DT and found that SVM was able to divide TNBC and non-TNBC with less errors compared to others. Zhang et al. (2020) classified glioblastoma subtypes using SVM and RF with methylation data.

2.4.2 Anti Cancer Drug Response Prediction

The complexity of the tumor and its microenvironment lead to partial or no response to anti cancer drugs. Therefore, finding the relationship between drug response and molecular features of cancer cells or their microenvironment will be helpful in the identification of novel diagnostic/predictive biomarkers and evaluating drug response to guide personalized medicine. Miranda et al. (2021) used DNA methylation profiles at global scale from several cancer cell lines in the Genomics of Drug Sensitivity in Cancer (GDSC) database to predict eight anti cancer drug's cytotoxic responses by machine learning algorithms. Here, authors used RF, SVM, gradient boosting machines, and KNN for both classification and regression. The predictions made by the RF classifier were significantly correlated with Temozolomide drug responses for low-grade gliomas. Bomane et al. (2019) assessed ML algorithms, namely RF, XGB, LGBM, logistic regression LR, classification and regression tree (CART) on six molecular profiles (CPG and CGI DNA methylation, mRNA expression, miRNA profiles, isomiR expression, CNV) of breast tumors to predict paclitaxel response. A study found that DNA methylation and miRNA profiles out of six molecular profiles were the most informative overall.

2.4.3 Survival Prediction

Survival is the time during which a patient survives after disease diagnosis. Survival analysis is crucial in cancer patient management because of *tumor heterogeneity*. Integration of multi omics data and ML algorithms holds promise for improving the

survival of cancer patients. Mitchel et al. (2019) developed ML workflow using decision-level integration of multi omics tumor data to predict the overall survival of breast cancer patients. This study predicted the survival with an accuracy of 85% and area under the curve (AUC) of 87% with multi omics data and identified best integrated classification combination as methylation, miRNA, and gene expression. Recently, an auto-encoder was used to integrate and reduce the dimensions of pancreatic cancer patients' microRNA expression and DNA methylation data (Baek and Lee 2020). Machine learning models like SVM, RF, and LR and L2 regularized logistic regression were implemented to combine the clonal expansion of DNA mutations and multi omics data to predict cancer recurrence and survival within five years. This study revealed that mutated genes with low cellular prevalence (CP) values (i.e., mutated in smaller clones) were not significantly associated with recurrence and survival. However, the topmost CP value genes which usually mutated in the initial stages of tumor development were significantly related to poor prognosis in pancreatic cancer.

2.4.4 Metastasis Prediction

Cancer metastasis contributes to cancer-related mortality. Early prediction of it can improve prognosis. Most of the time, the metastasis prediction models use gene expression data. Recently, miRNA expression levels and DNA methylation patterns have also been explored for metastasis prediction. Tuo et al. (2018) used the SVM-based classifier on gene expression profiles to predict whether the breast cancer samples were metastatic or non metastatic, and prediction accuracy was evaluated by training and validating the model on TCGA data, an independent dataset. The mRNA- and miRNA-specific classifiers were used to differentiate cross-cancer tissue samples as primary or metastatic (Lee et al. 2019a). This study used three classification algorithms (LASSO, RF, and SVM) with bootstrap cross-validation method to determine how accurately mRNA and miRNA biomarkers can classify metastasis.

2.4.5 Biomarker Prediction

Cancer biomarkers may evaluate the risk of cancer development or progression in a specific tissue or therapeutic response. Thus, to decide appropriate therapy for cancer patients, the identification of cancer biomarkers is essential and may be useful for patients' survival. Tabl et al. (2019) used a machine learning multi class approach, the one-versus-rest technique to identify potential biomarkers which can increase breast cancer patients' survival. In this study, gene expression profiles of cancer patients who received different treatments like surgery, hormone therapy, and radiotherapy were used and also considered their status as living or deceased. The classifiers, namely random forest, SVM, and Naive Bayes, were implemented and found that random forest outperformed the others and showed a better classification

power for the hierarchical model. In another study, microRNA expression data were used for validating clinically selected miRNAs as breast cancer biomarkers using several machine learning classifiers (Rehman et al. 2019). The feature selection methods like information gain (IG), chi-squared (CHI2) and least absolute shrinkage and selection operation (LASSO) were implemented to rank the miRNAs by their importance and concluded that not all miRNAs carry equal weightage to act as a cancer biomarker, even among those clinically selected ones.

2.5 Conclusion and Future Directions

In the upcoming decade, advancement in artificial intelligence and proper implementation of machine learning methods in cancer genomics will reveal crucial aspects in oncology. Recent studies showed that the utilization of diverse omics data and their combination enhanced the cancer prediction performance of the machine learning models. Several challenges still exist, like data collections, pre processing, and storage. The collaboration between clinicians and bioinformaticians will be beneficial to get organized or structured data from various sources and at numerous scales. Recently, a study combined multi omics data with drug data to predict overall survival and different subtypes (at pathological, histological, and molecular levels) of glioma patients (Saurabh et al. 2020). This kind of comprehensive analysis can be helpful for doctors and clinicians in early diagnosis as well as in deciding the correct and personalized therapeutic strategies for individual cancer patients. In recent years, several studies integrated genomics data with pathological image data for identifying distinct cellular subtypes, prognostic biomarkers, mutational status, therapeutic strategies, and clinical outcomes. Applying a deep learning classifier on point mutation, copy number alteration, and gene expression data, Qu et al. (2021) predicted driver mutations and signaling pathways activity from histopathological whole slide images (WSI) of breast carcinoma patients. Recently, a new approach, radiogenomics, has emerged in the area of personalized medicine which integrates genetic and radiomic data for monitoring genetic variations in patients through medical images and can act as a better substitute for painful mediation (Shui et al. 2020; Gullo et al. 2020). Lee et al. (2019b) used machine learning classifiers on quantitative radiomic data of glioblastoma patients obtained from magnetic resonance images (MRI) and targeted sequencing of the IDH1 gene to predict its mutation status from images. The study showed a good (~80%) predictive power for IDH1 mutation by training 31 features of MRI and also observed good (i.e., 66.3–83.4%) accuracy through validation on an external set (Lee et al. 2019b). Such an integrative machine learning approach has an important role in improving diagnosis and prognosis. Nowadays, model explainability is gaining importance in the machine learning field as it explains the backend process of the model prediction, i.e., which particular features mainly contribute to the model prediction. The success of ML solutions may provide a handful of clinically relevant tools for cancer patient's treatment and management.

Acknowledgements CM acknowledges the Department of Science and Technology (DST) for the support as an Inspire Faculty (Award No. IFA19-PH248).

References

Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. Mach Learn 6(1):37–66

Ahmed N, Greening D, Samardzija C, Escalona RM, Chen M, Findlay JK et al (2016) Unique proteome signature of post-chemotherapy ovarian cancer ascites-derived tumor cells. Sci Rep 6(1):1–13

Alfardus H, de los Angeles Estevez-Cebrero CM, Rowlinson J, Aboalmaaly A, Lourdusamy A, Abdelrazig S et al (2021) Intratumour heterogeneity in microRNAs expression regulates glioblastoma metabolism. Sci Rep 11(1):1–14

Baek B, Lee H (2020) Prediction of survival and recurrence in patients with pancreatic cancer by integrating multi-omics data. Sci Rep 10(1):1–11

Bam M, Chintala S, Fetcko K, Williamsen BC, Siraj S, Liu S et al (2021) Genome wide DNA methylation landscape reveals glioblastoma's influence on epigenetic changes in tumor infiltrating CD4+ T cells. Oncotarget 12(10):967

Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. Cell Res 21(3):381–395

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M et al (2012) NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 41(D1):D991–D9D5

Belkin M, Niyogi P, Sindhwani V (2006) Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. J Mach Learn Res 7(11):2399–2434

Bhattacharya A, Ziebarth JD, Cui Y (2013) SomamiR: a database for somatic mutations impacting microRNA function in cancer. Nucleic Acids Res 41(D1):D977–DD82

Bomane A, Gonçalves A, Ballester PJ (2019) Paclitaxel response can be predicted with interpretable multi-variate classifiers exploiting DNA-methylation and miRNA data. Front Genet 10: 1041

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Breiman L, Friedman JH, Olshen RA, Stone CJ (2017) Classification and regression trees. Routledge, New York

Cantero D, Rodríguez de Lope Á, Moreno De La Presa R, Sepúlveda JM, Borrás JM, Castresana JS et al (2018) Molecular study of long-term survivors of glioblastoma by gene-targeted next-generation sequencing. J Neuropathol Exp Neurol 77(8):710–716

Cardoso JG, Andersen MR, Herrgård MJ, Sonnenschein N (2015) Analysis of genetic variation and potential applications in genome-scale metabolic modeling. Front Bioeng Biotechnol 3:13

Chapelle O, Scholkopf B, Zien A (2009) Semi-supervised learning (Chapelle, O. et al., Eds.; 2006) [book reviews]. IEEE Trans Neural Netw 20(3):542

Clark P, Niblett T (1989) The CN2 induction algorithm. Mach Learn 3(4):261–283

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Dagogo-Jack I, Shaw AT (2018) Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol 15(2):81–94

Davis AR, Stone SL, Oran AR, Sussman RT, Bhattacharyya S, Morrissette JJ et al (2021) Targeted massively parallel sequencing of mature lymphoid neoplasms: assessment of empirical application and diagnostic utility in routine clinical practice. Mod Pathol 34(5):904–921

Dietz S, Lifshitz A, Kazdal D, Harms A, Endris V, Winter H et al (2019) Global DNA methylation reflects spatial heterogeneity and molecular evolution of lung adenocarcinomas. Int J Cancer 144(5):1061–1072

Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A (2008) Misc functions of the Department of Statistics (e1071), TU Wien. R Pack 1:5–24

Dimov R, Feld M, Kipp DM, Ndiaye DA, Heckmann DD. Weka: Practical machine learning tools and techniques with java implementations. AI tools Seminar University of Saarland WS University of Waikato, Hamilton 2007;6(07)

Ditterrich T (1997) Machine learning research: four current direction. Artif Intell Mag 4:97–136

Flynn WF, Namburi S, Paisie CA, Reddi HV, Li S, Karuturi RKM et al (2018) Pan-cancer machine learning predictors of primary site of origin and molecular subtype. bioRxiv 2018:333914

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO et al (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6(269):pl1–ppl

Gao X, Xia X, Li F, Zhang M, Zhou H, Wu X et al (2021) Circular RNA-encoded oncogenic E-cadherin variant promotes glioblastoma tumorigenicity through activation of EGFR–STAT3 signalling. Nat Cell Biol 23(3):278–291

Garrett A-M, Lastakchi S, McConville C (2020) The personalisation of glioblastoma treatment using whole exome sequencing: a pilot study. Gene 11(2):173

Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A et al (2020) Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol 38(6):675–678

Gullo RL, Daimiel I, Morris EA, Pinker K (2020) Combining molecular and imaging metrics in cancer: radiogenomics. Insights Imaging 11(1):1–17

Guo M, Peng Y, Gao A, Du C, Herman JG (2019) Epigenetic heterogeneity in cancer. Biomark Res 7(1):1–19

Gupta S, Chatterjee S, Mukherjee A, Mutsuddi M (2017) Whole exome sequencing: uncovering causal genetic variants for ocular diseases. Exp Eye Res 164:139–150

Hagan MT, Demuth HB, Beale M (1997) Neural network design. PWS Publishing, Boston

Hass R, von der Ohe J, Ungefroren H (2020) Impact of the tumor microenvironment on tumor heterogeneity and consequences for cancer cell plasticity and stemness. Cancer 12(12):3716

Hasty P, Montagna C (2014) Chromosomal rearrangements in cancer: detection and potential causal mechanisms. Mol Cell Oncol 1(1):e29904

He X, Chang S, Zhang J, Zhao Q, Xiang H, Kusonmano K et al (2007) MethyCancer: the database of human DNA methylation and cancer. Nucleic Acids Res 36:D836–DD41

Hinton G, Roweis ST (2002) Stochastic neighbor embedding. NIPS, Toronto

Hira ZM, Gillies DF (2015) A review of feature selection and feature extraction methods applied on microarray data. Adv Bioinform 2015:198363

Huang H-Y, Li J, Tang Y, Huang Y-X, Chen Y-G, Xie Y-Y et al (2021) MethHC 2.0: Information repository of DNA methylation and gene expression in human cancer. Nucleic Acids Res 49 (D1):D1268–D1D75

Ilango S, Paital B, Jayachandran P, Padma PR, Nirmaladevi R (2020) Epigenetic alterations in cancer. Front Biosci 25(1):1058–1109

Kang Y, Pantel K (2013) Tumor cell dissemination: emerging biological insights from animal models and cancer patients. Cancer Cell 23(5):573–581

Kontomanolis EN, Koutras A, Syllaios A, Schizas D, Mastoraki A, Garmpis N et al (2020) Role of oncogenes and tumor-suppressor genes in carcinogenesis: a review. Anticancer Res 40(11): 6009–6015

Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: a review of classification techniques. Emerg Artif Intell Appl Comput Eng 160(1):3–24

Kramer O (2016) Scikit-learn. Machine learning for evolution strategies. Springer, Cham, pp 45–53

Lee JV, Berry CT, Kim K, Sen P, Kim T, Carrer A et al (2018) Acetyl-CoA promotes glioblastoma cell adhesion and migration through Ca^{2+}-NFAT signaling. Genes Dev 32(7–8):497–511

Lee SC, Quinn A, Nguyen T, Venkatesh S, Quinn TP (2019a) A cross-cancer metastasis signature in the microRNA–mRNA axis of paired tissue samples. Mol Biol Rep 46(6):5919–5930

Lee MH, Kim J, Kim S-T, Shin H-M, You H-J, Choi JW et al (2019b) Prediction of IDH1 mutation status in glioblastoma using machine learning technique based on quantitative radiomic data. World Neurosurg 125:e688–ee96

Legendre C, Gooden GC, Johnson K, Martinez RA, Liang WS, Salhia B (2015) Whole-genome bisulfite sequencing of cell-free DNA identifies signature associated with metastatic breast cancer. Clin Epigenetics 7(1):1–10

Li C, Wu S, Yang Z, Zhang X, Zheng Q, Lin L et al (2017) Single-cell exome sequencing identifies mutations in KCP, LOC440040, and LOC440563 as drivers in renal cell carcinoma stem cells. Cell Res 27(4):590–593

Lightbody G, Haberland V, Browne F, Taggart L, Zheng H, Parkes E et al (2019) Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. Brief Bioinform 20(5):1795–1811

Liu Y, Huang R, Liu Y, Song W, Wang Y, Yang Y et al (2018) Insights from multidimensional analyses of the pan-cancer DNA methylome heterogeneity and the uncanonical CpG–gene associations. Int J Cancer 143(11):2814–2827

Lu Y, Chan Y-T, Tan H-Y, Li S, Wang N, Feng Y (2020) Epigenetic regulation in human cancer: the potential role of epi-drug in cancer therapy. Mol Cancer 19(1):1–16

Martinez-Gutierrez AD, Catalan OM, Vázquez-Romo R, Porras Reyes FI, Alvarado-Miranda A, Lara Medina F et al (2019) miRNA profile obtained by next-generation sequencing in metastatic breast cancer patients is able to predict the response to systemic treatments. Int J Mol Med 44(4): 1267–1280

Marusyk A, Janiszewska M, Polyak K (2020) Intratumor heterogeneity: the Rosetta stone of therapy resistance. Cancer Cell 37(4):471–484

McQuerry JA, Chang JT, Bowtell DD, Cohen A, Bild AH (2017) Mechanisms and clinical implications of tumor heterogeneity and convergence on recurrent phenotypes. J Mol Med 95(11):1167–1178

Meacham CE, Morrison SJ (2013) Tumour heterogeneity and cancer cell plasticity. Nature 501(7467):328–337

Meienberg J, Bruggmann R, Oexle K, Matyas G (2016) Clinical sequencing: is WGS the better WES? Hum Genet 135(3):359–362

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C-C et al (2019) Package 'e1071'. R J 2019:1071

Miranda SP, Baião FA, Fleck JL, Piccolo SR (2021) Predicting drug sensitivity of cancer cells based on DNA methylation levels. PLoS One 16(9):e0238757

Mitchel J, Chatlin K, Tong L, Wang MD (2019) A translational pipeline for overall survival prediction of breast cancer patients by decision-level integration of multi-omics data. In: 2019 IEEE International conference on bioinformatics and biomedicine (BIBM). IEEE, Piscataway, NJ

Mueller JJ, Schlappe BA, Kumar R, Olvera N, Dao F, Abu-Rustum N et al (2018) Massively parallel sequencing analysis of mucinous ovarian carcinomas: genomic profiling and differential diagnoses. Gynecol Oncol 150(1):127–135

Murtaza M, Dawson S-J, Pogrebniak K, Rueda OM, Provenzano E, Grant J et al (2015) Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. Nat Commun 6(1):1–6

Nanda JS, Kumar R, Raghava GP (2016) dbEM: a database of epigenetic modifiers curated from cancerous and normal genomes. Sci Rep 6(1):1–6

O'Geen H, Echipare L, Farnham PJ (2011) Using ChIP-seq technology to generate high-resolution profiles of histone modifications. Epigenetics protocols. Springer, Cham, pp 265–286

Oliveira AM, Ross JS, Fletcher JA (2005) Tumor suppressor genes in breast cancer: the gatekeepers and the caretakers. Pathol Patterns Rev 124(suppl_1):S16–S28

Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H et al (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 344(6190): 1396–1401

Pearson K (1901) LIII. On lines and planes of closest fit to systems of points in space. Lond Edinb Dublin Philos Mag J Sci 2(11):559–572

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
Perera RM, Bardeesy N (2012) On oncogenes and tumor suppressor genes in the mammary gland. Cold Spring Harb Perspect Biol 4(6):a013466
Qu H, Zhou M, Yan Z, Wang H, Rustgi VK, Zhang S et al (2021) Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning. NPJ Precis Oncol 5(1):1–11
Rehman O, Zhuang H, Muhamed Ali A, Ibrahim A, Li Z (2019) Validation of miRNAs as breast cancer biomarkers with a machine learning approach. Cancer 11(3):431
Ren L, Li J, Wang C, Lou Z, Gao S, Zhao L et al (2021) Single cell RNA sequencing for breast cancer: present and future. Cell Death Dis 7(1):1–11
Rish I (2001) An empirical study of the Naive Bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence. SAGE, Montreal
Sakthikumar S, Roy A, Haseeb L, Pettersson ME, Sundström E, Marinescu VD et al (2020) Whole-genome sequencing of glioblastoma reveals enrichment of non-coding constraint mutations in known and novel genes. Genome Biol 21:1–22
Saurabh R, Nandi S, Sinha N, Shukla M, Sarkar RR (2020) Prediction of survival rate and effect of drugs on cancer patients with somatic mutations of genes—an AI based approach. Chem Biol Drug Des 96:1005–1019
Shui L, Ren H, Yang X, Li J, Chen Z, Yi C et al (2020) Era of radiogenomics in precision medicine: an emerging approach for prediction of the diagnosis, treatment and prognosis of tumors. Front Oncol 10:3195
Sood A, Miller AM, Brogi E, Sui Y, Armenia J, McDonough E et al (2016) Multiplexed immunofluorescence delineates proteomic cancer cell states associated with metabolism. JCI Insight 1(6):e87030
Tabl AA, Alkhateeb A, ElMaraghy W, Rueda L, Ngom A (2019) A machine learning approach for identifying gene biomarkers guiding the treatment of breast cancer. Front Genet 10:256
Tomczak K, Czerwińska P, Wiznerowicz M (2015) The cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol 19(1A):A68
Torgerson W (1952) The first major MDS breakthrough. Psychometrika 17:401–419
Tuo Y, An N, Zhang M (2018) Feature genes in metastatic breast cancer identified by MetaDE and SVM classifier methods. Mol Med Rep 17(3):4281–4290
Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10(1):57–63
Wang L-H, Wu C-F, Rajasekaran N, Shin YK (2018) Loss of tumor suppressor gene function in human cancer: an overview. Cell Physiol Biochem 51(6):2647–2693
Wang N, Zheng J, Chen Z, Liu Y, Dura B, Kwak M et al (2019) Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. Nat Commun 10(1):1–12
Weigelt B, Bi R, Kumar R, Blecua P, Mandelker DL, Geyer FC et al (2018) The landscape of somatic genetic alterations in breast cancers from ATM germline mutation carriers. J Natl Cancer Inst 110(9):1030–1034
Welch DR, Hurst DR (2019) Defining the hallmarks of metastasis. Cancer Res 79(12):3011–3027
Witten IH, Frank E (2000) Data mining: practical machine learning tools and techniques with Java implementations. Morgan Kaufmann, Burlington
Witten IH, Frank E, Trigg LE, Hall MA, Holmes G, Cunningham SJ (1999) Weka: Practical machine learning tools and techniques with Java implementations. ACM Sigmod Rec 31(1): 76–77
Wu J, Hicks C (2021) Breast cancer type classification using machine learning. J Pers Med 11(2):61
Wu H, Xu T, Feng H, Chen L, Li B, Yao B et al (2015) Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. Nucleic Acids Res 43(21):e141

Xi Y, Shi J, Li W, Tanaka K, Allton KL, Richardson D et al (2018) Histone modification profiling in breast cancer cell lines highlights commonalities and differences among subtypes. BMC Genomics 19(1):1–11

Yates LR, Knappskog S, Wedge D, Farmery JH, Gonzalez S, Martincorena I et al (2017) Genomic evolution of breast cancer metastasis and relapse. Cancer Cell 32(2):169–184

Zhang X, Zhang M, Hou Y, Xu L, Li W, Zou Z et al (2016) Single-cell analyses of transcriptional heterogeneity in squamous cell carcinoma of urinary bladder. Oncotarget 7(40):66069

Zhang M, Yang D, Gold B (2019a) Origin of mutations in genes associated with human glioblastoma multiform cancer: random polymerase errors versus deamination. Heliyon 5(3):e01265

Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H et al (2019b) The international cancer genome consortium data portal. Nat Biotechnol 37(4):367–369

Zhang Y-H, Li Z, Zeng T, Pan X, Chen L, Liu D et al (2020) Distinguishing glioblastoma subtypes by methylation signatures. Front Genet 11:1482

Zhao EY, Jones M, Jones SJ (2019) Whole-genome sequencing in cancer. Cold Spring Harb Perspect Med 9(3):a034579

Zhao Z, Zhang K-N, Wang Q, Li G, Zeng F, Zhang Y et al (2021) Chinese Glioma Genome Atlas (CGGA): a comprehensive resource with functional genomic data from Chinese gliomas. Genomics Proteomics Bioinformatics 19(1):1–2

Cancer Biomarkers in the Era of Systems Biology

3

Shazia Firdous, Sunil Kumar Srivastava, and Sudipto Saha

Abstract

Cancer biomarkers allow diagnosis, risk assessment, monitoring disease progression, and prediction of therapeutic response in oncology. Different forms and types of cancer biomarkers exist and it covers a broad range of biochemical entities including DNA, miRNA, cirRNA, and proteins. The advances in high-throughput technologies including genomics, transcriptomics, proteomics, and metabolomics have generated great opportunities for the discovery of new and effective cancer biomarkers. Due to the heterogeneous nature of cancer cells, it is difficult to identify the clinically useful precise cancer biomarkers. Multiomics data analyses using a systems approach play a vital role in the discovery of cancer biomarkers. In this chapter, a brief classification system for cancer biomarkers has been provided according to their biochemical nature and based on clinical utility along with the application of recent high-throughput approaches used in cancer biomarker discovery. Several databases and bioinformatics tools applied in cancer biomarker discovery have been mentioned. In summary, different cancer biomarker types, omics approaches used for cancer biomarker discovery, and dedicated cancer-related databases and tools have been discussed.

S. Firdous · S. Saha (✉)
Division of Bioinformatics, Bose Institute, Kolkata, West Bengal, India
e-mail: shazia@jcbose.ac.in; ssaha4@jcbose.ac.in

S. K. Srivastava
Department of Microbiology, Swami Shraddhanand College, University of Delhi, New Delhi, Delhi, India
e-mail: sksrivastava@ss.du.ac.in

S. Singh (ed.), *Systems Biomedicine Approaches in Cancer Research*,
https://doi.org/10.1007/978-981-19-1953-4_3

3.1 Introduction

Cancer, a heterogeneous group of diseases, represents the second leading cause of death globally and is responsible for an estimated ten million deaths worldwide in 2020 (Sung et al. 2021). For advancement and better treatment in cancer, patient's early diagnosis and accurate detection of cancer are the crucial steps. In the past few decades, the biomarker field along with the improved quality of medical services and technologies has transformed the ability of cancer researchers to easily diagnose and classify cancer at the molecular level and resulted in improved drug development and clinical trial design (Hu and Dignam 2019; Parker et al. 2021; Goossens et al. 2015; Goyal et al. 2021). The definition of biomarker has been given differently by different groups. Some of the definitions limit the scope of biomarkers up to biological molecule or biochemical features; on the other hand, a broader definition of biomarker provided by the Biomarker Consortium (Foundation of National Institute of Health) increases the probability of discovering new biomarkers in the ever-changing era of research biology (Califf 2018; Wu and Qu 2015; Strimbu and Tavel 2010). According to United Nations, World Health Organization (WHO) biomarkers can be defined as any measurable substance, structure, or process or its product that can predict the incidence of disease outcome (World Health Organization 2001). In simple words, any measurable indicator of disease condition and treatment response indicator can be considered as a biomarker (Zare Jeddi et al. 2021; Lassere 2008). In a study in 1848, light chain of immunoglobulin was identified as the first-ever cancer biomarker in the myeloma patient urine sample (Solomon 1980). Since then, numerous cancer biomarkers have been identified such as alfa-fetoprotein, carcinoembryonic antigen (CEA), and prostate-specific antigen (PSA). Every era of cancer biomarker discovery has been closely associated with the new and powerful technology (Tatarinov 1964; Xu et al. 2021; Campos-da-Paz et al. 2018; Cózar et al. 2021; Wang et al. 1979).

In the past few decades, high-throughput technologies such as genomics, metagenomics, transcriptomics, proteomics, and metabolomics have generated a significant amount of data for cancer biology. For example, a proteomics-based study identified DNA-dependent protein kinase catalytic subunit (DNA-PKcs) as a potential diagnostic biomarker for clinical outcomes in breast cancer (Asleh et al. 2021). A combined study of transcriptomics with metabolomics revealed a panel of important serum biomarkers such as 2-hydroxybutyric acid and 4-hydroxybutyric acid as candidate diagnostic biomarkers for lung cancer proliferation through the Ca^{2+} signaling pathway (Zheng et al. 2021). In addition, miRNA-194 was identified as a favorable prognostic biomarker for gastric cancer (Wang et al. 2021a). Moreover, the machine learning-based computational algorithm helps to classify the complex pattern of cancer research outcomes generated by a plethora of high-throughput experiments (Echle et al. 2021). As a result, these omics approaches when coupled with bioinformatics and computational methods provide great opportunities for biomarker discovery and facilitate therapeutic development for cancer (Menyhárt and Győrffy 2021; Yan et al. 2016).

As with time, the information about cancer biomarkers has expanded, and so also the complexities of tumor biology evolve adding challenges to the development of efficient biomarkers. More powerful tools, cross-validation techniques, and system biology approaches should be deployed in future to increase the yield of biomarkers in cancer therapeutics (Ileana Dumbrava et al. 2018; Sheng et al. 2020; Louie et al. 2021). In this chapter, we have highlighted three aspects of cancer biomarkers. The first part deals with the cancer biomarker classification according to the molecular types and the potential role they play in clinical oncology. The second part relates to the omics approaches that are nowadays considered as a powerful strategy to untangle the complex biological behavior of cancer cells. These high-throughput technologies help to discover molecular biomarkers with prognostic, diagnostic, and targeted therapeutic values. The third part discusses some important bioinformatics tools and software related to cancer biomarker discovery. The multiomics data integration and molecular regulatory network-based analysis tool can revolutionize the biomarker field for the effective treatment of cancer.

3.2 Categorization of Cancer Biomarkers

In biomedical science, cancer biomarker identification is one of the major multidisciplinary areas and their categorization should be considered contextual. Further, due to the enhancement in recent advanced technologies, enormous cancer biomarkers have been reported thereby it is difficult to categorize cancer biomarkers by considering only one aspect (Henry and Hayes 2012). However, according to contemporary findings, several attempts have been applied to classify cancer biomarkers (Nguyen et al. 2020). A basic schematic representation for the classification of cancer biomarkers is shown in Fig. 3.1.

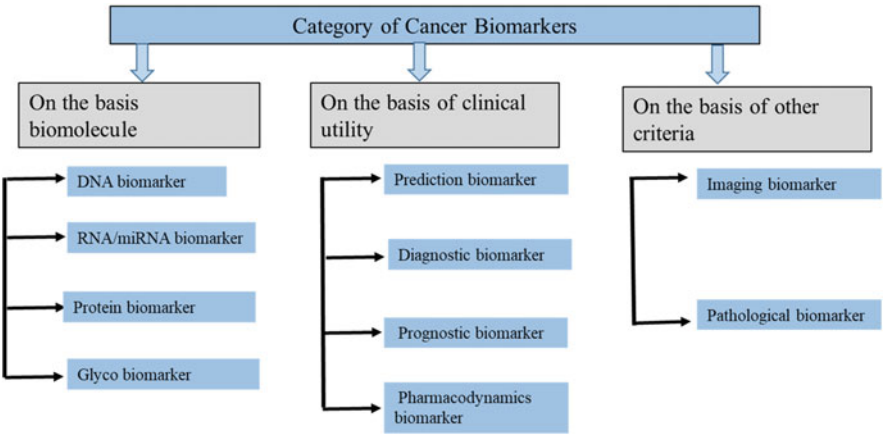


Fig. 3.1 Classification of cancer biomarkers based on biomolecules, clinical utility, and other criteria

3.2.1 Classification of Cancer Biomarkers Based on Biomolecules

3.2.1.1 DNA Cancer Biomarkers

Genetic alteration, gene rearrangement, and point mutation are responsible for cancer progression (Housman et al. 2014; Torgovnick and Schumacher 2015; Li et al. 2021a). The most common cancer DNA biomarker includes single nucleotide polymorphisms (SNPs). In a recent study, SNP–SNP interactions have been identified as a potential indicator of prostate cancer aggressiveness (Lin et al. 2021). Another study demonstrated m6A-associated functional SNPs (PLEKHA8, SMUG1, CDC123, RMI2, ACSM5) as major functional variants for thyroid cancer (Ruan et al. 2021). Moreover, researchers have defined circulating tumor DNA in localized nonsmall cell lung cancer as a prognostic biomarker and predicted the survival rate (Peng et al. 2020). Epigenetic CpG methylation also provides a broad range for early cancer detection and can be used as a cancer biomarker (Locke et al. 2019).

3.2.1.2 RNA Cancer Biomarkers

Quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR), microarray, Serial Analysis of Gene Expression (SAGE), differential display, microfluid card, and bead-based methods are commonly used to detect RNA or miRNA cancer biomarkers (Xi et al. 2017; Yamashita et al. 2008). Various identified mRNA and miRNA act as effective diagnostic and prognostic cancer biomarkers (Paramasivam 2021; Wang et al. 2021b; Bautista-Sánchez et al. 2020). In blood-based liquid biopsies, platelet RNA has been detected as an early diagnostic biomarker for cancer (Wurdinger et al. 2020). A recent article has indicated that miRNA plays an important role in identification of drug and decision-making of drug delivery in cancer therapeutics (Paramasivam 2021). Clinical researches reported that circRNAs expression indicates the cancer prognosis in different types of cancer, for example in a recent study circRNA expression in peripheral blood has been found to be correlated with cancer size (Li and Han 2019; Pan et al. 2019; Chen et al. 2017).

3.2.1.3 Protein Cancer Biomarkers

The most commonly used techniques used to identify protein-based biomarkers include Polyacrylamide Gel Electrophoresis (PAGE) and Two-Dimensional fluorescence Difference Gel Electrophoresis (2D-DIGE) (Issaq and Veenstra 2008). The high-throughput method includes proteomics study based on Mass Spectroscopy (MS), Surface-Enhanced Laser Absorption Desorption Ionization Time of Flight (SELDI-TOF), and Matrix-Associated Laser Absorption Desorption Ionization Time of Flight (MALDI-TOF) and discovered various protein and peptide biomarkers for ovarian and breast cancer (Zeidan et al. 2009; Liu 2011; Swiatly et al. 2017; Lv et al. 2019). Quantitative proteomics has also been applied to identify potential cancer protein biomarkers in different types of cancers (Kwon et al. 2021). To discover secreted protein-mediated interaction between the cancer cell and nonmalignant stroma, Stable Isotope Labeling with Amino Acids in Cell culture

(SILAC) was performed and further in pancreatic cancer, Wntless homolog protein (WLS) and Myristoylated Alanine-rich C-Kinase Substrate (MARCKS) found to be associated with oxaliplatin resistance (Kim et al. 2021; Wang et al. 2018). Isobaric Tags for Relative and Absolute Quantitation (iTRAQ), a quantitative proteomics approach was applied to find differentially expressed protein in metformin-treated cervical cancer cells and reported that metformin increases tumor suppressor gene expression IGFBP7 (Xia et al. 2020). Liquid Chromatography-Mass Spectrometry/Mass Spectrometry (LC-MS/MS) and antibody arrays are used in lung breast and colon cancer to find a panel of potential protein biomarkers (Wang et al. 2016a; Huang and Zhu 2017). Protein-based biomarkers are considered as a more valuable biomarker as compared to DNA- and RNA-based biomarkers as they are involved in functional molecular pathways and determine the disease initiation and progression state (Zhang et al. 2019). In a recent study, PINK1 protein has been reported as a prognostic biomarker for cancer (Zhu et al. 2020).

3.2.1.4 Carbohydrate Cancer Biomarkers

Cancer progression is often associated with changes in the expression of surface carbohydrates such as N-linked and O-linked glycans (Leney et al. 2017). The glycobiomarkers [glycoprotein, glycolipid, and proteoglycan] serve as candidate epidemiological cancer biomarkers (Lan et al. 2016; Daniotti et al. 2013). Mass spectrometry such as MALDI-TOF and Electrospray Ionization (ESI) are generally used for profiling of N- and O-linked glycosylation at serine and threonine residue of candidate protein molecules in human sera and cancer cell lines (Dube and Bertozzi 2005; An et al. 2006, 2010; Drake et al. 2017). Glycan microarray analysis has been found a good biomarker identification method for the diagnosis of breast cancer (Wang et al. 2008). Further in hepatocellular carcinoma, cancer-associated carbohydrate antigens (DSGG, fucosyl GM1, and Gb2 of CACAs) have been reported as potential biomarkers for early detection of cancer (Wu et al. 2012).

3.2.2 Classification of Cancer Biomarkers Based on Clinical Utility

Based on the putative application, cancer biomarkers can be classified under the following categories. Although some biomarkers are overlapping in nature, for example, the grading and staging cancer biomarker is also used as a prediction and screening biomarker (Ludwig and Weinstein 2005).

3.2.2.1 Prediction Cancer Biomarker

Predictive biomarkers predict the response and efficacy of the treatment and also help to determine the optimal dose of the drug at the initial treatment stages (Alves Martins et al. 2019; Bai et al. 2020). As cancer is a heterogeneous disease and the same cancer type responds differently to a drug thereby these types of biomarkers help in selecting a successful treatment process and minimizing the drug toxicity. A common predictive biomarker is overexpression of HER2, which predicts breast cancer's response to drugs like trastuzumab (Jørgensen and Hersom 2016). A high

level of circulating IFN-γ predicts the response of immunotherapies to immune checkpoint blockade in melanoma and nonsmall cell lung cancer patients (Karachaliou et al. 2018). A more recent study predicts that overexpression of excision repairs cross-complementation group 1 (ERCC1) increases DNA excision repair and imparts resistance to platinum-based drugs (Chung 2021). Additionally, in colorectal cancer, the mutation in MAPK pathway genes serves as a predictive biomarker for EGFR therapy and indicates resistance to cetuximab drug (Boussios et al. 2019).

3.2.2.2 Detection/Diagnostic Cancer Biomarker

The screening or detection of cancer biomarkers is the real indicator of the presence of cancer. These biomarkers help to identify benign cancer before metastasis. The tumor cells produce several immune factors, serum proteins, and circulating free DNA and RNA and these molecules can serve as cancer detection biomarkers (Parker et al. 2018). Diagnostic biomarkers play an important role in classifying patients into subtypes and also detecting the presence of the disease. Prostate-specific antigen (PSA) is the best-known cancer biomarker for prostate cancer early detection (Welch and Albertsen 2009). Cancer Antigen 19-9 (CA-19-9) is a diagnostic serum biomarker for pancreatic ductal carcinoma (Poruk et al. 2013). Further, another cancer antigen CA 125 is also used as a classical biomarker for the detection of ovarian cancer (Felder et al. 2014). The utility of cytokines as diagnostic biomarkers is increasing rapidly although further validation is required. IL-6 and VEGF serve as possible diagnostic biomarkers for ovarian and gastric cancer (Liang et al. 2015; Monastero and Pentyala 2017). Diagnostic biomarkers are often used in conjunction with other specific biomarkers to increase the specificity and diagnosis in the general population (Califf 2018).

3.2.2.3 Prognostic Cancer Biomarkers

Prognostic biomarkers allow to monitor the disease status, detect the recurrence rate, and provide an idea about the overall patient survival rate, independent of therapy (Sechidis et al. 2018; Ruberg and Shen 2015). It allows estimating the risk of the disease. In colon, lung, and breast cancer, patient's carcinoembryonic antigen (CEA) signifies a poor survival rate (Boonpipattanapong and Chewatanakornkul 2006; Su et al. 2012). Some diagnostic biomarkers such as Cancer antigen 19-9 (CA-19-9) and cancer antigen (CA 125) have also prognostic values and their presence can predict the survival rate in pancreatic ductal carcinoma and ovarian cancer, respectively (Poruk et al. 2013; Felder et al. 2014). Other prognostic biomarkers include miR-155 which suggests a poor clinical prognosis in hepatocellular carcinoma (Nalejska et al. 2014). A recent study reported five signature miRNAs having prognostic values in colon cancer (Lv et al. 2020). Further for breast cancer, circulating tumor cells serve as a prognostic indicator in nonmetastatic breast cancer as their presence is correlated with metastasis (Lucci et al. 2012).

3.2.2.4 Pharmacodynamics Cancer Biomarkers

This type of biomarkers is the new classification-based biomarker that determines the degree of the drug response (Sarker and Workman 2007). Pharmacodynamic biomarkers provide an idea about the interaction between a drug and its suspected target and whether the drug exerted a cellular response or not, and thereby guide treatment decision-making plans in real-time (Jackson 2012). For example, in nonsmall cell, lung cancer patient measurement of Mitogen-Activated Protein Kinase (MAPK) pathway inhibition receiving BRAF inhibitors can suggest a direct interaction between drug and target genes (Gainor et al. 2014). Further, Ki67 acts as a biomarker for cell proliferation, its expression after treatment with endocrine therapy serves as a pharmacodynamic response, and indicates target drug effects (Kelloff et al. 2005). Another example of pharmacodynamic biomarker example is monitoring the activity of PARP enzyme in white blood cells for the development of anticancer drug Olaparib (Dick et al. 2021).

3.2.3 Classification of Cancer Biomarkers Based on Other Criteria

3.2.3.1 Imaging Cancer Biomarkers

X-ray, Positron Emission Tomography (PET), Computed Tomography (CT), ultrasound, radionuclide imaging, and Magnetic Resonance Imaging (MRI) are the imaging techniques that are routinely used in clinical oncology for diagnosis, screening, and staging of cancer (Dregely et al. 2018; O'Connor et al. 2017). In oncology, imaging biomarkers are cost-effective noninvasive tools that easily allow identifying the disease state including assessment of the drug response. Several attempts have been made to do a regular assessment to reduce the risk of cancer development. For example, colonoscopy and mammography have been found to reduce the risk of developing colon cancer and breast cancer, respectively (Bischoff 2014).

3.2.3.2 Pathological Cancer Biomarkers

Various types of infectious agents such as viruses and bacteria constitute 15–20% of all human cancers (Srivastava et al. 2005; McLaughlin-Drubin and Munger 2008). The presence of pathogenic agents within the tumor cell makes them attractive pathogenic cancer biomarkers. The presence of HPV is associated with cervical cancer (Burd 2003). Further, Epstein Bair Virus (EBV) is closely associated with lymphoma and nasopharyngeal carcinoma (Pagano 1999). *Helicobacter pylori* is an established biomarker for gastric cancer (Wroblewski et al. 2010). Besides, various cancer pathogen detection methods have been revolutionized recently for rapid biomarker detection in the complex biological sample including Bioluminescence Resonance Energy Transfer (BRET), Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-based biosensors, and ELISA (Wu and Qu 2015).

3.3 Omics Approaches in Cancer Biomarker Research

"Omics" studies have been characterized by high-throughput technologies which help to investigate the genome, transcriptome, epigenome, proteome, and metabolome of cancer cells. These omics approaches facilitate the understanding of carcinogenesis at the molecular level. Figure 3.2 depicts a schematic representation of omics approaches that are commonly used to study the cellular behavior of cancer cells and some important biomarkers identified by using these omics approaches.

3.3.1 Genomics for Cancer Biomarkers

In oncology, numerous technologies such as Next-Generation Sequencing (NGS), Whole-Genome Sequencing (WGS), Comparative Genome Hybridization (CGH), and Fluorescence in situ Hybridization (FISH) have been widely used to analyze cancer-specific mutational changes (Hu et al. 2018; Zhao et al. 2019). Additionally, genomic studies primarily focus on the analysis of copy number variation and identification of chromosomal abnormality to characterize cancer cells at the molecular endpoint (Nogrady 2020). Further, the advances in sequencing technology helped the decision-making for personalized treatment strategy instead of based on

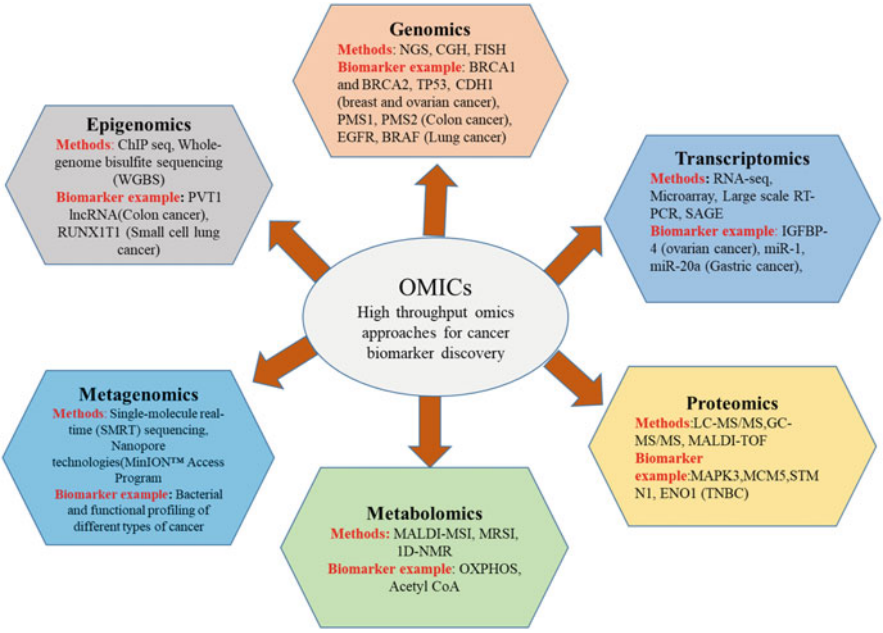


Fig. 3.2 Schematic representation of different omics approaches used for cancer biomarker discovery

cancer type. For example, FDA-approved biomarker EGFR mutation signifies the effectiveness of EGFR inhibitors like gefitinib (Tsimberidou et al. 2020). BRCA1 and BRCA2 mutations have been identified as hereditary markers for breast and ovarian cancer syndrome (Narod and Salmena 2011; Savanevich et al. 2021). Moreover, genomic profiling of pancreatic cancer identified centromere protein F as a novel therapeutic target and proved to be responsible for cancer progression (Chen et al. 2021).

3.3.2 Transcriptomics for Cancer Biomarkers

Transcriptomics studies are engaged in quantification, detection, and identification of altered mRNA, miRNA, and lncRNA in cancer cell populations (Chakraborty et al. 2018). Tools like RNA-seq, and microarray are commonly used to study the transcriptome. Expression Quantitative Trait Loci (eQTL) is a new approach for the analysis of functional variation sequence that leads to changes in gene expression (Hong et al. 2020; Geeleher et al. 2018). Various prognostic and predictive gene signatures have been identified in lung, breast, colon, and other tumor types (Vishnubalaji et al. 2019; Xiong et al. 2020; Sheng et al. 2019; Fang et al. 2021; Li et al. 2019). The transcriptomic study-based microarray technology has also been applied in precision oncology trials and aids in the clinical classification of breast cancer, colon cancer, and gastric cancer (Salem et al. 2017; Guinney et al. 2015; Lin et al. 2015). In breast cancer, transcriptomic data with bioinformatics study reveal BRIP1 as a noteworthy prognostic biomarker and its expression found to be correlated with various clinical features of breast cancer (Khan and Khan 2021). Furthermore, from RNA-seq data, a differential gene expression pattern has been revealed for different cancer tissue and their normal counterpart which will uncover the complex molecular pattern of cancer cells (Li et al. 2017). Different types of RNA serve as independent cancer biomarkers for instance in nonsmall cell lung cancer (NSCLC), the expression profiling of snoRNAs serves as an early diagnostic cancer biomarker (Liao et al. 2010). For renal cancer, hepatic cancer and glioblastoma piRNAs may serve as diagnostic and prognostic biomarkers (Busch et al. 2015; Liu et al. 2019; Rizzo et al. 2016). In addition, lncRNAs, such as XIST reported as a potential candidate biomarker for gastric cancer through transcriptomic study (Lu et al. 2017).

3.3.3 Proteomics for Cancer Biomarkers

Proteomic-based studies are considered as one of the innovative and dynamic high-throughput technology for determining the cellular function and location of main mediators of proteins (Olivier et al. 2019). The proteome of a defined entity be it a cell, an organelle, a tissue allows for better biomarker identification and to better understand the cancer surveillance mechanisms (Sallam 2015). Several studies reported that the proteomics approach can be used to identify the drug resistance

nature of cancer cells and treatment resistance biomarkers; for example through mass spectrometry, PYCR1 and ALDH18A1 expressions have been identified to be significantly associated with drug resistance in breast cancer (Shenoy et al. 2020). The drug resistance of cancer is associated with stemness and by applying the proteomics approach, new specific cancer biomarkers and therapeutic targets have been identified in the breast cancer stem cell population (Koh et al. 2020). Protein profiling of patients receiving immunotherapy is necessary to monitor the therapeutic response and thereby proteomic study can help to discover the potential prognostic biomarkers for cancer therapeutics (Chae et al. 2020; Harel et al. 2019).

3.3.4 Metabolomics for Cancer Biomarkers

Cancer affects intracellular metabolism and results in the inappropriate proliferation of cells (Vander Heiden and DeBerardinis 2017; Pavlova and Thompson 2016). Metabolomics is the study of altered metabolites that are produced by cellular processes mediated by proteins and thus it is a direct assessment of phenotype. Plasma or serum samples from patients are the major focus for metabolomic analysis of cancer cells. The methodologies that are used for metabolomic studies for biomarker detection include mass spectrometry and Nuclear Magnetic Resonance (NMR)-based imaging techniques, such as Magnetic Resonance Spectroscopic Imaging (MRSI) which can use both tissue/cell or biopsies samples for detection (Schmidt et al. 2021). Some putative metabolite biomarkers are altered carbohydrates in acute myeloid leukemia and unsaturated free fatty acids in colon cancer (Chaturvedi et al. 2013; Zhang et al. 2016). Other metabolite biomarkers include changes in citric acid, branched-chain amino acid for prostate cancer and pancreatic cancer (Giskeødegård et al. 2013; Mayers et al. 2014). Bladder cancer biomarkers detected from urinary metabolic profiling and 27 differentially metabolites have been detected (Li et al. 2021b).

3.3.5 Epigenomics for Cancer Biomarker

Epigenomics can be defined as the study of genome-wide chemical modification such as acetylation and methylation of DNA. The epigenetic modifications play an important role in uncovering the important genetic marker as these modifications regulate cellular interactions (Piunti and Shilatifard 2016). ChIP seq and Whole-Genome Bisulfite Sequencing (WGBS) are the two important powerful techniques for the identification of DNA-binding sites of transcription factors and to detect the methylated part in the sequences respectively (Chakraborty et al. 2018; Raj et al. 2017). MBD-isolated Genome Sequencing (MiGS) is another novel technique that allows the analysis of whole-genome sequencing patterns (Serre et al. 2010). For predicting the risk of head and neck cancer, DNA methylation in saliva has been found to be a potential epigenetic biomarker (Rapado-González et al. 2021). Chromatin immunoprecipitation (ChIP) studies demonstrated RUNX1T1 as an epigenetic

regulator of Small Cell Lung Cancer (SCLC) (He et al. 2021). A recent finding suggests that PD-L1 methylation in CpG loci can be considered as a valuable diagnostic biomarker for gastric cancer (Amini et al. 2021).

3.4 Bioinformatics Analytical Tools for Cancer Biomarker Discovery

The emerging high-throughput technologies result in the exponential growth of cancer biomarker data set from various resources. Thereby biologists face difficulty in extracting useful information from the available repositories as these contain various types of cancer-related information. The Cancer Genome Atlas (TCGA), a resource of multiomics cancer data platform that integrates genomics, epigenomics, and transcriptomics data of more than 30 human tumor types (Wang et al. 2016b). This aims to provide publicly available comprehensive atlas for molecular alteration in cancer cell. Pan-Cancer initiative is the new version of TCGA atlas and it is dedicated for comparison and analysis of molecular alteration found in different tumor types (Cancer Genome Atlas Research Network 2013). Recently published bioinformatics cancer-related database MarkerDB provides molecular cancer biomarker information along with the clinical significance such as diagnostic marker or prognostic marker (Wishart et al. 2021). Another database, OncoMX is a knowledge base; it integrates data for cancer mutation gene signatures, differential expression genes for cancer (Dingerdissen et al. 2020). Similarly, CIViCmine is another recently published database that provides list of curative clinically relevant cancer biomarkers information (Lever et al. 2019). Different machine learning and statistical approaches allow to the identification of biomolecules of interest from the large dataset with quantitative measurements. BioPlat is a software package for cancer biomarker discovery that allows high-throughput data filtering, gene expression calculation in silico (Butti et al. 2014). Another recently developed software Q omics that enables the analysis of patient survival, gene expression, and mutation of cancer-driven data set (Lee et al. 2021). There is an R-based tool available for cancer data analysis named CAncer bioMarker Prediction Pipeline (CAMPP), a standardized framework for the analysis of quantitative biological data (Terkelsen et al. 2020). Overall, several dedicated databases and tools are available for the storage and discovery of cancer biomarkers (Table 3.1).

3.5 Future Challenges

The future of biomarkers in oncology is potentially associated with the diagnostic, predictive, and prognostic cancer biomarkers. Despite the explosion of new technologies, a number of hurdles are associated with the identification of potential cancer biomarkers to be considered in clinical trials. The major challenges for cancer biomarker discovery can be considered at three levels (Henry and Hayes 2012; Teutsch et al. 2009). First is analytic validity which can be described as pre- and

Table 3.1 List of some important cancer-related databases and computational tools/software for cancer biomarkers storage and discovery

Name	Type	Description	URL
The Cancer Genome Atlas (TCGA) (Wang et al. 2016b)	Publically available database	TCGA database catalyze the high-throughput generated data characterization in the field of oncology. It integrates various bioinformatics and analytical tools such as TCPA which allows analyzing proteomics generated data. cBioportal for genomics data analysis and Funseq allow to annotate somatic variations	https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/using-tcga/tools
MarkerDB (Wishart et al. 2021)	An online database for cancer biomarkers	A bioinformatics database that contains four different molecular categories of cancer biomarkers information such chemical, protein, DNA, and karyotypes. Also, it provides information for diagnostic, predictive, prognostic, and exposure cancer biomarkers	https://markerdb.ca
OncoMX (Dingerdissen et al. 2020)	An online database for comparing cancer patient biomarker data in the context of a healthy person	A bioinformatics comparative tool that contains cancer-related mutation information along with differential gene expression data	http://data.oncomx.org
CIViCmine (Lever et al. 2019)	Clinically relevant cancer biomarkers database	CIViCmine is a cancer knowledgebase, provides literature-based information for cancer biomarkers. It will be helpful in precision oncology for identifying diagnostic and prognostic cancer biomarkers	http://bionlp.bcgsc.ca/civicmine/
BioPlat (Butti et al. 2014)	A bioinformatics software for cancer biomarker discovery	This software allows biologists to identify potential predictive and	http://www.cancergenomics.net

(continued)

Table 3.1 (continued)

Name	Type	Description	URL
		prognostic cancer biomarkers or gene signatures from high-throughput data. It offers various in silico biomarker validation and annotation tools	
Q-Omics (Lee et al. 2021)	A bioinformatics software for assisting in cancer research and therapeutics	This software integrates data of cancer mutation, gene expression, immune score, patient survival, and drug screening data from various bioinformatics resources including TCGA, GDSC, NCI, and DepMap databases. Thereby, simplifying the biomarker discovery process for cancer biomarkers	http://qomics.sookmyung.ac.kr
CAncer bioMarker prediction pipeline (CAMPP) (Terkelsen et al. 2020)	A bioinformatics software for high-throughput data analysis	This R-based pipeline allows users to normalize the obtained high-throughput cancer data. It performs various important functions for potential cancer biomarker discovery including differential expression/abundance analysis, correlation, and co-expression network analyses, survival analysis	https://github.com/ELELAB/CAncer-bioMarker-Prediction-Pipeline-CAMPP

postanalytical evaluation of biomarker detection assay. It determines the specificity and sensitivity of the technical aspects (Hayes 2015). The second is clinical validity, it determines the diagnostic accuracy of biomarkers by dividing the population of interest into two groups such as patient and reference group (Bossuyt 2010). The third is clinical utility, which relates to making a clinical decision with a high level of evidence to improve cancer treatment outcomes (Hayes 2021). The aim of cancer research is earlier cancer diagnosis and get better clinical outcomes for precision oncology. Nevertheless, multiomics approaches offer great advantages for translational cancer research over monogenic markers. The rapid development of omics

biomarkers increases the specificity of targeted therapeutic approach and leads to enhance predictive, preventive, and personalized medicine (PPPM) practice in clinical oncology (Lu and Zhan 2018).

References

Alves Martins BA et al (2019) Biomarkers in colorectal cancer: the role of translational proteomics research. Front Oncol 9:1284

Amini M, Hejazi M, Ghorban K, Mokhtarzadeh A, Baradaran B (2021) Identification of functional methylated CpG loci in PD-L1 promoter as the novel epigenetic biomarkers for primary gastric cancer. Gene 772:145376

An HJ et al (2006) Profiling of glycans in serum for the discovery of potential biomarkers for ovarian cancer. J Proteome Res 5:1626–1635

An Z et al (2010) Integrated ionization approach for RRLC-MS/MS-based metabonomics: finding potential biomarkers for lung cancer. J Proteome Res 9:4071–4081

Asleh K et al (2021) Proteomics-derived basal biomarker DNA-PKcs is associated with intrinsic subtype and long-term clinical outcomes in breast cancer. NPJ Breast Cancer 7:114

Bai R, Lv Z, Xu D, Cui J (2020) Predictive biomarkers for cancer immunotherapy with immune checkpoint inhibitors. Biomark Res 8:34

Bautista-Sánchez D et al (2020) The promising role of miR-21 as a cancer biomarker and its importance in RNA-based therapeutics. Mol Ther Nucleic Acids 20:409–420

Bischoff M (2014) Tumor markers, mammography and colon cancer screening. Prevention in oncology—what makes sense, what does not? MMW Fortschr Med 156:27

Boonpipattanapong T, Chewatanakornkul S (2006) Preoperative carcinoembryonic antigen and albumin in predicting survival in patients with colon and rectal carcinomas. J Clin Gastroenterol 40:592–595

Bossuyt PMM (2010) Clinical validity: defining biomarker performance. Scand J Clin Lab Investig 70:46–52

Boussios S et al (2019) The developing story of predictive biomarkers in colorectal cancer. J Pers Med 9:E12

Burd EM (2003) Human papillomavirus and cervical cancer. Clin Microbiol Rev 16:1–17

Busch J et al (2015) Piwi-interacting RNAs as novel prognostic markers in clear cell renal cell carcinomas. J Exp Clin Cancer Res 34:61

Butti MD et al (2014) BioPlat: a software for human cancer biomarker discovery. Bioinformatics 30:1782–1784

Califf RM (2018) Biomarker definitions and their applications. Exp Biol Med (Maywood) 243: 213–221

Campos-da-Paz M, Dórea JG, Galdino AS, Lacava ZGM, de Fatima Menezes Almeida Santos M (2018) Carcinoembryonic antigen (CEA) and hepatic metastasis in colorectal cancer: update on biomarker for clinical and biotechnological approaches. Recent Pat Biotechnol 12:269–279

Cancer Genome Atlas Research Network et al (2013) The cancer genome atlas pan-cancer analysis project. Nat Genet 45:1113–1120

Chae YK et al (2020) Mass spectrometry-based serum proteomic signature as a potential biomarker for survival in patients with non-small cell lung cancer receiving immunotherapy. Transl Lung Cancer Res 9:1015–1028

Chakraborty S, Hosen MI, Ahmed M, Shekhar HU (2018) Onco-multi-OMICS approach: a new frontier in cancer research. Biomed Res Int 2018:9836256

Chaturvedi A et al (2013) Mutant IDH1 promotes leukemogenesis in vivo and can be specifically targeted in human AML. Blood 122:2877–2887

Chen J et al (2017) Circular RNA profile identifies circPVT1 as a proliferative factor and prognostic marker in gastric cancer. Cancer Lett 388:208–219

Chen H et al (2021) Centromere protein F is identified as a novel therapeutic target by genomics profile and contributing to the progression of pancreatic cancer. Genomics 113:1087–1095
Chung C (2021) Predictive and prognostic biomarkers with therapeutic targets in colorectal cancer: a 2021 update on current development, evidence, and recommendation. J Oncol Pharm Pract 28(4):850–869. https://doi.org/10.1177/10781552211005525
Cózar JM, Hernández C, Miñana B, Morote J, Alvarez-Cubero MJ (2021) The role of prostate-specific antigen in light of new scientific evidence: an update in 2020. Actas Urol Esp (Engl Ed) 45:21–29
Daniotti JL, Vilcaes AA, Torres Demichelis V, Ruggiero FM, Rodriguez-Walker M (2013) Glycosylation of glycolipids in cancer: basis for development of novel therapeutic approaches. Front Oncol 3:306
Dick P, Jos H, B. (2021) Analytical validation of quantitative pharmacodynamic methods used in clinical cancer studies. Int Arch Clin Pharmacol 7:26
Dingerdissen HM et al (2020) OncoMX: a knowledgebase for exploring cancer biomarkers in the context of related cancer and healthy data. JCO Clin Cancer Inform 4:210–220
Drake RR et al (2017) MALDI mass spectrometry imaging of N-linked glycans in cancer tissues. Adv Cancer Res 134:85–116
Dregely I et al (2018) Imaging biomarkers in oncology: basics and application to MRI. J Magn Reson Imaging 48:13–26
Dube DH, Bertozzi CR (2005) Glycans in cancer and inflammation—potential for therapeutics and diagnostics. Nat Rev Drug Discov 4:477–488
Echle A et al (2021) Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer 124:686–696
Fang Z, Xu S, Xie Y, Yan W (2021) Identification of a prognostic gene signature of colon cancer using integrated bioinformatics analysis. World J Surg Oncol 19:13
Felder M et al (2014) MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. Mol Cancer 13:129
Gainor JF, Longo DL, Chabner BA (2014) Pharmacodynamic biomarkers: falling short of the mark? Clin Cancer Res 20:2587–2594
Geeleher P et al (2018) Cancer expression quantitative trait loci (eQTLs) can be determined from heterogeneous tumor gene expression data by modeling variation in tumor purity. Genome Biol 19:130
Giskeødegård GF et al (2013) Spermine and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness. PLoS One 8:e62375
Goossens N, Nakagawa S, Sun X, Hoshida Y (2015) Cancer biomarker discovery and validation. Transl Cancer Res 4:256–269
Goyal B et al (2021) Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer. Biochim Biophys Acta Rev Cancer 1875:188502
Guinney J et al (2015) The consensus molecular subtypes of colorectal cancer. Nat Med 21:1350–1356
Harel M et al (2019) Proteomics of melanoma response to immunotherapy reveals mitochondrial dependence. Cell 179:236–250.e18
Hayes DF (2015) Biomarker validation and testing. Mol Oncol 9:960–966
Hayes DF (2021) Defining clinical utility of tumor biomarker tests: a clinician's viewpoint. JCO 39:238–248
He T et al (2021) Identification of RUNX1T1 as a potential epigenetic modifier in small-cell lung cancer. Mol Oncol 15:195–209
Henry NL, Hayes DF (2012) Cancer biomarkers. Mol Oncol 6:140–146
Hong M et al (2020) RNA sequencing: new technologies and applications in cancer research. J Hematol Oncol 13:166
Housman G et al (2014) Drug resistance in cancer: an overview. Cancers (Basel) 6:1769–1792

Hu C, Dignam JJ (2019) Biomarker-driven oncology clinical trials: key design elements, types, features, and practical considerations. JCO Precis Oncol 3:1–12. https://doi.org/10.1200/PO.19.00086
Hu T et al (2018) Forward and reverse mutations in stages of cancer development. Hum Genom 12:40
Huang Y, Zhu H (2017) Protein array-based approaches for biomarker discovery in cancer. Genom Proteom Bioinform 15:73–81
Ileana Dumbrava E, Meric-Bernstam F, Yap TA (2018) Challenges with biomarkers in cancer drug discovery and development. Expert Opin Drug Discov 13:685–690
Issaq H, Veenstra T (2008) Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. BioTechniques 44:697–698
Jackson RC (2012) Pharmacodynamic modelling of biomarker data in oncology. ISRN Pharmacol 2012:590626
Jørgensen JT, Hersom M (2016) Companion diagnostics-a tool to improve pharmacotherapy. Ann Transl Med 4:482
Karachaliou N et al (2018) Interferon gamma, an important marker of response to immune checkpoint blockade in non-small cell lung cancer and melanoma patients. Ther Adv Med Oncol 10:1758834017749748
Kelloff GJ et al (2005) Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development. Clin Cancer Res 11:2785–2808
Khan U, Khan MS (2021) Prognostic value estimation of BRIP1 in breast cancer by exploiting transcriptomics data through bioinformatics approaches. Bioinform Biol Insights 15:11779322211055892
Kim YE et al (2021) SILAC-based quantitative proteomic analysis of oxaliplatin-resistant pancreatic cancer cells. Cancers 13:724
Koh E-Y, You J-E, Jung S-H, Kim P-H (2020) Biological functions and identification of novel biomarker expressed on the surface of breast cancer-derived cancer stem cells via proteomic analysis. Mol Cells 43:384–396
Kwon YW et al (2021) Application of proteomics in cancer: recent trends and approaches for biomarkers discovery. Front Med (Lausanne) 8:747333
Lan Y et al (2016) Serum glycoprotein-derived N- and O-linked glycans as cancer biomarkers. Am J Cancer Res 6:2390–2415
Lassere MN (2008) The biomarker-surrogacy evaluation schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Stat Methods Med Res 17:303–340
Lee J et al (2021) Q-omics: smart software for assisting oncology and cancer research. Mol Cells 44:843–850
Leney AC, El Atmioui D, Wu W, Ovaa H, Heck AJR (2017) Elucidating crosstalk mechanisms between phosphorylation and O-GlcNAcylation. Proc Natl Acad Sci USA 114:E7255–E7261
Lever J et al (2019) Text-mining clinically relevant cancer biomarkers for curation into the CIViC database. Genome Med 11:78
Li S, Han L (2019) Circular RNAs as promising biomarkers in cancer: detection, function, and beyond. Genome Med 11:15
Li M, Sun Q, Wang X (2017) Transcriptional landscape of human cancers. Oncotarget 8:34534–34551
Li H et al (2019) Transcriptomic analysis and identification of prognostic biomarkers in cholangiocarcinoma. Oncol Rep 42(5):1833–1842. https://doi.org/10.3892/or.2019.7318
Li L, Guan Y, Chen X, Yang J, Cheng Y (2021a) DNA repair pathways in cancer therapy and resistance. Front Pharmacol 11:629266
Li J et al (2021b) Bladder cancer biomarker screening based on non-targeted urine metabolomics. Int Urol Nephrol 54(1):23–29. https://doi.org/10.1007/s11255-021-03080-6

Liang B et al (2015) Circulating VEGF as a biomarker for diagnosis of ovarian cancer: a systematic review and a meta-analysis. Onco Targets Ther 8:1075–1082
Liao J et al (2010) Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. Mol Cancer 9:198
Lin X, Zhao Y, Song W, Zhang B (2015) Molecular classification and prediction in gastric cancer. Comput Struct Biotechnol J 13:448–458
Lin H-Y et al (2021) KLK3 SNP–SNP interactions for prediction of prostate cancer aggressiveness. Sci Rep 11:9264
Liu C (2011) The application of SELDI-TOF-MS in clinical diagnosis of cancers. J Biomed Biotechnol 2011:1–6
Liu Y et al (2019) The emerging role of the piRNA/piwi complex in cancer. Mol Cancer 18:123
Locke WJ et al (2019) DNA methylation cancer biomarkers: translation to the clinic. Front Genet 10:1150
Louie AD, Huntington K, Carlsen L, Zhou L, El-Deiry WS (2021) Integrating molecular biomarker inputs into development and use of clinical cancer therapeutics. Front Pharmacol 12:747194
Lu M, Zhan X (2018) The crucial role of multiomic approach in cancer research and clinically relevant outcomes. EPMA J 9:77–102
Lu Q et al (2017) Potential lncRNA diagnostic biomarkers for early gastric cancer. Mol Med Rep 16:9545–9552
Lucci A et al (2012) Circulating tumour cells in non-metastatic breast cancer: a prospective study. Lancet Oncol 13:688–695
Ludwig JA, Weinstein JN (2005) Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 5:845–856
Lv P et al (2019) Exploratory study on application of MALDI-TOF-MS to detect serum and urine peptides related to small cell lung carcinoma. Mol Med Rep 21(1):51–60. https://doi.org/10.3892/mmr.2019.10794
Lv Y, Duanmu J, Fu X, Li T, Jiang Q (2020) Identifying a new microRNA signature as a prognostic biomarker in colon cancer. PLoS One 15:e0228575
Mayers JR et al (2014) Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. Nat Med 20:1193–1198
McLaughlin-Drubin ME, Munger K (2008) Viruses associated with human cancer. Biochim Biophys Acta 1782:127–150
Menyhárt O, Győrffy B (2021) Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. Comput Struct Biotechnol J 19:949–960
Monastero RN, Pentyala S (2017) Cytokines as biomarkers and their respective clinical cutoff levels. Int J Inflamm 2017:4309485
Nalejska E, Mączyńska E, Lewandowska MA (2014) Prognostic and predictive biomarkers: tools in personalized oncology. Mol Diagn Ther 18:273–284
Narod SA, Salmena L (2011) BRCA1 and BRCA2 mutations and breast cancer. Discov Med 12:445–453
Nguyen TTH et al (2020) Salivary biomarkers in oral squamous cell carcinoma. JKAOMS 46:301–312
Nogrady B (2020) How cancer genomics is transforming diagnosis and treatment. Nature 579:S10–S11
O'Connor JPB et al (2017) Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 14:169–186
Olivier M, Asmis R, Hawkins GA, Howard TD, Cox LA (2019) The need for multi-omics biomarker signatures in precision medicine. Int J Mol Sci 20:E4781
Pagano JS (1999) Epstein-Barr virus: the first human tumor virus and its role in cancer. Proc Assoc Am Physicians 111:573–580
Pan B et al (2019) Identification of serum exosomal hsa-circ-0004771 as a novel diagnostic biomarker of colorectal cancer. Front Genet 10:1096

Paramasivam G (2021) Micro-RNA (miRNA): a biomarker to identify novel compounds in drug discovery and delivery for cancer therapy. Curr Drug Discov Technol 18:e130921188092
Parker LA et al (2018) Diagnostic biomarkers: are we moving from discovery to clinical application? Clin Chem 64:1657–1667
Parker JL et al (2021) Does biomarker use in oncology improve clinical trial failure risk? A large-scale analysis. Cancer Med 10:1955–1963
Pavlova NN, Thompson CB (2016) The emerging hallmarks of cancer metabolism. Cell Metab 23: 27–47
Peng M et al (2020) Circulating tumor DNA as a prognostic biomarker in localized non-small cell lung cancer. Front Oncol 10:561598
Piunti A, Shilatifard A (2016) Epigenetic balance of gene expression by Polycomb and COMPASS families. Science 352:aad9780
Poruk KE et al (2013) The clinical utility of CA 19-9 in pancreatic adenocarcinoma: diagnostic and prognostic updates. Curr Mol Med 13:340–351
Raj U, Aier I, Semwal R, Varadwaj PK (2017) Identification of novel dysregulated key genes in breast cancer through high-throughput ChIP-Seq data analysis. Sci Rep 7:3229
Rapado-González Ó et al (2021) Salivary DNA methylation as an epigenetic biomarker for head and neck cancer. Part II: a cancer risk meta-analysis. J Pers Med 11:606
Rizzo F et al (2016) Specific patterns of PIWI-interacting small noncoding RNA expression in dysplastic liver nodules and hepatocellular carcinoma. Oncotarget 7:54650–54661
Ruan X et al (2021) Genome-wide identification of m6A-associated functional SNPs as potential functional variants for thyroid cancer. Am J Cancer Res 11:5402–5414
Ruberg SJ, Shen L (2015) Personalized medicine: four perspectives of tailored medicine. Stat Biopharm Res 7:214–229
Salem H, Attiya G, El-Fishawy N (2017) Classification of human cancer diseases by gene expression profiles. Appl Soft Comput 50:124–134
Sallam RM (2015) Proteomics in cancer biomarkers discovery: challenges and applications. Dis Mark 2015:1–12
Sarker D, Workman P (2007) Pharmacodynamic biomarkers for molecular cancer therapeutics. Adv Cancer Res 96:213–268
Savanevich A, Ashuryk O, Cybulski C, Lubiński J, Gronwald J (2021) BRCA1 and BRCA2 mutations in ovarian cancer patients from Belarus: update. Hered Cancer Clin Pract 19:13
Schmidt DR et al (2021) Metabolomics in cancer research and emerging applications in clinical oncology. CA Cancer J Clin 71:333–358
Sechidis K et al (2018) Distinguishing prognostic and predictive biomarkers: an information theoretic approach. Bioinformatics 34:3365–3376
Serre D, Lee BH, Ting AH (2010) MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res 38: 391–399
Sheng M, Xie X, Wang J, Gu W (2019) A pathway-based strategy to identify biomarkers for lung cancer diagnosis and prognosis. Evol Bioinform 15:117693431983849
Sheng KL et al (2020) An integrated approach to biomarker discovery reveals gene signatures highly predictive of cancer progression. Sci Rep 10:21246
Shenoy A et al (2020) Proteomic patterns associated with response to breast cancer neoadjuvant treatment. Mol Syst Biol 16:e9443
Solomon A (1980) Monoclonal immunoglobulins as biomarkers of cancer. In: Sell S (ed) Cancer markers. Humana Press, Totowa, pp 57–87. https://doi.org/10.1007/978-1-4612-6117-9_3
Srivastava S, Verma M, Gopal-Srivastava R (2005) Proteomic maps of the cancer-associated infectious agents. J Proteome Res 4:1171–1180
Strimbu K, Tavel JA (2010) What are biomarkers? Curr Opin HIV AIDS 5:463–466
Su C-H et al (2012) The carcinoembryonic antigen as a potential prognostic marker for neuroendocrine carcinoma of the breast. Anticancer Res 32:183–188

Sung H et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J Clin 71:209–249
Swiatly A et al (2017) MALDI-TOF-MS analysis in discovery and identification of serum proteomic patterns of ovarian cancer. BMC Cancer 17:472
Tatarinov IS (1964) Detection of embryo-specific ALPHA-globulin in the blood serum of a patient with primary liver cancer. Vopr Med Khim 10:90–91
Terkelsen T, Krogh A, Papaleo E (2020) CAncer bioMarker prediction pipeline (CAMPP)—a standardized framework for the analysis of quantitative biological data. PLoS Comput Biol 16: e1007665
Teutsch SM et al (2009) The evaluation of genomic applications in practice and prevention (EGAPP) initiative: methods of the EGAPP working group. Genet Med 11:3–14
Torgovnick A, Schumacher B (2015) DNA repair mechanisms in cancer development and therapy. Front Genet 6:157
Tsimberidou AM, Fountzilas E, Nikanjam M, Kurzrock R (2020) Review of precision cancer medicine: evolution of the treatment paradigm. Cancer Treat Rev 86:102019
Vander Heiden MG, DeBerardinis RJ (2017) Understanding the intersections between metabolism and cancer biology. Cell 168:657–669
Vishnubalaji R, Sasidharan Nair V, Ouararhni K, Elkord E, Alajez NM (2019) Integrated transcriptome and pathway analyses revealed multiple activated pathways in breast cancer. Front Oncol 9:910
Wang MC, Valenzuela LA, Murphy GP, Chu TM (1979) Purification of a human prostate specific antigen. Investig Urol 17:159–163
Wang C-C et al (2008) Glycan microarray of Globo H and related structures for quantitative analysis of breast cancer. Proc Natl Acad Sci USA 105:11661–11666
Wang H et al (2016a) The clinical impact of recent advances in LC-MS for cancer biomarker discovery and verification. Expert Rev Proteom 13:99–114
Wang Z, Jensen MA, Zenklusen JC (2016b) A practical guide to the cancer genome atlas (TCGA). Methods Mol Biol 1418:111–141
Wang X et al (2018) SILAC-based quantitative MS approach for real-time recording protein-mediated cell-cell interactions. Sci Rep 8:8441
Wang J et al (2021a) miRNA-194 predicts favorable prognosis in gastric cancer and inhibits gastric cancer cell growth by targeting CCND1. FEBS Open Bio 11:1814–1826
Wang H et al (2021b) Circular RNA TMEM87A promotes cell proliferation and metastasis of gastric cancer by elevating ULK1 via sponging miR-142-5p. J Gastroenterol 56:125–138
Welch HG, Albertsen PC (2009) Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986–2005. J Natl Cancer Inst 101:1325–1329
Wishart DS et al (2021) MarkerDB: an online database of molecular biomarkers. Nucleic Acids Res 49:D1259–D1267
World Health Organization (2001) Biomarkers in risk assessment: validity and validation. World Health Organization, Geneva
Wroblewski LE, Peek RM, Wilson KT (2010) *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. Clin Microbiol Rev 23:713–739
Wu L, Qu X (2015) Cancer biomarker detection: recent achievements and challenges. Chem Soc Rev 44:2963–2997
Wu C-S et al (2012) Cancer-associated carbohydrate antigens as potential biomarkers for hepato-cellular carcinoma. PLoS One 7:e39466
Wurdinger T, In't Veld SGJG, Best MG (2020) Platelet RNA as pan-tumor biomarker for cancer detection. Cancer Res 80:1371–1373
Xi X et al (2017) RNA biomarkers: frontier of precision medicine for cancer. Noncoding RNA 3:E9
Xia C, Yang F, He Z, Cai Y (2020) iTRAQ-based quantitative proteomic analysis of the inhibition of cervical cancer cell invasion and migration by metformin. Biomed Pharmacother 123:109762
Xiong Y, Feng Y, Qiao T, Han Y (2020) Identifying prognostic biomarkers of non-small cell lung cancer by transcriptome analysis. Cancer Biomark 27:243–250

Xu Y, Guo Q, Wei L (2021) The emerging influences of alpha-fetoprotein in the tumorigenesis and progression of hepatocellular carcinoma. Cancers (Basel) 13:5096

Yamashita T, Honda M, Kaneko S (2008) Application of serial analysis of gene expression in cancer research. Curr Pharm Biotechnol 9:375–382

Yan W, Xue W, Chen J, Hu G (2016) Biological networks for cancer candidate biomarkers discovery. Cancer Inform 15:1–7

Zare Jeddi M et al (2021) Towards a systematic use of effect biomarkers in population and occupational biomonitoring. Environ Int 146:106257

Zeidan BA et al (2009) SELDI-TOF MS proteomics in breast cancer. Clin Proteom 5:133–147

Zhang Y et al (2016) Serum unsaturated free fatty acids: a potential biomarker panel for early-stage detection of colorectal cancer. J Cancer 7:477–483

Zhang X, Sun X-F, Shen B, Zhang H (2019) Potential applications of DNA, RNA and protein biomarkers in diagnosis, therapy and prognosis for colorectal cancer: a study from databases to AI-assisted verification. Cancers (Basel) 11:E172

Zhao EY, Jones M, Jones SJM (2019) Whole-genome sequencing in cancer. Cold Spring Harb Perspect Med 9:a034579

Zheng Y et al (2021) Combined metabolomics with transcriptomics reveals important serum biomarkers correlated with lung cancer proliferation through a calcium signaling pathway. J Proteome Res 20:3444–3454

Zhu L et al (2020) Pan-cancer analysis of the mitophagy-related protein PINK1 as a biomarker for the immunological and prognostic role. Front Oncol 10:569887

4 The Biology and Chemistry of Microsomal Prostaglandin E Synthase (mPGES) - I Inhibitors for Cancer Biomedicine

Kathiravan M K, Swethan Babu, Chitra Vellapandian, Vigneshwaran Namasivayam, and Senthilkumar Palaniappan

Abstract

According to WHO, Cancer is the leading cause of death worldwide, accounting for nearly 10 million deaths in 2020. The predicted global cancer burden is expected to be 28.4 million cases in 2040. Recent scientific evidences suggests that the overexpression of COX-2 and mPGES-1 through COX/mPGES-1/PGE2 pathway in cancer has resulted in a decreased survival rate. The currently available COX inhibitors reduces the production of PGE2 along with other prostanoids which are required for basic cellular functions. Due to the severe side effects of COX-2 inhibitors especially on the gastrointestinal and cardiovascular systems, mPGES-1 could be a better and safe target. The selective mPGES-1 inhibitors exploited in inflammation and multiple tumor types have opened new avenues and are emerging as a new therapeutic approach. Till date, a variety of

M. K. Kathiravan and Swethan Babu contributed equally with all other contributors.

K. M K · S. Babu
Department of Pharmaceutical Chemistry, SRM College of Pharmacy, SRM IST Kattankulathur, Kattankulathur, Tamil Nadu, India

209, Dr APJ Abdul Kalam Research Lab, SRM College of Pharmacy, SRM IST Kattankulathur, Kattankulathur, Tamil Nadu, India

C. Vellapandian
Department of Pharmacology, SRM College of Pharmacy, SRM IST Kattankulathur, Kattankulathur, Tamil Nadu, India

V. Namasivayam
Pharmaceutical Chemistry, Pharmaceutical Institute, University of Bonn, Bonn, Germany
e-mail: vnamasiv@uni-bonn.de

S. Palaniappan (✉)
Faculty of Pharmacy, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
e-mail: drsenthilkumar.p@kahedu.edu.in

S. Singh (ed.), *Systems Biomedicine Approaches in Cancer Research*,
https://doi.org/10.1007/978-981-19-1953-4_4

chemically diversified synthetic scaffolds such as imidazole's, substituted ureas, derivatives of amides and acids along with natural product inhibitors have been reported as effective and selective mPGES-1 inhibitors.

4.1 Introduction

Cancer is defined as an abnormal and uncontrolled cell proliferation which can spread throughout the body. According to WHO, cancer is the second leading cause of death worldwide with one in six patients die due to cancer and accounts for 10 million deaths in 2020 alone. Breast cancer, lung cancer, prostate cancer, colon cancer, and nonHodgkin lymphoma are the most common types of cancers (World Health Organization (WHO) n.d.). By the end of 2022 in United States alone, 1.9 million new cases of cancer might occur and lead to around 0.6 million deaths. Further study suggests that the global cancer burden is expected to be 28.4 million cases in 2040 with an increase of 47%. (World Health Organization (WHO) 2018). More populations are vulnerable to develop cancer, due to the dramatic change in lifestyle (e.g., tobacco and alcohol consumption, stress, obesity, improper diet, lack of physical activity etc.), sociocultural, and environment. The treatment option is radical radiotherapy for an early stage of cancer whereas most traditional treatment of chemotherapy is for advanced stages. This is currently the best effective treatment as most of the chemotherapeutic agents travel throughout the body and kills cancer cells thereby inhibiting the spread of cancer to other parts of the body (World Health Organization (WHO) n.d.). The prevention and treatment for different types of cancers remain one of the greatest challenges in the current medical treatment. Among the different strategies, chemotherapy based on systemic administration of a single or a combination of drugs remains the major therapeutic approach for cancer treatment currently (American Cancer Society 2015).

Chemotherapy has several drawbacks of which most pronounced is its long-term side effects that include early menopause, cardiac problems, respiratory problems, numbness, bone, and joint problems. Apart from killing the cancer cells, the chemotherapeutic agents can also damage the surrounding healthy tissues causing anemia and extreme fatigue. An inflammation related to cancer makes it more complicated due to the interconnection between cancer cells and stromal cells (World Health Organization (WHO) n.d.). Several studies suggest that high-risk tumors possess pro-inflammatory features in an immunosuppressive microenvironment of neuroblastoma (Hanahan and Weinberg 2011) with acute as well as chronic inflammation promoting the progression of cancer and metastasis (Aggarwal 2004). Inflammation is a proven attribute of cancer, the role of Prostaglandin E2 (PGE2), an inflammatory and oncogenic lipid mediator is coupled with various biochemical pathways leading to increased tumor cell proliferation, angiogenesis, and immunosuppression.

Prostaglandins (PGs) is synthesized from arachidonic acid, (a fatty acid obtained from phospholipid bilayer as a result of phospholipase A2 (PLA2)), plays a significant role in response to an inflammation. Oxidation of Arachidonic acid by Prostaglandin G/H synthase 1 & Prostaglanding G/H synthase 2 leads to the formation of Prostaglandin G2 and Prostaglandin PTGS2 respectively. PGH_2 is then transformed to Thromboxane A2 (TXA2), PGE2, PGF2α, PGI2, and PGD2 by either cytosolic prostaglandin E Synthase (cPGES), microsomal PGES (mPGES)- 1 or -2 (Ricciotti and FitzGerald 2011). mPGES-1 along with COX-2, induced by pro-inflammatory stimuli, increases PGE2 levels (Murakami et al. 2002; Samuelsson et al. 2007; Tai 2011). Prostaglandin E2 (PGE2) is a multifaceted bioactive lipid mediator of inflammation and cancer progression. PGE2 is mediated through four G-PCR (E prostanoid GPCR), namely EP1 (Gq), EP2 (Gs), EP3 (Gi), and EP4 (Gs) GPCR, referred to as EP receptors. The EP receptor 1-4 coupled to heterotrimeric G protein GαS and Gαi modulates the level of calcium, cAMP, and IP3, thereby activating divergent signaling pathways (O'Callaghan et al. 2015). The interlinkage between PGE2 and EP receptors depends on the nature of cell, tissue type, and its location. The EP receptors influence cell response to PGE2 in cancer cells. The activation of the EP receptor leads to EP1-dependent tumor cell migration and invasion. The EP2-induced angiogenesis along with the suppression of antitumor immune response is followed by EP4-related tumor cell migration and metastasis (Fig. 4.1). However, the characteristic features of EP3 are still not clear. mPGES-2 and cPGES are basically expressed and form the basis for the production of PGE2. The expression of mPGES-1 is relatively low in most tissues but in response to acute and chronic inflammatory stimuli, mPGES-1 is upregulated and couples with COX-2 to mediate inflammatory PGE2 production (Ma and Brusselaers 2018). The human mPGES-1 gene is restricted to chromosome 9q34.3 with three exons and spans possessing 152 amino acid residues of which 80% are similar to the enzymes in mouse and rat (Veettil et al. 2017). Selective inhibitors of COX-2 posseses various side effects including hypertension, edeme formation and congestive heart failure upon long term usage. Therefore, mPGES-1 can serve as a therapeutic target for inflammatory and other related disorders (Donnini et al. 2014).

4.2 Role of mPGES-1 in Cancer

The increased expression of COX-2 and mPGES-1 through COX/mPGES-1/PGE2 pathway in cancer has resulted in decreased survival rate (de Groot et al. 2007; Seo et al. 2009; Kim et al. 2019; Larsson et al. 2015). Studies on PGE2, use of EP antagonist, genetic deletion, and pharmacological inhibition of COX or mPGES-1 have suggested that PGE2 has a pro-tumorigenic role (Nakanishi and Rosenberg 2013). PGE2 increases the proliferation of cancer cells which promotes tumor-favoring M2 polarization of tumor-associated macrophages (TAMs). Further, it attracts immunosuppressive myeloid-derived suppressor cells and enhances the immunity inhibitory function of regulatory T cells. This leads to decreased amount and maturation of infiltrating antigen-presenting dendritic cells, inhibits antitumor

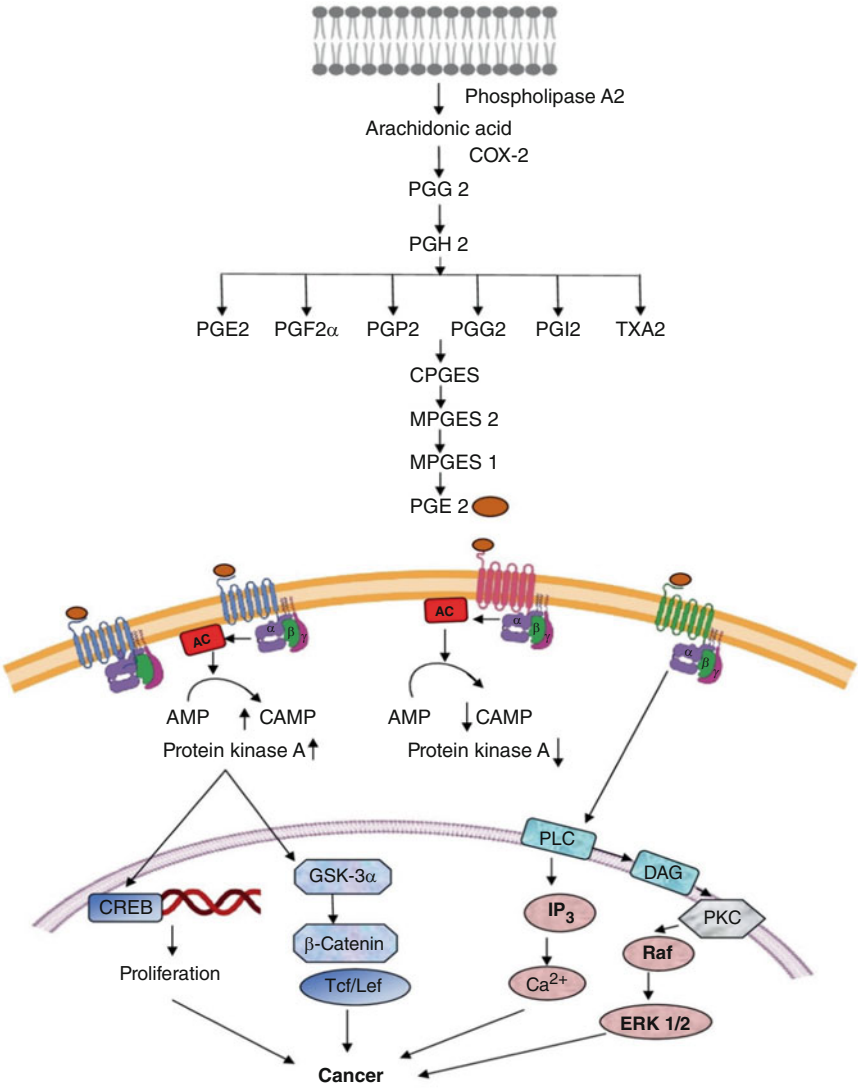


Fig. 4.1 Role of mPGES-1 pathway in cancer

activity of NK cells, cytotoxic T cells, and promotes inflammatory functions of Th17 cells (Nakanishi and Rosenberg 2013; Kalinski 2012). Moreover studies have proved that mice lacking mPGES-1 have slow growth of tumor when compared to transgenic mice with overexpression of COX-2 and mPGES-1 developed tumors in the gastric tract (Pierce et al. 1999).

The current clinical usage of COX inhibitors such as nonsteroidal anti-inflammatory drugs (NSAIDs) and COX-2 inhibitors (Coxibs) generally reduces

the production of PGE2 along with other prostanoids which are required for basic cellular functions. Due to the existing drawback with severe side effects of COX inhibitors especially on the gastrointestinal and cardiovascular systems, targeting mPGES-1 could be better and safe (McGraw et al. 2006). Inhibiting the production of PGE2 without disturbing other prostaglandins including PGI2 can be achieved by targeting the terminal mPGES-1 which cannot be achieved in COX-1/2 inhibition. Prostaglandin E2 (PGE2)-driven inflammation promotes tumor growth, immune suppression, angiogenesis, and resistance to established cancer therapies (Wilson et al. 2007).

EP2 subtype promotes the invasion of the tumor, metastasis, and further literature suggests that EP2 receptor activation by PGE2 predominantly promotes hepatocellular carcinoma invasion of cells (Cheng et al. 2014; Hsu et al. 2017a). The PI3K signaling pathway has a beneficial role in the regulation of cell proliferation, differentiation, trafficking, and migration (Sobhani et al. 2018) (Fig. 4.2). The PI3K/Akt cell survival is increased by EP2 and EP4 activations (Ma and St-Jacques 2018; Regan 2003), thus leading to upregulation of matrix metalloproteinases observed in several cancer types and regulating various types of therapies (Hsu et al. 2017b). EP2 receptor contributes more in breast cancer related to the metabolism and hence alters the growth factor (Allison et al. 2015). Nevertheless, several events like tumorigenesis, genetic and epigenetic begin converting from TGF-β from a tumor suppressor to a promoter of cell growth, metastasis, and invasion (Tian and Schiemann 2010). The attribution to the altered response of TGF-β in the suppression of TGF-β-induced Smad2/3 nuclear localization and signaling by PGE2 is followed by uncoupling TGF-β from activating Smad3. Additionally, EP2 plays a major role in regulating metastasis by downregulation of solute carrier family 19 member 3 in triple-negative breast cancer (Cheuk et al. 2015). Ablation of EP2 suppresses the tumor development of skin by limiting angiogenesis, promoting apoptosis (Kim et al. 2016; Rundhaug et al. 2011; Rundhaug and Fischer 2008), EP2 accelerates invasion of prostate tumor cells controlled by EP2 antagonist TG4-155 (Singh et al. 2011). The upregulation in the expression of EP2 is found in tumors of laryngeal carcinoma which is detected by a deeper invasion of submucosa or cartilage (Rogers et al. 1999; Otsnu 2013; Koeberle and Werz 2015; Nakanishi et al. 2010).

4.2.1 Binding Mechanism

PGE2 binding to GPCR leads to activation of trimeric G-protein (α, β, γ). The α and γ subunits of G proteins are anchored to membrane lipids covalently (Voss et al. 1993). The β subunit is bonded with γ subunit, forming heterodimeric structure $G_{\beta,\gamma}$. Activated receptor causes conformational changes in G_{α}, triggering dissociation of Guanosine diphosphate (GDP) (Kubota and Wakamatsu 2008). This results from the exchange of GDP with Guanosine triphosphate (GTP) causing the α subunit to dissociate from β and γ. Gs (EP2 and EP4) mediates signal transduction, Gs_{α} subunit binds and hydrolyses GTP and binds to plasma membrane-bound adenylyl cyclase

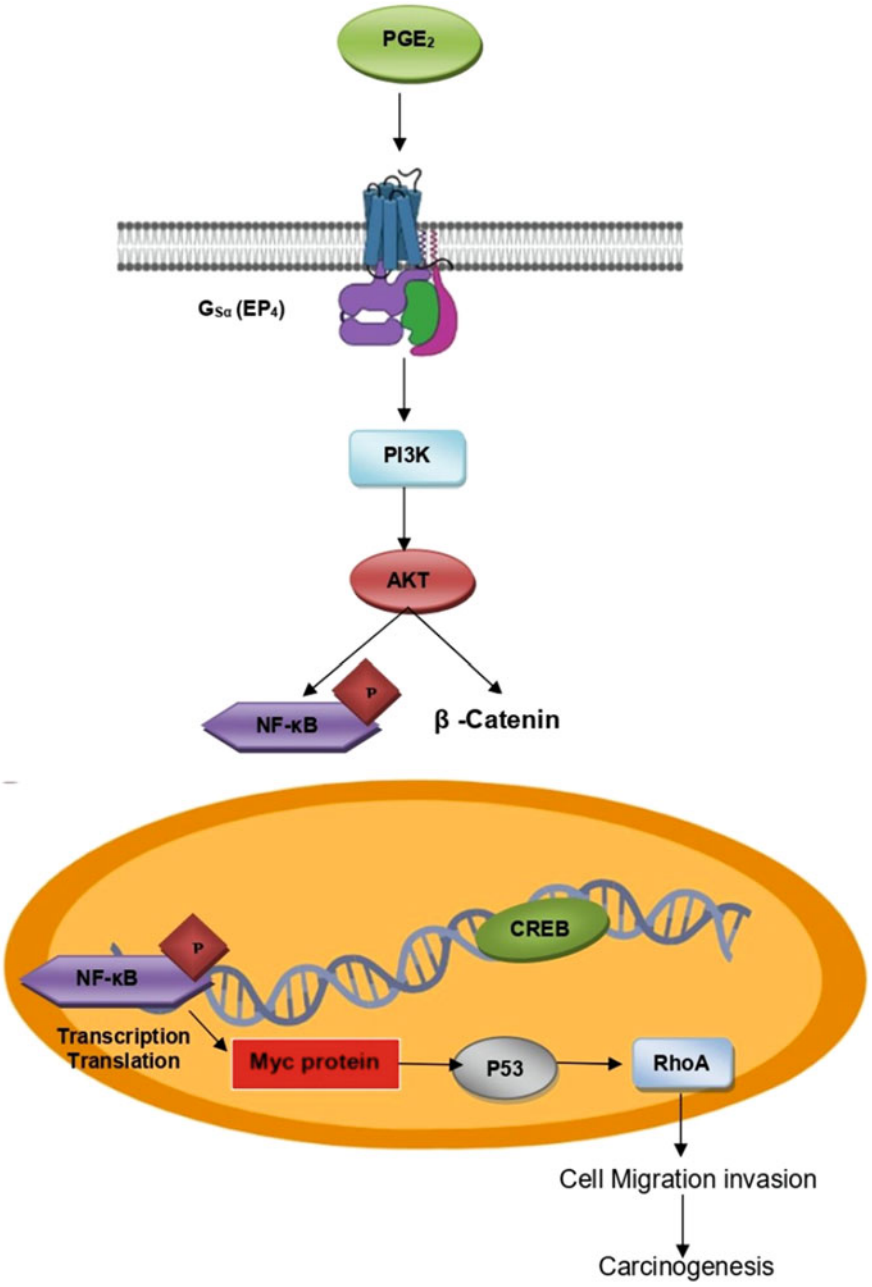


Fig. 4.2 Role of AKT-β catenin pathway in cancer development

(AC) (Morou and Georgoussi 2005). Activated AC catalyzes the synthesis of cAMP from ATP which activates protein kinase A (PKA/ cAMP-dependent protein kinase). cAMP binds to the regulatory subunits and alters its conformation causing dissociation of catalytic and regulatory subunits (Mochocki et al. 2015). The freed catalytic subunits get translocated into the nucleus and bind to transcription factor camp responsive element-binding protein, activated by phosphorylation. Once the phosphorylated cAMP response element-binding protein (CREB) recruits transcriptional co-activator CBP (CREB-binding protein) in stimulating the transcription of the gene for several proteins that plays a crucial role in the progression of cancer (Xiao et al. 2010). The second pathway that enhances the proliferation of cells and neoplasia is the β-catenin pathway (Dorsam and Gutkind 2007). Glycogen synthase kinase 3 β is phosphorylated and inactivated by protein kinase A and on the contrary β-catenin has been activated through the protein kinase phosphorylation (Rhee 2001). Upon activation of β-catenin by phosphorylation, it stimulates transcription of genes from various proteins which promotes proliferation of cell and cancer by binding to the transcription factor (Lef/TCF superfamily) in the DNA (Hino et al. 2005). The activation of inhibitory G protein (Gi) leads to inhibition of AC by α subunit leading to a lower level of cAMP and exerting the effect opposite to that of stimulatory G protein (Gs) EP1-mediated Signal (Gq). The EP1-mediated signal is exerted by Gq sub protein which initiates its effect on membrane-bound enzyme phospholipase-C-β and cleaves phosphatidyl-4,5-bisphosphate (Ng et al. 1999). After the cleavage of PIP_2, secondary messengers like Inositol triphosphate (IP_3) and Diacylglycerol (DAG) are released where IP_3 being the minor. The released IP_3 gets released into the cytosol and acts on the sarcoplasmic reticulum and generates calcium ion release leading to enhanced intracellular calcium levels. Simultaneously, DAG pairs with the released calcium and activates protein kinase C involved in regulation of cancer (Masur et al. 2001) (Fig. 4.2).

4.2.2 Crystal Structures of mPGES

Three different subtypes of PGES are available of which mPGES-1 is one of the potential targets for pain. Currently, there are 17 crystal structures reported for mPGES-1 in the Protein Data Bank (Table 4.1).

4.3 Small Molecule Inhibitors

The selective mPGES-1 inhibitors exploited in inflammation and multiple tumor types provides new avenues and emerge as a new therapeutic approach. Till date, a variety of chemically diversified synthetic scaffolds such as imidazoles, substituted urea, derivatives of amides and acids have been discovered as effective and selective mPGES-1 inhibitors. Apart from the synthetic inhibitor, some of the natural products from plants possessing anti-inflammatory activity and exhibiting mPGES-1 inhibition such as myrtucommulone A, hyperforin, arzanol, epigallocatechin-3-gallate

Table 4.1 List of crystal structures published for mPGES (https://www.rcsb.org/)

PDB ID	Protein	Species	Release year	Scaffold	Ligand
2PBJ	mPGES	Monkey	2008	–	Glutathione-heme
4YL3	mPGES-1	Human	2015	Imidazopyridine	5-[4-bromo-2-(2-chloro-6-fluorophenyl)-1H-imidazol-5-yl]-2-{[4-(trifluoromethyl)phenyl]ethynyl}pyridine
4YL1	mPGES-1	Human	2015	Indole 2-carboxylic acid	5-(4-tert-butylphenyl)-1-[4-(propan-2-yloxy)phenyl]-1H-indole-2-carboxylic acid
4YL0	mPGES-1	Human	2015	Phenanthrene imidazole	2-(9-chloro-1H-phenanthro[9,10-d]imidazol-2-yl)benzene-1,3-dicarbonitrile
4YK5	mPGES-1	Human	2015	Indole propionic acid	3-[1-(4-chlorobenzyl)-5-(2-fluoro-2′-methylbiphenyl-4-yl)-3-methyl-1H-indol-2-yl]-2,2-dimethylpropanoic acid
1Z9H	mPGES-2	Monkey	2005	–	–
6VL4	mPGES-1	Human	2020	Phenylacetic acid	(2R)-cyclopentyl{4-[(quinolin-2-yl)methoxy]phenyl}acetic acid
5BQI	mPGES-1	Human	2016	Pyridine 3-carboxmide	2-(difluoromethyl)-5-{[(2-methylpropanoyl)amino]methyl}-N-{5-methyl-4-[4-(trifluoromethyl)phenyl]-1H-imidazol-2-yl}pyridine-3-carboxamide
5BQH	mPGES-1	Human	2016	Benzamide	N-[4-(4-chlorophenyl)-1H-imidazol-2-yl]-2-(difluoromethyl)-5-{[(2-methylpropanoyl)amino]methyl}benzamide
5BQG	mPGES-1	Human	2016	Benzamide	2-chloro-N-(4-phenyl-1,3-thiazol-2-yl)benzamide
4AL1	mPGES-1	Human	2013	–	Analog of glutathione
4AL0	mPGES-1	Human	2013	–	Glutathione
5T37	mPGES-1	Human	2017	Benzamide	2-chloro-5-{[(2,2-dimethylpropanoyl)amino]methyl}-N-(1H-imidazol-2-yl)benzamide
5T36	mPGES-1	Human	2017	2-Amino benzoic acid	4-chloro-2-[({(1S,2S)-2-[(2,2-imethylpropanoyl)amino]cyclopentyl}methyl)amino]benzoic acid
5K0I	mPGES-1	Human	2016	Quinoline	1,5-anhydro-2,3,4-trideoxy-3-{[(4S)-3,3-dimethyl-1-(8-methylquinolin-2-yl)piperidine-4-carbonyl]amino}-D-erythro-hexitol
5TL9	mPGES-1	Human	2017	Benzoic acid	

					2-{2-[(1S,2S)-2-{[1-(8-methylquinolin-2-yl)piperidine-4-carbonyl]amino}cyclopentyl]ethyl}benzoic acid
4WAB	Fused mPGES-1 with LTC4S	Human	2014	Benzimidazole	2-[[2,6-bis(chloranyl)-3-[(2,2-dimethylpropanoylamino)methyl]phenyl]amino]-1-methyl-6-(2-methyl-2-oxidanyl-propoxy)-N-[2,2,2-tris(fluoranyl)ethyl]benzimidazole-5-carboxamide
4BPM	Fused mPGES-1 with LTC4S	Human	2014	Benimidazole	2-[[2,6-bis(chloranyl)-3-[(2,2-dimethylpropanoylamino)methyl]phenyl]amino]-1-methyl-6-(2-methyl-2-oxidanyl-propoxy)-N-[2,2,2-tris(fluoranyl)ethyl]benzimidazole-5-carboxamide

(EGCG), curcumin, carnosol, carnosic acid, tetra- or pentacyclic triterpene acids are also discussed. The advantages of such small bioactive molecules reported by research groups and pharmaceutical industries worldwide as mPGES-1 inhibitors are summarized (Bergqvist et al. 2019; Waltenberger et al. 2011; Hamza et al. 2011; Jin et al. 2015; Koeberle et al. 2016).

4.3.1 Imidazoles

4.3.1.1 Phenanthrene Imidazoles

Through high-throughput screening (HTS) campaign, a series of phenanthrene imidazole derivatives as mPGES-1 inhibitors were discovered by Merck Frosst et al. in 2006. Among the series, four compounds showed good IC_{50} values at micro and nanomolar concentrations. Compounds **1** and **2** (Merk Frosst Canada Ltd 2006, 2007) with 2,6-dicyano phenyl ring at the second position of the imidazole showed IC_{50} 0.7 nM and 0.9 nM respectively, in a cell-free assay. However, compound **3** with chloro-substitution at sixth position (Cote et al. 2007a) has shown good potency, selectivity, and was orally active with an IC_{50} value of 0.42 μM in A549 whole cell assay and IC_{50} of 1.3 μM in human whole blood assay. In 2009, Giroux et al. through SAR studies identified that a para-fluoro substitution on the biscyanophenyl ring led to phenanthrene imidazole compound **4** (Giroux et al. 2009) with good in vivo efficacy in the LPS-induced hyperalgesia guinea pig model (ED_{50} 14 mg/kg). The compound **4** (Fig. 4.3) had half-life of 2.3 h in rat, higher degree of metabolism in rat (32%) and human (19%) hepatocytes, quicker absorption, acceptable bioavailability (68%). Based on the results of compound **3** and **4**, it can be concluded that phenanthrene-fused imidazole lead needs further optimization to develop as a potent and selective inhibitor against mPEGS-1 (Cote et al. 2007b; Fei and Zhou 2014).

4.3.1.2 2,4-Biarylimidazoles

Wu et al. through HTS technique found 34 novel biarylimidazole derivatives with a wide range of substitutions as mPGES-1 inhibitors. Among the compounds, compound **5** (Wu et al. 2010) showed moderate activity with IC_{50} 600 nM in human mPGES-1 enzymatic assay with an IC_{50} of 3100 nm and 7000 nm respectively (IL-1β stimulated A549 epithelial lung carcinoma cells with 2% and 50% FBS). The SAR analysis of biarylimidazole suggests that compounds with substitution like mono-ortho-cyanophenyl and bis-ortho-cyanofluoro phenyl at second position of imidazole as well as modifications on the central imidazole ring reduces the binding affinity against mPGES-1. However, substitution at the fourth position of imidazole showed moderate inhibitory activity for four compounds at micromolar concentrations. Subsequent derivatization with a triple bond extended from the 4-pyridyl group and nonpolar substituents on the cyclohexyl alkynes resulted in good mPGES-1 inhibition such as in compound **6** (Wu et al. 2010), with IC_{50} 33 nM in human mPGES-1 and IC_{50} 620 nM in A549 whole cell assay, respectively. Further, the cyclohexenyl alkyne compound **7** (Wu et al. 2010) showed IC_{50}

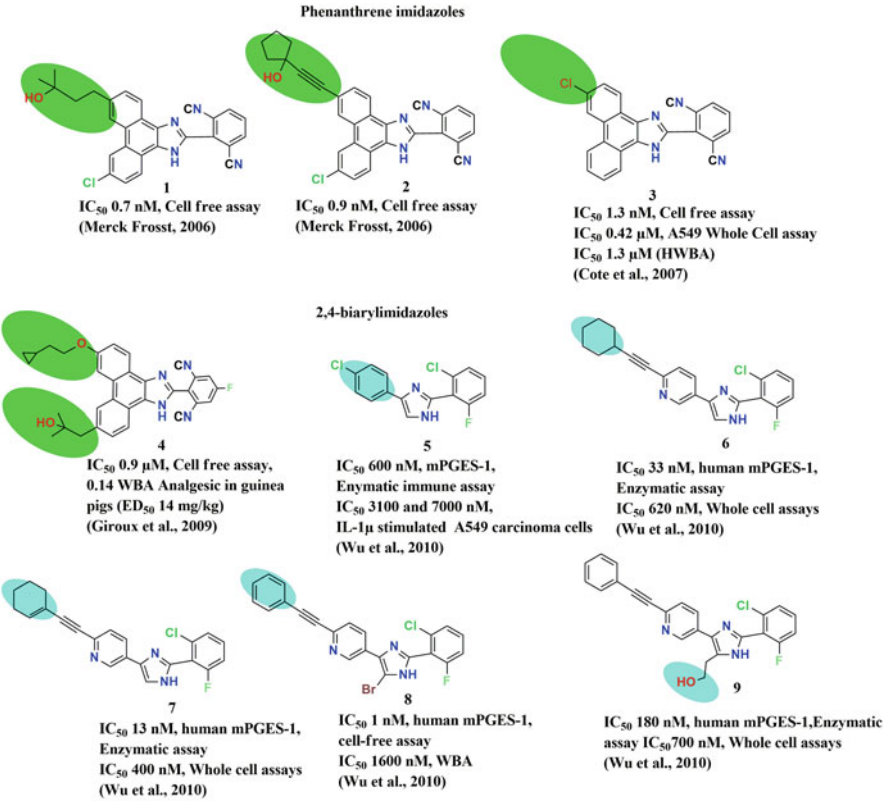


Fig. 4.3 Phenanthrene imidazole and benzimidazole as inhibitors of mPGES-1

13 nM in human mPGES-1 enzymatic assay and IC_{50} 400 nM in A549 whole cell assay. The presence of phenyl alkyne at fourth position and electronegative halogen such as bromine at fifth position of imidazole ring increased the potency in compound **8** (Wu et al. 2010) with an IC_{50} of 1 nM in human mPGES-1 cell-free assay and whole blood assay exhibited IC_{50} 1600 nM. Compound **9** (Wu et al. 2010) with ethyl hydroxyl group at fifth position showed a good inhibitory response in human mPGES-1 enzymatic assay (IC_{50} 180 nM) and in A549 whole cell assay (IC_{50} 700 nM) (Fig. 4.3). The study concludes that biarylimidazole derivatives can also serve as an excellent scaffold and required further optimization to develop into selective mPGES-1 inhibitors.

4.3.1.3 2-Amino Benzimidazoles

Researchers from Boehringer Ingelheim, identified 2-arylamino benzimidazole carboxamides derivatives as mPGES-1 inhibitors. Substitution at the second position (compound **10)** (Boehringer Ingelheim International GmbH 2010a) showed IC_{50} of 1 nm in the enzymatic assay. The preliminary success led to the screening of a wide

range of substitutions at the second position of benzimidazole such as 2,4-dichlorobenzyl pivalamide in compounds **11** and **12** (Boehringer Ingelheim International GmbH 2012a) with IC_{50} 1 nM and less than 1 nM in the A549 cell-based assay, respectively. Upon further exploration of benzimidazole, devoid of carboxamide residue, it subsequently led to compound **13** (Boehringer Ingelheim International GmbH 2012b) with an IC_{50} 1.3 nM in A549 cell-based assay.

Larsson et al. reported five new 2-aminobenzimidazole derivatives as inhibitors of human and rodent mPGES-1 against various models for inflammatory activity. Among the five molecules, compound **14** (Larsson et al. 2019) showed IC_{50} 0.024 μM in human mPGES-1, 0.17 μM in rat mPGES-1, 3.7 μM in whole blood assay. The compound **15** (Larsson et al. 2019) with trifluoro derivatives showed IC_{50} 0.023 μM in human mPGES-1, 0.078 μM in rat mPGES-1, and 2.5 μM in whole blood assay. However these two compounds at 10 μM had no inhibition against COX-1, PGIS, L-PGDS or H-PGDS. These interesting results needs further investigation in clinical research for identification of possible inhibitors against mPGES-1.

4.3.1.4 2-Amino Imidazoles

Chandrasekhar et al. identified 2-amino imidazoles as a novel and selective mPGES-1 inhibitors. Among the investigated compounds, mono and trisubstituted imidazoles, compound **16** and compound **17** (Chandrasekhar et al. 2016), showed good potency when compared with celecoxib. These compounds showed 100% inhibition in a concentration-dependent inhibition against mPGES-1 enzyme with IC_{50} 0.241 μM and 0.00094 μM respectively (Fig. 4.4) but showed poor inhibition in rat mPGES-1. Further through a rapid dilution approach, mPGES-1 was incubated with various concentrations of the inhibitors that were tenfold higher than their respective IC_{50} values and then diluted to 100-fold with the substrate solution to get an inhibitor concentration of 1/10 of their IC_{50} values. The study concludes that the compounds were of reversible inhibitor type.

4.3.1.5 Imidazopyridines

Another study from Boehringer Ingelheim reported novel Imidazopyridine carboxamides derivatives as a potent inhibitor of mPGES-1 (Boehringer Ingelheim International GmbH 2010b). The introduction of a carboxamide substituted at fifth position of the imidazopyridine scaffold with a nitrogen atom in the benzene ring led to the discovery of imidazopyridine-6-carboxamides derivatives. All the compounds were evaluated at 10 μM concentration against mPGES-1 in a cell-free assay. Further investigation of the compound **18** (Boehringer Ingelheim International GmbH 2010b) showed 100% inhibition at 10 and 1 μM, while compound **19** (Boehringer Ingelheim International GmbH 2012c) had IC_{50} 1 nM against mPGES-1.

4.3.2 Piperidine Carboxamides

A series of 304 piperidinyl benzoimidazole derivatives from NovaSAID were reported as mPGES-1 (35 compounds belong to piperidine-4-carboxamide

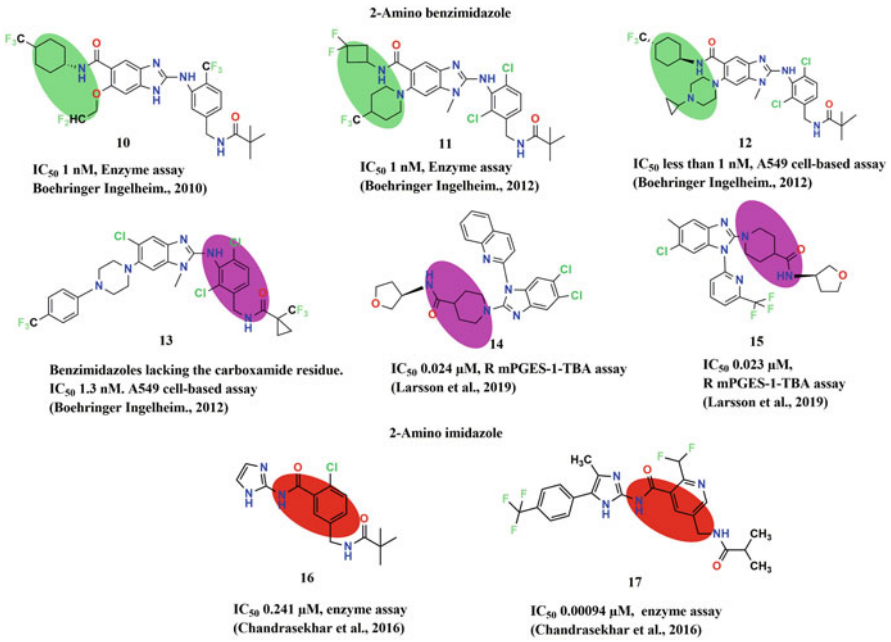


Fig. 4.4 2-amino imidazoles as potent inhibitors of mPGES-1

derivatives). Compound **20** (Fig. 4.5) (Novasaid 2011) with a naphthalene substitution had an IC_{50} 0.011 µM against human mPGES-1, IC_{50} 0.037 µM against rat mPGES-1, and IC_{50} 0.042 µM in the A549 cell-based assay. Compound **21** (Leclerc et al. 2013) with cyclopentyl substitution showed IC_{50} values of IC_{50} 0.9 µM and 0.09 µM in human and rat recombinant mPGES-1. Further the compound **20** showed potent inhibition when tested with human as well as murine cellular assays and hence inhibiting PGE_2 in A549 cells. Pfizer also developed a series of novel benzoxazole piperidine carboxyamides derivatives of which compound **22** (Arhancet et al. 2013) showed an IC_{50} 3 nM (enzymatic assay) and IC_{50} 109 nM (whole blood assay).

4.3.3 Trisubstituted Ureas

In 2011, Chiasson et al. reported a novel scaffold as mPEGS-1 inhibitors through in-house data collection available at Merck Frosst Center for Therapeutic Research. Among the series tested, compound **23**, with a trisubstituted urea (Chiasson et al. 2011), showed moderate inhibition of human recombinant mPGES-1 with 88% inhibition at 10 µM. Further refinement by SAR studies in trisubstituted urea analog suggests that modification of the ethylpyridyl and N-cyclopentyl moieties with benzyl and isopropyl functions led to compound **24** (Chiasson et al. 2011) with IC_{50}

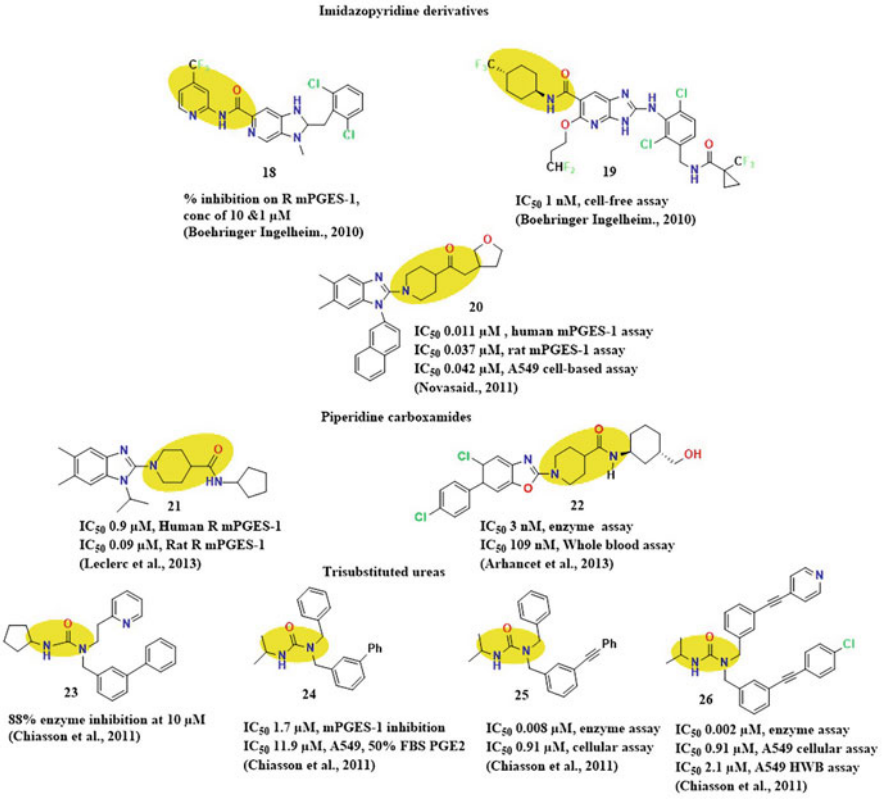


Fig. 4.5 Imidazopyridine, piperidine carboxamides, and trisubstituted urea as mPGES-1 inhibitors

from 10 μM to 1.7 μM. The detailed investigation reveals that *ortho*- and *para*-substituted regioisomeric analogs were not active. When the *meta*-position was substituted with an amide or ether, the compounds found optimum activity. The introduction of sulfonamide at *meta*-position showed moderate intrinsic potency with loss of activity in cell-based assay. Addition of alkyne linker provide a tolane analog compound **25** (Chiasson et al. 2011) with 300-fold (IC_{50} 0.008 μM) increased potency in an enzymatic assay whereas in cellular assay resulted in a tenfold increase with IC_{50} 0.91 μM. Substitution of isopropyl group with phenyl or a benzyl group resulted in a loss of activity. In subsequent trials, isosteric replacement with guanidine analogs led to complete loss of activity. Good potency and selectivity were achieved with substitution of a terminal pyridyl group and a phenyl substitution having electron-withdrawing group in the upper and lower tolane resulted in compound **26** (Chiasson et al. 2011) with mPGES-1 $IC_{50} = 0.002$ μM; A549 $IC_{50} = 0.34$ μM; HWB $IC_{50} = 2.1$ μM (Fig. 4.5).

4.3.4 Benzamides

Over the past decade, structurally diversified 180 substituted benzamide derivatives were explored by pharmaceutical companies in search of novel mPGES-1 inhibitors. Among them, two compounds **27** and **28** (Boehringer Ingelheim International GmbH 2011) showed good inhibition with an IC_{50} 1 nM in the HTRF recombinant enzyme assay. Glenmark pharmaceuticals identified about 40 bicyclic (quinazoline) benzamide series as mPGES-1 inhibitors. Compound **29** a benzamide derivative, (Glenmark Pharmaceuticals SA 2013a) showed 100% inhibition at 1 μM ($IC_{50} < 50$ nM) (Fig. 4.6). Further modifications of the benzamide by introduction of an amide and cyclic structure resulted in compound **30** (Glenmark Pharmaceuticals SA 2014) with similar inhibition. Benzoyl derivatives of 3-aminocarbazole were reported by Acraf et al. and described substitutions like halogen, methyl group, trihalomethyl group, nitro and cyano over CF_3 group as in compound **31** (**A.C.R.A.F. S.P.A.: 2009**). Linear or branched hydroxyalkyl group as well as carbonyl alkyl group comprising from 1 to 8 carbon atoms were also synthesized and tested. Of the different studies, compound **31** with CF_3 group (Aziende Chimiche Riunite Angelini Francesco 2009) showed activity (IC_{50} 2.55 μM and pIC_{50} 5.59) against human recombinant mPGES-1, as well (IC_{50} 0.438 μM and pIC50 6.36) against microsomal mPGES-1. Compound **31** affected PGE2-associated tumor growth with a decrease of PGE2 levels in A431 tumor cells. This compound also decreased A431 tumor volume and resulted in 100% inhibition of the tumor growth at 20 mg/kg in a dose-dependent manner.

4.3.5 Pirinixic Acids

Werz research group reported a new class of mPGES-1 inhibitors obtained from pirinixic acid. Substitution with sterically bulky lipophilic group at α-substitution such as n-hexyl, n-octyl, or naphthyl showed mPGES-1 inhibition better than pirinixic acid. Compound **32** (Fig. 4.6) (Chang and Meuillet 2011) also named as YS121 with α-(n-hexyl)-substitution inhibited mPGES-1 in cell-free assays with IC_{50} 3.4 μM. Compound **32** reduced PGE2 formation without influencing other prostanoid levels in human whole blood in a dose-dependent manner (EC_{50} 2 μM). Compound **32** with a dose of 1.5 mg/kg intraperitoneally inhibited exudate formation and leukocyte infiltration after 4 h and decreased the pleural levels of PGE2 and LTB4 with 36% and 48% inhibition, respectively. Furthermore compound **32** also suppressed the generation of 6-keto PGF1α (45% reduction) in the exudates. The possible mechanism of action could be peroxisome proliferator-activated receptors (PPAR)-α/γ agonism shown to downregulate COX-2 expression because pirinixic acid derivatives are known for its dual agonists against PPAR-α and -γ. The conversion of acid derivative into ester resulted in the loss of activity but bulky lipophilic substituents such as biphenyl-4-methane amine moiety at C6 position of

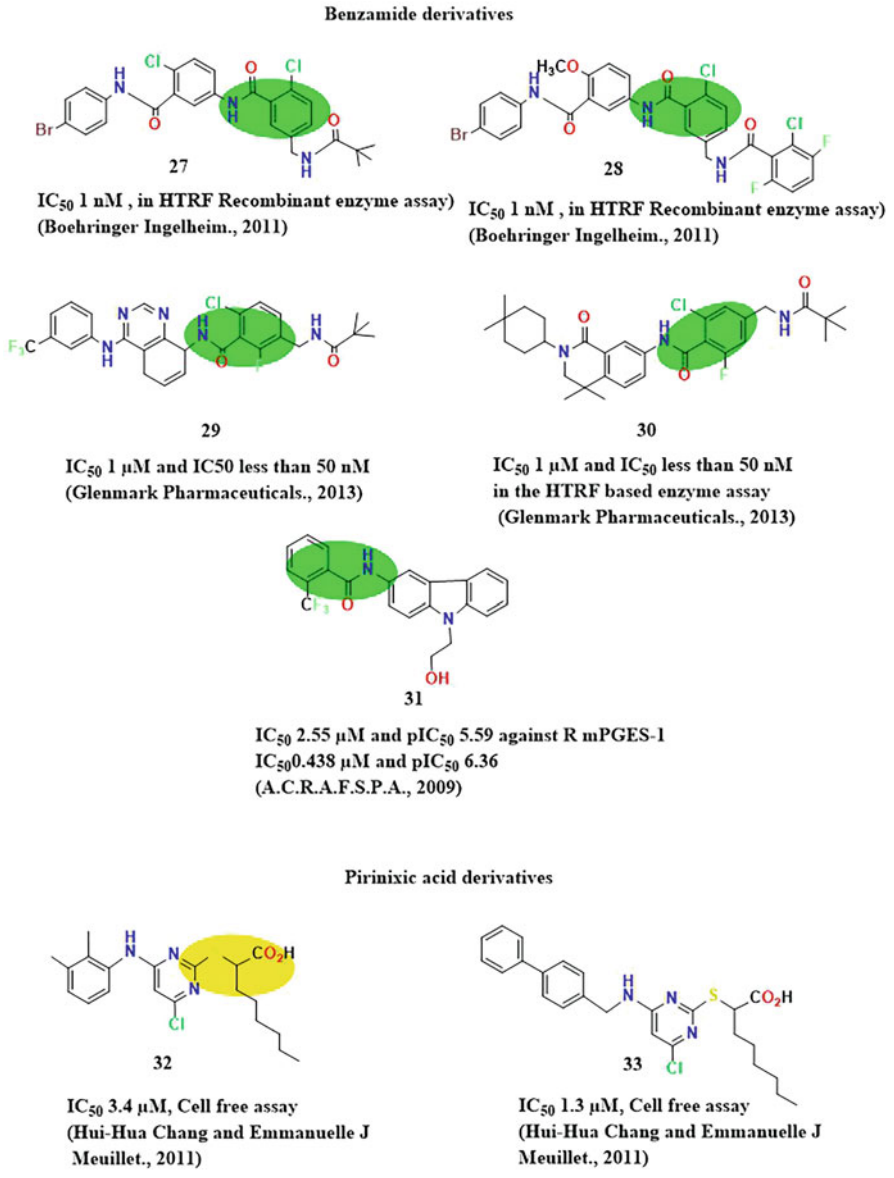


Fig. 4.6 Benzamide and pirinixic acid derivatives as potent inhibitors of mPGES-1

the pyrimidine ring improved the efficacy of mPGES-1 inhibition. This resulted in the identification of compound **33** (Chang and Meuillet 2011), with an IC_{50} value of 1.3 and 2.0 µM in cell-free assays for mPGES-1 (Hanke et al. 2013; Medeon Pharmaceuticals GmbH and University of Tubingen 2009).

4.3.6 Triterpene Acids

Verhoff et al. reported a new class of compounds such as tetra- and pentacyclic triterpene acid derivatives of Boswellia acid as potent mPGES-1 inhibitors. The acidic fractions (containing lipophilic acidic ingredients) of gum resins were analyzed for inhibition of mPGES-1 activity in a cell-free assay (microsomes of A549 IL-1β-stimulated cells) using MK-886 (10 μM; IC_{50} 2.4 μM) as reference compound. Four fractions showed potential inhibition against mPGES-1 with IC_{50} values of 1.9, 2.8, 1.6, and 0.4 μg/mL, respectively. These fractions were obtained from the gum resins of *B. serrata*, *B. sacra*, *B. carterii*, and *B. Papyrifera* respectively. The neutral fraction containing essential oil and mucilage fraction (10 μg/mL) did not show inhibition against mPGES-1. The acidic fraction of *B. papyrifera* gum showed 92% inhibition at 30 μg/mL when compared with MK-886 (10 μM = 0.49 μg/mL, 79% inhibition, IC_{50} 2.4 μM). Siemoneit et al. reported acidic fractions from frankincense gum containing boswellic acid resins for mPGES-1 inhibition activity (Siemoneit et al. 2010). In total, 17 known triterpene acids were isolated from different *Boswellia spp*. out of which compound **34** (Verhoff et al. 2014) and **35** (Verhoff et al. 2014) (Fig. 4.7) showed an IC_{50} 0.4 μM, and IC_{50} 1.2 μM respectively. It was concluded that various lipophilic extracts of gum resins from this *Boswellia spp*. possess bioactive molecules responsible for mPGES-1 inhibition.

4.3.7 Indole-Based Carboxylic Acids

An indole-based carboxylic acid derivative showed moderate inhibition of the human mPGES-1 enzyme with an IC_{50} of 1.6 μM. However, it was identified that the potency of this compound varied when tested in cell-based assays in the presence of fetal bovine serum (FBS) and found that this shift is due to the high degree of plasma protein binding. Compound **36** (Psarra et al. 2017) serves as a lead molecule and further optimization of the molecule is necessary to develop into potent mPGES-1 inhibitors (Fig. 4.7).

4.3.8 Aminobenzothiazoles

Chini et al. in 2020 reported a novel class of aminobenzothiazole scaffold as mPGES-1 inhibitor through a combinatorial approach from the available inhibitory activity of PGE2. The protected aminobenzothiazole nucleus was acylated leading to the formation of compound **37** (Chini et al. 2020) with IC_{50} 1.4 μM, compound **38** (Chini et al. 2020) with IC_{50} 0.7 μM. Compound **39** with (hydroxymethyl) phenyl substitution at position 5 of the benzothiazole scaffold-like (Chini et al. 2020) showed higher inhibitory activity, with an IC_{50} value of 2.6 μM. 4-Fluoro-2-(trifluoromethyl) phenyl moiety such as in compound **40** (Chini et al. 2020), showed better inhibition of mPGES-1 with IC_{50} 1.7 μM (Fig. 4.7).

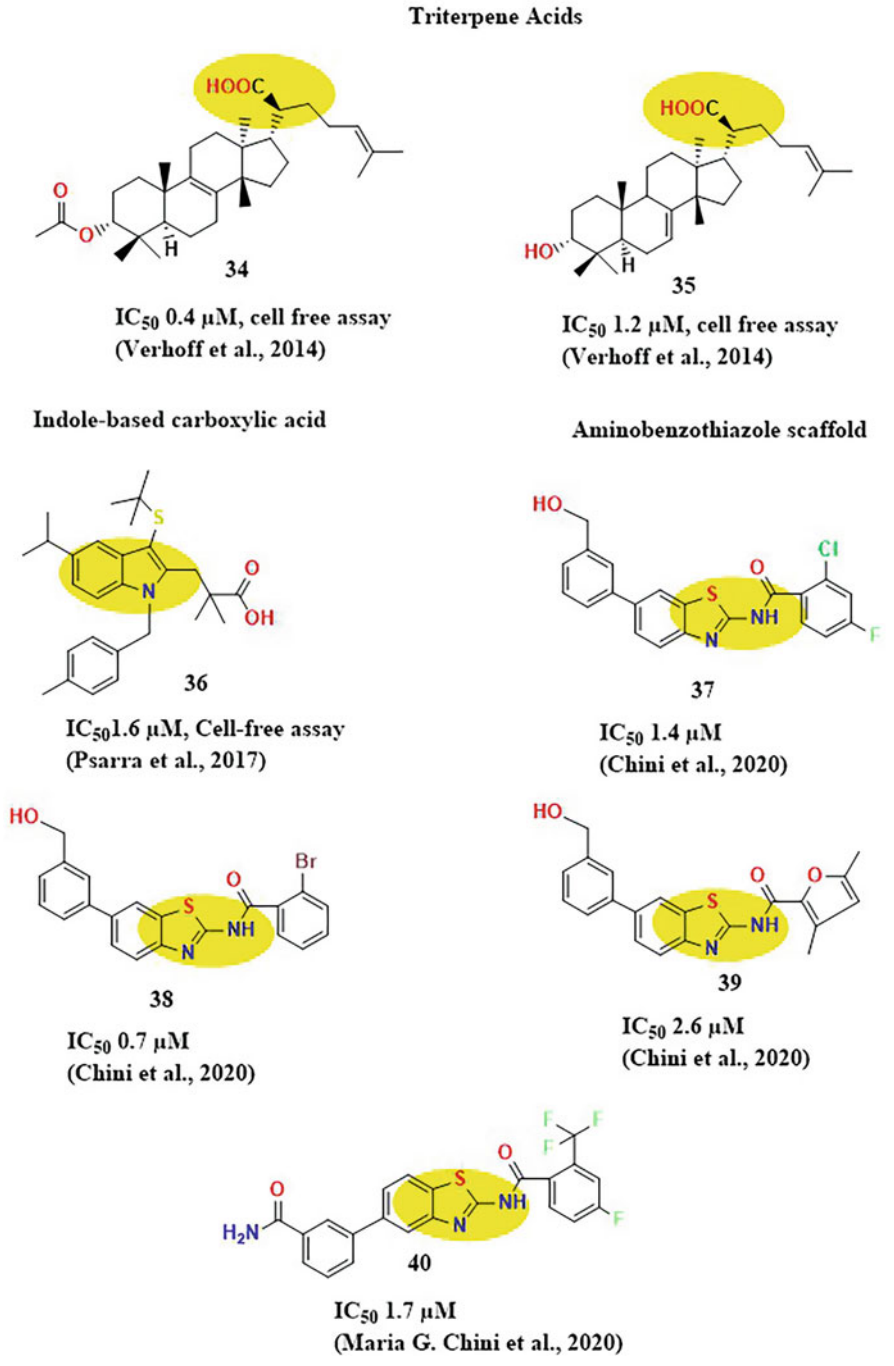


Fig. 4.7 Triterpene acid and amino thiazole as inhibitors of mPGES-1

4.3.9 Sulfomyl Phenylacetamides

In 2018, Shekfeh et al. reported novel inhibitors of human mPGES-1 using a multistep virtual screening approach such as molecular docking, fingerprints-based clustering with diversity-based selection, and molecular dynamics (MD) simulations. The generated hits were analyzed for stable interactions in the binding pocket of mPGES-1. These two compounds **41** (Shekfeh et al. 2018) and **42** (Shekfeh et al. 2018) showed mPGES-1 inhibitory activity with IC_{50} 1.2 μM and 1.3 μM, respectively, in a cell-free assay (microsomes from IL-1β-activated human A549 cells). On further screening, it was found that compounds with a benzothiazole ring showed better mPGES-1 inhibition similar to compound **43** (Shekfeh et al. 2018) with IC_{50} 2.3 μM, than the benzoxazole substitution compound **44** (Shekfeh et al. 2018) with IC_{50} 7.0 μM. Compound **45** (Shekfeh et al. 2018) with a benzothiazole ring showed significant inhibition of mPGES-1 activity with IC_{50} 0.6 μM better than benzoxazole and benzimidazole substitution (Fig. 4.8). The study suggests that the presence of polar aromatic functional group and benzothiazole ring are responsible for potent mPGES-1 inhibition.

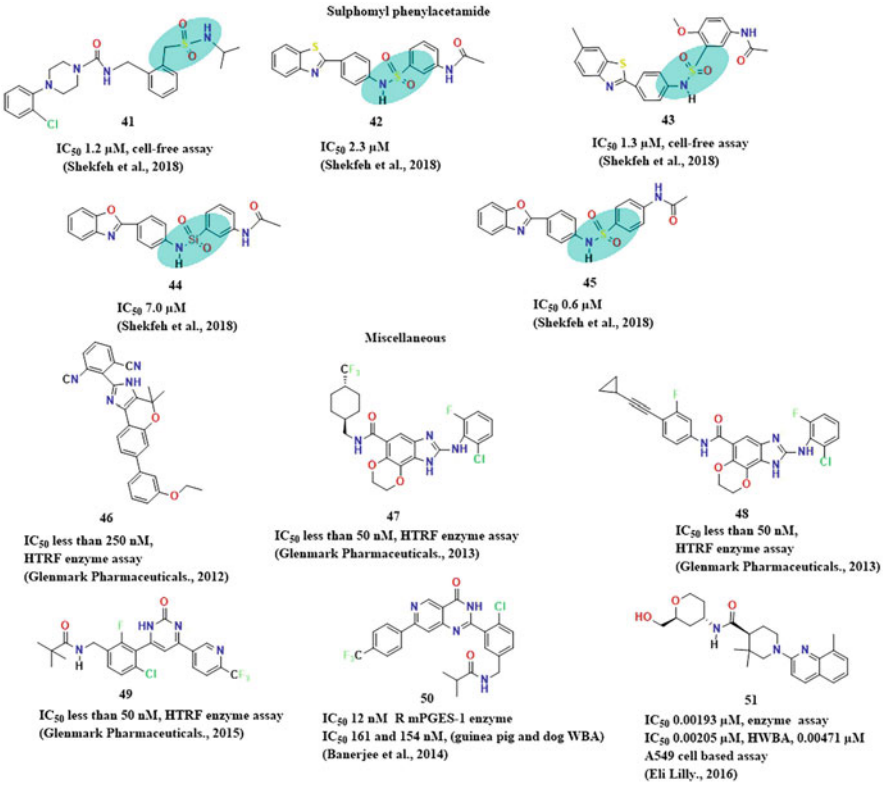


Fig. 4.8 Sulfomyl phenylacetamide with other diversified scaffold as mPGES-1 inhibitor

4.3.10 Other Scafolds

The researchers from Glenmark pharmaceuticals identified substituted bicyclic and tricyclic molecules with good inhibition against mPGES-1 (Glenmark Pharmaceuticals SA 2012). Compound **46** showed the highest in vitro inhibition (HTRF enzyme assay) with $IC_{50} < 250$ nM. Further, extensive work on these cyclic compounds (Glenmark Pharmaceuticals SA 2013b) with a wide range of substitutions on the benzamide residue such as in compounds **47** and **48** had an $IC_{50} < 50$ nM. Structures including substituted pyrimidine compounds were also claimed by Glenmark similar to compound **49** (Glenmark Pharmaceuticals SA 2015) with $IC_{50} < 50$ nM. Compound **50** (GRC27864) showed potent inhibition against recombinant guinea pig mPGES-1 enzyme with an IC_{50} 12 nM (Banerjee et al. 2014). The compound **50** strongly inhibited the formation of PGE2 in LPS-induced guinea pig and dog whole blood assays with IC_{50} 161±36.66 nM and 154±35.06 nM, respectively. Further the compound **50** showed selectivity (>1000-fold) over COX-1, COX-2, mPGES-2, PGES, PGI_2, PGD_2, and TXA_2 synthases. Compound **50** also entered clinical trials for further evaluation of safety and kinetic parameter in healthy subjects. Recently, Eli Lilly discovered novel carboxylic acid derivatives possessing methyl-piperidine and quinoline as mPGES-1 inhibitors. Compound **51** (Eli Lilly and Company 2016) (Fig. 4.8) showed an IC_{50} of 0.00193 μM in the enzymatic assay and IC_{50} of 0.00205 μM in a human whole blood assay and IC_{50} of 0.00471 μM in an A549 cell-based assay.

4.4 Conclusion

The reported studies suggests that mPGES-1 is a potential target for inflammatory as well as cancer to overcome the problems associated with the current drug treatment. Recently, various research groups and scientists have started working on mPGES-1 and attempted for creating hybrid scaffolds or new scaffolds as mPGES-1 inhibitors. Numerous efforts are being made to explore various therapeutic areas such as pain, inflammation, rheumatoid arthritis, and osteoarthritis. Further it is clearly understood and proven in the literature that mPGES-1 is responsible for the overproduction of PGE2 in different tumor cells. The current ongoing research of new mPGES-1 inhibitors, along with several extensive studies (Samuelsson et al. 2007; Sasaki et al. 2015; Hanaka et al. 2009; Sheng et al. 2001; Lauro et al. 2017; Korniluk et al. 2017) suggests that mPGES-1 is a key target for the treatment of wide variety of tumors as a new and safer therapeutic strategy.

References

Aggarwal BB (2004) Nuclear factor-kappaB: the enemy within. Cancer Cell 6(3):203–208

Allison SE, Petrovic N, Mackenzie PI, Murray M (2015) Pro-migratory actions of the prostacyclin receptor in human breast cancer cells that overexpress cyclooxygenase-2. Biochem Pharmacol 96:306–314

American Cancer Society (ed) (2015) Global cancer facts & figures, 3rd edn. American Cancer Society, Atlanta

Arhancet GB, Walker DP, Metz S et al (2013) Discovery and SAR of PF-4693627, a potent, selective and orally bioavailable mPGES-1 inhibitor for the potential treatment of inflammation. Bioorg Med Chem Lett 23:1114–1119

Aziende Chimiche Riunite Angelini Francesco (2009) A.C.R.A.F. S.P.A. 3-Aminocarbazole compound, pharmaceutical composition containing it and preparation method thereof. WO2009138376A1

Banerjee A, Pawar MY, Patil S et al (2014) Development of 2-aryl substituted quinazolin-4(3H)-one, pyrido[4,3-d]pyrimidin-4(3H)-one and pyrido [2,3-d]pyrimidin-4(3H)-one derivatives as microsomal prostaglandin E(2) synthase-1 inhibitors. Bioorg Med Chem Lett 24:4838–4844

Bergqvist F, Morgenstern R, Johan JP (2019) A review on mPGES-1 inhibitors: from preclinical studies to clinical applications. Prostaglandins Other Lipid Mediat 147:106383

Boehringer Ingelheim International GmbH (2010a) 1H-Benzimidazole-5-carboxamides as anti inflammatory agents. WO2010034797A1

Boehringer Ingelheim International GmbH (2010b) 3H-Imidazo[4,5-c]pyridine-6-carboxamides as anti-inflammatory agents. WO2010034798A1

Boehringer Ingelheim International GmbH (2011) Inhibitors of microsomal prostaglandin E2 synthase-1. WO2011048004A1

Boehringer Ingelheim International GmbH (2012a) 6-Amino-2-phenylamino-1H-benzimidazole-5-carboxamide derivatives and their use as microsomal prostaglandin E2 synthase-1 inhibitors. WO2012076673A1

Boehringer Ingelheim International GmbH (2012b) Novel compounds. US2012309738A1

Boehringer Ingelheim International GmbH (2012c) 2-(Arylamino)-3H-imidazo[4,5-b]pyridine-6-carboxamide derivatives and their use as mPGES-1 inhibitors. WO2012022792A1

Chandrasekhar S, Harvey AK, Xiao-Peng Y, Chambers MG, Oskins JL, Lin C, Seng TW, Thibodeaux SJ, Norman BH et al (2016) Lilly research laboratories. Eli Lilly and Company, Indianapolis

Chang H-H, Meuillet EJ (2011) Future. Med Chem 3(15):1909–1934

Cheng SY, Zhang H, Zhang M, Xia SK, Bai XM, Zhang L, Ma J, Rong R, Wang YP, Du MZ et al (2014) Prostaglandin E2 receptor EP2 mediates Snail expression in hepatocellular carcinoma cells. Oncol Rep 31:2099–2106

Cheuk IW, Shin VY, Siu MT, Tsang JY, Ho JC, Chen J, Tse GM, Wang X, Kwong A (2015) Association of EP2 receptor and SLC19A3 in regulating breast cancer metastasis. Am J Cancer Res 5:3389–3399

Chiasson JF, Boulet L, Brideau C et al (2011) Trisubstituted ureas as potent and selective mPGES-1 inhibitors. Bioorg Med Chem Lett 21:1488–1492

Chini MG, Giordano A, Potenza M, Terracciano S et al (2020) Targeting mPGES-1 by a combinatorial approach: identification of the aminobenzothiazole scaffold to suppress PGE2 levels. ACS Med Chem Lett 11(5):783–789

Cote B, Boulet L, Brideau C et al (2007a) Substituted phenanthrene imidazoles as potent, selective, and orally active mPGES-1 inhibitors. Bioorg Med Chem Lett 17:6816–6820

Cote B, Boulet L, Brideau C et al (2007b) Substituted phenanthrene imidazoles as potent, selective Identification of a novel, potent and selective inhibitor of mPGES-1 activity, and orally active mPGES-1 inhibitors. Bioorg Med Chem Lett 17(24):6816–6820

de Groot DJ, de Vries EG, Groen HJ, de Jong S (2007) Non-steroidal anti-inflammatory drugs to potentiate chemotherapy effects: from lab to clinic. Crit Rev Oncol 61:52–69

Donnini S, Finetti F, Terzuoli E, Bazzani L, Ziche M (2014) Targeting PGE2 signaling intumor progression and angiogenesis. For Immunopathol Dis Therap 1:1

Dorsam RT, Gutkind JS (2007) G-Protein-coupled receptors and cancer. Nat Rev Cancer 7:79–94

Eli Lilly and Company (2016) Novel carboxylic acid compounds useful for inhibiting microsomal prostaglandin E2 synthase-1. WO2016069374A1

Fei F, Zhou ZM (2014) New substituted benzimidazole derivatives: a patent review (2010-2012). Expert Opin Ther Pat 23:1157–1179

Giroux A, Boulet L, Brideau C et al (2009) Discovery of disubstituted phenanthrene imidazoles as potent, selective and orally active mPGES-1 inhibitors. Bioorg Med Chem Lett 19:5837–5841

Glenmark Pharmaceuticals SA (2012) Tricyclic compounds as mPGES-1 inhibitors. WO2012110860

Glenmark Pharmaceuticals SA (2013a) Tricyclic compounds as mPGES-1 inhibitors. US2013210844A1

Glenmark Pharmaceuticals SA (2013b) Tricyclic compounds as mPGES-1 inhibitors. WO2013153535

Glenmark Pharmaceuticals SA (2014) Substituted bicyclic compounds as mPGES-1 inhibitors. WO2014167444A1

Glenmark Pharmaceuticals SA (2015) Substituted pyrimidine compounds as mPGES-1 inhibitors. WO2015059618A1

Hamza A, Zhao X, Tong M, Tai HH, Zhan CG (2011) Novel human mPGES-1 inhibitors identified through structure-based virtual screening. Bioorg Med Chem 19:6077–6086

Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144(5):646–674

Hanaka H, Pawelzik SC, Johnsen JI, Rakonjac M, Terawaki K, Rasmuson A, Sveinbjornsson B, Schumacher MC, Hamberg M, Samuelsson B, Jakobsson PJ, Kogner P, Radmark O (2009) Microsomal prostaglandin E synthase 1 determines tumor growth in vivo of prostate and lung cancer cells. Proc Natl Acad Sci U S A 106(44):18757–18762

Hanke T, Dehm F, Liening S, Popella SD, Maczewsky J, Pillong M, Kunze J, Weinigel C, Barz D, Kaiser A, Wurglics M, Lammerhofer M, Schneider G, Sautebin L et al (2013) Aminothiazole-featured pirinixic acid derivatives as dual 5-lipoxygenase and microsomal prostaglandin E2 synthase-1 inhibitors with improved potency and efficiency in vivo. J Med Chem 56:9031–9044

Hino S, Tanji C, Nakayama KI, Kikuchi A (2005) Phosphorylation of beta-catenin by cyclic AMP-dependent protein kinase stabilizes beta-catenin through inhibition of its ubiquitination. Mol Cell Biol 25:9063–9072

Hsu HH, Chen MC, Day CH, Lin YM, Li SY, Tu CC, Padma VV, Shih HN, Kuo WW, Huang CY (2017a) Thymoquinone suppresses migration of LoVo human colon cancer cells by reducing prostaglandin E2-induced COX-2 activation. World J Gastroenterol 23:1171–1179

Hsu HH, Lin YM, Shen CY, Shibu MA, Li SY, Chang SH, Lin CC, Chen RJ, Viswanadha VP et al (2017b) Prostaglandin E2-induced COX-2 expressions via EP2 and EP4 signalling pathways in human LoVo colon cancer cells. Int J Mol Sci 18:1132

Jin Y, Smith CL, Hu L, Campanale KM, Stoltz R, Huffman LG, McNearney TA, Yang XY, Ackermann BL, Dean R, Regev A, Landschulz W (2015) Pharmacodynamic comparison of LY3023703: a novel microsomal prostaglandin E synthase 1 inhibitor, with celecoxib. Clin Pharmacol Ther 99(3):274–284

Kalinski P (2012) Regulation of immune responses by prostaglandin E2. J Immunol 188:21–28

Kim KM, Im AR, Kim SH, Hyun JW, Chae S (2016) Timosaponin AIII inhibits melanoma cell migration by suppressing COX-2 and in vivo tumor metastasis. Cancer Sci 107:181–188

Kim SH, Roszik J, Cho SN, Ogata D, Milton DR, Peng W, Menter DG, Ekmekcioglu S, Grimm EA (2019) The COX2 effector microsomal PGE2 synthase 1 is a regulator of immunosuppression in cutaneous melanoma. Clin Cancer Res 25:1650–1663

Koeberle A, Werz O (2015) Perspective of microsomal prostaglandin E synthase-1 as drug target in inflammation-related disorders. Biochem Pharmacol 98:1–5

Koeberle A, Laufer SA, Werz O (2016) Design and development of microsomal prostaglandin E2 synthase-1 inhibitors: challenges and future directions. J Med Chem 59(13):5970–5986

Korniluk A, Koper O, Kemona H, Dymicka- Piekarska V (2017) From inflammation to cancer. Ir J Med Sci 186(1):57–62

Kubota M, Wakamatsu K (2008) Peptide fragment of the m3 muscarinic acetylcholine receptor activates G_q but not G_{i2}. J Pept Sci 14(8):998–1002

Larsson K, Kock A, Idborg H, Arsenian Henriksson M, Martinsson T, Johnsen JI, Korotkova M, Kogner P, Jakobsson PJ (2015) COX/mPGES-1/PGE2 pathway depicts an inflammatory dependent high-risk neuroblastoma subset. Proc Natl Acad Sci U S A 112:8070–8075

Larsson K, Steinmetz J, Bergqvist F et al (2019) Biological characterization of new inhibitors of microsomal PGE synthase-1 in preclinical models of inflammation and vascular tone. Br J Pharmacol 176:4625–4638

Lauro G, Manfra M, Pedatella S, Fischer K, Cantone V, Terracciano S, Bertamino A, Ostacolo C, Gomez-Monterrey I, De Nisco M, Riccio R, Novellino E, Werz O, Campiglia P, Bifulco G (2017) Identification of novel microsomal prostaglandin E2 synthase-1 (mPGES-1) lead inhibitors from fragment virtual screening. Eur J Med Chem 125:278–287

Leclerc P, Idborg H, Spahiu L et al (2013) Characterization of a human and murine mPGES-1 inhibitor and comparison to mPGES-1 genetic deletion in mouse models of inflammation. Prostaglandins Other Lipid Mediat 107:26–34

Ma Y, Brusselaers N (2018) Maintenance use of aspirin or other non-steroidal anti-inflammatory drugs (NSAIDs) and prostate cancer risk. Prostate Cancer Prostatic Dis 21:147–152

Ma W, St-Jacques B (2018) Signalling transduction events involved in agonist-inducedPGE2/EP4 receptor externalization in cultured rat dorsal root ganglion neurons. Eur J Pain 22:845–861

Masur K, Lang K, Niggemann B, Zanker KS, Entschladen F (2001) High PKCalpha and low E-cadherin expression contribute to high migratory activity of colon carcinoma cells. Mol Biol Cell 12:1973–1982

McGraw DW, Mihlbachler KA, Schwarb MR, Rahman FF, Small KM, Almoosa KF, Liggett SB (2006) Airway smooth muscle prostaglandin-EP1 receptors directly modulate beta2-adrenergic receptors within a unique heterodimeric complex. J Clin Invest 116:1400–1409

Medeon Pharmaceuticals GmbH and University of Tubingen (2009) Pirinixic acid derivatives as prostaglandin E2 synthesis inhibitors for treating inflammatory diseases. WO2009117985A1

Merk Frosst Canada Ltd (2006) 2-(Phenyl or heterocyclic)-1H-phenantrho[9,10-d] imidazoles as mPGES-1 inhibitors. WO2006063466A1

Merk Frosst Canada Ltd (2007) Methods for treating or preventing neoplasias WO2007124589A1

Mochocki M, Morawski P, Kopta R, Brzezinska-Blaszczyk E, Stasikowska O, Lewy-Trenda I, Starska K (2015) Expression of prostaglandin E2 prostanoid receptor EP2 and interleukin-1β in laryngeal carcinoma-preliminary study. Contemp Oncol 19:113–119

Morou E, Georgoussi Z (2005) Expression of the third intracellular loop of the δ-opioid receptor inhibits signaling by opioid receptors and other G protein-coupled receptors. J Pharmacol Exp Ther 315(3):1368–1379

Murakami M, Nakatani Y, Tanioka T, Kudo I, Synthase PE et al (2002) Molecular biology of the arachidonate cascade. Prostaglandins Other Lipid Mediat 68–69:383–399

Nakanishi M, Rosenberg DW (2013) Multifaceted roles of PGE2 in inflammation and cancer. Semin Immunopathol 35:123–137

Nakanishi M, Gokhale V, Meuillet EJ, Rosenberg DW (2010) mPGES-1 as a target for cancer suppression: a comprehensive invited review Phospholipase A2 and lipid mediators. Biochimie 92(6):660–664

Ng T, Shima D, Squire A, Bastiaens PI, Gschmeissner S, Humphries MJ et al (1999) PKCalpha regulates beta1 integrin-dependent cell motility through association and control of integrin traffic. EMBO J 18:3909–3923

Novasaid AB (2011) Microsomal prostaglandin E synthase-1 (mPGES1) inhibitors. WO2011023812A1

O'Callaghan G, Houston A et al (2015) Prostaglandin E2 and the EP receptors in malignancy: possible therapeutic targets? Br J Pharmacol 172:5239–5250

Otsnu H (2013) Preparation of compounds as inhibitors of microsomal prostaglandin E synthase-1 (mPGES-1). PCT Int Appl 2013:024898

Pierce KL, Fujino H, Srinivasan D, Regan JW (1999) Activation of FP prostanoid receptor isoforms leads to Rho-mediated changes in cell morphology and in the cell cytoskeleton. J Biol Chem 274:35944–35949

Psarra A, Nikolaou A, Kokotou MG, Limnios D, et al (2017) Microsomal prostaglandin E2 synthase-1 inhibitors: a patent review. Expert Opinion on Therapeutic Patents

Regan JW (2003) EP2 and EP4 prostanoid receptor signaling. Life Sci 74:143–153

Rhee SG (2001) Regulation of phosphoinositide-specific phospholipase C. Annu Rev Biochem 70: 281–312

Ricciotti E, FitzGerald GA (2011) Prostaglandins and inflammation. Vasc Biol 31(5):986–1000

Rogers JH, Tamirisa P, Kovacs A, Weinheimer C, Courtois M, Blumer KJ, Kelly DP, Muslin AJ (1999) RGS4 causes increased mortality and reduced cardiac hypertrophy in response to pressure overload. J Clin Invest 104:567–576

Rundhaug JE, Fischer SM (2008) Cyclo-oxygenase-2 plays a critical role in UV-induced skin carcinogenesis. Photochem Photobiol 84:322–329

Rundhaug JE, Simper MS, Surh I, Fischer SM (2011) The role of the EP receptors for prostaglandin E2 in skin and skin cancer. Cancer Metastasis Rev 30:465–480

Samuelsson B, Morgenstern R, Jakobsson J et al (2007) Membrane prostaglandin E synthase-1: a novel therapeutic target. Pharmacol Rev 59:207–224

Sasaki Y, Nakatani Y, Hara S (2015) Role of microsomal prostaglandin E synthase-1 (mPGES-1)-derived prostaglandin E2 in colon carcinogenesis. Prostaglandins Other Lipid Mediat 121:42–45

Seo T, Tatsuguchi A, Shinji S, Yonezawa M, Mitsui K, Tanaka S, Fujimori S, Gudis K et al (2009) Microsomal prostaglandin E synthase protein levels correlate with prognosis in colorectal cancer patients. Virchows Arch 454:667–676

Shekfeh S, Çalışkan B, Fischer K, Yalçın T, Garscha U, Werz O et al (2018) A multi-step virtual screening protocol for the identification of novel non-acidic microsomal prostaglandin E_2 synthase-1 (mPGES-1) inhibitors. ChemMedChem 4(2):273–281. https://doi.org/10.1002/cmdc.201800701

Sheng H, Shao J, Washington MK, DuBois RN (2001) Prostaglandin E2 increases growth and motility of colorectal carcinoma cells. J Biol Chem 276(21):18075–18081

Siemoneit U, Koeberle A, Rossi A, Dehm F, Verhoff M, Reckel S, Maier TJ, Jauch J, Northoff H, Bernhard F, Doetsch V, Sautebin L, Werz O (2010) Inhibition of microsomal prostaglandin E2 synthase-1 as a molecular basis for the anti-inflammatory actions of boswellic acids from frankincense. Br J Pharmacol 16:1476–5381

Singh T, Vaid M, Katiyar N, Sharma S, Katiyar SK (2011) Berberine, an isoquinoline alkaloid, inhibits melanoma cancer cell migration by reducing the expressions of cyclo-oxygenase-2, prostaglandin E2 and prostaglandin E2 receptors. Carcinogenesis 32:86–92

Sobhani N, Roviello G, Corona SP, Scaltriti M, Ianza A, Bortul M, Zanconati F, Generali D (2018) The prognostic value of PI3K mutational status in breast cancer: a meta-analysis. J Cell Biochem 119:4287–4292

Tai HH (2011) Prostaglandin catabolic enzymes as tumor suppressors. Cancer Metastasis Rev 30: 409–417

Tian M, Schiemann WP (2010) PGE2 receptor EP2 mediates the antagonistic effect of COX-2 on TGF-beta signaling during mammary tumorigenesis. FASEB J 24:1105–1116

Veettil SK, Lim KG, Ching SM, Saokaew S, Phisalprapa et al (2017) Effects of aspirin and non-aspirin nonsteroidal anti-inflammatory drugs on the incidence of recurrent colorectal adenomas: a systematic review with meta-analysis and trial sequential analysis of randomized clinical trials. BMC Cancer 17:763

Verhoff M, Seitz S, Paul M et al (2014) Tetra- and pentacyclic triterpene acids from the ancient anti inflammatory remedy frankincense as inhibitors of microsomal prostaglandin E2 synthase 1. American Chemical Society, Washington

Voss T, Wallner E, Czernilofsky AP, Freissmuth M (1993) Amphipathic α-helical structure does not predict the ability of receptor-derived synthetic peptides to interact with guanine nucleotide-binding regulatory proteins. J Biol Chem 268(7):4637–4642

Waltenberger B, Wiechmann K, Bauer J, Markt P, Noha SM, Wolber G, Rollinger JM, Werz O, Schuster D, Stuppner H (2011) Pharmacophore modeling and virtual screening for novel acidic inhibitors of microsomal prostaglandin E(2) synthase-1 (mPGES-1). J Med Chem 54:3163–3174

Wilson SJ, McGinley K, Huang AJ, Smyth EM (2007) Heterodimerization of the alpha and beta isoforms of the human thromboxane receptor enhances isoprostane signaling. Biochem Biophys Res Commun 352:397–403

World Health Organization (WHO) (2018) India source: Globocan 2018. Available at https://gco.iarc.fr/today/data/factsheets/populations/356-india-fact-sheets.pdf

World Health Organization (WHO) (n.d.). Available at https://www.who.int/news-room/fact-sheets/detail/cancer

Wu TY, Juteau H, Ducharme Y et al (2010) Biarylimidazoles as inhibitors of microsomal prostaglandin E2 synthase-1. Bioorg Med Chem Lett 20:6978–6982

Xiao X, Li BX, Mitton B, Ikeda A, Sakamoto KM (2010) Targeting CREB for cancer therapy: friend or foe. Curr Cancer Drug Targets 10:384–391

Emerging Role of Structural and Systems Biology in Anticancer Therapeutics

5

Manzar Alam, Ahmad Abu Turab Naqvi, and Md. Imtaiyaz Hassan

Abstract

Structural biology methods presently play a significant role in the development of new therapeutic drugs such as approaches for cancer therapies. Structural biology is fundamental for recognizing how proteins and genes function and provides us with the necessary clues to design effective cancer therapies. X-ray crystallography has been established to be a dominant instrument in an essential method for the design and development of new compounds with improved affinity and specificity. It can provide delicately complete structural information concerning the interaction of a ligand with a drug or pharmacological target. Fragment-Based Screening has emerged with X-ray Crystallography has turned into an influential screening technology, capable of providing structural information in complexes that involve low-molecular-weight compounds, although with weak binding affinities. The current drug discovery process is extremely complex and requires multidisciplinary efforts and action in cancer therapeutics. Cancer drug development and discovery are leading the way in utilizing molecular biological and genetic information for developing medicine. In this chapter, we discuss the role of structural biology, including using X-ray crystallography and its role in drug development, with a special focus on the status of the development and discovery of cancer therapeutics. We further discussed how structural biology and systems biology are integrated and give rise to a relatively new domain called "structural systems biology."

M. Alam · M. I. Hassan (✉)
Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India
e-mail: mihassan@jmi.ac.in

A. A. T. Naqvi
Department of Computer Science, Jamia Millia Islamia, New Delhi, India

S. Singh (ed.), *Systems Biomedicine Approaches in Cancer Research*,
https://doi.org/10.1007/978-981-19-1953-4_5

Abbreviations

HTS	High-throughput screening
FBDD	Fragment-based drug discovery
SPR	Surface plasmon resonance
PDB	Protein DataBase
PKB	Protein kinase B
CLL	Chronic lymphocytic leukemia
TSG	Tumor suppressor genes
MEK1	MAP-kinase kinase 1
mTOR	Mammalian target of rapamycin
PI3K	Phosphatidylinositol-3 kinase
HDAC	Histone deacetylase

5.1 Introduction

Over the past decade, scientists and researchers have performed demonstration analyses of successful drug movements to establish "rules" to select novel target proteins (Surade and Blundell 2012). Thoughts about the employment of X-ray crystallography in drug development more than 40 years ago, like the initial 3D structures of proteins, have been established. However, these thoughts comprised the synthesis of ligands of hemoglobin for reducing sickling (Beddell et al. 1976; Goodford et al. 1980), the chemical alteration of insulins for enhancing half-lives in circulation (Blundell et al. 1972), and the design of inhibitors of serine proteases for controlling blood clotting. However, apart from an early endeavor in 1975 (Beddell et al. 1976), nearly all pharmaceutical companies believed X-ray crystallography was also costly. Though structures of relevant drug targets have generally not been openly accessible from X-ray crystallography, relative models based on homologs have been proven helpful in significant topographies of corresponding surfaces of ligands and protein targets, which began to be utilized in lead optimization in the 1980s (Blundell et al. 1983; Blundell 1996; Campbell 2000).

The structural information of the drugs and drug applicants has been obtained from X-ray crystallography. In recent years, NMR-based plans have become more influential and broadly utilized in drug development and discovery. However, the lead molecules that led to elarofiban have been designed to mimic the conformation of a part of fibrinogen while bound to the receptor. Those confirmations have been obtained from NMR analyses of fibrinogen peptides. The structural information could not clearly show which atoms in the molecular scaffold would not be alternated during optimization to conserve interactions with the target (Congreve et al. 2008; Hardy and Malikayil 2003). This difference between NMR and crystallographic moves toward requires that not be the case normally. Hence, NMR may now honestly probe-target protein-small molecule interfaces in constructive cases.

Hence, the impact of NMR techniques was delayed due to the extended intrinsic time needed for protein structures to be established by NMR (Carr et al. 2005; Hardy and Malikayil 2003; Scott et al. 2012).

Efforts to elucidate the molecular basis of cancer are not the latest. The recognition of the primary tumor-causing oncogenes and TSGs in the 1970s and 1980s and the finding of the ways tumor genes undermine signal transduction pathways (Varmus 2006). Cancer drug development has squeezed molecular cancer as a resource of disease-causing targets for mechanism-based drug discovery (Workman 2005a, b). Structural biology is fundamental to identifying how proteins and genes function, and provides us with the necessary clues to design effective cancer therapies.

Modern advances in tumor biology and genetics are conducted in new cancer drug development and treatment (Anwar et al. 2020; Beg et al. 2019; Garcia-Diaz and Kunkel 2006; Gupta et al. 2019a, b, 2020; Thomas et al. 2017). Hence, natural products and structural analogs have traditionally contributed to pharmacotherapy, especially for cancer (Atanasov et al. 2021; Mohammad et al. 2019, 2020; Naz et al. 2017, 2018a, b, 2019; Shamim Jairajpuri et al. 2021). Clinical trials are currently underway for several drugs derived from structure-based design methods (Hardy and Malikayil 2003). Protein structure may influence drug discovery at each stage in the design procedure. Protein structure might be utilized in target identification as well as selection (Gulzar et al. 2019; Hassan and Ahmad 2011; Hassan et al. 2013; Khan et al. 2016, 2017, 2019; Naqvi et al. 2018). Conventionally, this has involved homology recognition supported via information about protein structure; although now structural genomics programs are looking to define representative protein structures. X-ray crystallography was used to assist in the recognition of hits through virtual screening and, frankly, the screening of chemical fragments. However, the key functions of bioinformatics and structural biology in lead optimization continue to be as significant as ever (Lombardino and Lowe 2004; Whittle and Blundell 1994). Currently, an appreciation of the 3D structure of the compounds and their targets is an ingredient of each drug-discovery scheme. Hence, this target structure may be experimentally decided, a model created based on a virtual model of the receptor constructed based on the chemical structure of the identified active compound (Hillisch et al. 2004; Hubbard 2005).

5.2 Early Development of Structure-Guided Drug Discovery

John Kendrew and Max Perutz, their collaborators in Cambridge, have explained the initial protein structures of hemoglobin and myoglobin in the 1950s and 1960s (Bodo et al. 1959; Perutz 1997; Perutz et al. 1951, 1960). They have previously been aware of the significance of the work for medicine. However, the impacts of alterations in oxygen affinity and subunit cooperatively in abnormal hemoglobins that affect sickle-cell disease have been identified as the main objective. Hence, Dorothy Hodgkin's Oxford laboratory worked together with Jørgen Schlichtkrull of Novo to understand how diverse insulin's crystalline appearances might be utilized

as slow-acting therapeutics for diabetes treatment (Schlichtkrull 1958). This became a factual opportunity when insulin structure was explained (Adams et al. 1969; Blundell et al. 1971), as numerous insulin sequences have been described in Cambridge in Fred Sanger's laboratory (Sanger 1988). Structures and sequences motivated thoughts concerning insulin storage and receptor attaching and constructing more effectual therapeutics. Ideas regarding drug design have been induced via the determination of the enzyme structures-lysozyme, trypsin, and chymotrypsin and a promising set of interactions, which guided to the selectivity of enzyme substrate binding (Beddell et al. 1976). The clinically significant drug targets, including aspartic protease renin in the 1970s and 1980s (Atkinson et al. 1980; Schelling et al. 1980) that cleave angiotensinogen to form angiotensin I, have been modeled on less exciting enzymes like fungal pepsins (James et al. 1977; Subramanian et al. 1977). However, the utilization of protein crystallography in drug development accelerated in the 1980s, principally through utilizing a combination of interactive computer graphics and protein structure, including the Evans and Sutherland machines (Tickle et al. 1984). Hence, the renin model (Blundell et al. 1983) has been exploited broadly in structure-guided drug design in pharma production. However, the high-resolution X-ray structures of apo-enzymes and complexes of renin and its close-up homologs were pursued later (Dhanaraj et al. 1992; Rahuel et al. 1991). Over the last several years, numerous new approaches have been initiated that utilize information about the architecture of targets and screening of chemical libraries. Therefore, one of the most prominent was the development of structure-guided-FBDD (Thomas et al. 2017).

5.3 Structural Biology and Cancer

The determination of the structure of a protein target, possibly complexes for partner proteins, nucleic acid, lipids, or substrate, might provide a comprehensible imminent into the mechanism of the action of a protein that in turn may frequently be linked to its biological and therapeutic function (Capila and Linhardt 2002; Lee and Yaffe 2016). However, modern structural biology, predominantly protein crystallography, produces the structure of an increasing number of therapeutically significant targets (Hubbard 2005; Van Montfort and Workman 2017). The main problems limiting the number of structures are the capability of producing adequate quantities of pure, functional, soluble, homogenous proteins for crystallization trials and the capacity of the protein for making regular crystals appropriate for diffraction experiments (Gavira 2016; Hubbard 2005). However, this combination of limitations frequently denotes a structure not accessible for the entire therapeutic target. Even the structure of individual domains may be enough to have an actual impact on a discovery assignment and provide a framework for understanding the overall role of the protein (Davis et al. 2003; Hubbard 2005). Hence, the structure of the ligand-binding domain is adequate against which featured structure-based design may effectively design selective ligands (Congreve et al. 2011; Hillisch et al. 2004). The subtleties of

the receptor role in the cell might be understood in terms of the relationship between the diverse domains, which guides the receptor activity (Congreve et al. 2011).

Structural biology methods presently play an essential role in the progress of new therapeutic drugs such as cancer therapies (Garratt 2013; Li et al. 2016). X-ray crystallography is a predominantly influential instrument for discovering and developing new agents. However, X-ray diffraction solves the problem and improves the final structure (Garratt 2013; Zheng et al. 2015). The methods' limitations and prospective errors found in protein structure dropped in the PDB (Read et al. 2011). Theoretical docking and fragment-based ligand discovery are two approaches that structural biologists can use to focus their weapons in drug design problems (Yuriev and Ramsland 2013). In many of the applications illustrated in the literature, inhibitors of HSP90 and PKB are used as examples of how improved precise protein-ligand interactions can be used to improve binding and pharmacokinetic roles (Congreve et al. 2008; Hopkins et al. 2014). However, the increasing number of targeted therapeutics, abrasion rates for cancer drugs in the clinic are inferior to those for disease regions (Kola and Landis 2004). Cancer treatment assessments could benefit from statistical investigation with genetic and structural studies. Simultaneously with experimental approaches, statistical, genetic, and clinical trials, and basic theory, computational structural biology may assist in coining and identifying novel paradigms for elucidating the origin of cancer and treatments (Alam and Mishra 2021; Nussinov et al. 2019).

5.4 Protein X-ray Crystallography in Drug Discovery

X-ray crystallography has been and will continue to be the primary source of experimental structural biology statistics used in drug discovery (Zheng et al. 2015). With the initiation of structural biology in drug development and discovery procedures, medicinal chemists achieved the prospect of using full structural information consecutively for progressing screening hits into drug aspirants (Congreve et al. 2005; Maveyraud and Mourey 2020). X-ray crystallography was established as an important tool in this admiration, as it is capable of providing delicately inclusive structural information concerning the interface of a ligand with a drug or pharmacological target (Manzoni et al. 2018; Maveyraud and Mourey 2020). However, the potential of X-ray crystallography has been previously proven to permit the unambiguous structure determination of penicillin (Hodgkin 1949). For example, X-ray diffraction is regularly utilized in pharmaceutical companies for drug characterization and polymorphism (Aitipamula et al. 2018; Thakral et al. 2018).

The information of a protein structure might assist the design of particular ligands that is currently a broadly recognized obviousness, emerged in 1976 and the launch of the PDB in 1971 (Beddell et al. 1976; Bernstein et al. 1977), then developed in significance to the point that the "rational drug design cycle" has been detailed (Hol 1986). Hence, a drug may be realistically designed and optimized utilizing the information offered through the structure determination by the macromolecular target. However, the interface of the optimized compounds by drug target is

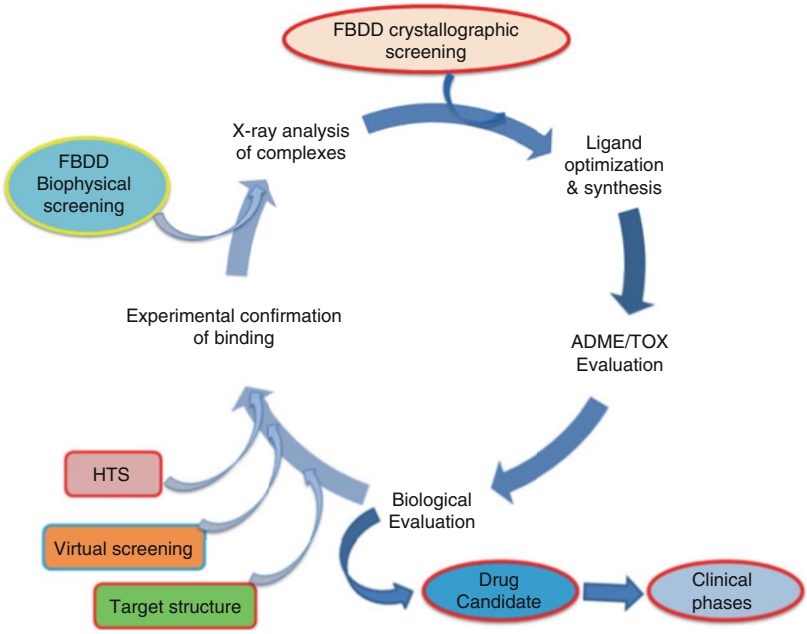


Fig. 5.1 Schematic to demonstrate the theory of the drug discovery cycle. Several techniques and tools are involved

structurally distinguished, permitting the subsequent cycle of chemical optimization. Therefore, the cycle's core is being utilized currently, though with more sophistication (Fig. 5.1). Like throughput augmented, X-ray crystallography improved from protein target structure determination, probably in the existence of a few prerecognized ligands, for structure-action association determination, wherever numerous structures of complexes are established sequentially for guide ligand optimization (Lesuisse et al. 2002).

The primary method for determining the 3D structures of proteins, viruses, and nucleic acids was macromolecular X-ray crystallography. However, X-ray crystallography has some basic limitations. A few can be conquered and balanced using promising techniques in other regions of structural biology (Milroy et al. 2014; Zheng et al. 2015). Hence, other structural biology methods help overcome the main limitation of X-ray crystallography by providing complementary structural information, which is valuable in drug development and discovery (Vénien-Bryan et al. 2017; Zheng et al. 2015). The information of 3D structures of proteins promises to accelerate drug discovery, but current developments in genome sequencing, bioinformatics, and robotics have drastically transformed the chances. However, several new protein targets were recognized from genome investigations and studied via X-ray study or NMR spectroscopy (Congreve et al. 2005).

5.5 Protein Crystallography, FBDD, and Cancer

Crystallography is the main tool for structure-driven drug design, as it permits knowledge of the 3D structures of protein targets and protein-ligand complexes. The path for crystal structure determination engages various steps; some might hinder its high-throughput use. However, current endeavors have generated considerable advances in computational and experimental tools and protocols (Caliandro et al. 2013). The recognition of chemical guides against the targets is the main step of the drug discovery procedure. Hence, starting points for chemical leads include natural products, HTS of large chemical libraries, and FBDD. However, a technique has developed over the past several years to create high-affinity ligands to serve as initial points for discovering drug aspirants (Larsson et al. 2011). The FBDD approach exploits compounds' molecular weight (<~300 Da) compared to those utilized in HTS. The beginnings of FBDD are debatable. Although was reported (Davies and Tickle 2011), X-ray crystallography was initially utilized for mapping the interactions of small-molecule organic solvents on protein surfaces (English et al. 1999; Fitzpatrick et al. 1993). Surprisingly, approaches utilizing FBDD were successful when large HTS screens have been unsuccessful, such as in the development of β-secretase inhibitors (Wyss et al. 2011).

FBDD has appeared as an influential tool to discover drug leads. The approach initially recognizes starting points: fragments concerning half the size of typical drugs (Erlanson 2011). Early conducted experiments exploited ligand-based NMR (Harner et al. 2013) and X-ray crystal screening (Blundell et al. 2002; Murray et al. 2012) documented at Astex primarily through the utilization of high-throughput studies of cocktails of 6–10 fragments soaked in apo-protein crystals. The resultant fragment hits have attained high-binding effectiveness per atom and frequently superior physicochemical functions in contrast to those from the HTS move toward that utilizes much bigger libraries of ~106 (Murray et al. 2012; Scott et al. 2012). FBDD diverges with admiration from the more recognized HTS in numerous features. Hence, as ligands become highly complex, the possibility of detecting related interactions falls noticeably in a given library (Hann et al. 2001).

Early FBDD assignments exploited crystallography (Lesuisse et al. 2002) or NMR (Fejzo et al. 1999) techniques as principal screening techniques. However, validation of hits is a fundamental constituent of the FBDD approach and must include a technique for estimating binding affinity. The comparatively fewer affinities mean that combined biophysical, biochemical, and structural methods should be utilized to monitor hit recognition, validation, and following explanation of lead molecules, as summarized in Fig. 5.1. The choice of techniques depends on factors including the accessibility of a perceptive biochemical test, the stability and solubility of the protein, the subsistence of crystals of apo-protein, and so on. However, numerous sets utilize a two-stage plan of HTS of fragment libraries utilizing fluorescence-based thermal shift measurements (Niesen et al. 2007; Scott et al. 2012), ligand-based NMR, SPR, and progressively with robotized screening amenities accessible on synchrotron beamlines and X-ray crystallographic screening. Hence, the fragment hits general between the methods are then confirmed

via optimization of the resolution of structures through X-ray diffraction and structure determination through NMR. The combination of these methods and others provides assurance of the quality of the hit. Hence, the validated fragment hits are then detailed via rising for a bigger molecular weight utilizing structure-guided methods (Renaud et al. 2016; Thomas et al. 2017).

FBDD has been extensively documented and has increased the pharmaceutical company's attractiveness as an influential option, complementing traditional HTS approaches for hit recognition (Turnbull and Boyd 2012). The initial fragment-derived drug, Vemurafenib, has been permitted, targeting a mutant type of BRAF and enlarging the lives of skin cancer patients. The drug Venetoclax, discovered by AbbVie and Genentech, binds with Bcl-2 and inhibits its interface with other proteins, which the US FDA has approved for CLL (Lampson and Davids 2017; Roberts et al. 2017). Ribociclib, by Novartis and Astex, has been approved to target Cdk4 and has been utilized in combination with letrozole for the treatment of advanced breast cancer (Peplow 2017; Yap et al. 2020). However, all of these campaigns attributed a significant function to structure-guided plans with a combination of action in comparatively small companies, frequently established through academics with an interest in protein structure, and identified as tumor therapeutics with the sturdy scientific and financial participation of big pharma (Atanasov et al. 2021; Thomas et al. 2017). Hence, natural products and structural analogs have traditionally made up the main contribution to pharmacotherapy, especially for cancer (Atanasov et al. 2021). However, several compounds/inhibitors are utilized for cancer therapy (Alam et al. 2021a, b, c, 2022).

5.6 Structure-Based Approaches in Cancer Therapeutics

The achievement of new tumor drug development depends on the inventive interaction between biology and chemistry. The traditional drug discovery procedure is an iterative cycle connecting chemical synthesis and biological assessments. However, the introduction of new methods has been essential mainly different high-throughput genomic plans for targeting and biomarker finding (Dalton and Friend 2006), HTS for hit identification (Clemons 2004; Wesche et al. 2005), as well as a structure-based design (Noble et al. 2004). However, numerous individual developments are increasing in speed and quality, emphasizing that the majority of clinical achievements to date have resulted from the close integration of various types of equipment and disciplines (biology, medicine, and chemistry), as well as the application of "joined-up thinking," particularly for addressing the problems that lead to drug failure in the clinic (Kola and Landis 2004). However, chemical biology assists assessment of compounds on a genome-broad range by interaction screens, which observe several biological systems concurrently in a fine-described way, as in the explication of kinase inhibitor selectivity outlines (Becker et al. 2004; Fabian et al. 2005); the recognition of attractive polypharmacy and combined therapies by the finding of synthetic lethality (Farmer et al. 2005; Morgan-Lappe et al. 2006).

Efforts to elucidate the molecular basis of tumors are not the latest. In the 1970s and 1980s, it was discovered that the discovery of tumor-causing oncogenes and TSGs undermined signal transduction pathways (Varmus 2006). Cancer drug development has squeezed molecular cancer as a resource of disease-causing targets for mechanism-based drug discovery (Workman 2005a, b). However, the initial generation of effective tumor drugs has been the cytotoxics that form the basis of most management and treatment regimens (Workman 2005b). However, the development and improvement of molecular cancer potential therapeutics has established that genes are involved. The procedure of utilizing cancer genes for developing molecular therapeutics and biomarkers is currently well recognized. The incorporation of these types is the basis for the progress of cancer medicine (Workman 2005a, b). Hence, oncogene activation and inactivation of TSGs are-frequently assisted via inactivation of DNA repair genes that affect genetic instability and lead to hijacking of signal transduction pathways and thus to the variety of well-identified phenotypic hallmark features of tumor (Vogelstein and Kinzler 2004). Oncogene products can be excellent targets; however, proteins downstream of the pathway might be suitable, including MEK1 and MEK2 in the MEK-ERK signaling and mTOR/PI3K signaling. However, the use of chemical probes has revealed that oncogenic support procedures including protein chaperoning (HSP90) and chromatin regulation (HDAC) may provide important drug targets (McDonald et al. 2006; Minucci and Pelicci 2006).

Medicinal chemistry persists as the stepwise cycles of drug design, biological evaluation, and chemical synthesis and recognizes SARs. The examination of the interface of ligands and proteins by co-crystallography updates and speeds up the procedure and was incredibly successful in kinase, HSP90, and HDAC inhibitor design. However, crystallographic recognition of promising small-molecule attaching positions surrounded by the bigger surfaces, which generally induce protein-protein interactions, was essential in determining pro-apoptotic agents that block MDM2-p53 and Bcl-2 (Alam et al. 2017, 2019; Fry and Vassilev 2005). Hence, structural biology has discovered numerous attaching modes for protein kinase inhibitors (Liu and Gray 2006; Noble et al. 2004). The sequence and structural relationship of kinases have looked like a potential liability; hence, they are identified as druggable targets (Cohen 2002). The binding position includes conserved features, which allow hydrogen bonding for adenosine, ribose, and phosphate constituents of ATP. MEK1 and MEK2 inhibitor PD318088 attach distantly and allosterically block enzyme action. Inhibition can engage attaching to locations that induce protein-protein interactions, as observed with ligands of the Akt pleckstrin homology domain (Barnett et al. 2005) and mTOR inhibitor rapamycin (Choi et al. 1996). The 3D structure is more apparently conserved among activated kinases than inactive forms. Hence, inhibitors that target stabilization of inactive enzymes may be highly selective (Liu and Gray 2006; Noble et al. 2004). Gatekeeper and mutational hot spots are documented invariants of Gleevec-resistant BCR-ABL (Gorre et al. 2001) and Iressa-resistant EGFR kinase (Paez et al. 2004). However, making a gatekeeper mutant of EGFR insensible to Iressa paralleled the

recognition of an EGFR T790M Iressa-resistant mutant in clinical studies (Blencke et al. 2003).

The result of kinase prevention on signaling might be probed via kinase mutants, which are susceptible to chemically orthogonal inhibitors not influencing the wild type (Shokat and Velleca 2002). Structure-based design was important in optimizing and identifying HSP90 inhibitors. However, the structure of natural products geldanamycin and radicicol bound to the N-terminal ATPase domain of HSP90 exposed the existence of an exclusive folding model in ATP attaching position, which comprises a network of firmly bound water molecules (Roe et al. 1999). However, co-crystallization of arylpyrazole HTS hit CCT018159 was revealed; the resorcinol motif of the small molecule utilizes the water network in a way similar to that of radicicol (Cheung et al. 2005). The initial synthetic inhibitors of HSP90 ATPase, a series of purines, have been identified from modeling based on HSP90-ATP (Chiosis et al. 2001). Following co-crystallography of the inhibitor, PU3, an unexpected change in conformation was discovered, forming a novel attaching pocket (Wright et al. 2004). However, iterative research of ligand-protein structures for the period of lead optimization is greatly expensive to understand the conformational changeability of the target and its outcome for drug design.

5.7 Structural Systems Biology in Cancer Therapeutics

The previous sections discussed structural biology applications, with an emphasis on X-ray crystallography. In this section, we will look at how structural and system biology come together to form a new field known as "structural systems biology." A large part of systems biology focuses on predicting the behavior of biological systems based on the molecules involved. As a result, understanding the interactions between these molecules is critical to these efforts. Although thousands of interactions are known, only a small fraction of them have precise molecular details. Because experimentally determining atomic structures for interacting proteins is difficult, predictive methods are essential for progress. In the end, structural details can transform abstract system representations into models that better reflect biological reality.

The amalgamation of structural and systems biology mainly focuses on the elucidation of protein-protein interactions as a part of significant biological pathways and structural details of the proteins involved in those pathways (Aloy and Russell 2006; Beltrao et al. 2007; Murray et al. 2021). Integration of systems biology with structural biology largely benefits from the power of the former in studying molecular entities as systems which function together to perform various complex tasks. This capability of systems biology improves the use of structural information of macromolecules involved in biological pathways, allowing researchers to get around the restrictions of examining such entities as isolated parts. Proteins, for example, are important macromolecules that play a role in complex regulatory networks that guide cell function. The majority of these proteins form complexes and interaction networks with other proteins, i.e., protein-protein interaction networks. Systems

biology plays a critical role in identifying and analyzing disease-related protein-protein interaction networks. Furthermore, protein-mediated expression regulation is a critical mechanism in a variety of disorders, and understanding these processes in depth allows for more effective disease management and the creation of better therapies. These interactions are crucial in the genesis and progression of cancer.

As previously discussed, the introduction of structural biology techniques and the expansion of protein structures have resulted in the production of more organized structural information, such as the origins of the Protein Data Bank. Systems biology makes extensive use of structural data to predict novel protein complexes (using protein-protein docking techniques), protein-protein interaction networks, and molecular pathways. Protein-protein docking (Vakser 2014) is a useful method for predicting protein complexes. These complexes perform vital biological functions such as transcription, DNA replication, translation, and other tasks. Irregularities in such cellular processes are associated with disease formation. Predicting and analyzing protein complexes can help researchers better understand the underlying causes of diseases like cancer. So far, several successful techniques and prediction systems for predicting protein-protein complexes based on the three-dimensional structure of the proteins have been created.

A key application of systems biology is the identification, analysis, and curation of biological pathways. As previously stated, proteins, as part of molecular pathways and interaction networks, regulate major cellular activities. As a result of various underlying circumstances, these pathways become disrupted in illness states. Therefore, a thorough understanding of these molecular pathways and their mechanisms of action is critical for disease therapeutic intervention. Incorporating structural information with these routes has numerous advantages and can aid in the comprehension of complex regulatory systems. When pathways are combined with structural information, they become more valuable for systems biology. It is easier to determine the affinity of an association when the nature of the interaction is understood.

5.8 Conclusion and Future Prospects

Crystallography is involved in multidisciplinary science, including biophysics, biology, medicinal chemistry, physics, mathematics, and earth sciences. Structural biology methods currently play a considerable role in the development of new cancer therapeutic drugs. However, X-ray crystallography has been recognized as the dominant instrument for developing new compounds. FBDD appeared with X-ray crystallography and has turned into potent screening equipment. Information about the 3D structures of protein targeting is nowadays the main function in drug discovery. However, its place in lead optimization is recognized with big groups of structural biologists employed in all the main pharmaceutical industries.

Structural biology and bioinformatics explain several key targets for cancer drug discovery: multidomain and multiprotein complexes. Drug discovery is an extremely complex and multidisciplinary activity involved in cancer therapeutics. A theory of tumor biology that is likely to be essential for targeted cancer

therapeutics is that of cancer heterogeneity, particularly the existence of cancer stem cells. However, a progressive advance in the development and improvement of truthfully personalized tumor medicine may be expected over the next few years.

Acknowledgments MA extends sincere thanks to the Indian Council of Medical Research for financial support (Grant No. 45/6/2020-DDI/BMS).

Declaration of Competing Interest The authors state that they have no known competing interests that could have appeared to influence the work recognized in this chapter.

Data Availability Statement The data that help the findings of this study are available in this chapter.

Author's Contribution Statement Manzar Alam: Conceptualization, Writing- Original draft preparation, Data curation, Investigation, Methodology. Ahmad Abu Turab Naqvi: Data curation, Investigation, Methodology. Md. Imtaiyaz Hassan: Conceptualization, Writing- Original draft preparation, Investigation, Supervision, project administration.

Funding This work is funded and supported by the Indian Council of Medical Research (Grant No. 45/6/2020-DDI/BMS) and (ISRM/12(22)/2020).

References

Adams MJ, Blundell TL, Dodson EJ, Dodson GG, Vijayan M, Baker EN, Harding MM, Hodgkin DC, Rimmer B, Sheat S (1969) Structure of rhombohedral 2 zinc insulin crystals. Nature 224: 491–495

Aitipamula S, Wong AB, Kanaujia P (2018) Evaluating suspension formulations of theophylline cocrystals with artificial sweeteners. J Pharm Sci 107:604–611

Alam M, Mishra R (2021) Bcl-xL expression and regulation in the progression, recurrence, and cisplatin resistance of oral cancer. Life Sci 2021:119705

Alam M, Kashyap T, Pramanik KK, Singh AK, Nagini S, Mishra R (2017) The elevated activation of NFκB and AP-1 is correlated with differential regulation of Bcl-2 and associated with oral squamous cell carcinoma progression and resistance. Clin Oral Investig 21:2721–2731

Alam M, Kashyap T, Mishra P, Panda AK, Nagini S, Mishra R (2019) Role and regulation of proapoptotic Bax in oral squamous cell carcinoma and drug resistance. Head Neck 41:185–197

Alam M, Ali S, Ahmed S, Elasbali AM, Adnan M, Islam A, Hassan M, Yadav DK (2021a) Therapeutic potential of ursolic acid in cancer and diabetic neuropathy diseases. Int J Mol Sci 22:12162

Alam M, Ali S, Mohammad T, Hasan GM, Yadav DK, Hassan M (2021b) B cell lymphoma 2: a potential therapeutic target for cancer therapy. Int J Mol Sci 22:10442

Alam M, Hasan GM, Hassan MI (2021c) A review on the role of TANK-binding kinase 1 signaling in cancer. Int J Biol Macromol 183:2364–2375

Alam M, Ali S, Ashraf GM, Bilgrami AL, Yadav DK, Hassan MI (2022) Epigallocatechin 3-gallate: from green tea to cancer therapeutics. Food Chem 379:132135

Aloy P, Russell RB (2006) Structural systems biology: modelling protein interactions. Nat Rev Mol Cell Biol 7:188–197

Anwar S, Shamsi A, Kar RK, Queen A, Islam A, Ahmad F, Hassan MI (2020) Structural and biochemical investigation of MARK4 inhibitory potential of cholic acid: Towards therapeutic implications in neurodegenerative diseases. Int J Biol Macromol 161:596–604

Atanasov AG, Zotchev SB, Dirsch VM, Supuran CT (2021) Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov 20:200–216

Atkinson A, Brown J, Fraser R, Leckie B, Lever A, Mackay A, Morton J, Robertson J, Pickering G (1980) The therapeutics of hypertension. Academic and The Royal Society of Medicine, London, pp 29–61

Barnett SF, Bilodeau MT, Lindsley CW (2005) The Akt/PKB family of protein kinases: a review of small molecule inhibitors and progress towards target validation. Curr Top Med Chem 5:109–125

Becker F, Murthi K, Smith C, Come J, Costa-Roldan N, Kaufmann C, Hanke U, Degenhart C, Baumann S, Wallner W et al (2004) A three-hybrid approach to scanning the proteome for targets of small molecule kinase inhibitors. Chem Biol 11:211–223

Beddell C, Goodford P, Norrington F, Wilkinson S, Wootton R (1976) Compounds designed to fit a site of known structure in human haemoglobin. Br J Pharmacol 57:201–209

Beg A, Khan FI, Lobb KA, Islam A, Ahmad F, Hassan MI (2019) High-throughput screening, docking, and molecular dynamics studies to identify potential inhibitors of human calcium/calmodulin-dependent protein kinase IV. J Biomol Struct Dyn 37:2179–2192

Beltrao P, Kiel C, Serrano L (2007) Structures in systems biology. Curr Opin Struct Biol 17:378–384

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 112:535–542

Blencke S, Ullrich A, Daub H (2003) Mutation of threonine 766 in the epidermal growth factor receptor reveals a hotspot for resistance formation against selective tyrosine kinase inhibitors. J Biol Chem 278:15435–15440

Blundell TL (1996) Structure-based drug design. Nature 384:23–26

Blundell TL, Cutfield JF, Cutfield SM, Dodson EJ, Dodson GG, Hodgkin DC, Mercola D, Vijayan M (1971) Atomic positions in rhombohedral 2-zinc insulin crystals. Nature 231:506–511

Blundell T, Dodson G, Hodgkin D, Mercola D (1972) Insulin: the structure in the crystal and its reflection in chemistry and biology. Adv Protein Chem 26:279–402

Blundell T, Sibanda BL, Pearl L (1983) Three-dimensional structure, specificity and catalytic mechanism of renin. Nature 304:273–275

Blundell TL, Jhoti H, Abell C (2002) High-throughput crystallography for lead discovery in drug design. Nat Rev Drug Discov 1:45–54

Bodo G, Dintzis H, Kendrew JC, Wyckoff H (1959) The crystal structure of myoglobin, V. A low-resolution three-dimensional Fourier synthesis of sperm-whale myoglobin crystals. Proc R Soc Lond Ser A Math Phys Sci 253:70–102

Caliandro R, Belviso DB, Aresta BM, de Candia M, Altomare CD (2013) Protein crystallography and fragment-based drug design. Future Med Chem 5:1121–1140

Campbell SF (2000) Science, art and drug discovery: a personal perspective. Clin Sci 99:255–260

Capila I, Linhardt RJ (2002) Heparin–protein interactions. Angew Chem Int Ed 41:390–412

Carr RA, Congreve M, Murray CW, Rees DC (2005) Fragment-based lead discovery: leads by design. Drug Discov Today 10:987–992

Cheung KM, Matthews TP, James K, Rowlands MG, Boxall KJ, Sharp SY, Maloney A, Roe SM, Prodromou C, Pearl LH et al (2005) The identification, synthesis, protein crystal structure and in vitro biochemical evaluation of a new 3,4-diarylpyrazole class of Hsp90 inhibitors. Bioorg Med Chem Lett 15:3338–3343

Chiosis G, Timaul MN, Lucas B, Munster PN, Zheng FF, Sepp-Lorenzino L, Rosen N (2001) A small molecule designed to bind to the adenine nucleotide pocket of Hsp90 causes Her2 degradation and the growth arrest and differentiation of breast cancer cells. Chem Biol 8:289–299

Choi J, Chen J, Schreiber SL, Clardy J (1996) Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. Science 273:239–242

Clemons PA (2004) Complex phenotypic assays in high-throughput screening. Curr Opin Chem Biol 8:334–338
Cohen P (2002) Protein kinases–the major drug targets of the twenty-first century? Nat Rev Drug Discov 1:309–315
Congreve M, Murray CW, Blundell TL (2005) Keynote review: structural biology and drug discovery. Drug Discov Today 10:895–907
Congreve M, Chessari G, Tisi D, Woodhead AJ (2008) Recent developments in fragment-based drug discovery. J Med Chem 51:3661–3680
Congreve M, Langmead CJ, Mason JS, Marshall FH (2011) Progress in structure based drug design for G protein-coupled receptors. J Med Chem 54:4283–4311
Dalton WS, Friend SH (2006) Cancer biomarkers–an invitation to the table. Science 312:1165–1168
Davies TG, Tickle IJ (2011) Fragment screening using X-ray crystallography. Top Curr Chem 317: 33–59
Davis AM, Teague SJ, Kleywegt GJ (2003) Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. Angew Chem Int Ed 42:2718–2736
Dhanaraj V, Dealwis C, Frazao C, Badasso M, Sibanda B, Tickle I, Cooper J, Driessen H, Newman M, Aguilar C (1992) X-ray analyses of peptide–inhibitor complexes define the structural basis of specificity for human and mouse renins. Nature 357:466–472
English AC, Done SH, Caves LS, Groom CR, Hubbard RE (1999) Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. Proteins 37: 628–640
Erlanson DA (2011) Introduction to fragment-based drug discovery. Top Curr Chem 317:1–32
Fabian MA, Biggs WH, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, Carter TA, Ciceri P, Edeen PT, Floyd M et al (2005) A small molecule-kinase interaction map for clinical kinase inhibitors. Nat Biotechnol 23:329–336
Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB, Santarosa M, Dillon KJ, Hickson I, Knights C et al (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 434:917–921
Fejzo J, Lepre CA, Peng JW, Bemis GW, Murcko MA, Moore JM (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. Chem Biol 6:755–769
Fitzpatrick PA, Steinmetz A, Ringe D, Klibanov AM (1993) Enzyme crystal structure in a neat organic solvent. Proc Natl Acad Sci 90:8653–8657
Fry DC, Vassilev LT (2005) Targeting protein-protein interactions for cancer therapy. J Mol Med 83:955–963
Garcia-Diaz M, Kunkel TA (2006) Mechanism of a genetic glissando: structural biology of indel mutations. Trends Biochem Sci 31:206–214
Garratt R (2013) Structural biology and cancer. Paper presented at: BMC proceedings
Gavira JA (2016) Current trends in protein crystallization. Arch Biochem Biophys 602:3–11
Goodford P, St-Louis J, Wootton R (1980) The interaction of human haemoglobin with allosteric effectors as a model for drug-receptor interactions. Br J Pharmacol 68:741
Gorre ME, Mohammed M, Ellwood K, Hsu N, Paquette R, Rao PN, Sawyers CL (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science 293:876–880
Gulzar M, Ali S, Khan FI, Khan P, Taneja P, Hassan MI (2019) Binding mechanism of caffeic acid and simvastatin to the integrin linked kinase for therapeutic implications: a comparative docking and MD simulation studies. J Biomol Struct Dyn 37:4327–4337
Gupta P, Mohammad T, Dahiya R, Roy S, Noman OMA, Alajmi MF, Hussain A, Hassan MI (2019a) Evaluation of binding and inhibition mechanism of dietary phytochemicals with sphingosine kinase 1: Towards targeted anticancer therapy. Sci Rep 9:18727
Gupta P, Mohammad T, Khan P, Alajmi MF, Hussain A, Rehman MT, Hassan MI (2019b) Evaluation of ellagic acid as an inhibitor of sphingosine kinase 1: a targeted approach towards anticancer therapy. Biomed Pharmacother 118:109245

Gupta P, Khan S, Fakhar Z, Hussain A, Rehman M, AlAjmi MF, Islam A, Ahmad F, Hassan M (2020) Identification of potential inhibitors of calcium/calmodulin-dependent protein kinase IV from bioactive phytoconstituents. Oxidative Med Cell Longev 2020(2020):2094635

Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. J Chem Inf Comput Sci 41:856–864

Hardy LW, Malikayil A (2003) The impact of structure-guided drug design on clinical agents. Curr Drug Discov 3:15–20

Harner MJ, Frank AO, Fesik SW (2013) Fragment-based drug discovery using NMR spectroscopy. J Biomol NMR 56:65–75

Hassan MI, Ahmad F (2011) Structural diversity of class I MHC-like molecules and its implications in binding specificities. In: Advances in protein chemistry and structural biology. Springer, New York, pp 223–270

Hassan MI, Waheed A, Grubb JH, Klei HE, Korolev S, Sly WS (2013) High resolution crystal structure of human beta-glucuronidase reveals structural basis of lysosome targeting. PLoS One 8:e79687

Hillisch A, Pineda LF, Hilgenfeld R (2004) Utility of homology models in the drug discovery process. Drug Discov Today 9:659–669

Hodgkin DC (1949) The X-ray analysis of the structure of penicillin. Adv Sci 6:85–89

Hol WG (1986) Protein crystallography and computer graphics—toward rational drug design. Angew Chem Int Ed 25:767–778

Hopkins AL, Keserü GM, Leeson PD, Rees DC, Reynolds CH (2014) The role of ligand efficiency metrics in drug discovery. Nat Rev Drug Discov 13:105–121

Hubbard RE (2005) 3D structure and the drug-discovery process. Mol BioSyst 1:391–406

James MN, Hsu I-N, Delbaere LT (1977) Mechanism of acid protease catalysis based on the crystal structure of penicillopepsin. Nature 267:808–813

Khan FI, Wei D-Q, Gu K-R, Hassan MI, Tabrez S (2016) Current updates on computer aided protein modeling and designing. Int J Biol Macromol 85:48–62

Khan P, Rahman S, Queen A, Manzoor S, Naz F, Hasan GM, Luqman S, Kim J, Islam A, Ahmad F et al (2017) Elucidation of dietary polyphenolics as potential inhibitor of microtubule affinity regulating kinase 4: in silico and in vitro studies. Sci Rep 7:9470

Khan P, Queen A, Mohammad T, Smita, Khan NS, Hafeez ZB, Hassan MI, Ali S (2019) Identification of alpha-mangostin as a potential inhibitor of microtubule affinity regulating kinase 4. J Nat Prod 82:2252–2261

Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 3:711–716

Lampson BL, Davids MS (2017) The development and current use of BCL-2 inhibitors for the treatment of chronic lymphocytic leukemia. Curr Hematol Malig Rep 12:11–19

Larsson A, Jansson A, Åberg A, Nordlund P (2011) Efficiency of hit generation and structural characterization in fragment-based ligand discovery. Curr Opin Chem Biol 15:482–488

Lee MJ, Yaffe MB (2016) Protein regulation in signal transduction. Cold Spring Harb Perspect Biol 8:005918

Lesuisse D, Lange G, Deprez P, Bénard D, Schoot B, Delettre G, Marquette J-P, Broto P, Jean-Baptiste V, Bichet P (2002) SAR and X-ray. A new approach combining fragment-based screening and rational drug design: application to the discovery of nanomolar inhibitors of Src SH2. J Med Chem 45:2379–2387

Li W, Zhang H, Assaraf YG, Zhao K, Xu X, Xie J, Yang D-H, Chen Z-S (2016) Overcoming ABC transporter-mediated multidrug resistance: molecular mechanisms and novel therapeutic drug strategies. Drug Resist Updat 27:14–29

Liu Y, Gray NS (2006) Rational design of inhibitors that bind to inactive kinase conformations. Nat Chem Biol 2:358–364

Lombardino JG, Lowe JA (2004) The role of the medicinal chemist in drug discovery—then and now. Nat Rev Drug Discov 3:853–862

Manzoni F, Wallerstein J, Schrader TE, Ostermann A, Coates L, Akke M, Blakeley MP, Oksanen E, Logan DT (2018) Elucidation of hydrogen bonding patterns in ligand-free, lactose-and glycerol-bound galectin-3C by neutron crystallography to guide drug design. J Med Chem 61:4412–4420

Maveyraud L, Mourey L (2020) Protein X-ray crystallography and drug discovery. Molecules 25: 1030

McDonald E, Workman P, Jones K (2006) Inhibitors of the HSP90 molecular chaperone: attacking the master regulator in cancer. Curr Top Med Chem 6:1091–1107

Milroy L-G, Grossmann TN, Hennig S, Brunsveld L, Ottmann C (2014) Modulators of protein–protein interactions. Chem Rev 114:4695–4748

Minucci S, Pelicci PG (2006) Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. Nat Rev Cancer 6:38–51

Mohammad T, Khan FI, Lobb KA, Islam A, Ahmad F, Hassan MI (2019) Identification and evaluation of bioactive natural products as potential inhibitors of human microtubule affinity-regulating kinase 4 (MARK4). J Biomol Struct Dyn 37:1813–1829

Mohammad T, Siddiqui S, Shamsi A, Alajmi MF, Hussain A, Islam A, Ahmad F, Hassan MI (2020) Virtual screening approach to identify high-affinity inhibitors of serum and glucocorticoid-regulated kinase 1 among bioactive natural products: Combined molecular docking and simulation studies. Molecules 25:823

Morgan-Lappe S, Woods KW, Li Q, Anderson MG, Schurdak ME, Luo Y, Giranda VL, Fesik SW, Leverson JD (2006) RNAi-based screening of the human kinome identifies Akt-cooperating kinases: a new approach to designing efficacious multitargeted kinase inhibitors. Oncogene 25: 1340–1348

Murray CW, Verdonk ML, Rees DC (2012) Experiences in fragment-based drug discovery. Trends Pharmacol Sci 33:224–232

Murray D, Petrey D, Honig B (2021) Integrating 3D structural information into systems biology. J Biol Chem 296:100562

Naqvi AA, Mohammad T, Hasan GM, Hassan M (2018) Advancements in docking and molecular dynamics simulations towards ligand-receptor interactions and structure-function relationships. Curr Top Med Chem 18:1755–1768

Naz H, Khan P, Tarique M, Rahman S, Meena A, Ahamad S, Luqman S, Islam A, Ahmad F, Hassan MI (2017) Binding studies and biological evaluation of beta-carotene as a potential inhibitor of human calcium/calmodulin-dependent protein kinase IV. Int J Biol Macromol 96:161–170

Naz F, Khan FI, Mohammad T, Khan P, Manzoor S, Hasan GM, Lobb KA, Luqman S, Islam A, Ahmad F (2018a) Investigation of molecular mechanism of recognition between citral and MARK4: a newer therapeutic approach to attenuate cancer cell progression. Int J Biol Macromol 107:2580–2589

Naz H, Tarique M, Khan P, Luqman S, Ahamad S, Islam A, Ahmad F, Hassan MI (2018b) Evidence of vanillin binding to CAMKIV explains the anti-cancer mechanism in human hepatic carcinoma and neuroblastoma cells. Mol Cell Biochem 438:35–45

Naz H, Tarique M, Ahamad S, Alajmi MF, Hussain A, Rehman MT, Luqman S, Hassan MI (2019) Hesperidin-CAMKIV interaction and its impact on cell proliferation and apoptosis in the human hepatic carcinoma and neuroblastoma cells. J Cell Biochem 120:15119–15130

Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2:2212–2221

Noble ME, Endicott JA, Johnson LN (2004) Protein kinase inhibitors: insights into drug design from structure. Science 303:1800–1805

Nussinov R, Tsai C-J, Shehu A, Jang H (2019) Computational structural biology: successes, future directions, and challenges. Molecules 24:637

Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ et al (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304:1497–1500

Peplow M (2017) Astex shapes CDK4/6 inhibitor for approval. Nat Biotechnol 35:395–396

Perutz D (1997) Structure of haemoglobin. Science 4:169
Perutz M, Liquori A, Eirich F (1951) X-ray and solubility studies of the haemoglobin of sickle-cell anaemia patients. Nature 167:929–931
Perutz M, Rossmann M, Cullis A, Muirhead H, Will G, North A (1960) Structure of hemoglobin. Paper presented at: Brookhaven Symp Biol
Rahuel J, Priestle JP, Grütter MG (1991) The crystal structures of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. J Struct Biol 107:227–236
Read RJ, Adams PD, Arendall WB III, Brunger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Lütteke T, Otwinowski Z (2011) A new generation of crystallographic validation tools for the protein data bank. Structure 19:1395–1412
Renaud J-P, Chung C-W, Danielson UH, Egner U, Hennig M, Hubbard RE, Nar H (2016) Biophysics in drug discovery: impact, challenges and opportunities. Nat Rev Drug Discov 15: 679–698
Roberts AW, Stilgenbauer S, Seymour JF, Huang DC (2017) Venetoclax in patients with previously treated chronic lymphocytic leukemia. Clin Cancer Res 23:4527–4533
Roe SM, Prodromou C, O'Brien R, Ladbury JE, Piper PW, Pearl LH (1999) Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. J Med Chem 42:260–266
Sanger F (1988) Sequences, sequences, and sequences. Annu Rev Biochem 57:1–29
Schelling P, Ganten U, Sponer G, Unger T, Ganten D (1980) Components of the renin-angiotensin system in the cerebrospinal fluid of rats and dogs with special consideration of the origin and the fate of angiotensin II. Neuroendocrinology 31:297–308
Schlichtkrull J (1958) Insulin crystals. Munksgaard, Copenhagen
Scott DE, Coyne AG, Hudson SA, Abell C (2012) Fragment-based approaches in drug discovery and chemical biology. Biochemistry 51:4990–5003
Shamim Jairajpuri D, Hussain A, Nasreen K, Mohammad T, Anjum F, Rehman T, Mustafa Hasan G, Alajmi MF, Hassan I (2021) Identification of natural compounds as potent inhibitors of SARS-CoV-2 main protease using combined docking and molecular dynamics simulations. Saudi J Biol Sci 28(4):2423–2431
Shokat K, Velleca M (2002) Novel chemical genetic approaches to the discovery of signal transduction inhibitors. Drug Discov Today 7:872–879
Subramanian E, Swan I, Liu M, Davies D, Jenkins J, Tickle I, Blundell T (1977) Homology among acid proteases: comparison of crystal structures at 3A resolution of acid proteases from Rhizopus chinensis and Endothia parasitica. Proc Natl Acad Sci 74:556–559
Surade S, Blundell TL (2012) Structural biology and drug discovery of difficult targets: the limits of ligandability. Chem Biol 19:42–50
Thakral NK, Zanon RL, Kelly RC, Thakral S (2018) Applications of powder X-ray diffraction in small molecule pharmaceuticals: achievements and aspirations. J Pharm Sci 107:2969–2982
Thomas SE, Mendes V, Kim SY, Malhotra S, Ochoa-Montaño B, Blaszczyk M, Blundell TL (2017) Structural biology and the design of new therapeutics: from HIV and cancer to mycobacterial infections: a paper dedicated to John Kendrew. J Mol Biol 429:2677–2693
Tickle I, Sibanda B, Pearl L, Hemmings A, Blundell T (1984) Protein crystallography, interactive computer graphics & drug design X-ray. In: Crystallography Drug Design. Clarendon Press, Oxford, pp 427–444
Turnbull PA, Boyd MS (2012) Targeting cancer using fragment based drug discovery. Anti Cancer Agents Med Chem 12:40–48
Vakser IA (2014) Protein-protein docking: from interaction to interactome. Biophys J 107:1785–1793
Van Montfort RL, Workman P (2017) Structure-based drug design: aiming for a perfect fit. Essays Biochem 61:431–437
Varmus H (2006) The new era in cancer research. Science 312:1162–1165

Vénien-Bryan C, Li Z, Vuillard L, Boutin JA (2017) Cryo-electron microscopy and X-ray crystallography: complementary approaches to structural biology and drug discovery. Acta Crystallogr Sect F 73:174–183

Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. Nat Med 10:789–799

Wesche H, Xiao SH, Young SW (2005) High-throughput screening for protein kinase inhibitors. Comb Chem High Throughput Screen 8:181–195

Whittle PJ, Blundell TL (1994) Protein structure-based drug design. Annual review of biophysics and biomolecular structure. Annu Rev Biophys Biomol Struct 23:349–375

Workman P (2005a) Drugging the cancer kinome: progress and challenges in developing personalized molecular cancer therapeutics. Cold Spring Harb Symp Quant Biol 70:499–515

Workman P (2005b) Genomics and the second golden era of cancer drug development. Mol BioSyst 1:17–26

Wright L, Barril X, Dymock B, Sheridan L, Surgenor A, Beswick M, Drysdale M, Collier A, Massey A, Davies N et al (2004) Structure-activity relationships in purine-based inhibitor binding to HSP90 isoforms. Chem Biol 11:775–785

Wyss DF, Wang Y-S, Eaton HL, Strickland C, Voigt JH, Zhu Z, Stamford AW (2011) Combining NMR and X-ray crystallography in fragment-based drug discovery: discovery of highly potent and selective BACE-1 inhibitors. Top Curr Chem 317:83–114

Yap YS, Chiu J, Ito Y, Ishikawa T, Aruga T, Kim SJ, Toyama T, Saeki T, Saito M, Gounaris I et al (2020) Ribociclib, a CDK 4/6 inhibitor, plus endocrine therapy in Asian women with advanced breast cancer. Cancer Sci 111:3313–3326

Yuriev E, Ramsland PA (2013) Latest developments in molecular docking: 2010–2011 in review. J Mol Recognit 26:215–239

Zheng H, Handing KB, Zimmerman MD, Shabalin IG, Almo SC, Minor W (2015) X-ray crystallography over the past decade for novel drug discovery–where are we heading next? Expert Opin Drug Discovery 10:975–989

Computational Tools and Databases for Fusion Transcripts: Therapeutic Targets in Cancer

6

Aditya Narayan, Bhavya Pahwa, and Shailesh Kumar

Abstract

In recent years, a rapid expansion in the field of RNAomics has led to a steep rise in data regarding expressed genes. This expansion in data has necessitated a consequent increase in the breadth and depth of tools which may be used for the study of RNA types. Gene fusions are considered hallmarks of many cancer types and may occur through chromosomal rearrangement or through noncanonical mechanisms in which chimeric RNA forms without rearrangement of the genome. To more effectively identify, validate, and understand the function of these novel RNA molecules, we present this chapter as a resource. In it, we discuss the role of fusion transcripts, identification of fusion transcripts, relevant software packages, and databases.

6.1 Introduction

Gene fusions are often considered to be a common feature present in cancer cells and present with rare cytogenetic signatures which may offer applications for disease identification, characterization, and treatment. Gene fusions are genes that possess DNA sequences from two different parental genes and may be created through

A. Narayan
University of Virginia, MR6 (Carter-Harrison Research Building), Charlottesville, VA, USA
e-mail: aditnara@stanford.edu

B. Pahwa
University College of Medical Sciences and GTB Hospital, New Delhi, Delhi, India

S. Kumar (✉)
Bioinformatics Lab, National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi, India
e-mail: shailesh@nipgr.ac.in

S. Singh (ed.), *Systems Biomedicine Approaches in Cancer Research*,
https://doi.org/10.1007/978-981-19-1953-4_6

several mechanisms including chromosomal translocations, inversions, deletions, and duplications. This may lead to proteins with domains derived from two genes in a novel fusion protein, a shift in reading frame, rearrangements of gene regulatory elements, and so on.

There has been a wide array of efforts to understand their prevalence, mechanism of creation, and function which have in turn lead to improvements in the ability to study several cancer subtypes. Broadly, well-studied examples of such gene fusions in cancer cells are described below:

- The first identified chromosome abnormality which was found to be strongly associated with cancer is BCR-ABL1, or the Philadelphia chromosome. This fusion of the BCR and ABL1 genes resulting from a reciprocal translocation event leads to the creation of a constitutively active tyrosine kinase (Ren 2005).
- In Burkitt's lymphoma, an aggressive mature B-cell neoplasia, chromosomal rearrangement leads to the creation of the IGH-MYC fusion and subsequent overexpression of the c-myc oncogene, a transcription factor which in turn leads to lymphomagenesis as well as accumulation of double-strand breaks in DNA (Yan et al. 2007).

While recurrent fusion genes are often associated with cancer phenotypes, fusion events are not necessarily limited to oncogenic processes. The formation of fusion genes in normal, noncancerous cells has been identified and has been shown to contribute to the development of more complex, multidomain proteins. This, in turn, contributes to protein evolution over longitudinal time scales.

While fusion genes are defined by the combination of DNA sequences, their precursor, chimeric RNAs, are hybrid RNA transcripts which contain nucleotides from different parental genes. These chimeric RNAs are not necessarily produced through the creation of the fusion of genes at the genomic level, and instead refer more broadly to any hybrid transcript based on gene annotations (Elfman et al. 2020). A critical reason for this distinction is that many means of chimeric RNA production in which there are no changes to the corresponding genome have been elucidated.

- Chimeric RNAs may be produced through the process of intergenic splicing. This most commonly occurs through a read-through of genes which lie in cis to create a hybrid mRNA. This is referred to as cis-splicing of adjacent genes (cis-SAGe) and has been found to be the primary way in which chimeric RNA production forms in noncancer cells (Singh et al. 2020).
- Chimeric RNA may form from parental genes which are found on different chromosomes. This process may be referred to as trans-splicing and is theorized to take place through splicing of precursor mRNAs (Jia et al. 2016).
- Parental genes may be separated by large linear distances on the same chromosomes.
- Trans-splicing of sense and antisense transcripts may occur between sense and antisense transcripts of a single gene.

- "Back-splicing" in which a downstream gene is transcribed prior to an upstream gene, leading to the creation of a circular fusion chimera (Wu et al. 2019).

Broadly, due to their ubiquity, varying mechanisms of formation, and diversity, chimeric RNAs are seen as a way in which the body expands the functional genome. Our understanding of chimer RNA identification and function has been expanded through large-scale analysis of datasets (GTEx, TCGA, Ensembl, etc.). The creation of such datasets has rapidly expanded in recent years given the advent of novel sequencing technologies which have improved researchers' access to critical insights. Despite this veritable explosion of databases and associated software tools, our knowledge of chimeric RNAs remains incomplete. Challenges which arise in the study of chimeric RNA are numerous, but not insurmountable. These are as follows:

- Relatively low levels of chimeric RNA expression may lead to biased statistical analyses leading to over or underestimation of certain sequences.
- The possibility of chimeric RNA developing from template switching events during RT-PCR.
- The only unique sequence in the chimeric RNA lies at the relatively small junction between the two parent sequences.
- Homology between chimeric RNA and parental genes causes bioinformatic predictive tools, biochemical techniques such as sequence-targeted assays, and the like to be particularly challenging.

Despite the barriers to studying this novel class of molecules, chimeric RNA is a fertile field of study for geneticists at all levels. To better support future generations of researchers in exploring this field, this chapter will explore software tools which may be applied to effectively identify and characterize chimeric RNA.

6.1.1 Identification of Chimeric RNA

A wide array of chimeric RNA prediction tools exists to support researchers in their search for potential candidates. These tools employ RNA-Seq datasets as purely genomic DNA datasets do not encompass the full potential for chimeric RNA production. This is since genomic instability is a hallmark of cancer cells. Further, RNA seq captures only the expressed parts of the genome (exome) which are transcriptionally active. This reduces the cost of the entire process to detect fusion transcripts in cancer. Datasets may be obtained from online databases such as TCGA. TCGA offers the GDC Data Transfer Tool to download raw sequencing data and we suggest that readers attempt to apply these tools for themselves. Raw data must first be processed by applying basic software tools to ensure their usage for various detection software.

- First, it is necessary to engage in quality control of raw sequencing reads. Example tools for this purpose are described below:
- ClinQC: This a highly accessible pipeline which may be used for converting between raw data formats, quality control, and trimming of raw sequencing data from both Sanger and NGS sequencing platforms. Data are converted to FASTQ format, which is among the most accepted formats for chimeric RNA prediction software which will be described later. The software prepares a quality control report which further facilitates downstream analysis (Pandey et al. 2016).
- NGSQC Toolkit: The NGSQC toolkit allows for highly efficient processing of NGS data with filtration of high-quality results as well as quality checking. It is an open-source application freely available online and implemented in perl. The toolkit offers high ease of use and is effective for sequencing data sourced from Rorche 454 and Illumina platforms (Patel and Jain 2012).
- FastQC: FastQC is one of the most applied tools to perform quality control screens on raw data from high-throughput sequencing methods. It allows for the import of data as BAM, SAM, or FastQ formats, creates summary graphics to assess data, and allows for export of data as various file types (Andrews n.d.).

Following quality control steps and conversion to file types appropriate for downstream analysis, it is possible to pass the data to fusion transcript prediction tools. There are over 35 software tools that are implicated in the identification of fusion transcripts. Fusion detection occurs in three stages, namely, (1) mapping and filtering, (2) fusion junction detection, and (3) fusion assembly and selection.

These mechanisms establish the categorical basis of division of tools for the identification of fusion transcripts.

1. *Mapping and filtering*: This is the initiation step in the identification of fusion transcripts and much software are based solely on this principle. After mapping is completed, pairs are evaluated for alignment and the irrelevant reads are removed. This is referred to as split mapping. Example software which applies split mapping are FusionMap and TopHat-Fusion. While some software, for example SnowShoes-FTD, utilizes spanning reads in which all mapped reads are preserved without filtration. Further incorrect reads are discarded by filtering techniques, as exemplified by FusionSeq with ten filters to remove illegitimate fusions. One such filter is when the fusion is intrachromosomal, such that the two genes are located on the same chromosome, and they can be recognized as a read-through transcript. This is applied by tools such FusionMap, FusionHunter, ShortFuse, SnowShoes-FTD, and TopHat-Fusion.
2. *Fusion junction detection*: This is the second step in fusion junction detection via "split read" mapping. It involves the independent alignment of first and last segments of the each "split read" that are generated by the discarding of unmapped reads in the previous stage. Alignment patterns are recognized, boundaries of the original fragments are adjusted, and realignment is performed to accurately identify fusion transcripts. Split read mapping is influenced by the size of partitioned segments. Small fragments not only sensitize the process but

are also more likely to provide false-positive results. To combat this, either the read is further split into two segments or a fixed "proposed" segment is utilized. Spanning reads facilitate the detection of fusion breakpoints followed by extraction of candidates by split read.

3. *Fusion assembly and selection*: In this step, mapped reads are referred to as "supporting reads". Owing to the presence of fusion junctions in the insert sequences, spanning reads also are good supporting modalities. Supporting reads are beneficial in the sense that they help in eliminating false candidates; however, the risk of true-positive results which are simply expressed at low levels being removed increases at the same rate. This problem is tackled by the availability of scoring functions in the tools. These functions are dependent on factors like, read depth, mapping quality, and number of supporting reads. Final scores are derived via empirical analysis (FusionSeq) or machine learning modalities (deFuse).

Here we describe the basic applications for commonly used tools.

6.1.1.1 FusionSeq

FusionSeq is a computational suite designed to detect candidate chimeric RNA/gene fusions through analysis of paired-end RNA seq data and offers high ease-of-use given that it is able to function irrespective of the mapping approach. The output of FusionSeq is a list of high confidence fusion candidates which are scored to provide for ease of follow-up validation studies. The results are accessible through a web browser. Drawbacks to the use of FusionSeq arise when considering the high CPU time and memory usage, particularly when analyzing large numbers of samples in parallel. This is because FusionSeq selects for all possible exons involved in the junction sequence and produces a junction library from all possible pairs of "tiles" which cover the exons and are each offset by one nucleotide. RNA-seq reads are mapped to these junctions but particularly with higher exon counts, this approach can be time consuming due to somewhat inefficient screening of false-positives (Sboner et al. 2010).

6.1.1.2 TopHat

TopHat is an algorithm to identify chimeric RNA transcripts representing fusion gene products. TopHat-Fusion is the most recent and updated version of this tool and offers the ability to align reads across fusion junctions. The software accepts and aligns RNA-seq reads but critically, does not rely on gene annotation. This is relevant as it allows for the tool to identify novel fusions which are derived from parental genes which are known, unknown, or unannotated variants of known genes (Kim and Salzberg 2011).

6.1.1.3 JAFFA

Frequently, methods used for the identification of chimeric RNA are designed for use with short read lengths. JAFFA is a software tool which compares cancer transcriptomes to references, as opposed to the genome and is optimized for read

lengths that are 100 bp or greater. The cancer transcriptome is inferred through long reads or de novo assembly of short reads.

JAFFA operates through a pipeline in which RNA-seq reads serve as the input and candidate fusion genes with breakpoint sequences are the output. Features include the presence of three modes which vary in appropriateness based on input read length: Assembly (wherein short reads are assembled de novo into contigs prior to detecting fusions), Direct (RNA-seq reads which do not map to known transcripts are employed), and Hybrid (combination of direct and assembly approaches) (Davidson et al. 2015).

6.1.1.4 EricScript

EricScript (chimERIC tranSCRIPT detection algorithm) is a tool for the detection of chimeric transcripts in paired-end RNA seq data (Benelli et al. 2012). This software differs from other prediction tools in that it is highly efficient due to its use of an exon junction reference which allows for reduced run times. Importantly, the package presents scores that allow for highly efficient detection of true from false-positive transcripts, which is a common challenge when distinguishing between potential fusions. For researchers, this scoring mechanism allows for efficient screening of potential output transcripts and allows for a reduced number of targets for data analyses. A study performed by Kumar et al. identified that EricScript was distinguished in its balance between time and memory requirements relative to sensitivity (Kumar et al. 2016).

6.1.1.5 SOAPfuse

SOAPfuse is an open-source tool that may be applied for the detection of fusion transcripts from paired-end RNA-seq data inputs. It can identify features of RNA-seq datasets such as insert size and read length, so full homogeneity of the dataset is less critical. This software was developed in perl and is limited in that it is only executable in Linux OS. SOAPfuse functions through alignment of RNA-seq paired-end reads against human reference sequences to detect candidate fusions. It employs both discordant mapping paired end reads as well as junction reads to confirm the sites. A junction library is constructed and is used to filter out false-positive fusions. The output of the program is a list of high likelihood fusions as well as their locations, junction sequences with single-nucleotide resolution, and diagrams displaying the varying location of reads relative to junction sequences and exon expression levels. This output data allows for effective follow-up analysis (Jia et al. 2013).

6.1.1.6 STARChip

A rapidly expanding field of research states that circular isoforms of RNA are expressed across the genome and may be correlated with disease. The value in detecting such nonlinear RNA alignments lies in the fact that it allows for more rapid detection of chromosomal rearrangements which are commonly associated with cancer. STAR Chimeric Post (STARChip) is a software package which applies the STAR aligner to chimeric alignments in order to produce annotated circRNA and

fusions. This tool is effective for high-dimensional datasets and offers high performance at relatively low computing time (Akers et al. 2018).

6.1.1.7 FuSeq

FuSeq is a fusion detection method which applies a recent quasi-mapping method for alignment which allows it to operate with far lower computational time than many other tools. The tool functions through a pipeline for mapped read-pairs and another junction split-reads. Following the process, false-positive results are minimized through application of a range of filters (Vu et al. 2018). Additional tools are summarized in Table 6.1.

6.2 Fusion Transcripts Databases

There are several fusion transcripts databases available for scientific community. Almost all these resources are freely available, harboring the information of fusion coordinates, tissue, condition, sample information, cancer type, etc. Our research group also developed a database of fusion transcripts for model plant *Arabidopsis thaliana*. Most popular fusion transcripts databases are mentioned in the Table 6.2.

6.3 Validation of Transcripts

Following the generation of potential fusion transcript lists, there is a wide range of possible approaches to validating the chimeric RNA and ensuring that they are not false-positive results. Some of the most readily applied approaches are described below:

- In-Silico Validation: By utilizing the predicted junction sequence at the breakpoint between parental gene sequences, it is possible to identify commonly expressed chimeric RNA. This is performed by searching for the junction sequence in the raw RNA sequencing reads using string-matching software.
- Validation Through Query of Online Databases: Databases containing chimeric expressed sequence tags and junction sequences and may be queried for certain sequences to determine if they have been previously validated. Table 6.2 describes some of these databases.
- Application of Wet Lab Approaches: Reverse-transcription polymerase chain reaction (RT-PCR) may be used to detect and measure the expression of chimeric RNA transcripts. After isolating RNA from a sample and creating cDNA, PCR may be applied specifically to the junction sequence to determine expression levels. Primers may be designed such that they flank this unique junction sequence and allow the researcher to amplify the sequence if present.

Table 6.1 Summarizes the tools available to identify fusion transcripts along with their methodologies

Method	Brief overview of methodology
Arriba (Uhrig 2019)	Arriba extracts gene fusions from the chimeric alignments reported by STAR (Dobin et al. 2013) by applying a collection of filters which recognize frequent types of artifacts found in RNA-Seq data
ChimeraScan (Iyer et al. 2011)	Identifies candidate fusions from discordant Bowtie (Langmead et al. 2009) genome alignments. Unmapped reads are trimmed and realigned. Junction breakpoint reads are resolved by aligning to candidate fused exons. Fusions are filtered based on an abundance of fusion-supporting reads
ChimPipe (Rodriguez-Martin et al. 2017)	The GEMtools RNA-seq pipeline (GEMTools 2019) and GEM alignment utility (Marco-Sola et al. 2012) are used to capture discordant and chimeric read alignments, and fusion candidates are filtered according to fusion evidence and additional gene-based filters
deFuse (McPherson et al. 2011)	Aligns reads to spliced and unspliced gene sequences using Bowtie (Langmead et al. 2009), resolves split read junctions using a novel dynamic programming algorithm, and uses an AdaBoost classifier to discriminate between likely true versus false fusions
EricScript (Benelli et al. 2012)	BWA (Li and Durbin 2009) is used to align reads to the genome. Discordant reads are used to identify candidate gene fusions. BLAT (Kent 2002) is then used in an iterative local alignment step to define precise fusion breakpoints by aligning to customized targets of fused exons. An AdaBoost classifier trained with synthetic data is used to score and rank fusion predictions
FusionCatcher (Nicorici et al. 2014)	Leverages a collection of alignment utilities including Bowtie (Langmead et al. 2009), Bowtie2 (Langmead and Salzberg 2012), BLAT (Kent 2002), and STAR (Dobin et al. 2013) with a collection of customized target databases to identify and characterize fusion candidates. Rigorous filtering of fusion predictions according to gene and fusion annotations is employed
FusionHunter (Li et al. 2011)	First uses Bowtie to align reads to the genome and identify candidate fusions based on discordant read pairs. Then creates a "pseudoreference" by positioning candidate fusion genes with canonical ordering, realigns reads using a custom algorithm, and identifies both split and spanning reads providing evidence for gene fusions
InFusion (Okonechnikov et al. 2016)	Reads are first aligned to the reference transcriptome using Bowtie2. Unaligned and discordantly aligned reads are further examined in the context of the genome and transcriptome to cluster evidence and define candidate fusions
JAFFA-Assembly (Davidson et al. 2015)	After removing intronic and intergenic region aligning reads defined by Bowtie genome alignments, the remaining reads are assembled using Oases (Schulz et al. 2012) and the assembled contigs are mapped directly to the transcriptome using BLAT. Chimeric BLAT alignments are further assessed as fusion candidates

(continued)

Table 6.1 (continued)

Method	Brief overview of methodology
JAFFA-Direct (Davidson et al. 2015)	After removing intronic and intergenic region aligning reads defined by Bowtie genome alignments, the remaining reads are mapped directly to the transcriptome using BLAT. Chimeric BLAT alignments are further assessed as fusion candidates
JAFFA-Hybrid (Davidson et al. 2015)	After removing intronic and intergenic region aligning reads defined by Bowtie genome alignments, the remaining reads are assembled using Oases. Both the assembled transcripts and the original reads that failed to map to the genome are then mapped directly to the transcriptome using BLAT. Chimeric BLAT alignments are further assessed as fusion candidates
MapSplice (Wang et al. 2010)	An RNA-seq aligner based on Bowtie similar to TopHat (Trapnell et al. 2009) and includes fusion-finding capabilities, although specific algorithmic details are lacking
nFuse (McPherson et al. 2012)	Designed for use with WGS-seq and RNA-seq but can be executed with RNA-seq only, leveraging its included deFuse with Bowtie2
Pizzly (Melsted et al. 2017)	Uses a k-mer-based strategy to examine reads that do not map to isoforms consistently via kallisto (Bray et al. 2016) pseudoalignment
PRADA (Torres-Garcia et al. 2014)	Reads are aligned to a combined genome and transcriptome reference using BWA. Discordant reads identify fusion candidates, and junction reads are identified by mapping to a database of all possible 5′-3′ chimeric exon junction database
SOAP-fuse (Jia et al. 2013)	The SOAP2 aligner (Hurgobin 2016) is used to map reads to genomes and spliced transcripts to identify fusion candidates
STARChip (Akers et al. 2018)	Uses chimeric reads reported by STAR aimed primarily at identifying circular RNAs but also reports fusion candidates
STAR-Fusion (Haas 2019a)	Uses chimeric read alignments reported by STAR in its Chimeric.out.junction file to identify candidate fusions followed by extensive filtering of likely artifacts
STAR-SEQR (STAR-SEQR 2019)	Uses chimeric reads reported by STAR to find fusions
TopHat-Fusion (Kim and Salzberg 2011)	A modified execution of the TopHat aligner (Trapnell et al. 2009; Kim et al. 2013) to examine initially unmapped reads as supporting fusion events
TrinityFusion-C (Haas 2019b)	De novo assembles only the chimeric reads defined by STAR using the Trinity assembler (Tomczak et al. 2015), and subsequently leverages GMAP (Jang et al. 2020; Kim and Zhou 2019) for chimera candidate detection
TrinityFusion-D (Haas 2019b)	De novo assembles all input reads using Trinity, and subsequently leverages GMAP for chimera candidate detection
TrinityFusion-UC (Haas 2019b)	De novo assembles both chimeric and unmapped reads defined by STAR using the Trinity assembler, and subsequently leverages GMAP for chimera candidate

Table 6.2 Different databases available to identify fusion transcripts along with their methodologies

Database	Brief overview of database
The Cancer Genome Atlas (TCGA) (Tomczak et al. 2015)	TCGA seeks to create a comprehensive profile of genomic alterations associated with cancers through profiling human tumor cohorts
ChimerDB (Jang et al. 2020)	ChimerDB is one of the most comprehensive databases available for the study of gene fusions. It includes deep sequencing data as well as information from publications
Fusion Gene Annotation Database (FusionGDB) (Kim and Zhou 2019)	FusionGDP provides functional annotations as well as information on protein structure, fusion transcript amino acid sequences, breakpoint mapping, and the like for a range of known fusion genes
FusionCancer (Wang et al. 2015)	FusionCancer is a database based on gene fusion identification from RNA-seq datasets in human cancers. This is a query engine with annotated information of cancer fusion genes and which offers high ease of use for researchers
FusionHub (Panigrahi et al. 2018)	This is a web platform which allows for querying of multiple gene fusion databases. It allows for multiple visualization approaches and allows for ease of annotation
AtFusionDB (Singh et al. 2019)	AtFusionDB is a comprehensive database which contains fusion transcript information specific to Arabidopsis thaliana. There are a variety of annotation tools, search modules, and visualization approaches which facilitate the study of plant genomes
ChiTaRS (Balamurali et al. 2020)	ChiTaRS is an incredibly comprehensive chimeric transcript database with annotated information from eight species' genomes. A number of features exist within the database including information on druggable fusion targets and transcripts with clinical correlates

6.4 Conclusion

In the years following the discovery of the Philadelphia chromosome, there has been an explosion of evidence supporting gene rearrangements as correlates and/or causative agents of oncogenesis. This has led to databases, software tools, and biochemical techniques which have allowed for increasingly efficient and effective analysis of the novel field of chimeric RNA production. While the majority chimeric RNA is of unclear functional significance, advances in genomic editing approaches may expand the potential for novel explorations of the functional genome.

References

Akers NK, Schadt EE, Losic B (2018) STAR chimeric post for rapid detection of circular RNA and fusion transcripts. Bioinformatics 34(14):2364–2370. https://doi.org/10.1093/bioinformatics/bty091

Andrews S (n.d.) FASTQC: a quality control tool for high-throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Balamurali D, Gorohovski A, Detroja R, Palande V, Raviv-Shay D (2020) Milana Frenkel-Morgenstern. ChiTaRS 5.0: the comprehensive database of chimeric transcripts matched with druggable fusions and 3D chromatin maps. Nucleic Acids Res 48(1):825–834. https://doi.org/10.1093/nar/gkz1025

Benelli M et al (2012) Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. Bioinformatics 28(24):3232–3239

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34(5):525–527

Davidson NM, Majewski IJ, Oshlack A (2015) JAFFA: high sensitivity transcriptome-focused fusion gene detection. Genome Med 7(1):43. https://doi.org/10.1186/s13073-015-0167-x

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1):15–21

Elfman J, Pham L-P, Li H (2020) The relationship between chimeric RNAS and gene fusions: Potential implications of reciprocity in cancer. J Genet Genomics 47(7):341–348

GEMTools (2019) GEMTools 2019. Available from http://gemtools.github.io/

Haas BJ (2019a) STAR-fusion code and documentation on GitHub. Available from https://github.com/STAR-Fusion/STAR-Fusion/wiki

Haas BJ (2019b) TrinityFusion - fusion and foreign transcript detection via RNA-seq de novo assembly. Available from https://github.com/trinityrnaseq/TrinityFusion/wiki

Hurgobin B (2016) Short read alignment using SOAP2. Methods Mol Biol 1374:241–252

Iyer MK, Chinnaiyan AM, Maher CA (2011) ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics 27(20):2903–2904

Jang YE, Jang I, Kim S, Cho S, Kim D, Kim K, Kim J, Hwang J, Kim S, Kim J, Kang J, Lee B, Lee S, Chimer DB (2020) 4.0: an updated and expanded database of fusion genes. Nucleic Acids Res 48(1):817–824

Jia W, Qiu K, He M et al (2013) SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol 14:12. https://doi.org/10.1186/gb-2013-14-2-r12

Jia Y, Xie Z, Li H (2016) Intergenically spliced chimeric RNAs in cancer. Trends Cancer 2(9):475–484. https://doi.org/10.1016/j.trecan.2016.07.006

Kent WJ (2002) BLAT–the BLAST-like alignment tool. Genome Res 12(4):656–664

Kim D, Salzberg SL (2011) TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol 12(8):72. https://doi.org/10.1186/gb-2011-12-8-r72

Kim P, Zhou X (2019) FusionGDB: fusion gene annotation DataBase. Nucleic Acids Res 47(1):994–1004. https://doi.org/10.1093/nar/gky1067

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14(4):R36

Kumar S, Vo A, Qin F et al (2016) Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. Sci Rep 6:21597. https://doi.org/10.1038/srep21597

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357–359

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):25

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760

Li Y, Chien J, Smith DI, Ma J (2011) FusionHunter: identifying fusion transcripts in cancer using paired-end RNA-seq. Bioinformatics 27(12):1708–1710

Marco-Sola S, Sammeth M, Guigo R, Ribeca P (2012) The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods 9(12):1185–1188

McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG et al (2011) deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput Biol 7(5):e1001138

McPherson A, Wu C, Wyatt AW, Shah S, Collins C, Sahinalp SC (2012) nFuse: discovery of complex genomic rearrangements in cancer using high-throughput sequencing. Genome Res 22(11):2250–2261

Melsted P, Hateley S, Joseph IC, Pimentel H, Bray N, Pachter L (2017) Fusion detection and quantification by pseudoalignment. bioRxiv. 2017:166322. https://doi.org/10.1101/166322

Nicorici D, Satalan M, Edgren H, Kangaspeska S, Murumagi A, Kallioniemi O, Virtanen S, Kilkku O et al (2014) FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv 011650. https://doi.org/10.1101/011650

Okonechnikov K, Imai-Matsushima A, Paul L, Seitz A, Meyer TF, Garcia-Alcalde F (2016) InFusion: advancing discovery of fusion genes and chimeric transcripts from deep RNA-sequencing data. PLoS One 11(12):e0167417

Pandey RV, Pabinger S, Kriegner A et al (2016) ClinQC: a tool for quality control and cleaning of Sanger and NGS data in clinical research. BMC Bioinf 17:56. https://doi.org/10.1186/s12859-016-0915-y

Panigrahi P, Jere A, Anamika K (2018) FusionHub: a unified web platform for annotation and visualization of gene fusion events in human cancer. PLoS ONE 13(5):e0196588. https://doi.org/10.1371/journal.pone.0196588

Patel RK, Jain M (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7(2):e30619. https://doi.org/10.1371/journal.pone.0030619

Ren R (2005) Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. Nat Rev Cancer 5(3):172–183. https://doi.org/10.1038/nrc1567

Rodriguez-Martin B, Palumbo E, Marco-Sola S, Griebel T, Ribeca P, Alonso G et al (2017) ChimPipe: accurate detection of fusion genes and transcription-induced chimeras from RNA-seq data. BMC Genomics 18(1):7

Sboner A, Habegger L, Pflueger D et al (2010) FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biol 11:R104. https://doi.org/10.1186/gb-2010-11-10-r104

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28(8):1086–1092

Singh A, Zahra S, Das D, Kumar S (2019) AtFusionDB: a database of fusion transcripts in Arabidopsis thaliana. Database 2019:135. https://doi.org/10.1093/database/bay135

Singh S, Qin F, Kumar S, Elfman J, Lin E, Pham LP, Yang A, Li H (2020) The landscape of chimeric RNAs in non-diseased tissues and cells. Nucleic Acids Res 48(4):1764–1778

STAR-SEQR (2019) STAR-SEQR code and documentation on GitHub 2019. Available from https://github.com/ExpressionAnalysis/STAR-SEQR

Tomczak K, Czerwińska P, Wiznerowicz M (2015) The cancer genome atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol 19(1):68–77. https://doi.org/10.5114/wo.2014.47136

Torres-Garcia W, Zheng S, Sivachenko A, Vegesna R, Wang Q, Yao R et al (2014) PRADA: pipeline for RNA sequencing data analysis. Bioinformatics 30(15):2224–2226

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25(9):1105–1111

Uhrig S (2019) Arriba - fast and accurate gene fusion detection from RNA-Seq data. Available from https://github.com/suhrig/arriba

Vu T, Deng W, Trac Q et al (2018) A fast detection of fusion genes from paired-end RNA-seq data. BMC Genomics 19:786. https://doi.org/10.1186/s12864-018-5156-1

Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL et al (2010) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res 38(18):178

Wang Y, Wu N, Liu J et al (2015) FusionCancer: a database of cancer fusion genes derived from RNA-seq data. Diagn Pathol 10:131. https://doi.org/10.1186/s13000-015-0310-4

Yan Y, Park SS, Janz S, Eckhardt LA (2007) In a model of immunoglobulin heavy-chain (IGH)/MYC translocation, the Igh 3′ regulatory region induces MYC expression at the immature stage of B cell development. Genes Chromosomes Cancer 46(10):950–959. https://doi.org/10.1002/gcc.20480

Wu K, Liao X, Gong Y et al (2019) Circular RNA F-circSR derived from SLC34A2-ROS1 fusion gene promotes cell migration in non-small cell lung cancer. Mol Cancer 18(1):98. https://doi.org/10.1186/s12943-019-1028-9

Understanding the Molecular Kinetics in NSCLC Through Computational Method

7

Prajakta Nimsarkar, Pooja Gulhane, and Shailza Singh

Abstract

Lung cancer is one of the most common causes of cancer-related death in both men and women worldwide. Over the years, NSCLC has been demonstrated to have a relatively high incidence and fatality rate. Non-small cell lung cancer (NSCLC) accounts for around 85% of all lung cancer cases. Computed tomography (CT) or positron emission tomography (PET) can aid in the early detection and correct staging of non-small cell lung cancer, as well as selecting the best course of treatment. Chemotherapy is beneficial for people with metastatic disease, and concurrent chemotherapy and radiation are indicated for stage III lung cancer patients; furthermore, surgical removal of the tumor remains the single most consistent and successful curative option. Suppression of angiogenesis, the advent of epidermal growth factor receptor inhibitors, and other novel anticancer drugs are transforming the current and future of lung cancer and will undoubtedly enhance the number of survivors. However, around 70% of patients with non-small cell lung cancer have locally advanced or metastatic disease at the time of diagnosis, which increases the mortality rate even after concurrent therapies are administered, as cancer quickly becomes uncontrollable, and mutations in tumor cells can lead to resistance to existing therapies. Potential targets must be identified to solve this challenge, and innovative therapies, such as inhibitor development or other novel techniques, must be used. Computational methods have benefited from seeing these targets, which might help limit or block tumor development and proliferation.

P. Nimsarkar · P. Gulhane · S. Singh (✉)
National Centre for Cell Science, Pune, India

S. Singh (ed.), *Systems Biomedicine Approaches in Cancer Research*,
https://doi.org/10.1007/978-981-19-1953-4_7

7.1 Introduction

Lung cancer has become one of the number one causes of death among cancers worldwide. It accounts for 85–90% of all lung cancers (Clark and Alsubait 2021). According to GLOBOCAN, lung cancer has been the most common cancer globally for several decades. GLOBOCAN estimated the worldwide incidence of lung cancer in 2020 to be 11.4% of the entire world population and that of India to be 5.5% of the entire population of India. GLOBOCAN estimated the global mortality rate of lung cancer in 2020 to be 18% and the mortality rate in India in 2020 to be 7.8%. According to the statistical analysis carried out by GLOBOCAN, the number of new cases reported worldwide in 2020 of males (all ages) with lung cancer was 14.3% of the entire male world population and of India was 8% of the total male population of India. However, there were very few to no new cases of lung cancer in females reported. Thus, as per the statistical data obtained from GLOBOCAN, the World Health Organization (WHO) ranked lung cancer among the top 5 most frequently occurring cancers worldwide; in males, it was ranked the first position and that in females at third position, following the second rank with an incidence rate in both sexes worldwide (Sung et al. 2021; Bray et al. 2018).

Out of the two types of lung cancers, viz. small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), NSCLC has shown to have a relatively high rate of incidence and mortality over the decades (Clark and Alsubait 2021). Non-small cell lung cancer accounts for approximately 85% of the total lung cancer cases. Non-small cell lung cancer has been further classified into various types among which the major three types are as follows: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Out of the three, adenocarcinoma occurs at a comparatively higher rate; it starts in cells in the outer part of the lungs which normally secrete mucus (Travis et al. 2015). The mutations most often defined in adenocarcinoma occur in the KRAS and EGFR genes. Adenocarcinoma is more easily found before it has spread. It is most commonly found in nonsmokers, and it tends to occur at a younger age than other types of lung cancers. Squamous cell carcinoma is related to the flat cells that line the inside of the airways of the lungs, and they are often linked with a history of smoking; these tend to be found in the central part of the lungs (bronchus). Approximately 80% of adenocarcinoma tumors exhibit overexpression of the EGFR gene and 30% of squamous cell carcinoma overexpress the HER2 gene. The genetic variations in squamous cell carcinomas are diverse, and no targeted therapies are focused against its genetic alterations. Apart from these, large cell carcinomas contribute only 3% to the total NSCLC cases. They appear in any part of the lung and tend to grow and spread quickly which makes it harder to treat them. Other subtypes of NSCLC include adenosquamous carcinoma and sarcomatoid carcinoma; these are much less common. A glimpse of cancer cell progressing by targeting various means and signaling pathways is shown Fig. 7.1.

The stage of the disease is directly correlated to survival and is a key determinant of treatment. Stages of NSCLC are based on a permutation of factors including the size and location of the tumor and whether it has spread to the lymph nodes or other parts of the body. There are five stages of NSCLC lining up from Stage 0 (zero) and

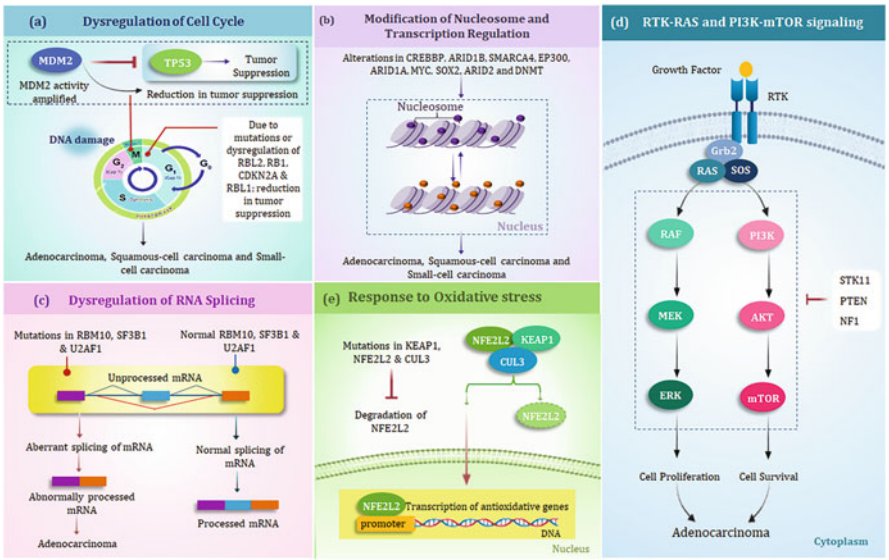


Fig. 7.1 Cancer progression due to mutations and abruptions in various cell processes signaling pathways

Stages I–IV. Stage 0 is referred to as in situ disease meaning the cancer is "in place" and has not grown into nearby tissues.

1. *Stage I lung cancer* is a small tumor that has not spread to any lymph nodes, making it possible for a surgeon to completely remove it; this stage is further divided into four substages based on the size of the tumor: stage IA1 where the cancer is no larger than 3 cm across and the part that has invaded into deeper lung tissues is no more than 1/2 cm across, IA2-where the tumor is larger than 1 cm but no longer than 2 cm across, IA3 where the tumor is larger than 2 cm but no longer than 3 cm across, and stage IB where the tumor is larger than 3cm but not larger than 4cm across.
2. *Stage II lung cancer* is divided into two substages: Stage IIA describes a tumor larger than 4cm but not larger than 5 cm across; it has grown into the visceral pleura (the membrane surrounding the lungs) and into the main bronchus, due to which the airways are partially clogged. Stage IIB cancer describes a tumor that is 5 cm or less in size and has spread to the lymph nodes, it can also be a tumor more than 5 cm wide which has not diffused to lymph nodes. Sometimes stage II tumors can be resected with surgery.
3. *Stage III lung cancers* are classified into stages IIIA, IIIB, or IIIC based on the size of the tumor and which lymph nodes the cancer has spread to. Stage III tumors may be difficult to remove with surgery as they may spread to the lymph nodes in the center of the chest outside the lungs or grow into nearby structures in

the lungs which make it less likely for the surgeon to completely remove cancer; it can thus be treated with systemic therapy and radiation therapy.

4. *Stage IV lung cancer* means cancer has metastasized to more than one area in the other lung including the fluid surrounding the lung or the heart or distinct parts of the body through the bloodstream. At this stage, once the cancer cells get into the blood, they can spread anywhere in the body. Non-small cell lung cancer is more likely to spread to the brain, bones, liver, and adrenal glands. Stage IV NSCLC is divided into two substages: Stage IVA where cancer spreads within the chest and/or has spread to one area outside the chest; Stage IVB where cancer has spread outside of the chest to more than one place in one organ or more than one organ.

In general, surgery is not recommended for stages III and IV lung cancer as it is impossible to remove the tumor since it tends to spread to the lymph nodes above the collarbone and into vital structures (heart, large blood vessels, and main breathing tubes leading to the lungs) within the chest (Goldstraw et al. 2011; Howington et al. 2013).

7.2 Modulation of Immunometabolism in Cancerous Cells

Cancer cells have been shown to exhibit metabolic characteristics that are significantly different from those of normal tissues. Cancer cells, unlike normal cells, reconfigure their cellular metabolic networks to meet their high needs for building blocks and energy generation, allowing them to proliferate and flourish indefinitely. Oncogenic mutations such as RAS, EGFR, MYC, and BRAF, which are common in cancer cells, might impact metabolic alterations in cancer. Cancer cells, unlike normal cells, reconfigure their cellular metabolic networks to meet their high needs for building blocks and energy generation, allowing them to proliferate and flourish indefinitely.

Immunometabolism is the interaction of metabolism and immunology and the metabolic regulation of immune function and metabolic regulation by immune system chemicals and cells. Based on the features of the tumor microenvironment, two adverse outcomes in cancer immunity can arise. Tumor-infiltrating immune cells (innate and adaptive immunity) can successfully inhibit tumor development and finally eliminate tumor cells on the one hand. Through immunoediting, tumor cells can gain the capacity to evade immunosurveillance and elimination.

Cancer cells have long been recognized to hijack cellular systems that control survival, development, and proliferation, resulting in the creation and progression of tumors. The genetic and epigenetic alterations that create stem cell-like features, such as unrestricted cell division and blocked differentiation, are the most well-known drivers of malignant transformation.

1. *Glucose metabolism*: The Warburg effect is the first known cancer-specific metabolic change. In this process, cancer cells rely on glycolysis for glucose

metabolism even when oxygen is present, resulting in high lactate levels and decreased usage of the tricarboxylic acid (TCA) cycle. This metabolic reprogramming has been suggested as an option to compensate for mitochondrial malfunction in cancer cells since the TCA cycle and subsequent oxidative phosphorylation create cellular energy more effectively than glycolysis. The glucose-6-phosphate also participates in the pentose phosphate pathway, which causes the tumor microenvironment to become acidic by increasing CO_2 generation in cancer cells. The stimulation of the pentose phosphate route also causes the nucleic acid synthesis pathway to become active, resulting in cancer cell growth (Schuurbiers et al. 2014).

Loss-of-function mutations in succinate dehydrogenase cause a significant accumulation of succinate in cancer cells, which functions as an oncometabolite and can cause epigenetic modifications by inhibiting ketoglutarate-dependent dioxygenases, ultimately leading to a malignant phenotype. The TCA cycle also regulates amino acid and fatty acid synthesis to meet cancer cells' growth, proliferation, and survival demands. Thus, cancer cells have been found to have increased oxidative phosphorylation, and reduction of mitochondrial DNA has been demonstrated to limit cancer cell tumorigenic capacity drastically. As a result, metabolic shift to aerobic glycolysis, in addition to ATP generation, appears to be a mechanism of supplying cancer cells with the precursors of proteins, lipids, amino acids, and nucleic acids for sustaining their increased proliferation and cellular structure.

Effector immune cells, such as activated cytotoxic T cells, undergo substantial metabolic reprogramming in order to execute effector roles such as killing cancer cells and secreting cytokines. T cells upregulate the key glucose transporter Glut1 in response to antigenic stimulation, followed by increased glucose uptake and glycolysis. In contrast, cancer cells maintain and eventually increase high glucose uptake and glycolysis, leading to a decrease in intratumoral glucose levels for T cells. Glucose deprivation can directly impede the production of IFN-γ, a critical T-cell effector molecule in tumor-infiltrating CD8+ T cells. Glyceraldehyde-3-phosphate dehydrogenase is devoted to its metabolic job when T cells can undertake high rates of glycolysis. Glyceraldehyde-3-phosphate dehydrogenase blocks IFN-γ translation in the presence of low glycolytic flux. T-cell hyporesponsiveness is thus caused by glucose restriction. On the other hand, rather than being oxidized in mitochondrial respiration, pyruvate, one of the terminal products of glycolysis, is mainly reduced to lactate in cancer cells, contributing to the acidity of the tumor microenvironment, which inhibits T-cell activity. As a result, tumor glucose metabolism might be thought of as a mechanism of tumor rejection resistance (Vanhove et al. 2019).

2. *Amino acid metabolism*: Aside from glucose metabolism, amino acid metabolism is also significant in cancer cell immune metabolism. The amino acid glutamine has been identified as a critical ingredient for T-lymphocytes' effector activity. T cells consume many arginine when antigens activate them, and tryptophan boosts the development of memory T cells by causing a metabolic transition from glycolysis to OXPHOS, which boosts antitumor activity. On the other hand,

cancer cells frequently overexpress the amino acid catabolic enzyme indoleamine-2,3-dioxygenase (IDO), which might result in tryptophan depletion outside the cell. Cyclooxygenase-2 (COX) and prostaglandin E2 are required for constitutive expression of indoleamine-2,3-dioxygenase. T-cell activity is inhibited when tryptophan is depleted because it activates general control nonderepressible 2 (GCN2), a stress–response kinase (Wei et al. 2021).

Furthermore, malignancies and myeloid-derived suppressor cells (MDSCs) degrade arginine through arginase 1 overexpression, resulting in lower CD3 ζ (zeta) chain expression, cell cycle arrest, and a weakened antigen-specific T-cell response. The elevated indoleamine-2,3-dioxygenase activity resulted in the accumulation of tryptophan metabolism byproducts, mostly kynurenines, which inhibits CD8+ T-cell proliferation and effector function via aryl hydrocarbon receptor (AHR). The interaction of kynurenines favors induction of regulatory phenotype in naive T cells with AHR (Lukey et al. 2017).

3. *Hypoxia and oxidative stress*: Tumors are frequently hypoxic, and hypoxia can act as a metabolic aid in promoting a malignant phenotype. Hypoxia can increase glucose absorption and glycolysis by inducing numerous glycolytic genes. Increased glycolysis is linked to long-term malignant development. TCR- and CD28-mediated T-cell activation is less effective in hypoxic environments. OXPHOS production and ROS production, both of which are required for normal T-cell effector function and antigen-specific proliferation, require oxygen. For appropriate T-cell signaling, low amounts of ROS are essential. ROS levels may be inadequate in the presence of hypoxia (Deben et al. 2018).

   High ROS levels, on the other hand, can be hazardous, and ROS produced in the tumor microenvironment can compromise immune cells by downregulating the CD3 (zeta) chain. Macrophages are sensitive to oxygen availability. It has been shown that anti-inflammatory M2-like tumor-associated macrophages (TAMs) cluster in hypoxic tumor locations. In contrast, pro-inflammatory M1-like TAMs dwell in normoxic tumor regions. Intratumoral hypoxia-induced semaphorin 3A phosphorylates vascular endothelial growth factor (VEGF) receptor 1 and recruits M2-like TAMs to hypoxic areas (Tafani et al. 2016).
4. *Nucleotide metabolism*: Hypoxia causes adenosine buildup in tumors by increasing adenine nucleotide breakdown via the 5′-nucleotidase pathway. The ectonucleotidases CD39 and CD73 convert ATP to AMP and AMP to adenosine, respectively, convert ATP to AMP and AMP to adenosine, and quickly degrade ATP to adenosine. T- and NK-cell activation and cytotoxic ability are both inhibited by adenosine (Valles et al. 2012).

7.3 Treatment Options Presently Offered

a. Surgery: If the tumor is confirmed to be resectable and the patient can endure surgery, patients with stages I, II, and IIIA NSCLC often receive surgery to remove the tumor. A lobe or part of the lung that contains the tumor may be removed by surgeons. Imaging investigations and biopsies are performed, as well

as an assessment of patient variables to evaluate if the tumor is resectable. Many surgeons now use video-assisted thoracoscopic surgery (VATS), which involves making a small incision in the chest and inserting a thoracoscope (Howington et al. 2013).

b. Chemotherapy: Cytotoxic combination chemotherapy is the first-line treatment for stage IV NSCLC, with histology, age vs. comorbidity, and performance status influencing treatment options (PS) (Ramalingam and Belani 2008). Platinum (cisplatin or carboplatin) with gemcitabine, paclitaxel, vinorelbine, docetaxel, pemetrexed, and irinotecan regimen is prescribed by the American Society of Clinical Oncology for patients with a PS of 0 or 1 (Masters et al. 2015). The patients in these investigations had an average overall survival of about 8–10 months (Kelly et al. 2001; Scagliotti et al. 2002; Schiller et al. 2002; Fossella et al. 2003). The particular combination is determined by the types and frequency of harmful effects and should be done on an individual basis. Pemetrexed may, nevertheless, be beneficial to adenocarcinoma patients. If cancer progresses or the condition is stable, but the medication does not decrease the tumors after four treatment cycles, therapy should be discontinued (Pisters et al. 2007; Scott et al. 2007). Patients with a PS of 3 are unlikely to benefit from cytotoxic chemotherapy since the risk of side effects might severely reduce their quality of life.

c. *Radiotherapy*: High-energy beams are used in radiotherapy to eliminate cancer cells by damaging their DNA. This treatment can aid in controlling or eliminating tumors in particular areas of the body. Radiotherapy can benefit patients not recommended for surgical resection and NSCLC with chest localization (Sebastian et al. 2018). For early-stage NSCLC patients with a single small nodule in the lung and no metastases to adjacent lymph nodes, a procedure known as stereotactic body radiation treatment (SBRT) is employed. This method employs an advanced coordinate system to precisely detect the tumor and guarantee that the tracking device is placed correctly (Grutters et al. 2010). SBRT was reported to have higher 2-year overall survival rates, cheaper costs, and improved patient convenience in a meta-analysis evaluating the efficacy of radiation with photons, protons, and carbon ions for NSCLC. High rates of local control in medically inoperable patients with stage 1 NSCLC were reported with receiving SBRT, according to 50-month outcomes from a prospective phase II trial of 70 medically inoperable patients who had SBRT (Fakiris et al. 2009). The toxicity and effectiveness of SBRT were investigated in phase III multicenter trial of individuals with NSCLC that was an early stage but medically inoperable. Patients who got SBRT had a three-year survival rate of 55.8% with substantial treatment-related morbidity, according to the 55 patients studied. SBRT has been reported to provide local control and results similar to surgical resection with reduced rates of treatment-related morbidity due to these and other trials (Timmerman et al. 2010; Lagerwaard et al. 2012).

d. *Testing for biomarkers*: Personalized therapy has improved survival in patients with NSCLC by focusing on the correct molecular targets in tumors. EGFR mutations and anaplastic lymphoma kinase (ALK) rearrangements have been successfully treated with targeted medicines. Other molecular abnormalities

discovered by genomic testing include gene rearrangements in the MET amplification, RET, and ROS1 genes and activating mutations in the KRAS, HER2, and BRAF and genes, which might be future therapeutic targets (Riely et al. 2008, 2009; Lynch et al. 2004; Brose et al. 2002; Soda et al. 2007).

Computed tomography or positron emission tomography can help in the early diagnosis and accurate staging of non-small cell lung cancer playing a crucial role in determining appropriate therapy. Chemotherapy is considered valuable for patients with metastatic disease, and the administration of concurrent chemotherapy and radiation is recommended for stage III lung cancer patients; also, the feasibility of surgically removing the tumor remains the single most consistent and successful option for cure. The suppression of angiogenesis, the introduction of epidermal growth factor receptor inhibitors, and other new anticancer agents are changing the current and future of this disease and will certainly increase the number of lung cancer survivors. An overview of flow of therapies adopted for NSCLC patients has been depicted in Fig. 7.3.

7.4 Immunotherapy

Immunotherapy is a groundbreaking oncology treatment that relies on the body's natural defense mechanism to combat cancer. Because specific cancer cells resemble healthy cells, the immune system cannot distinguish between normal and malignant cells in the body (Jemal et al. 2011; Siegel et al. 2015). Immunotherapy is thought to function by enhancing the immune system's ability to target cancer cells and stop or limit their development or by preventing cancer cells from spreading to other regions of the body. According to research, improved survival has been linked to a robust antitumor immune response. Patient survival is linked to more natural killer cells, CD8+ T cells, dendritic cells, and CD4+ T cells. Working mechanism of various therapies is represented in Figs. 7.2 and 7.3.

Inhibiting the inhibitory receptors cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) and programmed cell death 1 (PD-1) and its ligand, PD-L1, is part of this strategy. The immune system uses these immunological checkpoints to maintain self-tolerance and control the immune response in the body to protect tissues from harm as the immune system responds to a pathogen (Gajewski et al. 2013; Neurath and Finotto 2012; Zhang et al. 2015). Table 7.1 depicts the strategies adopted in immunotherapy targeting mechanisms modulating immune processes helping tumor cell survival.

7.5 Vaccine Development as a Technique of Immunotherapy

In NSCLC, the purpose of vaccination treatment is to bend the immunological balance in favor of activation, allowing the host to respond to tumor-associated antigens. Antigen-specific immunotherapy and tumor vaccines are two vaccine-

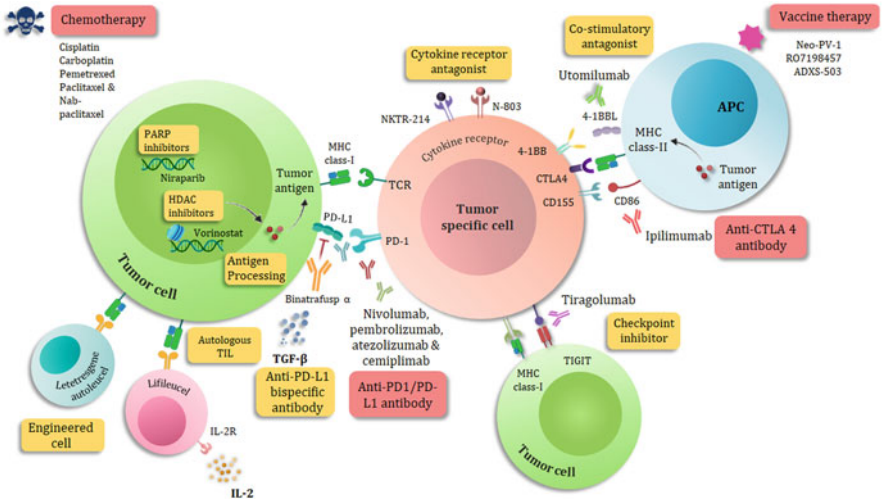


Fig. 7.2 Subsist and potential therapies for NSCLC: The immunological microenvironment of non-small cell lung cancer (NSCLC) is shown, with approved first-line therapy methods (red boxes) and attractive experimental targets (yellow boxes) underlined. Within each area, particular therapeutic drugs are mentioned as examples. *TAA* tumor-associated antigen, *TCR* T cell receptor, *TIGIT* T-cell immunoreceptor with Ig and ITIM domains, *TIL* tumor-infiltrating lymphocyte, *TKI* a tyrosine kinase inhibitor

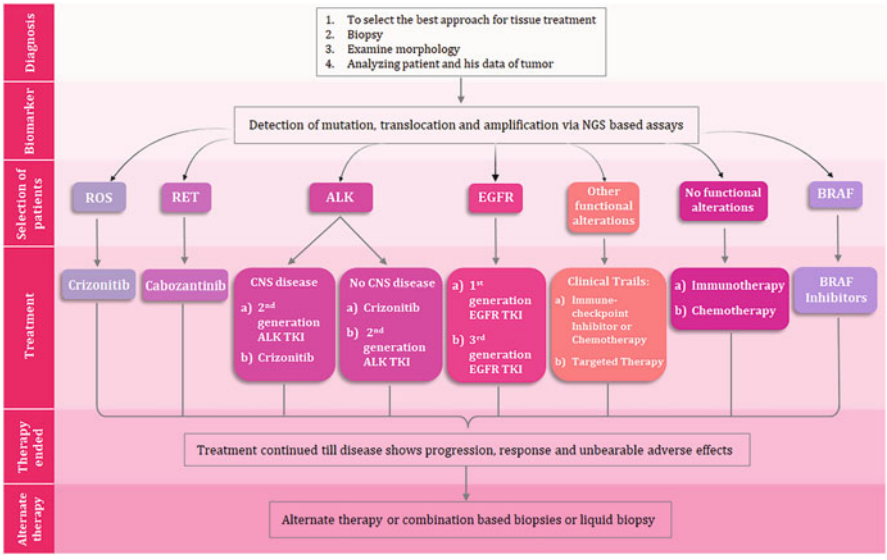


Fig. 7.3 Systematic flowchart of therapies given to patients of NSCLC

Table 7.1 Immunotherapy trials by vaccine development in NSCLC yield some promising outcomes

Types of immunotherapy	Agents	Phase	Outcomes
PD-1	Nivolumab vs. docetaxel	III	Demonstrated better overall survival in comparison to docetaxel and benefitted sustained survival over varied histology
PD-L1	MK-3475	III	Shows antitumor functionality with better tolerance in patients expressing PD-L1
Anti-CTLA-4	Ipilimumab	III	Compared to gp100 alone, ipilimumab with or without gp100 enhanced survival
	Ipilimumab + paclitaxel + carboplatin	II	Immune-related advancement was better when given in a phased manner

based treatments for NSCLC that are currently under development. Autologous or allogeneic tumor cells develop tumor (whole-cell) vaccines. The host's immune system is exposed to several tumor-associated antigens by these vaccinations. Antigen-specific immunotherapy entails antitumor immunity directed to target antigens expressed on tumor cells. Due to the fact that these vaccinations target a specific antigen, they may not be suitable for all individuals.

7.6 Whole-Cell Tumor Vaccine

7.6.1 Cell-Based Vaccine

7.6.1.1 Belagenpumatucel-L (Lucanix®)

The allogeneic tumor cell vaccine belagenpumatucel-L (Lucanix®, NovaRx Corp.) utilizes genetically engineered entire tumor cells. It consists of four irradiation NSCLC lines, two of which are adenocarcinoma lines, one squamous line, and one large cell carcinoma line (H520, RH2, H460, and SKLU-1, respectively), all of which have been gene-modified using the transforming growth factor TGF-β2 (Iyengar and Gerber 2013). Higher levels of TGF-β2 are linked to immune system suppression, which results in natural killer cell neutralization and dendritic cell suppression. Through the expression of antisense RNA, the vaccine's plasmid TGF-β2 transgene component may decrease tumor formation (Kong et al. 1999).

In the phase II trial, 75 NSCLC patients at stages II–IV were tested with belagenpumatucel-L. It positively responded to candidates with an increased antibody level and better tolerance. Patients without progression after frontline chemotherapy (phase IIIA, $n = 42$; phase IIIB/IV, $n = 490$) were randomized 1:1 (belagenpumatucel-L or placebo) between 4 and 17.4 weeks after the end of frontline chemotherapy (Nemunaitis et al. 2006; Giaccone et al. 2015). They were treated until disease progression or withdrawal in a phase III trial of belagenpumatucel-L.

The effectiveness of belagenpumatucel-L in enhancing overall survival was investigated in this study. The study's secondary objectives were progression-free survival, response rate, and safety. Adenocarcinoma was found in 57% of the included individuals (270 vaccination and 262 placeboes), whereas squamous cell carcinoma was found in 27%. The primary goal of this research was not met; median overall survival in the vaccination group was 20.3 months compared to 17.8 months in the placebo group [hazard ratio (HR) 0.94; $P = 0.594$]. A predetermined COX regression helped uncover predictive markers for an improved outcome. The vaccination group outlived the placebo group by a considerable and clinically relevant margin. For patients randomized within 12 weeks of finishing chemotherapy, overall survival was enhanced by 7.3 months in the vaccination group. With belagenpumatucel-L, the median overall survival was 20.7 months, compared to 13.7 months with placebo (HR 0.75; $P = 0.083$). Improved median overall survival was observed in patients who had previously received pretreatment radiation (Giaccone et al. 2015).

7.6.2 Recombinant Vaccine

7.6.2.1 TG4010 (MVA-MUC1-IL-2) Vaccine

TG4010 is a recombinant vaccine virus (modified virus of Ankara or MVA) that encodes human MUC1 and IL-2, targeting the MUC1 antigen on malignant cells. Two regimens of TG4010 coupled with first-line chemotherapy were evaluated in patients with stage IIIB/IV NSCLC in phase II randomized, open-label research. Patients in arm 1 ($n = 44$) received TG4010 plus cisplatin (100 mg/m^2 day 1) and vinorelbine (25 mg/m^2 days 1 and 8). In contrast, patients in arm 2 ($n = 21$) received TG4010 monotherapy until disease progression and then TG4010 with the same treatment as in arm 1. Arms 1 and 2 had median survival rates of 12.7 and 14.9 months, respectively (Trevor et al. 2001).

The researchers included 138 patients with stage IIIB/IV NSCLC who had MUC1 immunohistochemistry. Seventy-four patients got TG4010 with cisplatin and gemcitabine for up to six cycles, whereas the remaining 74 patients received just chemotherapy. Patients in the TG4010 with chemotherapy arm and those in the chemotherapy-only arm did not have substantially different progression-free survival rates at six months (43.2% (95 % CI: 33.4–53.5%) and 35.1% (95% CI: 25.9–45.3%), respectively, $P = 0.307$). Patients in the TG4010 group had a median overall survival of 10.7 months (95% confidence interval: 8.8–18.0), while those in the chemotherapy-only group had a median overall survival of 10.3 months (95% confidence interval: 8.3–12.5); however, these findings were not statistically significant (Ramlau et al. 2008; Quoix et al. 2011).

7.6.2.2 CIMAVax EGF

Human recombinant EGF coupled to the P64K Neisseria meningitides recombinant protein led to the development of CIMAVax EGF. Cuba, Peru, and Venezuela have all given their approval. The impact of cyclophosphamide in reducing inhibition of

T-suppressor cells was investigated in a pilot study. Serial antibody measurements of EGF were performed using an enzyme-linked immunosorbent assay (ELISA) throughout the trials and were stratified based on their results; a good antibody responder produced an antibody response to a titer greater than 1:4000, while a poor antibody responder produced an antibody response to a titer less than 1:4000. The outcomes of the pooled studies revealed that pretreatment with cyclophosphamide prior to EGF delivery had no meaningful effect on antibody responses. There was a substantial difference in antibody responders when ISA 51 was used instead of aluminum hydroxide. Anti-EGF antibody titers and immune response length appear to have a survival time connection (Neninger Vinageras et al. 2008).

Eighty patients with stage IIIB/IV NSCLC who had completed first-line chemotherapy took part in a phase II experiment examining the immunogenicity, safety, and effect on survival of an EGF-based cancer vaccine. The finest supportive care or EGF vaccines were given to patients at random. The median overall survival of good antibody responders who were immunized was 11.7 months, compared to 3.6 months for poor antibody responders who were not vaccinated. Vaccinated individuals with serum EGF levels less than 168 pg/mL had a longer median overall survival of 13 months compared to 5.6 months in those with serum EGF levels more than 168 pg/mL. The trial's study indicates a tendency toward improved overall survival for vaccinated patients, statistically significant in the subgroup of patients younger than 60 against those over 60 (11.57 and 5.33 months, respectively, $P = 0.124$). There was a direct link between serum EGF levels falling and survival and a link between antibody response and survival (Gonzalez et al. 2003).

7.6.2.3 GVAX®

Irradiated autologous or allogeneic tumor cells that have been genetically engineered to release recombinant granulocyte-macrophage colony-stimulating factors are used in GVAX (GM-CSF). GVAX has been found to promote CD8+ and CD4+ T-cell responses and antibody responses by inducing the infiltration of antigen-presenting dendritic cells into the vaccination site (Eager and Nemunaitis 2005). Melanoma, renal cell carcinoma, prostate cancer, and non-small cell lung cancer have all been examined using the GVAX platform. A successful autologous vaccination was developed for 34 patients with stages IIB-IV NSCLC in a phase I study (97%) (Salgia et al. 2003). The vaccination was given weekly for three weeks and biweekly until the patient was taken out of the research or the vaccine supply ran out. Local skin response at the immunization site was the most common adverse event reported. Five patients had stable cancer, one had a mixed response, and two patients who had previously undergone surgical resection had no signs of disease for more than 42 months (Geary and Salem 2013; Antonarakis and Eisenberger 2013).

A phase I/II research with GVAXNSCLC was performed after this trial. The research included 83 patients, 20 of whom were in the early stages (I/II) and 63 in the late stages (III/IV). Ten individuals with early-stage NSCLC and 33 patients with late-stage NSCLC received vaccinations. Patients received 5×10^6 to 100×10^6 vaccine cells each dose throughout 3–6 biweekly immunizations, followed by 6 months of monthly therapy. The toxicity of GVAXNSCLC was not dose-limiting,

with local responses at the immunization site being the most prevalent. Three of the late-stage patients with severe disease experienced long-term tumor regression. Two of them had total remission for more than five years. Despite some encouraging results in NSCLC, the disappointing VITAL studies for prostate cancer have dampened enthusiasm for phase III trials utilizing the GVAX platform at this time (Yu et al. 2016).

7.7 Antigen-Specific Vaccines

7.7.1 Peptide- and Protein-Based Vaccines

7.7.1.1 Liposome BLP25 Vaccine (Stimuvax®, Tecemotide)

L-BLP25 is a peptide-based vaccine (Vlad et al. 2004). The exposed core peptide of membrane-associated glycoprotein (MUC1) located on the apical surface of epithelial cells of the digestive, genitourinary, and respiratory systems is the target for L-BLP25. In a randomized phase IIb research in patients with stages IIIB and IV NSCLC, the effects of the L-BLP25 vaccine on survival and toxicity were investigated. According to revised survival analyses, patients in the L-BLP25 + BSC arm had a median survival time of 17.2 months and a 31% 3-year survival rate, compared to 13.0 months and 17% 3-year survival rate in the BSC alone group. Based on the phase IIb research findings, a phase III trial (START) was launched. Patients with unresectable stage III NSCLC who did not advance after initial chemoradiotherapy were randomly assigned to receive L-BLP25 or placebo ($n = 1513$). Although L-BLP25 was well tolerated, the START trial's primary effectiveness endpoint was not met. Patients treated with L-BLP25 with concomitant chemotherapy and radiation, on the other hand, had a considerable survival benefit (Butts et al. 2005, 2011, 2014).

7.7.1.2 Melanoma-Associated Antigen A3 (MAGE-A3)

MAGE-A3 is expressed mainly on tumor cells (35% of NSCLCs) but not on normal cells, except testicular germ cells and placental trophoblast, where greater expression is linked to advanced illness and a worse prognosis (Woo et al. 2002).

The tolerance and effectiveness of MAGE-A3 as a tumor-specific vaccination target in NSCLC were assessed in a multicenter, double-blinded phase II clinical study. One hundred eighty-two patients with totally resected MAGE-A3 (+), stage pIB or pII were randomly randomized to receive recombinant MAGE-A3 protein plus adjuvant or placebo after surgery in a 2:1 ratio. Patients received five immunization rounds every 3 weeks, followed by eight vaccination cycles every 3 months. The disease-free interval was the study's primary goal, with secondary endpoints including safety, disease-free survival, and overall survival. Disease-free survival and overall survivals were 0.73 (95% CI: 0.45–1.16) and 0.66 (95% CI: 0.36–1.20), respectively, after a median follow-up of 28 months. The study's findings revealed an excellent trend for MAGE-A3 activity in the treatment of NSCLC, with a 27% increase in the disease-free interval and disease-free survival. Although the

improvement was not statistically significant, the promising results inspired the MAGRIT phase III experiment (MAGE-A3 as adjuvant non-small cell lung cancer immunotherapy) (Linsley et al. 1991; Takagi et al. 1998).

The MAGRIT research examined disease-free and overall survival, as well as yearly disease-free survival from 2 to 5 years, lung cancer-specific survival, disease-free specific survival, and adverse events in patients with MAGE-A3-positive lung malignancies. There were 2270 patients in the MAGE-A3 or placebo groups, all in stages IB, II, or IIIA. In the overall population or individuals who did not undergo chemotherapy, MAGE-A3 as an adjuvant medication did not improve disease-free survival compared to placebo. The MAGRIT study confirms that vaccinations are well tolerated with moderate side effects and no apparent rise in immune-mediated diseases. However, the primary endpoint was not met (Vansteenkiste et al. 2016).

7.8 Molecular Biology of Non-small Cell Lung Cancer

The tumor initiation and progression involve an array of steps that encloses DNA damage or mutations induced by external risk factors. This field carcinogenesis and pre-neoplasia in airway epithelium for development of the tumor. Every step has a critical focus in intense clinical and laboratory investigations for cancer therapies (Larsen and Minna 2011). In non-small cell lung cancer malignancies, tumorigenesis relates to the activation of growth-promoting proteins (KRAS, EGFR, BRAF, MEK-1, HER2, MET, ALK and RET, PI3K) as well as inactivation of tumor suppressor genes (p53, phosphatase with tensin homology (PTEN), LKB-1).

Activation of growth-promoting oncogenes can occur by gene amplification or other genetic modification including point mutations and structural rearrangements leading to uncontrolled signaling through oncogenic pathways. Signaling pathways regulated by oncogenes and tumor suppressor genes are often related with cross-talk between pathways involved in carcinogenesis which adds to the complexity in the incidence of mutational tumor facilitating disease progression (Lynch et al. 2004; Luo et al. 2009). Non-small cell lung cancer can be defined by specific mutations in the majority of genes encoding EGFR, KRAS, MAPK, and PI3K signaling pathways. Mutations in these genes are mutually exclusive; EGFR and ALK mutations prevail in adenocarcinomas that develop in nonsmokers; KRAS and BRAF mutations are more common in smokers or former smokers (Ju et al. 2012). In NSCLC, EGFR mutations occur in the first four exons of the intracellular tyrosine kinase domain, most commonly Exon 19 in-frame deletion having over 20 variants, the commonest being "delE746-A750" (accounts for ~45%), and the next commonest EGFR mutations are missense mutations mainly in Leu 858R, a single nucleotide point mutation in exon 21 directing to single amino acid change from leucine to arginine at codon 858 (accounts for ~40%). In NSCLC, almost all EGFR mutations occur in adenocarcinoma although they may also be seen in adenosquamous carcinomas. These mutations are more frequently but not exclusively found in patients who are females, younger, and with no history of smoking (Kohno et al. 2012; Lipson et al. 2012). These studies give a complete picture of

genetic changes in lung malignancies; distinguishing physiologically significant driver mutations from the overwhelming number of passenger variants remains difficult. The scarcity of high-frequency recurrent mutations underlines the variability and complexity of lung cancer molecular biology, with shared pathways impacted by a variety of genetic changes, making personalized treatment challenging to achieve.

7.9 Tumor Suppressor Genes

Tumor suppressor genes are essential inhibitors of normal cell proliferation. As indicated in Knudson's two-hit theory, loss of tumor suppressor gene (TSG) activity is an effective mechanism of carcinogenesis that necessitates the inactivation of both gene alleles. Individual genes are frequently inactivated in one allele due to mutations, epigenetic silencing, or other aberrations. In contrast, the second allele is frequently inactivated due to loss of heterozygosity (LOH), which occurs when a chromosome region is lost due to deletion, nonreciprocal translocation, or mitotic recombination. TSGs such as TP53, retinoblastoma 1 (RB1), serine–threonine kinase 11 (STK11), CDKN2A, FHIT, RASSF1A, and PTEN are often inactivated in lung cancer. Individual genes are frequently inactivated in one allele due to mutations, epigenetic silencing, or other aberrations (Knudson 1993). In contrast, the second allele is frequently inactivated due to loss of heterozygosity (LOH), which occurs when a chromosome region is lost due to deletion, nonreciprocal translocation, or mitotic recombination. TP53, retinoblastoma 1 (RB1), serine–threonine kinase 11 (STK11), CDKN2A, FHIT, RASSF1A, and PTEN are TSGs that are often inactivated in lung cancer, and these genes map to chromosomal areas routinely discovered in LOH investigations. TSGs like TP53 (17p13), RB (13q12), p16 (9p21), and PTEN, for example, are commonly implicated in areas of lung cancer that show allelic loss (10q22) (Raso and Wistuba 2007). Mutations in many TSGs not previously recognized to have a substantial role in lung adenocarcinoma were discovered in research by Ding et al., including the TSG NF1 (associated in neurofibromatosis type 1), which was mutated in 13 tumors, and the TP53 regulator ATM, which was mutated in 13 patients.

1. *TP53*: The nuclear phosphoprotein TP53, which is encoded on chromosome 17p13, is a 53-kDa nuclear phosphoprotein that recognizes and binds to damaged DNA and works as a transcription factor directing the expression of a variety of genes (Mogi and Kuwano 2011). Damaged DNA or carcinogenic stress causes TP53 to be activated, resulting in cell cycle arrest and the production of cyclin-dependent kinase inhibitors, allowing for DNA repair or apoptosis. With the hemizygous deletion of 17p13, which contains the TP53 locus, occurring in 90% of small cell carcinomas and roughly 65% of NSCLC, TP53 inactivation is one of the most prominent genetic abnormalities in lung cancer. In 80–100% of small cell lung carcinomas, inactivating mutations in TP53 (mainly missense

mutations near the DNA-binding domain) have been observed (Wistuba et al. 2000).

A meta-analysis of TP53 in over 4000 NSCLC patients identified mutations or protein accumulation in just 46.8% of cases. SCC is more prevalent than ADC and is related to higher tumor stage, grade, and male gender (Wistuba et al. 2000; Tammemagi et al. 1999). In the Cancer Genome Atlas (TCGA) study, TP53 mutations were discovered in at least 81% of SCCs that underwent extensive genomic investigation. In 85 of 188 ADCs, Ding et al. discovered TP53 alterations (45%) (Husgafvel-Pursiainen et al. 2000). TP53 mutations are linked to a history of smoking or exposure to tobacco smoke in the environment in NSCLC. Smoking-related malignancies had a much greater frequency of G-to-T transversions than G-to-C transversions (believed to be generated by polycyclic aromatic hydrocarbons in tobacco smoke) and G to A transitions at CpG dinucleotides, which are more prevalent in never smokers. In lung NSCLC, abnormal p53 identified by protein expression or mutational analysis is an unfavorable prognostic factor, according to a meta-analysis of 74 studies. Treatment resistance has also been linked to mutations in the TP53 gene. Mutations in the TP53 gene can develop with mutations in the EGFR and KRAS genes (Takagi et al. 1998).

2. *LKB1 (STK11)*: LKB1 is a TSG that encodes a serine–threonine kinase that inhibits mTOR and has been linked to various biological processes, including cell cycle control, chromatin remodeling, cell polarity, and energy metabolism (Marignani 2005; Shaw et al. 2004). Thirty percent of ADCs have been shown to have deregulated mTOR pathway components (excluding KRAS mutations). Patients with Peutz–Jeghers syndrome have LKB1/STK11 gene mutations. LKB1 is suppressed in lung cancer by a variety of somatic mutations or deletions that result in shortened proteins, with inactivation of LKB1 occurring in 11–30% of lung ADC, making it the third most prevalent genetic aberration in lung ADC after TP53 and KRAS. Inactivation of LKB1 is more prevalent in lung ADCs than in SCCs. There is evidence of a relation between LKB1 mutations and male smoking history and a link with KRAS mutations (Sanchez-Cespedes et al. 2002; Koivunen et al. 2008a, b; Matsumoto et al. 2007; Onozato et al. 2007).
3. *PTEN*: On chromosome 10, PTEN encodes a lipid and protein phosphatase that dephosphorylates PI-(3,4,5)-triphosphate, inhibiting the PI3K/AKT/mTOR signaling cascade (Brognard et al. 2001). PTEN's TSG function is inactivated, resulting in unfettered activation of AKT/protein kinase B unaffected by ligand binding. PTEN mutations are found in roughly 5% of NSCLC cases, with SCC being more prevalent than ADC (10.2% vs. 1.7%) and linked to smoking history. On the other hand, reduced protein expression has been documented in roughly 75% of NSCLC cases (Jin et al. 2010; Marsit et al. 2005).
4. *The $p16^{INK4a}$-Cyclin D1-CDK4-RB Pathway*: The $p16^{INK4A}$/RB pathway regulates the cell cycle progression from G1 to S phase (Harbour et al. 1988). RB1 is a tumor suppressor gene that produces the RB protein, regulating the G1/S transition in the cell cycle by binding the transcription factor E2F1. RB1 was the first TSG identified in lung cancer. It is inactivated in roughly 90% of small cell

lung carcinomas but only 10–15% of non-small cell lung carcinomas. The pathway is primarily shut down in NSCLC because of changes in cyclin D1, CDK4, and the cyclin-dependent kinase inhibitor p16 (CDKN2A) (Brambilla et al. 1999; Raso and Wistuba 2007). $p16^{INK4A}$ prevents cell cycle progression across the G1/S checkpoint by inhibiting cyclin D1-dependent phosphorylation of RB protein. Inactivation of $p16^{INK4A}$ in roughly 80% of NSCLC was changed in 72% of lung SCCs studied by TCGA, primarily due to homozygous deletion, methylation, or inactivating mutations. Furthermore, nearly 40% of NSCLC patients have cyclin D1 overexpression caused by gene amplification or other causes (Brambilla et al. 1999; Otterson et al. 1994).

7.10 Genetic Mutations Leading to Cancer Development

1. *Epidermal growth factor (EGFR)*: EGFR mutations have a role in developing several cancers, including NSCLC. EGFR is a transmembrane tyrosine kinase that has an external ligand-binding domain and an intracellular tyrosine kinase domain. When the ligand epidermal growth factor binds to the receptor, it forms homodimer or heterodimer with other members of the EGFR family, and the tyrosine kinase domain is activated (Prenzel et al. 2001). The PI3K/AKT/mTOR, RAS/RAF/MAPK, and JAK/STAT signaling pathways involve EGFR-stimulated signaling. Survival, cell proliferation, neovascularization, differentiation, metastasis, and invasion are all regulated by EGFR. Constitutive tyrosine kinase activation and oncogenic transformation of lung epithelial cells in vitro are caused by activating mutations in the EGFR gene. Multiple lung ADC was developed in a transgenic mouse model with inducible expression of the most frequent EGFR mutations, which were susceptible to slight drug inhibition. Enhanced protein expression or gene copy numbers are two different strategies for increased EGFR signaling (Bethune et al. 2010; Yarden and Sliwkowski 2001; Sordella et al. 2004; Greulich et al. 2005).
EGFR mutations are found in the first four exons of the intracellular tyrosine kinase domain in NSCLC, with the most prevalent exon 19 in-frame deletions (45%). There are over 20 variations, with delE746-A750 being the most common. Missense mutations, notably L858R, a single nucleotide point mutation in exon 21 that results in a single amino acid shift from leucine to arginine at codon 858 (40%), are the second most prevalent EGFR alterations (Yip et al. 2013).

   EGFR mutations are seen nearly exclusively in ADC in lung cancer, although they can also be found in adenosquamous carcinomas. Patients who are female, younger, and have never smoked are more likely to have EGFR mutations; however, this is not always the case. In histologically thoroughly sampled pure SCCs, EGFR mutations occur only extremely rarely. EGFR mutations were found in two instances out of 188 SCCs, both with L861G mutations. While EGFR mutations are uncommon in SCCs, variant III mutations affect EGFR's extracellular domain, copy number increases, and protein overexpression are

more prevalent in SCCs than in ADCs (Kosaka et al. 2004; Shigematsu et al. 2005a, b; Marchetti et al. 2013; Wu et al. 2008; Heist et al. 2012).

In individuals who acquire resistance to EGFR-TKIs, secondary mutations in EGFR arise or are clonally chosen, the most frequent of which is the T790M activating point mutation in exon 20, which replaces a "bulkier" methionine for threonine, interfering with reversible TKI binding. T790M is detected in around 50% of tumors from individuals who develop TKI resistance. Exon 20 mutations, including T790M variants linked to EGFR-TKI treatment resistance, were found in 29% of patients with EGFR mutations in a therapy-naive cohort. EGFR-TKI resistance can also be caused by the activation of downstream pathways that circumvent EGFR inhibition, such as the PI3K pathway via MET amplification (Wu et al. 2008; Balak et al. 2006; Engelman et al. 2007).

2. *KRAS*: KRAS is a proto-oncogene which encodes a G protein that controls signal transduction pathways that govern cell proliferation, differentiation, and survival. It belongs to the RAS family of proto-oncogenes (KRAS, NRAS, and HRAS are all found in humans) (Downward 2003; Karnoub and Weinberg 2008). In typical quiescent cells, Ras proteins are linked to GDP and inactive. Following activation of upstream growth factor receptors, a transition to the activated guanosine triphosphate (GTP) bound form occurs. Ras-GTP that has been activated binds to and activates a variety of downstream pathways, including the mitogen-activated protein kinase (MAPK) pathway, the RAS/RAF/MEK/MAPK pathway, and the PI3-K [PI3K/AKT/mammalian target of rapamycin (mTOR)] pathways. KRAS is involved in downstream signaling generated by several growth factor receptors, including EGFR, and constitutive activation of the protein eliminates the need for growth factor signaling (Karnoub and Weinberg 2008). Activating mutations impair the protein's GTPase function, preventing the active RAS-GTP from being converted to GDP, resulting in enhanced signaling across various downstream growth-promoting pathways. The MAPK/RAF/RAS/MEK signal transduction cascade is implicated in numerous lung malignancies, with alterations in the system found in 132 of 188 tumors, with mutations in KRAS being the most frequent (Ding et al. 2008).

   The most prevalent oncogenic change in lung ADC is activating mutations in the KRAS oncogene, which occur in roughly 25–40% of patients. In contrast, HRAS and NRAS mutations are sporadic. KRAS mutations are more prevalent in Western countries than in Asian ones. They are more common in males and smokers; hence, differences in the prevalence of KRAS mutations in lung ADC are most likely due to different patient groups. KRAS mutations have been found in 0–15% of ADC in never smokers. KRAS mutations are also uncommon or nonexistent in SCCs and small cell cancers (Sequist et al. 2011; Ding et al. 2008; Yip et al. 2013). Only one KRAS mutation at codon 61 was discovered in a comprehensive genomic study of 188 SCCs. Single amino acid changes in hotspots situated predominantly in codon 12 and codons 13 and 61 are seen in KRAS mutations in lung cancer. In smokers, G-to-T transversions are the most prevalent mutations (84%). In contrast, never

smokers are more likely to have G to A transitions (Riely et al. 2008; Rodenhuis and Slebos 1992; Schmid et al. 2009).

KRAS mutations seldom occur in tandem with EGFR mutations, in line with their role as driving mutations. According to a meta-analysis, KRAS mutations cause constitutive activation of pathways downstream of EGFR, making KRAS mutant tumors resistant to EGFR tyrosine kinase inhibitors (TKIs) (Kosaka et al. 2004; Mao et al. 2010; Shigematsu et al. 2005a, b; Tam et al. 2006). Distinct KRAS mutant proteins have different clinical importance, according to data. In the BATTLE study (prospective phase II biomarker-integrated approaches of targeted therapy for lung cancer elimination), G12C or G12V mutant KRAS predicted shorter progression-free survival than other KRAS mutation wild-type KRAS. Furthermore, distinct amino acid changes were linked to activation of different pathways (MEK with Gly12Asp and PI3-K and mutant Gly12Val or Ral with Gly12Cys) due to divergent protein conformations resulting in changed capacity to bind with downstream protein mediators (Linardou et al. 2008; Ihle et al. 2012). This emphasizes the need to evaluate the clinical and therapeutic implications of individual genetic changes in lung cancer before using targeted treatments and designing clinical trials. The high incidence of KRAS mutations in lung cancer makes it an excellent therapeutic target. Nevertheless, clinical trials of targeted medicines have largely failed.

3. *MEK*: MEK1, also called as MAPK1, is a serine–threonine kinase that plays a crucial role as a RAS downstream target. MEK1 stimulates MAPK2 and MAPK3 in the BRAF pathway (Downward 2003). In NSCLC, two of 107 lungs ADC were discovered to have an activating mutation in exon two that did not implicate the kinase domain. The mutations were exclusive to other driver mutations and were linked to in vitro function gain (Marks et al. 2008).
4. *PI3K/AKT/mTOR*: The PI3K/AKT/mTOR signaling system regulates cell proliferation, survival, differentiation, adhesion, and motility. Both NSCLC and small cell carcinoma have been linked to changes in this pathway. EGFR, HER2, insulin-like growth factor receptor, vascular endothelial growth factor receptor, and platelet-derived growth factor receptor are among the membrane tyrosine kinase receptors that activate the pathway (Engelman et al. 2006; Cully et al. 2006). PI3K is recruited to the cell membrane by activated receptor tyrosine kinases, where it phosphorylates PIP2 to PIP3 (phosphatidylinositol 4,5-bisphosphate (PIP2) to phosphatidylinositol 3,4,5-triphosphate). PIP3 then recruits the serine–threonine kinase AKT to the membrane, phosphorylated by PI3 kinase and mTOR. AKT's downstream target, mTOR, is a serine/threonine kinase (Brognard et al. 2001). Tuberous sclerosis and Bcl-2 are linked to death promotors advancing to cell proliferation. Survival is among the numerous targets activated by activated AKT. Other pathways that interact with RAS/RAF/MEK (rat sarcoma/rapidly accelerated fibrosarcoma/MAPK or Erk kinase) include RAS/RAF/MEK (rat sarcoma/rapidly accelerated fibrosarcoma/MAPK or Erk kinase/MAPK or Erk kinase/MAPK or Erk kinase/MAPK or Erk).

In many tumors, including 50–70% of NSCLC, the PI3K/AKT/mTOR pathway is typically disrupted. The Cancer Genome Atlas research found significant changes in the PI3K pathway in 47% of SCCs. Activating mutations in EGFR, KRAS, PI3K, or AKT, and PIK3CA amplification or loss of negative regulation by the tumor suppressor gene PTEN all play a role in pathway activation in lung carcinogenesis (Papadimitrakopoulou 2012; Vivanco and Sawyers 2002).

Phosphatidylinositol 3-kinases (PI3Ks) are intracellular lipid kinases, and the PIK3CA gene encodes the primary catalytic subunit, the p110 alpha isoform. Constitutive ligand-independent pathway activation results from activating mutations and amplification of PIK3CA. PIK3CA mutations, which usually affect the catalytic domain, have been found in 1–3% of NSCLCs, with SCCs being more prevalent than ADCs. PIK3CA mutations, unlike other oncogenic driver mutations, can arise in combination with EGFR or KRAS mutations, suggesting that they are not genuine driver mutations (Vivanco and Sawyers 2002). In vitro investigations of lung cancer cell lines with PIK3CA mutations or copy number increases, on the other hand, reveal enhanced PI3 kinase activity responsive to small-molecule inhibition, while in vivo animal models with PIK3CA mutant expression develop multiple ADC, showing oncogenic potential. In NSCLC, especially in SCCs, PIK3CA may be amplified. An increased copy number of PIK3CA has been documented in 5% of small cell carcinoma cell lines. Although uncommon, AKT mutations have been identified in 0.5–2% of NSCLC, particularly SCCs, activating the PI3K/AKT/mTOR pathway (Vivanco and Sawyers 2002).

5. *MET*: On chromosome 7q21-q31, the proto-oncogene MET encodes a membrane tyrosine kinase receptor known as hepatocyte growth factor receptor. When the ligand hepatocyte growth factor binds to the receptor, it causes homodimerization, kinase activation, and signaling via downstream pathways such as RAS/RAF/MEK/MAPK, PI3K/AKT, and c-SRC kinase (Sadiq and Salgia 2013). MET gene amplification is identified in 1–7% of treatment-naive NSCLC patients; however, amplification was reported in 21% of patients in one research. Increased MET copy number is possibly more prevalent in SCC than in ADC. It is mutually exclusive with KRAS mutations. Overexpression of the MET protein occurs due to MET amplification, which activates downstream signaling pathways (Bean et al. 2007; Cappuzzo et al. 2009; Onozato et al. 2009; Go et al. 2010). Evidence of gene amplification related to constitutive receptor phosphorylation, activation of the PI3K/AKT pathway, and susceptibility to MET inhibition has been established in vitro. Secondary EGFR-TKI resistance is caused by MET amplification, which occurs in around 20% of individuals with acquired resistance. In this situation, MET amplification drives and maintains the PI3K/AKT pathway, circumventing TKI-mediated EGFR blockade, implying that concurrent MET inhibition might be a mechanism to overcome TKI resistance. About 3–5% of ADC has MET mutations (Beau-Faller et al. 2008).

6. *ROS1*: OS1 is a proto-oncogene that encodes a transmembrane tyrosine kinase receptor with substantial similarity to ALK in the protein kinase domain. It is found on chromosome 6q22 (Chin et al. 2012). The PI3K/AKT/mTOR, STAT3, and RAS/MAPK/ERK pathways are activated when ROS1 is activated. ROS1 fusion was found in an NSCLC cell line (1 of 41) and a patient sample (1 of 150) (SLC34A2-ROS1 and CD74-ROS1, respectively) in a large-scale phosphoproteomic search for tyrosine kinase activity in lung cancer in 2007 (Bergethon et al. 2012). Following that, employing whole-genome and transcriptome sequencing, a new KDELR2-ROS1 in-frame fusion was discovered in adenocarcinoma from a nonsmoker. ROS1 rearrangements were discovered in 18 of 694 ADCs (2.6 %) and 13 of 1116 ADCs in two significant investigations utilizing FISH (1.2%). KDELR2, FIG, SDC4, TPM3, EZR, LRIG3, EZR, CD74, and SLC34A2 are some of the 5′ fusion partners reported in the ROS1 gene rearrangements; however, it is unclear what role, if any, the partner plays in the fusion kinase's oncogenic activity (Takeuchi et al. 2012). Similarly, as ALK rearrangements, ROS1 rearrangements appear to be more prevalent in younger individuals who have never smoked or are of Asian origin. Lung tumors with ROS1 rearrangements are also responsive to kinase inhibitors, such as the ALK/MET inhibitor crizotinib, according to in vitro and early clinical data (Bergethon et al. 2012).
7. *ALK*: In a fraction of lung malignancies, fusions of the intracellular kinase domain with the amino-terminal end of echinoderm microtubule-associated protein-like 4 (EML4) occur (Koivunen et al. 2008a, b; Choi et al. 2008). A brief inversion causes the rearrangement on chromosome 2p, in which intron 13 of EML4 is joined to intron 19 of ALK [inv (2) (p21; p23)] in the most frequent variation. Exons 1–13 of EML4 joining exons 20–29 of ALK have been identified as the most prevalent EML4-ALK fusion variant. KIF5B (kinesin family member 5b), TFG (TRK-fused gene), and KLC-1 have all recently been discovered as partner genes in a small % age of ALK rearrangements (1% of cases) (kinesin light chain1) (Soda et al. 2007). In vitro, the oncogenic EML4-ALK fusion protein displays gain of function activity. In vivo mice models expressing EML4-ALK develop numerous lungs ADC sensitive to pharmacologic ALK inhibition. Through the RAS/RAF/MAPK1, PI3K/AKT, and JAK3-STAT3 signaling pathways, ALK activation is associated with cell growth and apoptosis suppression (Choi et al. 2008; Mao et al. 2010; Shaw et al. 2004). Although other studies have indicated a somewhat lower incidence, ALK rearrangements have been observed in about 4% of unselected NSCLC. They are more prevalent in ADCs from younger patients who do not smoke or smoke lightly, and they virtually always show up in ADCs (Shaw et al. 2004; Rikova et al. 2007; Takeuchi et al. 2009). While ALK rearrangements usually are mutually exclusive with EGFR and KRAS mutations, examples of coexisting EGFR mutations have been described, indicating that TKI resistance is possible (Solomon et al. 2009). Drug resistance develops with evidence of new ALK point mutations and activation of EGFR signaling involvement in some cases, even though ALK inhibition with the

tyrosine kinase inhibitor crizotinib provides dramatic responses (Sasaki et al. 2011).

8. *BRAF*: BRAF is a serine/threonine protein kinase, the downstream effector protein of KRAS. It stimulates the MAPK signal transduction pathway, which is involved in cell proliferation and survival regulation. BRAF activates downstream mediators MEK1 and MEK2, activating ERK1 and ERK2, which regulate growth-regulating proteins such as c-JUN and ELK1. Increased kinase activity and transforming activity in vitro result from activating mutations in BRAF (Schmid et al. 2009; Davies et al. 2002).
   In melanoma, activating BRAF mutations are prevalent, while only around 3% of NSCLC patients have them. NSCLC has a lower percentage of V600E mutations that disrupt the protein's kinase domain than melanoma and colorectal cancer. Exon 15 V600E mutations contribute to up to 50% of BRAF mutations in lung ADC, followed by exon 11 G469A and exon 15 D594G. Some BRAF mutations in NSCLC are found in the kinase domain (e.g., V600E, D594G, and L596R), whereas others are found in the G-loop of the activation domain (e.g., V600E, D594G, and L596R) (such as G465V and G468A) (Davies et al. 2002; Brose et al. 2002; Marchetti et al. 2011; Naoki et al. 2002; Paik et al. 2011). Because the BRAF and KRAS genes are part of the EGFR-mediated signaling system, mutations in these genes are virtually invariably mutually exclusive, consistent with a similar downstream route to transformation. ADC nearly invariably has BRAF mutations in lung cancer. BRAF variants that are not V600E have been linked to current or former smokers, whereas V600E mutations have been linked to female never smokers (Falchook et al. 2012). While BRAF mutations are infrequent, they are an attractive therapeutic target since targeted treatments for melanoma are currently in clinical use, albeit there is less data on how well this approach works in NSCLC.

   FGFR1: Several genes, including SOX, PDGFRA, and FGFR1, have exhibited somatic gene amplification in SCCs (Cancer Genome Atlas Research Network 2012). The MAPK and PI3K pathways are activated by FGFR1, a membrane receptor tyrosine kinase that governs cell growth (Tran et al. 2013). In vitro, FGFR1 amplification exerts an oncogenic impact responsive to small-molecule inhibition in NSCLC cell lines. Amplification of the FGFR1 gene has been found in around 20% of SCCs, while they are rare in ADCs (Tran et al. 2013; Dutt et al. 2011).
9. *RET*: RET is a receptor tyrosine kinase that plays a role in neural crest formation. It is found on chromosome 10q11.2 (Wells Jr and Santoro 2009). Although RET mutations have long been linked to papillary and medullary thyroid carcinoma, activation of RET by chromosomal rearrangement has only recently been discovered in a small percentage of lung malignancies (Lipson et al. 2012; Ju et al. 2012). The functional RET kinase domain from exons 12–20 is fused to KIF5B (kinesin family 5B gene), which is 10 Mb away from RET on chromosome 10 and encodes a coiled-coil domain involved in organelle trafficking (Ju et al. 2012; Kohno et al. 2012). KIF5B-RET fusions have been detected in 1–2% of lung ADC using massively parallel sequencing methods. They are

mutually exclusive of other driver mutations affecting EGFR, KRAS, or ALK. RET rearrangements were found in 10 of 159 lung ADC from never or light smokers who were known to be the wild type for other driver mutations (EGFR, KRAS, ALK, HER2, BRAF, and ROS1) (Lipson et al. 2012). Rearrangements of RET, like ALK and ROS1, appear to be linked to ADC in never smokers. Notably, numerous multi-kinase inhibitors are effective against RET, and cell lines producing KIF5B-RET fusions are susceptible to RET suppression in vitro (Lipson et al. 2012; Kohno et al. 2012).

10. *DDR2*: In SCCs, mutations in DDR2 were found in 3.8% of patients after a sequencing screen that included the complete tyrosine kinome. DDR2 is a collagen-binding membrane-bound receptor tyrosine kinase involved in cell proliferation and survival control. In vitro, DDR2 mutations are linked to carcinogenic activity, responsive to dasatinib suppression (Hammerman et al. 2011).
11. *HER2*: Along with EGFR, the human epidermal growth factor receptor 2 (HER2/ERBB2) genes encodes a membrane-bound receptor tyrosine kinase (Tzahar et al. 1996). It does not directly bind ligands, unlike other ERBB receptors. However, it can form heterodimers with other ligand-bound members of the receptor family. Signaling occurs through several signal transduction pathways, including PI3K, MAPK, and JAK/STAT. HER2 activation occurs in a limited percentage of lung tumors, with overexpression occurring in around 20% of cases, gene amplification in 2%, and activating mutations in 1.6–4% of NSCLC (Graus-Porta et al. 1997; Heinmöller et al. 2003; Stephens et al. 2004). Exon 20 in-frame insertions of 3–12 base pairs in length are HER2 activating mutations. Multiple adenosquamous carcinomas developed in a transgenic mouse model expressing mutant HER2 and were susceptible to small-molecule inhibition, indicating that HER2 had carcinogenic potential (Tomizawa et al. 2011). In studies, HER2 mutations are related to the female gender, Asian ethnicity, and nonsmoking status, similar to the clinical profile of EGFR mutant tumors. HER2 mutations occur predominantly in ADC, and mutations occur in tumors that are wild-type for EGFR and KRAS (Perera et al. 2009).

7.11 Mathematical Modeling of AKT Signaling Pathway for Cancer Development

Cancer is the world's leading cause of death, especially lung cancer. In order to create anticancer medications, a thorough understanding of oncogene and tumor suppressor signaling networks in cancer cells is essential. To govern cell development, cell division, cell death, and cell migration, various transcription factors work in concert. The AKT which has been proven to either prevent or stimulate tumor development is the subject of this research. The PI3K/AKT/mTOR pathway has been linked to carcinogenesis and disease progression in NSCLC patients. Several PI3K, AKT, and mTOR inhibitors are now being developed and tested in preclinical studies and early phase clinical trials for NSCLC. AKT is a protein kinase that

belongs to the AGC (PKA/PKG/PKC) family. It has three homologs: AKT1, AKT2, and AKT3 found on chromosomes 14q32, 19q13, and 1q44, respectively. Following AKT activation, a variety of downstream consequences are possible. BAD and BAX, two pro-apoptotic Bcl2 family members, may be inhibited as a consequence.

Mdm2 is phosphorylated by AKT, which inhibits p53-mediated apoptosis and forkhead transcription factors that create cell death promoters. The transcription factor nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) is essential in PI3K/AKT pathway activation. Apoptosis, cell cycle regulation, immunological modulation, cell survival, and cell adhesion and differentiation are all regulated by NF-κB. 18 AKTs stop the IB family, namely IκB, from negatively regulating NF-κB. IκB returns NF-κB to the cytoplasm after removing it from DNA. The activation of the protein kinase mTOR is another essential downstream pathway activated by AKT activation. TSC2 is phosphorylated, which activates Rheb, which drives the multiprotein complex mTORC1. mTORC1 promotes carcinogenesis, cell cycle regulation, and apoptosis suppression by activating the eIF4 complex further downstream. AKT activation is triggered by phosphorylation of serine 473 by another mTOR complex, mTORC2.

The AKT pathway is upregulated in a large number of patients with NSCLC. In a study of 110 NSCLC tumors, immunohistochemistry found that 51% had elevated AKT activation. There was also a link between AKT activation and enhanced mTOR and forkhead activity, which are essential AKT downstream targets. The disparity between levels of AKT overexpression and the existence of somatic mutations might point to coexisting mutations or amplifications that activate AKT. AKT activation has been seen in preclinical investigations in NSCLC cell lines, with loss of PTEN, EGFR or PIK3CA mutation, or HER2 amplification being implicated.

Mathematical modeling is essential for identifying underlying mechanisms in malignancies, interactions with other cells such as immune cells, cellular invasion, cancer treatment, apoptotic mechanisms, and particular signaling pathways such as JAK-STAT, MYC-p53, and microRNAs. The primary mechanism of AKT-mediated cancer cell death is yet unknown. To our knowledge, no mathematical research has looked at the fundamental processes of AKT's apoptosis mediation in cancer cells. A mathematical model of AKT-mediated apoptosis pathways in controlling tumor development and cancer cell death has been established. We look at the best anticancer medicine dosing schedule using optimal control theory. Figure 7.4 shows a mathematical model of AKT's influence on various pathways for tumor growth.

RTKs are the high-affinity cell surface receptors for many polypeptides growth factors, cytokines, and hormones. When the ligands such as growth factor bind to the RTKs, PI3K activation takes place through phosphorylation, which further activates the signal cascade ultimately activating AKT.

Once AKT is activated, it is translocated from the plasma membrane to the cytoplasm and nucleus, where many of its substrates reside. Phosphorylation by AKT can be inhibitory or stimulatory, either suppressing or enhancing the activity of target proteins. Depending on the target protein, AKT can regulate different cell functions.

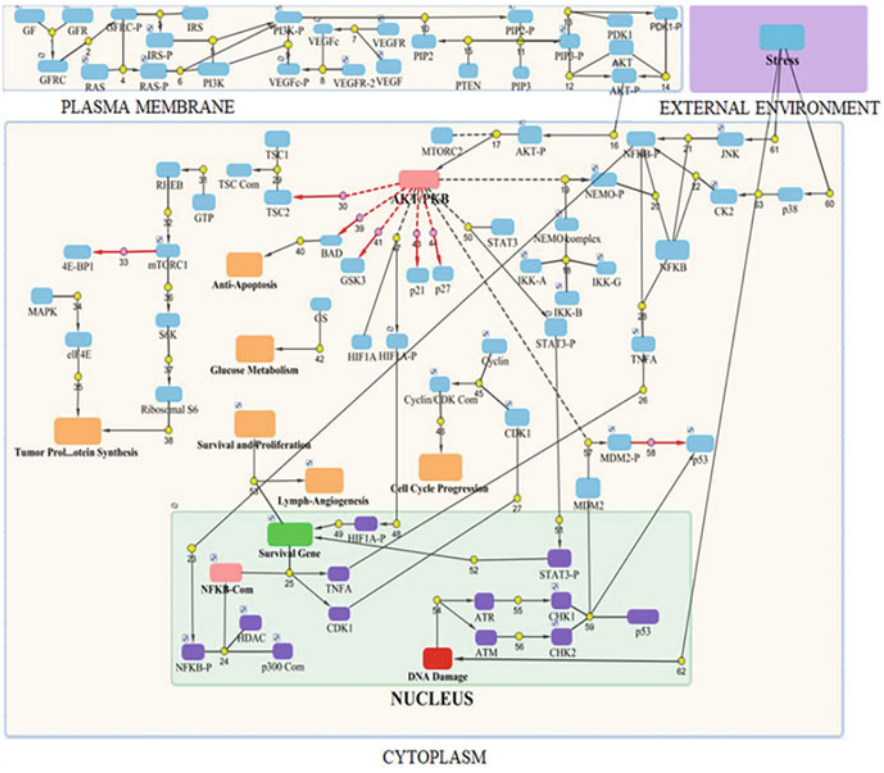


Fig. 7.4 Mathematical model representing AKT and its effect during NSCLC

AKT enhances the survival of cells by blocking the function of proapoptotic protein BAD (BCL2 associated agonist of cell death) and its downstream processes, leading to the antiapoptotic activity of the cancer cells to take place.

AKT activation through the VEGF results in AKT activating the transcription factors—HIF1α, Ets-1, NFKB, and STAT3, which help in enhancing the process of angiogenesis and lymphangiogenesis which facilitates the tumor in performing metastatic and invasive activities. AKT also promotes tumor survival and proliferation via mediating the NFKB pathway where it indirectly activates NEMO (NF-kappa-B essential modulator) which in turn activates the transcription factors responsible for tumor cell survival and proliferation. AKT also plays an important role in promoting protein synthesis, glucose metabolism, and cell cycle metabolism via inhibiting the TSC1/TSC2 complex, GSK3 protein, and p21and p27 tumor suppressors along with their downstream processes respectively as it is essential for the survival and proliferation of tumor cells.

AKT has been found to be mediating one of the most important tumor suppressor pathways, i.e., the p53 signaling pathway. p53 signaling pathway is activated due to DNA damage caused by extracellular factors such as UV, stress, hypoxia, and

genotoxic drugs. AKT inhibits the activity of tumor suppressor gene p53 indirectly via activating its feedback inhibitor MDM2.

References

Antonarakis ES, Eisenberger MA (2013) Phase III trials with docetaxel-based combinations for metastatic castration-resistant prostate cancer: time to learn from past experiences. J Clin Oncol 31(14):1709–1712. https://doi.org/10.1200/JCO.2013.48.8825

Balak MN, Gong Y, Riely GJ, Somwar R, Li AR, Zakowski MF, Chiang A, Yang G, Ouerfelli O, Kris MG, Ladanyi M, Miller VA, Pao W (2006) Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. Clin Cancer Res 12(21):6494–6501. https://doi.org/10.1158/1078-0432.CCR-06-1570

Bean J, Brennan C, Shih JY, Riely G, Viale A, Wang L, Chitale D, Motoi N, Szoke J, Broderick S, Balak M, Chang WC, Yu CJ, Gazdar A, Pass H, Rusch V, Gerald W, Huang SF, Yang PC, Miller V, Pao W (2007) MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. Proc Natl Acad Sci U S A 104(52):20932–20937. https://doi.org/10.1073/pnas.0710370104

Beau-Faller M, Ruppert AM, Voegeli AC, Neuville A, Meyer N, Guerin E, Legrain M, Mennecier B, Wihlm JM, Massard G, Quoix E, Oudet P, Gaub MP (2008) MET gene copy number in non-small cell lung cancer: molecular analysis in a targeted tyrosine kinase inhibitor naïve cohort. J Thorac Oncol 3(4):331–339. https://doi.org/10.1097/JTO.0b013e318168d9d4

Bergethon K, Shaw AT, Ou SH, Katayama R, Lovly CM, McDonald NT, Massion PP, Siwak-Tapp C, Gonzalez A, Fang R, Mark EJ, Batten JM, Chen H, Wilner KD, Kwak EL, Clark JW, Carbone DP, Ji H, Engelman JA, Mino-Kenudson M, Iafrate AJ (2012) ROS1 rearrangements define a unique molecular class of lung cancers. J Clin Oncol 30(8):863–870. https://doi.org/10.1200/JCO.2011.35.6345

Bethune G, Bethune D, Ridgway N, Xu Z (2010) Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update. J Thorac Dis 2(1):48–51

Brambilla E, Moro D, Gazzeri S, Brambilla C (1999) Alterations of expression of Rb, p16(INK4A) and cyclin D1 in non-small cell lung carcinoma and their clinical significance. J Pathol 188(4):351–360

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68(6):394–424. https://doi.org/10.3322/caac.21492

Brognard J, Clark AS, Ni Y, Dennis PA (2001) Akt/protein kinase B is constitutively active in non-small cell lung cancer cells and promotes cellular survival and resistance to chemotherapy and radiation. Cancer Res 61(10):3986–3997

Brose MS, Volpe P, Feldman M, Kumar M, Rishi I, Gerrero R, Einhorn E, Herlyn M, Minna J, Nicholson A, Roth JA, Albelda SM, Davies H, Cox C, Brignell G, Stephens P, Futreal PA, Wooster R, Stratton MR, Weber BL (2002) BRAF and RAS mutations in human lung cancer and melanoma. Cancer Res 62(23):6997–7000

Butts C, Murray N, Maksymiuk A, Goss G, Marshall E, Soulières D, Cormier Y, Ellis P, Price A, Sawhney R, Davis M, Mansi J, Smith C, Vergidis D, Ellis P, MacNeil M, Palmer M (2005) Randomized phase IIB trial of BLP25 liposome vaccine in stage IIIB and IV non-small-cell lung cancer. J Clin Oncol 23(27):6674–6681. https://doi.org/10.1200/JCO.2005.13.011

Butts C, Maksymiuk A, Goss G, Soulières D, Marshall E, Cormier Y, Ellis PM, Price A, Sawhney R, Beier F, Falk M, Murray N (2011) Updated survival analysis in patients with stage IIIB or IV non-small-cell lung cancer receiving BLP25 liposome vaccine (L-BLP25): phase IIB randomized, multicenter, open-label trial. J Cancer Res Clin Oncol 137(9):1337–1342. https://doi.org/10.1007/s00432-011-1003-3

Butts C, Socinski MA, Mitchell PL, Thatcher N, Havel L, Krzakowski M, Nawrocki S, Ciuleanu TE, Bosquée L, Trigo JM, Spira A, Tremblay L, Nyman J, Ramlau R, Wickart-Johansson G, Ellis P, Gladkov O, Pereira JR, Eberhardt WE, Helwig C (2014) Tecemotide (L-BLP25) versus placebo after chemoradiotherapy for stage III non-small-cell lung cancer (START): a randomised, double-blind, phase 3 trial. Lancet 15(1):59–68. https://doi.org/10.1016/S1470-2045(13)70510-2

Cancer Genome Atlas Research Network (2012) Comprehensive genomic characterization of squamous cell lung cancers. Nature 489(7417):519–525. https://doi.org/10.1038/nature11404

Cappuzzo F, Marchetti A, Skokan M, Rossi E, Gajapathy S, Felicioni L, Del Grammastro M, Sciarrotta MG, Buttitta F, Incarbone M, Toschi L, Finocchiaro G, Destro A, Terracciano L, Roncalli M, Alloisio M, Santoro A, Varella-Garcia M (2009) Increased MET gene copy number negatively affects survival of surgically resected non-small-cell lung cancer patients. J Clin Oncol 27(10):1667–1674. https://doi.org/10.1200/JCO.2008.19.1635

Chin LP, Soo RA, Soong R, Ou SH (2012) Targeting ROS1 with anaplastic lymphoma kinase inhibitors: a promising therapeutic strategy for a newly defined molecular subset of non-small-cell lung cancer. J Thorac Oncol 7(11):1625–1630. https://doi.org/10.1097/JTO.0b013e31826baf83

Choi YL, Takeuchi K, Soda M, Inamura K, Togashi Y, Hatano S, Enomoto M, Hamada T, Haruta H, Watanabe H, Kurashina K, Hatanaka H, Ueno T, Takada S, Yamashita Y, Sugiyama Y, Ishikawa Y, Mano H (2008) Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. Cancer Res 68(13):4971–4976. https://doi.org/10.1158/0008-5472.CAN-07-6158

Clark SB, Alsubait S (2021) Non small cell lung cancer. In: StatPearls. StatPearls Publishing, St. Petersburg

Cully M, You H, Levine AJ, Mak TW (2006) Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis. Cancer 6(3):184–192. https://doi.org/10.1038/nrc1819

Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Futreal PA (2002) Mutations of the BRAF gene in human cancer. Nature 417(6892):949–954. https://doi.org/10.1038/nature00766

Deben C, Deschoolmeester V, De Waele J, Jacobs J, Van den Bossche J, Wouters A, Peeters M, Rolfo C, Smits E, Lardon F, Pauwels P (2018) Hypoxia-induced cisplatin resistance in non-small cell lung cancer cells is mediated by HIF-1α and mutant p53 and can be overcome by induction of oxidative stress. Cancer 10(4):126. https://doi.org/10.3390/cancers10040126

Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, Fulton L, Fulton RS, Zhang Q, Wendl MC, Lawrence MS, Larson DE, Chen K, Dooling DJ, Sabo A, Hawes AC, Wilson RK (2008) Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455(7216):1069–1075. https://doi.org/10.1038/nature07423

Downward J (2003) Targeting RAS signalling pathways in cancer therapy. Cancer 3(1):11–22

Dutt A, Ramos AH, Hammerman PS, Mermel C, Cho J, Sharifnia T, Chande A, Tanaka KE, Stransky N, Greulich H, Gray NS, Meyerson M (2011) Inhibitor-sensitive FGFR1 amplification in human non-small cell lung cancer. PLoS One 6(6):e20351. https://doi.org/10.1371/journal.pone.0020351

Eager R, Nemunaitis J (2005) GM-CSF gene-transduced tumor vaccines. Mol Ther 12(1):18–27. https://doi.org/10.1016/j.ymthe.2005.02.012

Engelman JA, Luo J, Cantley LC (2006) The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. Genetics 7(8):606–619. https://doi.org/10.1038/nrg1879

Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, Lindeman N, Gale CM, Zhao X, Christensen J, Kosaka T, Holmes AJ, Rogers AM, Cappuzzo F, Mok T, Lee C, Johnson BE, Cantley LC, Jänne PA (2007) MET amplification leads to gefitinib resistance in lung cancer

by activating ERBB3 signaling. Science 316(5827):1039–1043. https://doi.org/10.1126/science.1141478

Fakiris AJ, McGarry RC, Yiannoutsos CT, Papiez L, Williams M, Henderson MA, Timmerman R (2009) Stereotactic body radiation therapy for early-stage non-small-cell lung carcinoma: four-year results of a prospective phase II study. Int J Radiat Oncol 75(3):677–682. https://doi.org/10.1016/j.ijrobp.2008.11.042

Falchook GS, Long GV, Kurzrock R, Kim KB, Arkenau TH, Brown MP, Hamid O, Infante JR, Millward M, Pavlick AC, O'Day SJ, Blackman SC, Curtis CM, Lebowitz P, Ma B, Ouellet D, Kefford RF (2012) Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumours: a phase 1 dose-escalation trial. Lancet 379(9829):1893–1901. https://doi.org/10.1016/S0140-6736(12)60398-5

Fossella F, Pereira JR, von Pawel J, Pluzanska A, Gorbounova V, Kaukel E, Mattson KV, Ramlau R, Szczesna A, Fidias P, Millward M, Belani CP (2003) Randomized, multinational, phase III study of docetaxel plus platinum combinations versus vinorelbine plus cisplatin for advanced non-small-cell lung cancer: the TAX 326 study group. J Clin Oncol 21(16):3016–3024. https://doi.org/10.1200/JCO.2003.12.046

Gajewski TF, Schreiber H, Fu YX (2013) Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol 14(10):1014–1022. https://doi.org/10.1038/ni.2703

Geary SM, Salem AK (2013) Prostate cancer vaccines: update on clinical development. Onco Targets Ther 2(5):e24523. https://doi.org/10.4161/onci.24523

Giaccone G, Bazhenova LA, Nemunaitis J, Tan M, Juhász E, Ramlau R, van den Heuvel MM, Lal R, Kloecker GH, Eaton KD, Chu Q, Dunlop DJ, Jain M, Garon EB, Davis CS, Carrier E, Moses SC, Shawler DL, Fakhrai H (2015) A phase III study of belagenpumatucel-L, an allogeneic tumour cell vaccine, as maintenance therapy for non-small cell lung cancer. Eur J Cancer 51(16):2321–2329. https://doi.org/10.1016/j.ejca.2015.07.035

Go H, Jeon YK, Park HJ, Sung SW, Seo JW, Chung DH (2010) High MET gene copy number leads to shorter survival in patients with non-small cell lung cancer. J Thorac Oncol 5(3):305–313. https://doi.org/10.1097/JTO.0b013e3181ce3d1d

Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG, Shepherd FA (2011) Non-small-cell lung cancer. Lancet 378(9804):1727–1740. https://doi.org/10.1016/S0140-6736(10)62101-0

Gonzalez G, Crombet T, Torres F, Catala M, Alfonso L, Osorio M, Neninger E, Garcia B, Mulet A, Perez R, Lage R (2003) Epidermal growth factor-based cancer vaccine for non-small-cell lung cancer therapy. Ann Oncol 14(3):461–466. https://doi.org/10.1093/annonc/mdg102

Graus-Porta D, Beerli RR, Daly JM, Hynes NE (1997) ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. EMBO J 16(7):1647–1655. https://doi.org/10.1093/emboj/16.7.1647

Greulich H, Chen TH, Feng W, Jänne PA, Alvarez JV, Zappaterra M, Bulmer SE, Frank DA, Hahn WC, Sellers WR, Meyerson M (2005) Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. PLoS Med 2(11):313. https://doi.org/10.1371/journal.pmed.0020313

Grutters JP, Kessels AG, Pijls-Johannesma M, De Ruysscher D, Joore MA, Lambin P (2010) Comparison of the effectiveness of radiotherapy with photons, protons and carbon-ions for non-small cell lung cancer: a meta-analysis. Radiother Oncol 95(1):32–40. https://doi.org/10.1016/j.radonc.2009.08.003

Hammerman PS, Sos ML, Ramos AH, Xu C, Dutt A, Zhou W, Brace LE, Woods BA, Lin W, Zhang J, Deng X, Lim SM, Heynck S, Peifer M, Simard JR, Lawrence MS, Onofrio RC, Salvesen HB, Seidel D, Zander T, Meyerson M (2011) Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. Cancer Discov 1(1):78–89. https://doi.org/10.1158/2159-8274.CD-11-0005

Harbour JW, Lai SL, Whang-Peng J, Gazdar AF, Minna JD, Kaye FJ (1988) Abnormalities in structure and expression of the human retinoblastoma gene in SCLC. Science 241(4863):353–357. https://doi.org/10.1126/science.2838909

Heinmöller P, Gross C, Beyser K, Schmidtgen C, Maass G, Pedrocchi M, Rüschoff J (2003) HER2 status in non-small cell lung cancer: results from patient screening for enrollment to a phase II study of herceptin. Clin Cancer Res 9(14):5238–5243

Heist RS, Sequist LV, Engelman JA (2012) Genetic changes in squamous cell lung cancer: a review. J Thorac Oncol 7(5):924–933. https://doi.org/10.1097/JTO.0b013e31824cc334

Howington JA, Blum MG, Chang AC, Balekian AA, Murthy SC (2013) Treatment of stage I and II non-small cell lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 143(5): 278–313. https://doi.org/10.1378/chest.12-2359

Husgafvel-Pursiainen K, Boffetta P, Kannio A, Nyberg F, Pershagen G, Mukeria A, Constantinescu V, Fortes C, Benhamou S (2000) p53 mutations and exposure to environmental tobacco smoke in a multicenter study on lung cancer. Cancer Res 60(11):2906–2911

Ihle NT, Byers LA, Kim ES, Saintigny P, Lee JJ, Blumenschein GR, Tsao A, Liu S, Larsen JE, Wang J, Diao L, Coombes KR, Chen L, Zhang S, Abdelmelek MF, Tang X, Papadimitrakopoulou V, Minna JD, Lippman SM, Hong WK, Powis G (2012) Effect of KRAS oncogene substitutions on protein behavior: implications for signaling and clinical outcome. J Natl Cancer Inst 104(3):228–239. https://doi.org/10.1093/jnci/djr523

Iyengar P, Gerber DE (2013) Locally advanced lung cancer: an optimal setting for vaccines and other immunotherapies. Cancer J 19(3):247–262. https://doi.org/10.1097/PPO.0b013e318292e51a

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. CA Cancer J Clin 61(2):69–90. https://doi.org/10.3322/caac.20107

Jin G, Kim MJ, Jeon HS, Choi JE, Kim DS, Lee EB, Cha SI, Yoon GS, Kim CH, Jung TH, Park JY (2010) PTEN mutations and relationship to EGFR, ERBB2, KRAS, and TP53 mutations in non-small cell lung cancers. Lung Cancer 69(3):279–283. https://doi.org/10.1016/j.lungcan.2009.11.012

Ju YS, Lee WC, Shin JY, Lee S, Bleazard T, Won JK, Kim YT, Kim JI, Kang JH, Seo JS (2012) A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Res 22(3):436–445. https://doi.org/10.1101/gr.133645.111

Karnoub AE, Weinberg RA (2008) Ras oncogenes: split personalities. Mol Cell Biol 9(7):517–531. https://doi.org/10.1038/nrm2438

Kelly K, Crowley J, Bunn PA Jr, Presant CA, Grevstad PK, Moinpour CM, Ramsey SD, Wozniak AJ, Weiss GR, Moore DF, Israel VK, Livingston RB, Gandara DR (2001) Randomized phase III trial of paclitaxel plus carboplatin versus vinorelbine plus cisplatin in the treatment of patients with advanced non-small-cell lung cancer: a Southwest Oncology Group trial. J Clin Oncol 19(13):3210–3218. https://doi.org/10.1200/JCO.2001.19.13.3210

Knudson AG (1993) Antioncogenes and human cancer. Proc Natl Acad Sci U S A 90(23): 10914–10921. https://doi.org/10.1073/pnas.90.23.10914

Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T, Sakamoto H, Tsuta K, Furuta K, Shimada Y, Iwakawa R, Ogiwara H, Oike T, Enari M, Schetter AJ, Okayama H, Haugen A, Skaug V, Chiku S, Yamanaka I, Shibata T (2012) KIF5B-RET fusions in lung adenocarcinoma. Nat Med 18(3):375–377. https://doi.org/10.1038/nm.2644

Koivunen JP, Kim J, Lee J, Rogers AM, Park JO, Zhao X, Naoki K, Okamoto I, Nakagawa K, Yeap BY, Meyerson M, Wong KK, Richards WG, Sugarbaker DJ, Johnson BE, Jänne PA (2008a) Mutations in the LKB1 tumour suppressor are frequently detected in tumours from Caucasian but not Asian lung cancer patients. Br J Cancer 99(2):245–252. https://doi.org/10.1038/sj.bjc.6604469

Koivunen JP, Mermel C, Zejnullahu K et al (2008b) EML4-ALK fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. Clin Cancer Res 14:4275–4283

Kong F, Jirtle RL, Huang DH, Clough RW, Anscher MS (1999) Plasma transforming growth factor-beta1 level before radiotherapy correlates with long term outcome of patients with lung carcinoma. Cancer 86(9):1712–1719

Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T (2004) Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. Cancer Res 64(24):8919–8923. https://doi.org/10.1158/0008-5472.CAN-04-2818

Lagerwaard FJ, Verstegen NE, Haasbeek CJ, Slotman BJ, Paul MA, Smit EF, Senan S (2012) Outcomes of stereotactic ablative radiotherapy in patients with potentially operable stage I non-small cell lung cancer. Int J Radiat Oncol 83(1):348–353. https://doi.org/10.1016/j.ijrobp.2011.06.2003

Larsen JE, Minna JD (2011) Molecular biology of lung cancer: clinical implications. Clin Chest Med 32(4):703–740. https://doi.org/10.1016/j.ccm.2011.08.003

Linardou H, Dahabreh IJ, Kanaloupiti D, Siannis F, Bafaloukos D, Kosmidis P, Papadimitriou CA, Murray S (2008) Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. Lancet 9(10):962–972. https://doi.org/10.1016/S1470-2045(08)70206-7

Linsley PS, Brady W, Grosmaire L, Aruffo A, Damle NK, Ledbetter JA (1991) Binding of the B cell activation antigen B7 to CD28 costimulates T cell proliferation and interleukin 2 mRNA accumulation. J Exp Med 173(3):721–730. https://doi.org/10.1084/jem.173.3.721

Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, Jarosz M, Curran JA, Balasubramanian S, Bloom T, Brennan KW, Donahue A, Downing SR, Frampton GM, Garcia L, Juhn F, Mitchell KC, White E, White J, Zwirko Z, Peretz T, Stephens PJ (2012) Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nat Med 18(3):382–384. https://doi.org/10.1038/nm.2673

Lukey MJ, Katt WP, Cerione RA (2017) Targeting amino acid metabolism for cancer therapy. Drug Discov Today 22(5):796–804. https://doi.org/10.1016/j.drudis.2016.12.003

Luo J, Solimini NL, Elledge SJ (2009) Principles of cancer therapy: oncogene and non-oncogene addiction. Cell 136(5):823–837. https://doi.org/10.1016/j.cell.2009.02.024

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350(21):2129–2139. https://doi.org/10.1056/NEJMoa040938

Mao C, Qiu LX, Liao RY, Du FB, Ding H, Yang WC, Li J, Chen Q (2010) KRAS mutations and resistance to EGFR-TKIs treatment in patients with non-small cell lung cancer: a meta-analysis of 22 studies. Lung Cancer 69(3):272–278. https://doi.org/10.1016/j.lungcan.2009.11.020

Marchetti A, Felicioni L, Malatesta S, Grazia Sciarrotta M, Guetti L, Chella A, Viola P, Pullara C, Mucilli F, Buttitta F (2011) Clinical features and outcome of patients with non-small-cell lung cancer harboring BRAF mutations. J Clin Oncol 29(26):3574–3579. https://doi.org/10.1200/JCO.2011.35.9638

Marchetti A, Ardizzoni A, Papotti M, Crinò L, Rossi G, Gridelli C, Barberis M, Maiorano E, Normanno N, Taddei GL, Scagliotti G, Clemente C, Pinto C (2013) Recommendations for the analysis of ALK gene rearrangements in non-small-cell lung cancer: a consensus of the Italian Association of Medical Oncology and the Italian Society of Pathology and Cytopathology. J Thorac Oncol 8(3):352–358. https://doi.org/10.1097/JTO.0b013e31827d5280

Marignani PA (2005) LKB1, the multitasking tumour suppressor kinase. J Clin Pathol 58(1):15–19. https://doi.org/10.1136/jcp.2003.015255

Marks JL, Gong Y, Chitale D, Golas B, McLellan MD, Kasai Y, Ding L, Mardis ER, Wilson RK, Solit D, Levine R, Michel K, Thomas RK, Rusch VW, Ladanyi M, Pao W (2008) Novel MEK1 mutation identified by mutational analysis of epidermal growth factor receptor signaling pathway genes in lung adenocarcinoma. Cancer Res 68(14):5524–5528. https://doi.org/10.1158/0008-5472.CAN-08-0099

Marsit CJ, Zheng S, Aldape K, Hinds PW, Nelson HH, Wiencke JK, Kelsey KT (2005) PTEN expression in non-small-cell lung cancer: evaluating its relation to tumor characteristics, allelic

loss, and epigenetic alteration. Hum Pathol 36(7):768–776. https://doi.org/10.1016/j.humpath.2005.05.006

Masters GA, Temin S, Azzoli CG, Giaccone G, Baker S Jr, Brahmer JR, Ellis PM, Gajra A, Rackear N, Schiller JH, Smith TJ, Strawn JR, Trent D, Johnson DH (2015) Systemic therapy for stage IV non-small-cell lung cancer: American Society of Clinical Oncology clinical practice guideline update. J Clin Oncol 33(30):3488–3515. https://doi.org/10.1200/JCO.2015.62.1342

Matsumoto S, Iwakawa R, Takahashi K, Kohno T, Nakanishi Y, Matsuno Y, Suzuki K, Nakamoto M, Shimizu E, Minna JD, Yokota J (2007) Prevalence and specificity of LKB1 genetic alterations in lung cancers. Oncogene 26(40):5911–5918. https://doi.org/10.1038/sj.onc.1210418

Mogi A, Kuwano H (2011) TP53 mutations in nonsmall cell lung cancer. J Biomed Biotechnol 2011:583929. https://doi.org/10.1155/2011/583929

Naoki K, Chen TH, Richards WG, Sugarbaker DJ, Meyerson M (2002) Missense mutations of the BRAF gene in human lung adenocarcinoma. Cancer Res 62(23):7001–7003

Nemunaitis J, Dillman RO, Schwarzenberger PO, Senzer N, Cunningham C, Cutler J, Tong A, Kumar P, Pappen B, Hamilton C, DeVol E, Maples PB, Liu L, Chamberlin T, Shawler DL, Fakhrai H (2006) Phase II study of belagenpumatucel-L, a transforming growth factor beta-2 antisense gene-modified allogeneic tumor cell vaccine in non-small-cell lung cancer. J Clin Oncol 24(29):4721–4730. https://doi.org/10.1200/JCO.2005.05.5335

Neninger Vinageras E, de la Torre A, Osorio Rodríguez M, Catalá Ferrer M, Bravo I, Mendoza del Pino M, Abreu Abreu D, Acosta Brooks S, Rives R, del Castillo Carrillo C, González Dueñas M, Viada C, García Verdecia B, Crombet Ramos T, González Marinello G, Lage Dávila A (2008) Phase II randomized controlled trial of an epidermal growth factor vaccine in advanced non-small-cell lung cancer. J Clin Oncol 26(9):1452–1458. https://doi.org/10.1200/JCO.2007.11.5980

Neurath MF, Finotto S (2012) The emerging role of T cell cytokines in non-small cell lung cancer. Cytokine Growth Factor Rev 23(6):315–322. https://doi.org/10.1016/j.cytogfr.2012.08.009

Onozato R, Kosaka T, Achiwa H, Kuwano H, Takahashi T, Yatabe Y, Mitsudomi T (2007) LKB1 gene mutations in Japanese lung cancer patients. Cancer Sci 98(11):1747–1751. https://doi.org/10.1111/j.1349-7006.2007.00585.x

Onozato R, Kosaka T, Kuwano H, Sekido Y, Yatabe Y, Mitsudomi T (2009) Activation of MET by gene amplification or by splice mutations deleting the juxtamembrane domain in primary resected lung cancers. J Thorac Oncol 4(1):5–11. https://doi.org/10.1097/JTO.0b013e3181913e0e

Otterson GA, Kratzke RA, Coxon A, Kim YW, Kaye FJ (1994) Absence of p16INK4 protein is restricted to the subset of lung cancer lines that retains wildtype RB. Oncogene 9(11):3375–3378

Paik PK, Arcila ME, Fara M, Sima CS, Miller VA, Kris MG, Ladanyi M, Riely GJ (2011) Clinical characteristics of patients with lung adenocarcinomas harboring BRAF mutations. J Clin Oncol 29(15):2046–2051. https://doi.org/10.1200/JCO.2010.33.1280

Papadimitrakopoulou V (2012) Development of PI3K/AKT/mTOR pathway inhibitors and their application in personalized therapy for non-small-cell lung cancer. J Thorac Oncol 7(8):1315–1326. https://doi.org/10.1097/JTO.0b013e31825493eb

Perera SA, Li D, Shimamura T, Raso MG, Ji H, Chen L, Borgman CL, Zaghlul S, Brandstetter KA, Kubo S, Takahashi M, Chirieac LR, Padera RF, Bronson RT, Shapiro GI, Greulich H, Meyerson M, Guertler U, Chesa PG, Solca F, Wong KK (2009) HER2YVMA drives rapid development of adenosquamous lung tumors in mice that are sensitive to BIBW2992 and rapamycin combination therapy. Proc Natl Acad Sci U S A 106(2):474–479. https://doi.org/10.1073/pnas.0808930106

Pisters KM, Evans WK, Azzoli CG, Kris MG, Smith CA, Desch CE, Somerfield MR, Brouwers MC, Darling G, Ellis PM, Gaspar LE, Pass HI, Spigel DR, Strawn JR, Ung YC, Shepherd FA (2007) Cancer Care Ontario and American Society of Clinical Oncology adjuvant chemotherapy

and adjuvant radiation therapy for stages I-IIIA resectable non small-cell lung cancer guideline. J Clin Oncol 25(34):5506–5518. https://doi.org/10.1200/JCO.2007.14.1226

Prenzel N, Fischer OM, Streit S, Hart S, Ullrich A (2001) The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification. Endocr Relat Cancer 8(1):11–31. https://doi.org/10.1677/erc.0.0080011

Quoix E, Ramlau R, Westeel V, Papai Z, Madroszyk A, Riviere A, Koralewski P, Breton JL, Stoelben E, Braun D, Debieuvre D, Lena H, Buyse M, Chenard MP, Acres B, Lacoste G, Bastien B, Tavernaro A, Bizouarne N, Bonnefoy JY, Limacher JM (2011) Therapeutic vaccination with TG4010 and first-line chemotherapy in advanced non-small-cell lung cancer: a controlled phase 2B trial. Lancet 12(12):1125–1133. https://doi.org/10.1016/S1470-2045(11)70259-5

Ramalingam S, Belani C (2008) Systemic chemotherapy for advanced non-small cell lung cancer: recent advances and future directions. Oncologist 13(1):5–13. https://doi.org/10.1634/theoncologist.13-S1-5

Ramlau R, Quoix E, Rolski J, Pless M, Lena H, Lévy E, Krzakowski M, Hess D, Tartour E, Chenard MP, Limacher JM, Bizouarne N, Acres B, Halluard C, Velu T (2008) A phase II study of Tg4010 (Mva-Muc1-Il2) in association with chemotherapy in patients with stage III/IV Non-small cell lung cancer. J Thorac Oncol 3(7):735–744. https://doi.org/10.1097/JTO.0b013e31817c6b4f

Raso MG, Wistuba II (2007) Molecular pathogenesis of early-stage non-small cell lung cancer and a proposal for tissue banking to facilitate identification of new biomarkers. J Thorac Oncol 2(7):128–135. https://doi.org/10.1097/JTO.0b013e318074fe42

Riely GJ, Kris MG, Rosenbaum D, Marks J, Li A, Chitale DA, Nafa K, Riedel ER, Hsu M, Pao W, Miller VA, Ladanyi M (2008) Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma. Clin Cancer Res 14(18):5731–5734. https://doi.org/10.1158/1078-0432.CCR-08-0646

Riely GJ, Marks J, Pao W (2009) KRAS mutations in non-small cell lung cancer. Proc Am Thorac Soc 6(2):201–205. https://doi.org/10.1513/pats.200809-107LC

Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan Z, Stokes M, Sullivan L, Mitchell J, Wetzel R, Macneill J, Ren JM, Yuan J, Bakalarski CE, Comb MJ (2007) Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell 131(6):1190–1203. https://doi.org/10.1016/j.cell.2007.11.025

Rodenhuis S, Slebos RJ (1992) Clinical significance of ras oncogene activation in human lung cancer. Cancer Res 52(9):2665–2669

Sadiq AA, Salgia R (2013) MET as a possible target for non-small-cell lung cancer. J Clin Oncol 31(8):1089–1096. https://doi.org/10.1200/JCO.2012.43.9422

Salgia R, Lynch T, Skarin A, Lucca J, Lynch C, Jung K, Hodi FS, Jaklitsch M, Mentzer S, Swanson S, Lukanich J, Bueno R, Wain J, Mathisen D, Wright C, Fidias P, Donahue D, Clift S, Hardy S, Neuberg D, Dranoff G (2003) Vaccination with irradiated autologous tumor cells engineered to secrete granulocyte-macrophage colony-stimulating factor augments antitumor immunity in some patients with metastatic non-small-cell lung carcinoma. J Clin Oncol 21(4):624–630. https://doi.org/10.1200/JCO.2003.03.091

Sanchez-Cespedes M, Parrella P, Esteller M, Nomoto S, Trink B, Engles JM, Westra WH, Herman JG, Sidransky D (2002) Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. Cancer Res 62(13):3659–3662

Sasaki T, Koivunen J, Ogino A, Yanagita M, Nikiforow S, Zheng W, Lathan C, Marcoux JP, Du J, Okuda K, Capelletti M, Shimamura T, Ercan D, Stumpfova M, Xiao Y, Weremowicz S, Butaney M, Heon S, Wilner K, Christensen JG, Jänne PA (2011) A novel ALK secondary mutation and EGFR signaling cause resistance to ALK kinase inhibitors. Cancer Res 71(18):6051–6060. https://doi.org/10.1158/0008-5472.CAN-11-1340

Scagliotti GV, De Marinis F, Rinaldi M, Crinò L, Gridelli C, Ricci S, Matano E, Boni C, Marangolo M, Failla G, Altavilla G, Adamo V, Ceribelli A, Clerici M, Di Costanzo F, Frontini L, Tonato M (2002) Phase III randomized trial comparing three platinum-based

doublets in advanced non-small-cell lung cancer. J Clin Oncol 20(21):4285–4291. https://doi.org/10.1200/JCO.2002.02.068

Schiller JH, Harrington D, Belani CP, Langer C, Sandler A, Krook J, Zhu J, Johnson DH (2002) Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N Engl J Med 346(2):92–98. https://doi.org/10.1056/NEJMoa011954

Schmid K, Oehl N, Wrba F, Pirker R, Pirker C, Filipits M (2009) EGFR/KRAS/BRAF mutations in primary lung adenocarcinomas and corresponding locoregional lymph node metastases. Clin Cancer Res 15(14):4554–4560. https://doi.org/10.1158/1078-0432.CCR-09-0089

Schuurbiers OC, Meijer TW, Kaanders JH, Looijen-Salamon MG, de Geus-Oei LF, van der Drift MA, van der Heijden EH, Oyen WJ, Visser EP, Span PN, Bussink J (2014) Glucose metabolism in NSCLC is histology-specific and diverges the prognostic potential of 18FDG-PET for adenocarcinoma and squamous cell carcinoma. J Thorac Oncol 9(10):1485–1493. https://doi.org/10.1097/JTO.0000000000000286

Scott WJ, Howington J, Feigenberg S, Movsas B, Pisters K (2007) Treatment of non-small cell lung cancer stage I and stage II: ACCP evidence-based clinical practice guidelines. Chest 132(3): 234–242. https://doi.org/10.1378/chest.07-1378

Sebastian NT, Xu-Welliver M, Williams TM (2018) Stereotactic body radiation therapy (SBRT) for early stage non-small cell lung cancer (NSCLC): contemporary insights and advances. J Thorac Dis 10(21):2451–2464. https://doi.org/10.21037/jtd.2018.04.52

Sequist LV, Heist RS, Shaw AT, Fidias P, Rosovsky R, Temel JS, Lennes IT, Digumarthy S, Waltman BA, Bast E, Tammireddy S, Morrissey L, Muzikansky A, Goldberg SB, Gainor J, Channick CL, Wain JC, Gaissert H, Donahue DM, Muniappan A, Dias-Santagata D (2011) Implementing multiplexed genotyping of non-small-cell lung cancers into routine clinical practice. Ann Oncol 22(12):2616–2624. https://doi.org/10.1093/annonc/mdr489

Shaw RJ, Bardeesy N, Manning BD, Lopez L, Kosmatka M, DePinho RA, Cantley LC (2004) The LKB1 tumor suppressor negatively regulates mTOR signaling. Cancer Cell 6(1):91–99. https://doi.org/10.1016/j.ccr.2004.06.007

Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, Fong KM, Lee H, Toyooka S, Shimizu N, Fujisawa T, Feng Z, Roth JA, Herz J, Minna JD, Gazdar AF (2005a) Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 97(5):339–346. https://doi.org/10.1093/jnci/dji055

Shigematsu H, Takahashi T, Nomura M, Majmudar K, Suzuki M, Lee H, Wistuba II, Fong KM, Toyooka S, Shimizu N, Fujisawa T, Minna JD, Gazdar AF (2005b) Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. Cancer Res 65(5):1642–1646. https://doi.org/10.1158/0008-5472.CAN-04-4235

Siegel RL, Miller KD, Jemal A (2015) Cancer statistics, 2015. CA Cancer J Clin 65(1):5–29. https://doi.org/10.3322/caac.21254

Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 448(7153):561–566. https://doi.org/10.1038/nature05945

Solomon B, Varella-Garcia M, Camidge DR (2009) ALK gene rearrangements: a new therapeutic target in a molecularly defined subset of non-small cell lung cancer. J Thorac Oncol 4(12): 1450–1454. https://doi.org/10.1097/JTO.0b013e3181c4dedb

Sordella R, Bell DW, Haber DA, Settleman J (2004) Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. Science 305(5687):1163–1167. https://doi.org/10.1126/science.1101637

Stephens P, Hunter C, Bignell G, Edkins S, Davies H, Teague J, Stevens C, O'Meara S, Smith R, Parker A, Barthorpe A, Blow M, Brackenbury L, Butler A, Clarke O, Cole J, Dicks E, Dike A, Drozd A, Edwards K, Stratton MR (2004) Lung cancer: intragenic ERBB2 kinase mutations in tumours. Nature 431(7008):525–526. https://doi.org/10.1038/431525b

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. https://doi.org/10.3322/caac.21660

Tafani M, Sansone L, Limana F, Arcangeli T, De Santis E, Polese M, Fini M, Russo MA (2016) The interplay of reactive oxygen species, hypoxia, inflammation, and sirtuins in cancer initiation and progression. Oxidative Med Cell Longev 2016:3907147. https://doi.org/10.1155/2016/3907147

Takagi Y, Osada H, Kuroishi T, Mitsudomi T, Kondo M, Niimi T, Saji S, Gazdar AF, Takahashi T, Minna JD, Takahashi T (1998) p53 mutations in non-small-cell lung cancers occurring in individuals without a past history of active smoking. Br J Cancer 77(10):1568–1572. https://doi.org/10.1038/bjc.1998.258

Takeuchi K, Choi YL, Togashi Y, Soda M, Hatano S, Inamura K, Takada S, Ueno T, Yamashita Y, Satoh Y, Okumura S, Nakagawa K, Ishikawa Y, Mano H (2009) KIF5B-ALK, a novel fusion oncokinase identified by an immunohistochemistry-based diagnostic system for ALK-positive lung cancer. Clin Cancer Res 15(9):3143–3149. https://doi.org/10.1158/1078-0432.CCR-08-3248

Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, Asaka R, Hamanaka W, Ninomiya H, Uehara H, Lim Choi Y, Satoh Y, Okumura S, Nakagawa K, Mano H, Ishikawa Y (2012) RET, ROS1 and ALK fusions in lung cancer. Nat Med 18(3):378–381. https://doi.org/10.1038/nm.2658

Tam IY, Chung LP, Suen WS, Wang E, Wong MC, Ho KK, Lam WK, Chiu SW, Girard L, Minna JD, Gazdar AF, Wong MP (2006) Distinct epidermal growth factor receptor and KRAS mutation patterns in non-small cell lung cancer patients with different tobacco exposure and clinicopathologic features. Clin Cancer Res 12(5):1647–1653. https://doi.org/10.1158/1078-0432.CCR-05-1981

Tammemagi MC, McLaughlin JR, Bull SB (1999) Meta-analyses of p53 tumor suppressor gene alterations and clinicopathological features in resected lung cancers. Cancer Epidemiol Biomark Prev 8(7):625–634

Timmerman R, Paulus R, Galvin J, Michalski J, Straube W, Bradley J, Fakiris A, Bezjak A, Videtic G, Johnstone D, Fowler J, Gore E, Choy H (2010) Stereotactic body radiation therapy for inoperable early stage lung cancer. JAMA 303(11):1070–1076. https://doi.org/10.1001/jama.2010.261

Tomizawa K, Suda K, Onozato R, Kosaka T, Endoh H, Sekido Y, Shigematsu H, Kuwano H, Yatabe Y, Mitsudomi T (2011) Prognostic and predictive implications of HER2/ERBB2/neu gene mutations in lung cancers. Lung Cancer 74(1):139–144. https://doi.org/10.1016/j.lungcan.2011.01.014

Tran TN, Selinger CI, Kohonen-Corish M, McCaughan BC, Kennedy CW, O'Toole SA, Cooper WA (2013) Fibroblast growth factor receptor 1 (FGFR1) copy number is an independent prognostic factor in non-small cell lung cancer. Lung Cancer 81(3):462–467. https://doi.org/10.1016/j.lungcan.2013.05.015

Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin J, Beasley MB, Chirieac LR, Dacic S, Duhig E, Flieder DB, Geisinger K, Hirsch FR, Ishikawa Y, Kerr KM, Noguchi M, Pelosi G, Powell CA, Tsao MS, Wistuba I (2015) The 2015 World Health Organization Classification of Lung Tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol 10(9):1243–1260. https://doi.org/10.1097/JTO.0000000000000630

Trevor KT, Hersh EM, Brailey J, Balloul JM, Acres B (2001) Transduction of human dendritic cells with a recombinant modified vaccinia Ankara virus encoding MUC1 and IL-2. Cancer Immunol Immunother 50(8):397–407. https://doi.org/10.1007/s002620100214

Tzahar E, Waterman H, Chen X, Levkowitz G, Karunagaran D, Lavi S, Ratzkin BJ, Yarden Y (1996) A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor. Mol Cell Biol 16(10):5276–5287. https://doi.org/10.1128/MCB.16.10.5276

Valles I, Pajares MJ, Segura V, Guruceaga E, Gomez-Roman J, Blanco D, Tamura A, Montuenga LM, Pio R (2012) Identification of novel deregulated RNA metabolism-related genes in

non-small cell lung cancer. PLoS One 7(8):e42086. https://doi.org/10.1371/journal.pone.0042086

Vanhove K, Graulus GJ, Mesotten L, Thomeer M, Derveaux E, Noben JP, Guedens W, Adriaensens P (2019) The metabolic landscape of lung cancer: new insights in a disturbed glucose metabolism. Front Oncol 9:1215. https://doi.org/10.3389/fonc.2019.01215

Vansteenkiste JF, Cho BC, Vanakesa T, De Pas T, Zielinski M, Kim MS, Jassem J, Yoshimura M, Dahabreh J, Nakayama H, Havel L, Kondo H, Mitsudomi T, Zarogoulidis K, Gladkov OA, Udud K, Tada H, Hoffman H, Bugge A, Taylor P, Altorki N (2016) Efficacy of the MAGE-A3 cancer immunotherapeutic as adjuvant therapy in patients with resected MAGE-A3-positive non-small-cell lung cancer (MAGRIT): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 17(6):822–835. https://doi.org/10.1016/S1470-2045(16)00099-1

Vivanco I, Sawyers CL (2002) The phosphatidylinositol 3-kinase AKT pathway in human cancer. Cancer 2(7):489–501. https://doi.org/10.1038/nrc839

Vlad AM, Kettel JC, Alajez NM, Carlos CA, Finn OJ (2004) MUC1 immunobiology: from discovery to clinical applications. Adv Immunol 82:249–293. https://doi.org/10.1016/S0065-2776(04)82006-6

Wei Z, Liu X, Cheng C, Yu W, Yi P (2021) Metabolism of amino acids in cancer. Front Cell Dev Biol 8:603837. https://doi.org/10.3389/fcell.2020.603837

Wells SA Jr, Santoro M (2009) Targeting the RET pathway in thyroid cancer. Clin Cancer Res 15(23):7119–7123. https://doi.org/10.1158/1078-0432.CCR-08-2742

Wistuba II, Berry J, Behrens C, Maitra A, Shivapurkar N, Milchgrub S, Mackay B, Minna JD, Gazdar AF (2000) Molecular changes in the bronchial epithelium of patients with small cell lung cancer. Clin Cancer Res 6(7):2604–2610

Woo EY, Yeh H, Chu CS, Schlienger K, Carroll RG, Riley JL, Kaiser LR, June CH (2002) Cutting edge: regulatory T cells from lung cancer patients directly inhibit autologous T cell proliferation. J Immunol 168(9):4272–4276. https://doi.org/10.4049/jimmunol.168.9.4272

Wu JY, Wu SG, Yang CH, Gow CH, Chang YL, Yu CJ, Shih JY, Yang PC (2008) Lung cancer with epidermal growth factor receptor exon 20 mutations is associated with poor gefitinib treatment response. Clin Cancer Res 14(15):4877–4882. https://doi.org/10.1158/1078-0432.CCR-07-5123

Yarden Y, Sliwkowski MX (2001) Untangling the ErbB signalling network. Mol Cell Biol 2(2):127–137. https://doi.org/10.1038/35052073

Yip PY, Yu B, Cooper WA, Selinger CI, Ng CC, Kennedy CW, Kohonen-Corish MR, McCaughan BC, Trent RJ, Boyer MJ, Kench JG, Horvath LG, O'Toole SA (2013) Patterns of DNA mutations and ALK rearrangement in resected node negative lung adenocarcinoma. J Thorac Oncol 8(4):408–414. https://doi.org/10.1097/JTO.0b013e318283558e

Yu TW, Chueh HY, Tsai CC, Lin CT, Qiu JT (2016) Novel GM-CSF-based vaccines: one small step in GM-CSF gene optimization, one giant leap for human vaccines. Hum Vaccin Immunother 12(12):3020–3028. https://doi.org/10.1080/21645515.2016.1221551

Zhang CY, Qi Y, Li XN, Yang Y, Liu DL, Zhao J, Zhu DY, Wu K, Zhou XD, Zhao S (2015) The role of CCL20/CCR6 axis in recruiting Treg cells to tumor sites of NSCLC patients. Biomed Pharmacother 69:242–248. https://doi.org/10.1016/j.biopha.2014.12.008