7. Summary and future works

Cancer is one of the most deadly disease of present time, which can be sense by the fact that its mortality rate is almost equal to the incidence rate (American Cancer Society, 2013b; WHO, 2014). Numbers of chemotherapies are available for the treatment of cancer, but none of them are up to the mark to cure cancer. Present cancer therapies face many challenges that affect its efficacy, out of which drug resistance is one of the major hurdles. In the initial phase cancer therapy works and control the tumour growth but at the later stages when cancer relapse (Meacham and Morrison, 2013), it fails because of the drug resistance. As discussed in chapter 1, various factors contribute to drug resistance, such as increased efflux, reduced uptake, mutation, increased metabolism and alteration in the apoptosis pathway.

Apoptosis is a mechanism of programmed cell death, which has two branches intrinsic and extrinsic. This pathway has critical roles in carcinogenesis from cancer progression to development of drug resistance (Wong, 2011). Apoptosis pathway consists of various proteins like, anti-apoptotic e.g. Bcl-2, Bcl-xL, MCL-1 etc. and pro-apoptotic e.g. Bid, Bad, Bax, Bim etc. Fine balance between these crucial proteins determines the apoptosis (Kodama et al., 2013). Proteins belong to the Inhibitor of Apoptosis (IAP) family are also important in determining the direction of apoptosis (Salvesen and Duckett, 2002). For example, X-linked inhibitor of apoptosis protein (XIAP) is highly expressed in most of the cancer, which inhibit apoptosis in cancer cells (Mizutani et al., 2007). Due to this fact, many apoptosis proteins e.g. XIAP, c-IAP1, Bcl-2 etc. have been explored as drug target for cancer therapy.

In the post-genomic era, high throughput studies have been generated a significant amount of cancer genomics data. Databases like CCLE and COSMIC provide large amount of genomic data (gene expression, copy number variation and mutation) of thousands of cancer cell lines and millions of tumour samples. Genomic of Drug Sensitivity in Cancer (GDSC) encompasses the pharmacological profiling of hundreds of anticancer drugs. These two kinds of data can be explored to draw substantial conclusion about cancer.

In our study, we have developed a database known as "CancerDR". This is a pharmacogenomics database comprises of information about 116 cancer drug targets and 148 anticancer drugs. This data is compiled from various resources like, CCLE, COSMIC, GDSC and various other databases at one platform. This database contains various associations between drug target mutations and subsequent drug resistance as we have shown in case of EGFR. CancerDR can help in the elucidation of such relationships between target mutations and associated drug resistance. This database contains data from cell lines; in future we will integrate the data from tumour samples to make this database more robust.

In one of our studies, we have developed QSAR models using chemical properties of anticancer drugs. These models were developed for the designing of promiscuous inhibitors against 16 pancreatic cancer cell lines. We used latest machine learning techniques to develop these models and we achieved decent performance for all the 16 models. We further showed the usefulness of these models for screening and validation of FDA approved drugs. Our models successfully recapitulate the drug-to-oncogene relation in pancreatic cancer cell lines. Further, we have integrated these models into a web-based platform called "DiPCell", which help users to design promiscuous inhibitors. In future, we will try to develop such models for the cell lines belong to the other tissues e.g. lung, breast, kidney etc.

As we have discussed in chapter 1, apoptosis plays a critical role in carcinogenesis and understanding of apoptosis proteins in the context of cancer can help in the development of anticancer therapies. Therefore, we have developed a database called "ApoCanD", which is dedicated to apoptosis proteins. In this database, we have collected and compiled the information about apoptosis proteins in cancer, such as their gene expression, copy number variation and mutation in cancer cell lines and tumours. We have also compiled the information from 1000 Genome project to develop a comparative benchmark of apoptosis proteins in normal and cancer tissues. This database is freely accessible at the URL: http://crdd.osdd.net/raghava/apocand.

To further explore apoptosis proteins as therapeutic targets, we have developed QSAR models against five apoptosis proteins namely, XIAP, c-IAP1, Bcl-2, Bcl-xL and MDM2. These five proteins play a critical role in the apoptosis and they are also considered as

potential drug targets for the development of anticancer therapy. We took bioassay for these proteins from PubChem and ChEMBL and features like chemical descriptors, fingerprints, principal components and docking energy were used to develop QSAR models for each protein. We adopted five machine-learning techniques with 10-fold cross-validations and achieved decent performance for all the models. We have also checked the robustness of these models by evaluating their performance on independent datasets. Further, we have used these models to screen libraries like FDA approved drugs and ZINC database and predict their activities. We have also developed a web server called "XIAP*in*" in which we have integrated models developed for XIAP protein. In future, we will integrate the models of other proteins also in similar web servers for their better use by the scientific community.

In the past, single gene biomarkers have been used for the development of personalized medicine (Glas et al., 2005; Jiang and Wang, 2010). To keep this tenet in mind, in our study, we tried to designate the apoptosis genes as single gene biomarkers for the determinants of drug response. We prioritize the apoptosis genes according to their correlation with drug response (IC_{50}) of 24 anticancer drugs used in CCLE. We have successfully elucidated the biomarkers for each drug as drug response determinant. In another study, we have shortlisted the drugs, which target apoptosis pathway and correlate their response of cancer cell lines with genomic features like gene expression, copy number variation and mutation. Pharmacological profiles of these drugs were obtained from GDSC and genomic features were obtained from CCLE. We used four machine-learning techniques to correlate and to develop predictive models for these drugs. We have successfully correlated their drug response with these genomic features and achieved maximum correlation (R) of 0.89 in case of BMS536924. For other drugs also, we achieved decent correlation using SVM algorithm. Hybrid features performed better as compared to the gene expression, copy number variation and mutation features alone. In future, we will integrate these models into a web-based platform for their maximal use by the research community.