Identification of therapeutically potential targets and their ligands for the treatment of OSCC
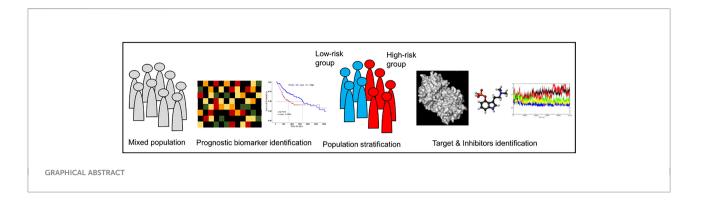
Pratima Kumari[1,2], Sugandh Kumar[1], Madhusmita Sethy[1], Shyamlal Bhue[1,2], Bineet Kumar Mohanta[1,2] and Anshuman Dixit[1]*

[1]Computational Biology and Bioinformatics Laboratory, Institute of Life Sciences, Bhubaneswar, India, [2]Regional Centre for Biotechnology (RCB), Faridabad, India

Recent advancements in cancer biology have revealed molecular changes associated with carcinogenesis and chemotherapeutic exposure. The available information is being gainfully utilized to develop therapies targeting specific molecules involved in cancer cell growth, survival, and chemoresistance. Targeted therapies have dramatically increased overall survival (OS) in many cancers. Therefore, developing such targeted therapies against oral squamous cell carcinoma (OSCC) is anticipated to have significant clinical implications. In the current work, we have identified drug-specific sensitivity-related prognostic biomarkers (*BOP1*, *CCNA2*, *CKS2*, *PLAU*, and *SERPINE1*) using gene expression, Cox proportional hazards regression, and machine learning in OSCC. Dysregulation of these markers is significantly associated with OS in many cancers. Their elevated expression is related to cellular proliferation and aggressive malignancy in various cancers. Mechanistically, inhibition of these biomarkers should significantly reduce cellular proliferation and metastasis in OSCC and should result in better OS. It is pertinent to note that no effective small-molecule candidate has been identified against these biomarkers to date. Therefore, a comprehensive *in silico* drug design strategy assimilating homology modeling, extensive molecular dynamics (MD) simulation, and ensemble molecular docking has been applied to identify potential compounds against identified targets, and potential molecules have been identified. We hope that this study will help in deciphering potential genes having roles in chemoresistance and a significant impact on OS. It will also result in the identification of new targeted therapeutics against OSCC.

KEYWORDS

chemotherapy, chemoresistance, prognosis, drug discovery, OSCC

GRAPHICAL ABSTRACT

Introduction

Oral squamous cell carcinoma (OSCC) constitutes a major subset of head and neck squamous cell carcinoma (HNSCC) and accounted for an estimated 0.37 million cases in the year 2020 (1). The high morbidity and mortality of OSCC pose a great challenge to its management (1, 2). The overall 5-year survival rate in OSCC is comparatively lower (~50%) than that in many other cancers (3, 4). For example, in India, 0.13 million new cases were detected, whereas 75,000 patients died in 2020 (1). Current treatment modalities include surgery, radiotherapy, chemotherapy, or their combinations. These are successful in patients with primary tumors, whereas patients with high-risk features (invasion/perineural invasion, metastasis, T3/T4 stage, or involvement of two or more lymph nodes) show less improvement. There is no clinical evidence to support the likely outcome in the case of high-risk oral cancer (5, 6).

Patient response to chemotherapy has been linked to tumor lineage and genetics. Changes in cellular gene expression in

response to small-molecule treatments can provide insights into cellular processes governing the clinical outcome. Identifying the responsive features and mechanism of action is of immense value in cancer therapy and can be critical for the development of novel medicines. Furthermore, the heterogeneous response of patients to cancer therapies and the frequent development of drug resistance highlight the importance of a therapeutic response (7, 8). However, an accurate prediction of a therapeutic response and the identification of new anticancer drugs have remained a challenging task.

In recent years, the increasing understanding of genomics and the advent of next-generation sequencing (NGS) with advancements in bioinformatics approaches have made it possible to identify potential molecular targets for the betterment of chemotherapy. Numerous cancer studies with the help of NGS were able to identify novel and rare somatic mutations in more efficient and accurate ways. In a variety of cancers such as bladder cancer, renal cell carcinoma, small-cell lung cancer, prostate cancer, acute myelogenous leukemia, and chronic lymphocytic leukemia, researchers were able to accurately identify genetic alterations. On the other hand, together with NGS, the bioinformatics approach was successful in exploiting the heterogeneous nature of cancer to develop cancer diagnostic, prognostic, and predictive markers (9). Gene-based approaches are being used for the development of new therapeutic agents. The targeted approaches have been proven highly useful in the development of therapies for many cancers (10).

Doxorubicin is known to inhibit the topoisomerase-II (TOP2) activity in eliciting its antineoplastic effect. It is one of the most effective anticancer drugs widely used in the treatment of several cancers including breast cancer, lung cancer, gastric cancer, ovarian cancer, thyroid cancer, non-Hodgkin's and Hodgkin's lymphoma, multiple myeloma, sarcoma, and pediatric cancer (11). High toxicity and early resistant phenotype have limited the use of doxorubicin (12). Thus, it is imperative to study the molecular changes associated with doxorubicin resistance. Therefore, in the current work, we aimed to explore gene expression changes in the response to

---

Abbreviations: ACC, adrenocortical carcinoma; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma; COAD, colon adenocarcinoma; DLBC, diffuse large B-cell lymphoma; ESCA, esophageal carcinoma; GBM, glioblastoma multiforme; HNSCC, head and neck squamous cell carcinoma; KICH, kidney chromophobe; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; AML, acute myeloid leukemia; LGG, brain lower-grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; READ, rectum adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumor; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma; UCS, uterine carcinosarcoma; MESO, mesothelioma; UVM, uveal melanoma; GEM, gemcitabine; GC, gastric cancer; OC, ovarian cancer, TMZ, temozolomide; NSCLC, non-small cell lung carcinoma.

doxorubicin treatment in OSCC using various datasets and the ways to reduce the chances of emergence of such resistance by molecularly targeted therapies. We have also identified potential ligands that may increase the effect of doxorubicin and/or can delay the progression of drug-related resistant phenotypes or sensitize oral cancer cells to chemotherapy.

Materials and methods

Data collection and differential gene expression analysis

In the present study, data were collected from two sources: The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) and Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). Since we wanted to analyze OSCC data, we further filtered out the OSCC samples according to the International Classification of Diseases (ICD) code. The ICD classifies diseases based on the site of disease occurrence (Supplementary Table S1). Finally, 319 OSCC and 44 normal adjacent tissue (NAT) samples were obtained. Additionally, two more datasets were obtained from GEO (13, 14): 1) mRNA expression profiling data of 27 OSCC patients (GEO ID-GSE23558) and 2) mRNA expression profiling data of the doxorubicin-treated SCC25 cell line (GSE58074). Two replicates of each (mock and treatment) were taken for expression analysis. The clinical details of TCGA patients are given in Supplementary Table S2.

The gene expression analysis was performed using R studio (http://www.rstudio.com/) version 3.4.4 using the limma-voom library for TCGA samples, whereas GEO samples were analyzed using GEO2R. Genes with |log2FC ≥1| and p-value<0.05 were considered significantly differentially expressed genes (DEGs). The ggplot2, complex heatmap, and circular library in R were used for volcano plot and heatmap. Common DEGs in all of the three datasets were considered genes of interest for further analysis.

Functional enrichment analysis

Gene Ontology (GO) and pathway enrichment of common genes present in all of the datasets were used to identify enriched biological events as a result of doxorubicin perturbation. We have used Reactome (https://reactome.org) (15, 16) and GO (http://geneontology.org/) (17, 18) online databases for enrichment analysis of the gene sets. Enriched biological pathways and GO terms were considered significant if p-value ≤0.05.

Cox proportional hazards regression

Cox proportional hazards regression (Coxph) is a semiparametric model used to predict the outcome of disease based on one or more predictors in survival time (time-to-event) through the hazard ratio (HR) function. It assumes that the effects of predictor variables have an additive effect on the hazard, i.e.,

$$h(t) \quad = \quad h_0(t) \quad \exp \quad (\beta_1 x_1 \quad + \cdots\cdots \quad + \quad \beta_n x_n) \quad (1)$$

where $h(t)$ is the hazard at time t for a subject with a set of predictors $x_1 \ldots x_n$, $h_0(t)$ is the baseline hazard function, and $\beta_1, \ldots, \beta_n$ are the model coefficients describing the effect of the predictors on the overall hazard. HR is used for the interpretation of the Cox model. The HR examination shows selective factors that influence the rate of an event happening (e.g., death) at a particular point in time. An HR above 1 indicates a predictor that is positively associated with the event probability (death) and thus negatively associated with the length of survival, indicating worse prognosis; a negative HR indicates a protective effect of the predictors with which it is associated; while an HR equal to 1 means no effect. The DEGs were subjected to a Cox regression analysis using the "survival" package in R (https://github.com/therneau/survival).

Identification of biomarker signature and validation of prognosis-related genes

Machine learning (ML) is becoming popular in cancer biology in identifying prognostic, diagnostic, and therapeutic biomarkers. Therefore, we implemented two commonly used machine learning algorithms, viz., random forest (RF) and partial least square (PLS) regression method, to identify prognostic targets. RF is a frequently used algorithm for the identification of prognostic biomarkers in many diseases including cancer (19, 20).

Survival analysis

To evaluate the reliability of the predicted prognosis signature, Kaplan–Meier (K-M) plots were generated to estimate the survival of the patients based on their median mRNA expression. An mRNA expression above the median value was considered high, whereas an expression below the median was counted as low expression. The R package "survival" was used to plot the patient's overall survival (OS), and significance was calculated based on the log-rank p-value. To further reflect the sensitivity and specificity of signature mRNAs, we employed a time-dependent receiver

operating characteristic (ROC) curve analysis *via* the R package "survivalROC."

Risk assessment and validation of the risk model

The risk score is an additive model of the mRNA expression level multiplied by their Cox regression coefficient. The risk score was calculated as follows:

Risk score

$$= \text{Coefficients of signature Gene A}$$
$$\times \text{ mRNA expression of A } + \dots\dots$$
$$\dots.$$
$$+ \text{ Coefficients of signature Gene N}$$
$$\times \text{ mRNA expression of N} \qquad (2)$$

The median risk score was used to divide the patients into high- and low-risk groups.

Use of the protein–protein interaction network to understand the potential signature biomarker

Next, we constructed the protein–protein interaction (PPI) network to understand the importance of key targets in the human interactome. The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (https://string-db.org/) was used to create a PPI network of identified genes with a cutoff score of >0.4, which was equivalent to medium confidence. Subsequently, a cluster analysis using MCODE (21) was done to decipher the modules in the created network. To understand the functional significance of the modules, pathway analysis was done using Reactome (www.reactome.org).

Identification of candidate small-molecule drugs

Protein structures

The X-ray crystal structures of SERPINE1 [Protein Data Bank (PDB): 4AQH], CCNA2 (PDB: 1H1R), PLAU (PDB: 1OWE), and CKS2 (PDB: 5HQ0) were used in the current studies. The three-dimensional (3D) structure of BOP1 is not available in the PDB.

Homology modeling

The 3D structure of BOP1 was modeled by a homology modeling approach using modeler v9.20. Delta-blast was used to identify suitable templates for modeling against the PDB database.

The template was selected based on the expected value (E-value), bit score, and percentage of query coverage and identity. The carboxy-terminal domain of Erb1 of *Escherichia coli* (PDB-ID: 4U7A chain A, identity 40.1%, coverage 83%) was finally selected as the template to model the structure of BOP1. Twenty models were generated for each protein, and the final model was selected based on the lowest molpdf score.

System setup and molecular dynamics simulation

In the current study, the preliminary topology and coordinates for all proteins were generated in VMD v1.93, whereas the simulations were run in NAMD v2.14. All protein structures (BOP1, PLAU, SERPINE1, CKS2, and CCNA2 after stripping cocrystallized ligands) were prepared and solvated in a rectangular water box (TIP3P water model) with a buffering distance of 10 Å. Ions were added to ensure the electroneutrality of the solvated system. The SETTLE algorithm was used for the water molecules. The associated system topology and coordinates were generated by applying charmm34 force field parameters for molecular dynamics (MD) simulation. Prior to the simulation, the system was properly minimized with a stepwise minimization protocol. Firstly, the water molecules and ions were minimized that was then followed by hydrogen atoms and the side chains of the complex. The side chains were minimized for 40,000 steps, whereas the backbone atoms and the bond lengths of hydrogen atoms were kept fixed. Thereafter, all of the atoms were allowed to relax freely, and the whole system was energy-minimized for 40,000 steps with nominal restraints on Cα atoms (10 kcal/mol) to prevent any abrupt change in the structure. Subsequently, an equilibration protocol was followed where the system was heated gradually from 0 to 310 K in steps of 30 K with a canonical ensemble [constant volume, constant temperature (NVT)]. At each step, a 20-picosecond (ps) simulation was run to allow the system to adjust to the temperature. Once the system attained 310 K, an isobaric and isothermic ensemble [constant pressure, constant temperature (NPT)] was applied for a period of 100 ps with a constant pressure of 1.0 bar using Langevin dynamics. Finally, the applied restraints on Cα atoms were removed, and the system was equilibrated for 1 ns at 310 K using the Langevin piston coupling algorithm. During the whole simulation, the Particle Mesh Ewald (PME) algorithm was used to calculate the long-range electrostatic interactions with fixed periodic boundary conditions. The covalent bond interactions involving hydrogen were constrained using the SHAKE algorithm. Once the system was simulated with a constant 310 K temperature and 1.0 bar pressure, then the production run was done for a time period of 100 ns. The analyses of the MD trajectories were performed to analyze the structure and dynamic behavior of all proteins during MD. The trajectories were analyzed for root mean

square deviation (RMSD) and radius of gyration (Rg); these analyses were performed using VMD and in-house Perl and Tcl scripts. Five additional equidistant frames were generated from each trajectory for ensemble docking studies.

To study the stability of the protein–ligand complex, another simulation was run for 300 ns with selected ligands docked in the protein targets. The ligand parameters were generated using the Antechamber module of AMBER12 molecular simulation package (www.ambermd.org). The complex was solvated in a box of water with 10 Å buffering distance. AMBER parameters and coordinate files were generated using the tleap module of AMBER12. Equilibration and simulation were done using NAMD v2.14 as described above. The trajectories were analyzed for RMSD, hydrogen bonds, and salt bridges to assess the stability and strength of the interactions between the ligand and the protein.

Molecular docking

The frames obtained from the MD simulation for the selected proteins were prepared using the protein preparation wizard module of Schrödinger molecular modeling software v9.3. OPLS-2005 force field was used for energy minimization, whereas Schrödinger's LigPrep module was employed to generate 3D conformers and energy minimization of the potential small-molecule drugs from the US Food and Drug Administration (US FDA) library. A maximum of 32 conformers were generated per ligand. Before docking, the grid box (active site) was defined as residues within 5 Å of
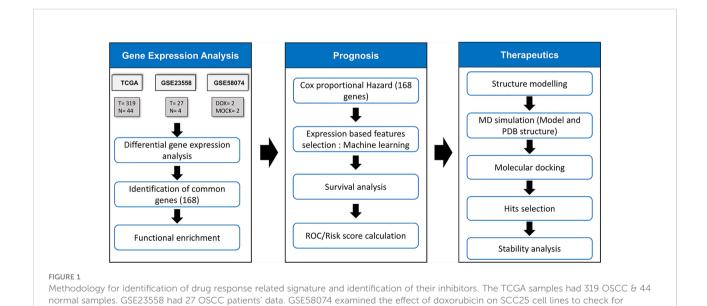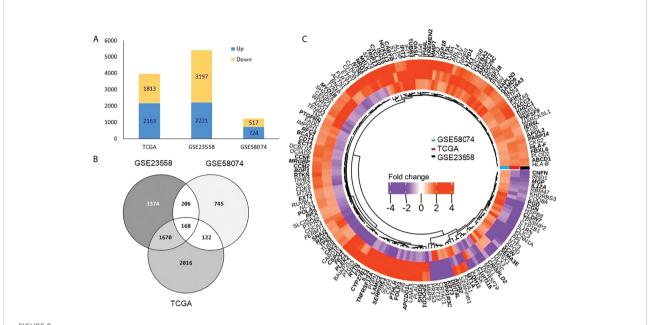
cocrystallized ligands. The active site of the modeled proteins or proteins without cocrystallized ligands was predicted through the sitemap algorithm as implied in Schrödinger. Docking was performed using standard precision (SP) mode with flexible ligand sampling in Schrödinger's Glide module. The average docking score was calculated based on the glide score in the five frames. The flowchart of methodology is given in Figure 1.

Results

Identification of differentially expressed genes

We identified 3,976, 5,418, and 1,241 DEGs in TCGA, GSE23558, and GSE58074 datasets, respectively. In TCGA samples, 2,163 genes were overexpressed and 1,813 genes were underexpressed, whereas 2,221 and 3,197 genes were found overexpressed and underexpressed, respectively, in GSE23558. In GSE58074, which contained doxorubicin-treated cell line data, 724 genes were overexpressed and 517 genes were underexpressed (Figure 2A and Supplementary Table S3). Among the three datasets, 168 common DEGs were identified. These were considered for further analysis (Figure 2B). DEGs are represented in a volcano plot with $\log_2$FC and -$\log_{10}$ p-value (Supplementary Figure S1). The heatmap represents the expression of common DEGs from the three datasets, with 134, 132, and 85 overexpressed genes and 34, 36, and 83 underexpressed genes in TCGA, GSE23558, and GSE58074,

FIGURE 1
Methodology for identification of drug response related signature and identification of their inhibitors. The TCGA samples had 319 OSCC & 44 normal samples. GSE23558 had 27 OSCC patients' data. GSE58074 examined the effect of doxorubicin on SCC25 cell lines to check for molecular markers underlying doxorubicin response..

FIGURE 2

Differentially expressed genes **(A)** Bar graph representing total DEGs both up (blue) and down (yellow) in three datasets. **(B)** Venn diagram showing number of DEGs (common and unique) among TCGA, GSE23558 and GSE58074. A total of 168 common DEGs were obtained. **(C)** Heatmap shows DEGs related to doxorubicin response (GSE58074) and non-treated samples in OSCC from TCGA and GSE23558. Three types of distinct expression pattern are discernible (1) 52 genes overexpressed (orange) in all the three data sets, (2) 76 genes underexpressed in GSE58074 (purple) while overexpressed in TCGA (orange) and GSE23558 (orange), and (3) 26 genes underexpressed in TCGA (purple) and GSE23558 (purple) while overexpressed in GSE58074 (orange).

respectively (Figure 2C and Supplementary Table S3). From the heatmap, it is clearly visible that genes that are overexpressed in cancer prior to any treatment get downregulated after treatment and *vice versa*, whereas some of the genes present do not respond to the treatment, as there are no changes in their expression patterns.

Functional enrichment analysis

The common DEGs (n = 168) were studied for functional enrichment to investigate their involvement in various cellular processes. The top 10 enriched GO functions and pathways are shown in Figure 3. Biological process (BP) terms related to immune response, biological stress, cell proliferation, and survival were found to be significantly enriched. Enriched cellular component (CC) terms include extracellular related components, protein complex, membrane-bound components, and cellular vesicles. Molecular function (MF) enriched terms included various signaling pathways involved in ligand–receptor complex, enzymatic functions, G protein–related function, kinases, and other protein-binding functions. Pathway
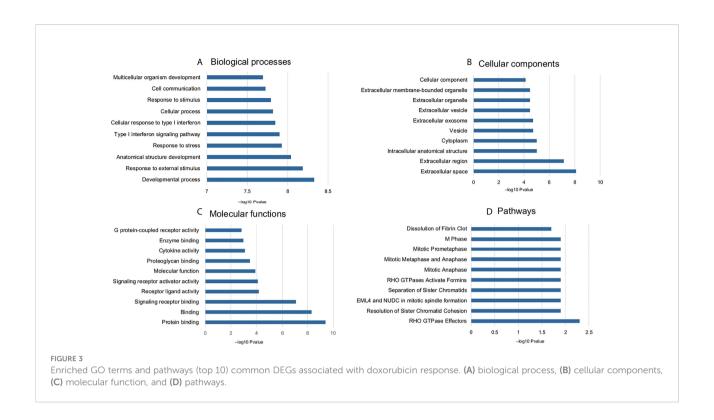
enrichment shows that most of the signaling related to cell cycle and its regulation such as G2/M transition, nucleotide salvage pathways, and DNA damage bypass. The detailed pathway and GO term information is provided in Supplementary Table S4 and Supplementary Figure S2.

Exploration of prognostic biomarkers

The prognostic significance of the selected DEGs (168 genes) was assessed through univariate Coxph regression analysis. A total of 59 genes were found to be significantly (p-value<0.05) associated with OS (Figure 4A and Supplementary Table S3). As indicated earlier, the higher values of HR (HR >1) indicate worse prognosis, whereas negative HR values (<-1) show favorable prognosis. Next, machine learning algorithms, *viz.*, RF and PLS regression, were employed to identify the potential predictive prognostic signature genes. We selected the top predictive genes based on the robustness of the prediction. A cutoff percentile score of ≥80 was used; nine genes were selected from RF, whereas seven genes were predicted by PLS (a total of 16) (Supplementary Table S5). Then, we performed survival

FIGURE 3

Enriched GO terms and pathways (top 10) common DEGs associated with doxorubicin response. **(A)** biological process, **(B)** cellular components, **(C)** molecular function, and **(D)** pathways.

analysis to identify the potential prognostic markers. We found that five genes, *viz.*, *BOP1*, *PLAU*, *SERPINE1*, *CCNA2*, and *CKS2*, were significantly related to the survival of the patients. The risk assessment score was calculated with the help of Cox regression analysis that resulted in the following equation (Eq. 2) and was used to determine the risk score (high and low):

$$
\begin{aligned}
\text{Risk score} \\
= \ & (3.207785 \ \times \ \text{expression}_{BOP1}) \\
& + (1.986074 \ \times \ \text{expression}_{PLAU}) \\
& - (1.962924 \ \times \ \text{expression}_{SERPINE1}) \\
& + (2.913185 \ \times \ \text{expression}_{CCNA2}) \\
& + (4.164371 \ \times \ \text{expression}_{CKS2}) \quad (2)
\end{aligned}
$$

Importantly, low-risk patients have a low expression of these mRNAs as compared with those of high-risk patients (Figure 4B). It can be seen clearly in Figures 4D–H that the low–risk score group had better OS; therefore, the expression of the identified mRNAs has significant impact on the OS of the patients. To further assess the accuracy of the prognostic model, we constructed an ROC curve to assess the impact of the expression of these genes on patients' OS for 1, 3, and 5 years. The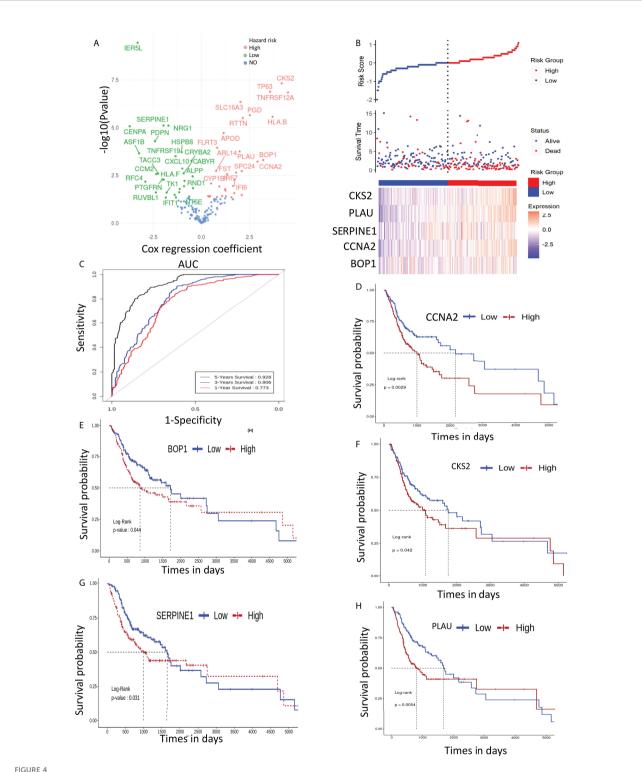 ROC curve is shown in Figure 4C, and the area under the curve (AUC) was 0.773, 0.806, and 0.928 for 1, 3, and 5 years, respectively.
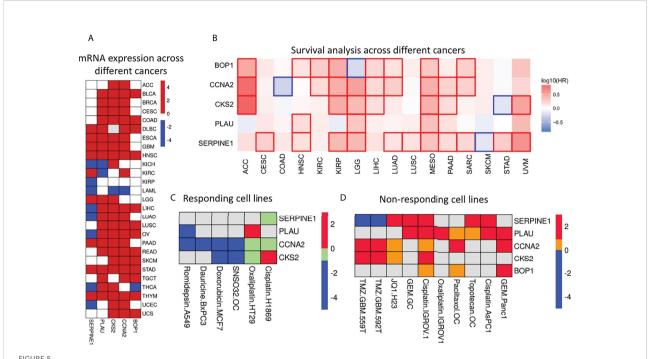
Assessment of the Five Prognostic Signature Genes Across Cancers

The Cox regression and Machine learning (ML) analysis showed that overexpression of the identified genes is related to poor prognosis. Furthermore, we investigated the expression of these genes across different cancers and OS of patients to assess their clinical importance. For this, we have used GEPIA2 database (22). GEPIA2 retrieves data from TCGA and the Genotype-Tissue Expression (GTEx) portal. The expression analysis showed that these five genes were significantly upregulated (|log2FC >1|, p-value<0.05) in most of the tumor tissues as compared with those in normal tissues (Figure 5A and Supplementary Table S6). Survival analysis across various cancers showed that a high expression of these five genes is related to poor survival in most of the cancers (Figure 5B). Moreover, we also checked the expression of these genes in response to drug treatment. For this, we analyzed the expression array data available in the NCBI-GEO dataset (Supplementary Table S7). Two types of analysis were done: 1) expression analysis in drug-sensitive cells and 2) expression analysis in drug-resistant cells (Figures 5C, D). We found that the mRNA

FIGURE 4
(A) Volcano plot of cox regression showing hazard ratio (HR) for 168 genes. **(B)** Patient stratification according to risk score to predict the survival time of patients with high- and low-expression level of prognostic genes. **(C)** ROC depicting the effect of selected genes on overall survival (1, 3, and 5 years). **(D–H)** depict the effect of signature gene expression on overall survival of OSCC patients..
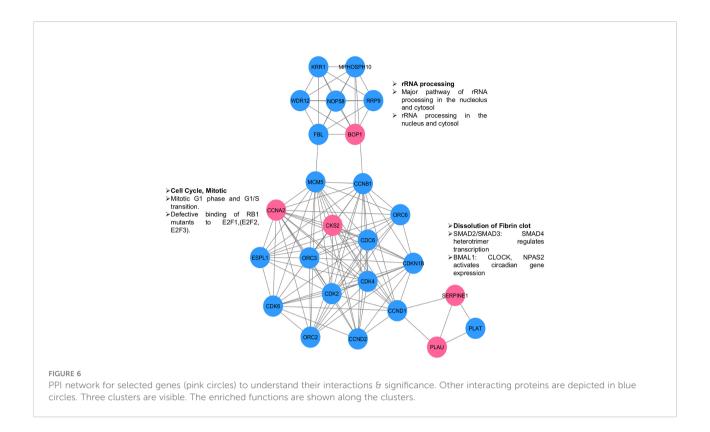
FIGURE 5

Expression & significance of selected prognostic genes. **(A)** Differential gene expression analysis across different cancers. Red, blue and white squares depict over, under and insignificant expression respectively. It is clear that these genes are upregulated in most of the cancers. **(B)** Survival map: Red and blue squares indicate poor survival due to over and under expression respectively. The figure clearly indicates the expression of these genes have significant effect on survival of the patients across cancers. **(C)** Heatmap represents expression level of these genes in drug sensitive cells. These genes were found to be significantly differentially expressed ($|\log_2 Fc| > 1$ and pvalue < 0.05). The down- (blue) and up- (red) regulated genes in response to drug; green square indicates down-regulated genes with significant pvalue but $|\log_2 Fc| < 1$. Grey square indicated no differential expression. **(D)** The expression of selected genes in drug resistant cells. Red and blue squares indicate significant ($|\log_2 Fc| > 1$ and pvalue < 0.05) up- and down-regulated genes; yellow square indicates up-regulated genes with significant pvalue ($|\log_2 Fc| < 1$); grey square indicated no differential expression of the genes.

expression level decreased for most of the genes upon drug treatment. For example, *CCNA2* mRNA levels were found to be decreased significantly with fold change -1.07, -4.33, -1.06, and -4.17 when treated with romidepsin, SNAO32, dauricine, and doxorubicin in A549, primary ovarian cancer, BxPC3, and MCF7 cell lines, respectively. Similarly, *PLAU*, *CKS2*, and *SERPINE1* mRNA levels were also found to be decreased in response to various drugs, whereas in the case of the drug-resistant cell line, an increase in mRNA expression levels of *PLAU*, *CKS2*, *CCNA2*, *BOP1*, and *SERPINE1* was observed. For example, PLAU showed increased mRNA expression with fold changes of 2.14, 1.8, 1.5, 1.4, and 1.2 in Panc1 gemcitabine-, MKN28 gemcitabine-, AsPc1 cisplatin-, IGROV1 cisplatin-, and IGROV1 oxaliplatin-resistant cell lines, respectively. Additionally, we analyzed the presence of genetic alterations (e.g., amplification, mutation, deletion, structural variant) in these five genes using CBioPortal (https://www.cbioportal.org/), which is an online consortium for cancer genomics. We investigated the TCGA Pan-Cancer Atlas dataset for the genomic alteration study (23, 24). It clearly showed that amplification is the most common alteration observed in *BOP1*,

PLAU, *SERPINE1*, and *CKS2* across the cancers, whereas in *CCNA2*, mutation and deep deletion are prevalent (Supplementary Figure S3).

Protein−protein interaction network

The PPIs are at the heart of various molecular mechanisms. Therefore, a PPI network was constructed to understand the interactions of selected proteins with other proteins in the human interactome for better understanding of their regulatory roles (Figure 6). We identified three distinct clusters in the network; the proteins in those clusters are involved in 1) the regulation of rRNA processing; 2) the regulation of the cell cycle; and 3) angiogenesis, growth factors, and transcription factors such as Suppressor of Mothers against Decapentaplegic (SMAD) SMAD2/SMAD3. The overall network analysis indicates that these genes are connected to many important genes, and targeting them will affect cellular processes playing critical roles in the pathogenesis of OSCC.

FIGURE 6

PPI network for selected genes (pink circles) to understand their interactions & significance. Other interacting proteins are depicted in blue circles. Three clusters are visible. The enriched functions are shown along the clusters.

Molecular dynamics simulations

The MD analysis was performed to assess the flexibility of the binding site that is not discernible through PDB structures. The generated structural ensembles were used for the identification of small-molecule binders. MD simulations were done for 100 ns on each protein (total 500 ns), and the stability of the simulation was evaluated using RMSD. The RMSD values reveal the structural changes that occurred during the MD. The RMSD plots for all proteins indicated that each system was stabilized quickly and then remained stable throughout the simulation time, as evidenced by the movement of the RMSD curve within 2 Å. These plots suggested that each system was quite stable for the docking study. Five equidistant frames for each protein at 20-ns distance (from 100-ns simulation), representing the dynamics of the protein structure, were extracted. Docking studies were then performed on these frames.

Screening of potential compounds from the US Food and Drug Administration−approved library

The binding affinity of US FDA-approved drugs with each of the proteins was assessed through molecular docking. The average glide docking score (average of five frames) was used to identify potential binders of the individual proteins. The top

20 compounds for each protein, based on their glide score, are given in Supplementary Table S8. Details for individual proteins are given below.

BOP1 **is a** RNA-binding protein involved in ribosome biogenesis, cell cycle, and cell proliferation (25). The docking analysis indicated saquinavir to be the best binder with an average docking score of -10.9 (Figure 7A). The active site of BOP1 is surrounded by multiple beta sheets forming a barrel-like structure. The molecule saquinavir is ensconced in a pocket lined by Trp182, Pro104, Pro229, Pro63, Thr181, Pro368, Gly184, Leu266, Val268, and Val309. Analysis of docking poses indicates that saquinavir has several hydrogen-bonding interactions with residues, *viz.*, Trp182, Val268, and Val309.

CCNA2 (cyclin A2) binds with both CDK2 and CDK1. It is required for entry into the S and M phases of the cell cycle. The overexpression of CCNA2 leads to cell growth and proliferation. Diacetolol was found to be the best binder, with an average docking score of -7.6. It is a beta-blocker used as an antihypertensive and antiarrhythmic agent. The docking study indicated that the ligand diacetolol binds in a cavity lined by residues His233, Gln337, Ser340, Tyr347, Tyr350, Pro352, Val354, Ile355, Ala356, Cys390, Asp393, and Leu394 of CCNA2. It makes HB interaction with Tyr347 (Figure 7B). It can be an attractive chemotherapeutic option for OSCC.

PLAU encodes a serine protease (uPA) that converts plasminogen to plasmin. uPA is involved in the degradation of the basement membrane and extracellular matrix (26). *PLAU* is
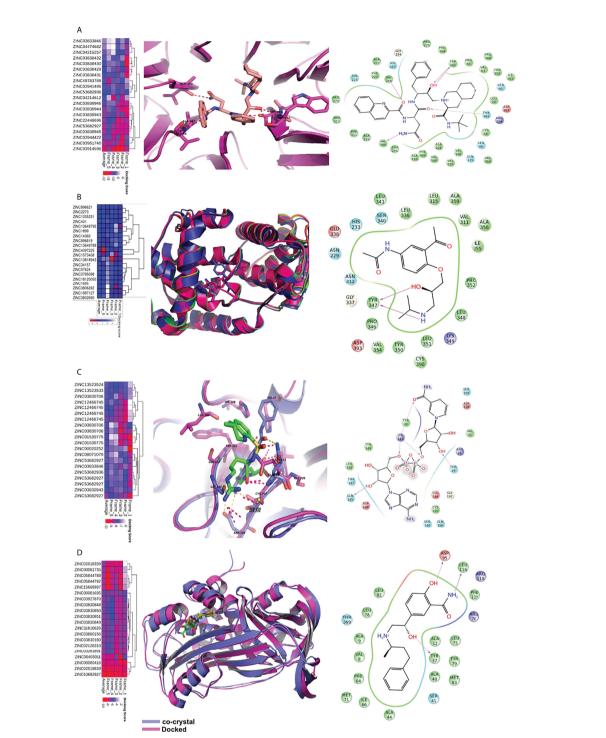
FIGURE 7

Virtual screening of FDA approved drugs against identified proteins. Top scoring 20 molecules for each of the five proteins are shown in heatmap (figures **A–D**). The first column depicts the heatmap of docking scores in five MD frames. The average docking score is also depicted for better understanding. In the second column, proteins are shown in cartoon while the ligands are shown in sticks. The dotted lines depict hydrogen bonds. The third column depicts the interaction map of the ligand receptor interactions. The pink arrows show hydrogen bonds. The arrowhead depicts the HB-acceptor molecule. Pi-pi stacking interactions are depicted by green line. **(A)** The docking of saquinavir with BOP1. It makes hydrogen bonding interactions with backbone of TRP-182, VAL-268, VAL-309, and sidechain of THR-181. **(B)** The binding of molecules in CCNA2. The diacetolol is shown in blue sticks. The ligand makes hydrogen bonds with Tyr347 **(C)** The binding of ligands in PLAU. NADH is shown in binding site of PLAU. It makes hydrogen bonding interactions with THR30, TYR31, LYS154, and SER157. **(D)** The binding of ligands in SERPINE1, and docked molecule labetalol. The drug forms hydrogen bonds with sidechain of ASP-95, TYR37.

one of the potential biomarkers for HNSCC and several other cancers. Nicotinamide adenine dinucleotide hydrogen (NADH), argatroban, diminazene (DIZE), and pentamidine are among the molecules that showed good binding affinity with PLAU. NADH binding to PLAU is shown in Figure 7C. It makes hydrogen-bonding interactions with Thr30, Tyr31, Lys154, and Ser157. NADH, due to its role in energy production, may help against wastage and weakness of cancer patients. It is also used to improve mental alertness. It can be given orally and has cleared clinical trials as a chemotherapeutic agent for other illnesses (27).

SERPINE1 (PAI-I) inhibits the plasminogen activator uPA/uPAR complex that promotes cell matrix degradation and cell migration. Overexpression of *SERPINE1* is highly associated with poor survival in primary tumor, lymph node, and head and neck cancer metastasis. The docking study showed that NADH, reproterol, and labetalol bind to *SERPINE1* with high affinity. The molecule labetalol makes hydrogen-bonding interactions with Tyr37, Ser41, Asp95, Phe117, and Arg118. The phenyl ring sits in the vicinity of hydrophobic residues such as Phe64, Ile66, and Phe117 (Figure 7D). We could not find a good binder for CKS2. The overall analysis indicates that the identified small molecules hold potential as possible therapeutics for OSCC and other cancers.

Stability analysis of the drug–ligand complex

Post docking MD simulation was run to check the stability of the ligands inside the binding cavity of the proteins. The stability of the simulation was analyzed by RMSD. The hydrogen bond interactions between ligand and protein were also calculated to check their strength.

The RMSD analysis of the MD simulation of the BOP1–saquinavir complex indicated that the simulation stabilized at about 10 ns. At about 40 ns, there is a slight change in ligand RMSD (<1 Å). Thereafter, the trajectory is very smooth, and both the protein and ligand RMSDs show a very stable trajectory (Figure 8A). The hydrogen bond analysis showed that saquinavir has two hydrogen-bonding interactions (with Thr149 and Trp150) with >50% occupancy (Table 1), which indicates that the HB interactions are strong and the BOP1–ligand complex is highly stable.

The RMSD during the MD simulation of the PLAU–NADH complex initially increases until about 10 ns; afterward, its movement is confined in a small window (<2 Å) (Figure 8B). This clearly indicated that the simulation is stable. The ligand RMSD was also calculated. It shows a close trend as that of the protein, again indicating good stability of the ligand inside the
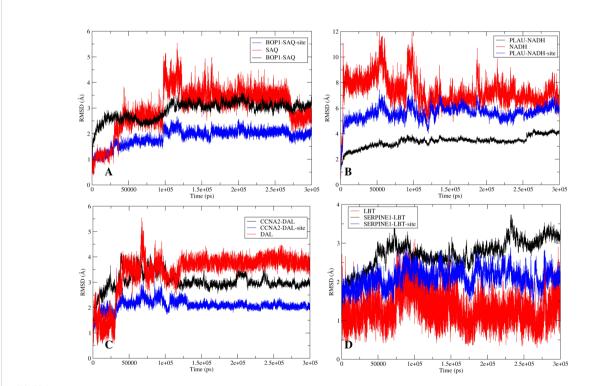

FIGURE 8
The stability analysis: The root mean square deviation (RMSD) analysis for the protein-ligand complexes. The RMSD of protein is shown in red while the RMSD of ligand is shown in blue. (A) RMSD for BOP1-Saquinavir complex. (B) RMSD for PLAU-NADH complex. (C) RMSD for CCNA2-Diacetolol complex. (D) RMSD for SERPINE1-Labetalol complex.

TABLE 1 Hydrogen bond analysis between the ligand and the protein.

Protein BOP	S. No.	Residue	Ligand	Occupancy*
	1	THR149	Saquinavir	63.43%
	2	TRP150	Saquinavir	63.51%
	3	PHE195	Saquinavir	31.31%
PLAU				
	1	LYS154	NADH	71.71%
	2	TYR31	NADH	22.49%
	3	SER157	NADH	16.89%
CCNA2				
	1	TYR176	Diacetolol	19.89%
	2	LYS175	Diacetolol	14.47%
SERPINE1				
	1	ASP95	Labetalol	60.24%
	2	TYR37	Labetalol	16.19%

*Defined as the HB interaction present in x% of frames out of a total of 100,000 frames.

binding cavity. Hydrogen bonds of the ligand with Lys154 and Tyr31 showed occupancy of about 71% and 23%, respectively (Table 1). All of these observations point to the high stability of the NADH–PLAU complex. The RMSD analysis of the MD simulation of the CCNA2–diacetolol complex indicated that the simulation stabilized quickly, as indicated by the movement of the RMSD in a narrow window. Both the protein and ligand RMSDs show a stable trajectory; however, we could not find a strong hydrogen bonding interaction between diacetolol and CCNA2 (Figure 8C and Table 1).

The MD simulation of Serpine1 indicated that after an initial increase, the RMSD gets stabilized quickly, as indicated by a rangebound (<2 Å) movement of the RMSD curve (Figure 8D). The ligand RMSD was also calculated, and it showed movement in a very narrow window, indicating the stability of the simulation and that of the complex. The HB analysis indicated that the hydrogen bond between the ligand and Asp95 is highly stable, as measured by an occupancy of >60% (Table 1). Overall analysis indicates that the Serpine1–labetalol complex is highly stable.

Discussion

In recent years, prognosis-based gene signature identification has been of immense interest for the prediction of outcome or for evaluation of the course of a disease (28, 29). Therapeutic biomarker prediction models are currently in focus to identify predictive factors of the response to chemotherapy (30, 31). Non-responding cancer cells are either refractory to chemotherapy or have acquired resistance during the course of the treatment. Both are strongly related to molecular alteration of the targets. The major cha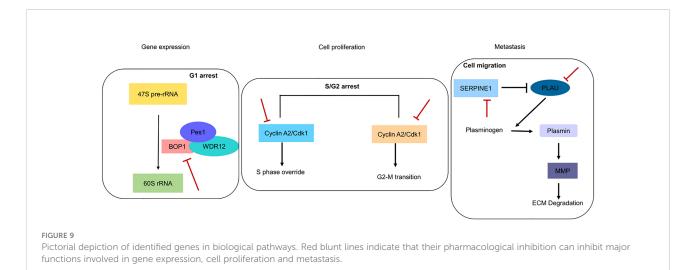llenge is acquired chemoresistance during treatment, which eventually leads to cancer regrowth, even if the tumor initially responds to the chemotherapeutic agent. Moreover, reports suggest changes in gene and protein expression levels in cancer tissues and cell lines after chemotherapy. For instance, a xenograft study has shown that drug treatment (5-fluorouracil and cisplatin) first reduces the expression level of drug-specific sensitivity-related genes followed by their upregulation in the regrowth phase and emergence of chemoresistance in esophageal cancer (32). Chemoresistance, either intrinsic or acquired, contributes to the low survival of the patients, necessitating the need for identification of drug-specific sensitivity-related markers and their modulators to enhance the sensitivity of cancer cells to subdue/delay chemoresistance. One such example is inhibition of the Phosphoinositode-3-kinase/AKT serine/threonine kinase (PI3K/AKT) pathway to increase the sensitivity and reverse acquired resistance of esophageal cancer cells to chemotherapeutic drugs (33). Thus, it is imperative to identify the therapeutic targets that not only are advantageous in predicting the clinical outcome but also can resolve the emerging chemotherapy-related issues as well.

In our study, using different genomic data, we examined the changes in gene expression before and after treatment to investigate the key genes, pathways, biomarkers, and risk gene signature. Interestingly, we found that most of the genes that were highly expressed in OSCC patients are found to be significantly downregulated after doxorubicin treatment. We also get another cluster of genes whose expression does not change significantly after doxorubicin treatment. In fact, the expression of some of them increases further. In total, 168 overlapping DEGs were found from the three datasets and were considered for further study. Five gene signatures (*SERPINE1*, *PLAU*, *BOP1*, *CKS2*, and *CCNA2*) were proposed in this study through Cox regression and machine learning. The overexpression of these genes increases the risk of adverse

outcome. Application of Cox regression resulted in an equation that can be used for risk stratification of OSCC patients. From the survival analysis, we were able to predict that increased expression of these genes is related to poor prognosis of the patients. A similar study found SMA and SERPINE1 to be significantly associated with prognosis in OSCC (34); SERPINE1, PLAU, and ACTA1 acted as both diagnostic and prognostic markers (35). Another study by Liu et al. (36) identified an eight-gene prognostic signature in HNSCC that included PLAU. CCNA2 is identified as an independent indicator of worse OS and may serve as a reliable biomarker to identify high-risk subgroups with poor prognosis in OSCC (37). Moreover, these five genes were highly expressed and reported to be associated with cancer progression in various cancers including HNSCC (35, 38), glioblastoma multiforme (GBM) (39), epithelial ovarian cancer (40, 41), gastric cancer (42, 43), breast cancer (44, 45), bladder cancer (46, 47), esophageal cancer (48), colorectal cancer (49, 50), hepatocellular carcinoma (51), melanoma (52), and non-small cell lung carcinoma (NSCLC) (53). In cancer pathogenesis, these genes perturbed numerous cellular mechanisms such as extracellular matrix (ECM) modulation, epithelial-to-mesenchymal transition (EMT), cell migration, and angiogenesis (39, 47–49, 54–59). Interestingly, from a genetic study, we also found that gene amplification in SERPINE1, PLAU, BOP1, and CKS2 is the primary cause of their overexpression, except in CCNA2, where mutation is the predominant cause in HNSCC. Experimental evidence showed that amplification of these genes occurs at both mRNA and protein levels in several cancers such as breast cancer (60, 61), prostate cancer (62), rectal cancer (63), hepatocellular cancer (51), NSCLC (64), gastric cancer (65), and tongue cancer (66).

Overexpression of these markers is indicative of poor prognosis and corresponds to chemotherapy resistance. We showed that initially during chemotherapy, there is downregulation of these genes in the responsive cancer (Figure 5C), but as treatment continues, these genes are upregulated to give rise to chemoresistant phenotypes (Figure 5D). We examined the mRNA expression of these five genes with different drug treatments, which include doxorubicin, cisplatin, oxaliplatin, gemcitabine, topotecan, temozolomide, paclitaxel, JQ1, romidepsin, dauricine, and SNSO32, in various cancers to support our outcome. All of the drugs that we used in our study have DNA as their target and inhibit DNA replication or transcription. In another analysis with the drug olaparib, which is a poly(ADP-ribose) polymerase 1 (PARP1) inhibitor (GSE165585), we found that our genes of interest were not significantly expressed (Supplementary Table S3). In support of our results, we found that PARP1 inhibitors help in the sensitization of temozolomide-resistant glioblastoma cancer (67). We found that after drug treatment, gene expression changes from high to low or low to high or remains unchanged. Further additional new gene expression was also

seen. We found that during the period of treatment responsiveness, our genes change their expression level from low to high again, which is the same as the prior treatment. Other studies have also determined the involvement of these genes in resistance generation, for example, SERPINE1 is upregulated in cisplatin-resistant oral cancer cell lines (SCC9, SCC4, and H357) (68) and paclitaxel-resistant breast cancer (44). In the context of therapeutic potential, SERPINE1 has been identified as a potential therapeutic target, as it acts as a pro-proliferative oncogenic factor (69). PLAU is found to be upregulated in cisplatin-resistant oral cancer cell lines (68). Both PLAU and SERPINE1 were found to be highly expressed in breast cancer patients with adjuvant endocrine therapy and related to shorter disease-free survival and OS (70). Whereas in vivo, BOP1 downregulation was reported to inhibit paclitaxel resistance and Cancer stem cells (CSC)-like phenotype in triple-negative breast cancer (TNBC) cells (71). In estrogen receptor–positive (ER+) breast cancer, tamoxifen resistance is correlated with the overexpression of CCNA2 (72). CKS2 in complex with SSPB1 regulates mitochondrion DNA replication in cervical cancer and can be indicative of chemoradioresistance (73). In the present study, we found that these genes are upregulated and are related to resistance, and they are involved in vital biological processes (Figure 9); therefore, targeting these genes can be of immense therapeutic benefit.

Thus, it is imperative to identify modulators that can check aberrant expression of these genes and increase the barrier toward the emergence of chemoresistance. In this milieu, we also identified small-molecule ligands that can enhance the efficacy/sensitivity of chemotherapeutic agents. A combination of molecules targeting proteins can provide a potent therapeutic option with reduced changes of emergence of chemoresistance. Many of the identified molecules are already reported to be effective against cancers such as colorectal, breast, and lung. Some of the selected proteins are involved in mutually exclusive pathways, as evident from the network analysis, and thus have different mechanisms of action. Simultaneously targeting them can be advantageous to both primary tumor and advanced metastatic tumors. Most of the top-scoring molecules for BOP1 were HIV protease inhibitors, e.g., saquinavir, indinavir, nelfinavir, and ritonavir (Norvir). They are reported to induce cell death in both the chemosensitive and chemoresistant ovarian cancer cell lines in a dose-dependent manner (74). The lopinavir/ritonavir combination is reported to have significant inhibition on cell growth and migration, whereas it enhanced radiosensitivity in HNSCC cell lines (75). The anticancer potential of protease inhibitors is already reported in various previous publications. Some of these molecules are in clinical trials. Therefore, they may present attractive options as targeted chemotherapeutic agents against chemosensitive and chemoresistant OSCC as well (76–78). Moreover, BOP1 inhibition provides additional advantage for non-cancerous cells by inducing a cytoprotective nucleolar stress response and reducing damage to normal tissues from anticancer

FIGURE 9
Pictorial depiction of identified genes in biological pathways. Red blunt lines indicate that their pharmacological inhibition can inhibit major functions involved in gene expression, cell proliferation and metastasis.

drugs such as camptothecin or methotrexate. It is reported that *BOP1* inhibition together with camptothecin results in selective killing of p53 null cells, producing a synergistic effect (79, 80). *PLAU* encodes uPA, commonly associated with cancer progression *via* apoptosis inhibition and breakdown of the ECM. It also promotes angiogenesis (81). PLAU, as a top identified molecule, inhibits various biological functions such as NADH due to its role in energy production and may help against wastage and weakness of cancer patients. It is also used to improve mental alertness. It can be given orally and has cleared clinical trials as a chemotherapeutic agent for other illnesses (27). Argatroban, which is an antithrombotic agent, inhibits the metastasis in breast cancer (82) and melanomas (83). It has significant antineoplastic effect on gliomas as well (84). It downregulates the MAPK/ERK and STAT phosphorylation that results in a reduction of interleukin (IL)-6, IL-12, and tumor necrosis factor (85). The antiparasitic pentamidine has been shown to be effective against various human cancers such as melanoma (86), breast cancer (87), lung cancer, ovarian cancer, and cervical cancer (88, 89). However, the mechanism of its antineoplastic action has not been elucidated fully. Overall, these results infuse great confidence in our analysis. Labetalol, which we have identified as a SERPINE1 binder, blocks both alpha and beta adrenoceptors and has been used for the treatment of hypertension. Alpha-blockers have been reported to increase recurrence-free survival (RFS) (90). Recently, it has been suggested that beta-blockers hinder mechanisms that initiate tumorigenesis, angiogenesis, and metastasis. Beta-blockers have shown good antineoplastic activity in various cancer cell lines. They are also reported to increase the effect of anticancer chemotherapy (91, 92). Therefore, labetalol, having a mix of alpha- and beta-blocker activities, can be a potential candidate for the treatment of OSCC. Reproterol is a β2-agonist used as an antiasthmatic drug. Thus, we suggest that targeting BOP1 and/or

PLAU can be advantageous against both primary and metastatic tumors. Their inhibitors can also be combined together with fluorouracil and methotrexate (93). We have also identified CCNA2 ligands, whereas no direct small-molecule inhibitors for CCNA2 are yet known. While targeting the CCNA2 function, we will be able to target the S/G2/G2-M phase of the cell cycle.

Conclusion

In this study, we have identified changes in the gene expression level as a result of treatment in OSCC. Applying machine learning techniques and Cox regression, we constructed a five-gene-based prognostic signature that can stratify the patients (high and low risk). It is evident that overexpression of these genes is related to poor prognosis and reduced survival in many cancers including OSCC. They are also related to the emergence of chemoresistance against many drugs. Changes in their expression levels, pretreatment, posttreatment, and resistant cell lines make them suitable for targeting. We have also identified potential molecules that can bind to these proteins with high affinity. Since the identified proteins are involved in disparate processes, a combination of molecules targeting them can provide a potent therapeutic option with reduced chances of chemoresistance. We hope that this study provides new avenues for the design of better chemotherapeutic agents especially against chemoresistance in OSCC.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found in the article/ Supplementary Material.

Author contributions

The work was designed and conceptualized by PK and AD. PK, SK, AD, SB, MS, and BM analyzed the data and performed the work. The manuscript was prepared and written by PK, SK, and AD. All authors read and approved the final manuscript.

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.910494/full#supplementary-material

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin* (2021) 7:1–9. doi: 10.3322/caac.21660

2. Jitender S, Sarika G, Varada HR, Omprakash Y, Mohsin K. Screening for oral cancer. *J Exp Ther Oncol* (2016) 11(4):303–7. doi: 10.1007/s12032-021-01548-0

3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590

4. Nocini R, Capocasale G, Marchioni D, Zotti F. A snapshot of knowledge about oral cancer in Italy: A 505 person survey. *Int J Environ Res Public Health* (2020) 17(13):4889. doi: 10.3390/ijerph17134889

5. Hartner L. Chemotherapy for oral cancer. *Dental Clinics North Am* (2018) 62 (1):87–97. doi: 10.1016/j.cden.2017.08.006

6. Bernier J, Domenge C, Ozsahin M, Matuszewska K, Lefèbvre JL, Greiner RH, et al. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *New Engl J Med* (2004) 350(19):1945–52. doi: 10.1056/NEJMoa032641

7. Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME, et al. Harnessing connectivity in a Large-scale small-molecule sensitivity dataset. *Cancer Discov* (2015) 5(11):1210–23. doi: 10.1158/2159-8290.CD-15-0235

8. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol* (2016) 12(2):109–16. doi: 10.1038/nchembio.1986

9. Guan YF, Li GR, Wang RJ, Yi YT, Yang L, Jiang D, et al. Application of next-generation sequencing in clinical oncology to advance personalized treatment of cancer. *Chin J Cancer* (2012) 31(10):463–70. doi: 10.5732/cjc.012.10216

10. Lee YT, Tan YJ, Oon CE. Molecular targeted therapy: Treating cancer with specificity. *Eur J Pharmacol* (2018) 834:188–96. doi: 10.1016/j.ejphar.2018.07.034

11. Thorn CF, Oshiro C, Marsh S, Hernandez-Boussard T, McLeod H, Klein TE, et al. Doxorubicin pathways: pharmacodynamics and adverse effects. *Pharmacogenet Genomics* (2011) 21(7):440–6. doi: 10.1097/FPC.0b013e32833ffb56

12. Sakagami H, Okudaira N, Masuda Y, Amano O, Yokose S, Kanda Y, et al. Induction of apoptosis in human oral keratinocyte by doxorubicin. *Anticancer Res* (2017) 37(3):1023–9. doi: 10.21873/anticanres.11412

13. Ambatipudi S, Gerstung M, Pandey M, Samant T, Patil A, Kane S, et al. Genome-wide expression and copy number analysis identifies driver genes in gingivobuccal cancers. *Genes Chromosomes Cancer* (2012) 51(2):161–73. doi: 10.1002/gcc.20940

14. Hazar-Rethinam M, de Long LM, Gannon OM, Topkas E, Boros S, Vargas AC, et al. A novel E2F/sphingosine kinase 1 axis regulates anthracycline response in squamous cell carcinoma. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2015) 21 (2):417–27. doi: 10.1158/1078-0432.CCR-14-1962

15. Wu G, Haw R. Functional interaction network construction and analysis for disease discovery. *Methods Mol Biol (Clifton NJ)* (2017) 1558:235–53. doi: 10.1007/978-1-4939-6783-4_11

16. Sidiropoulos K, Viteri G, Sevilla C, Jupe S, Webber M, Orlic-Milacic M, et al. Reactome enhanced pathway visualization. *Bioinf (Oxford England)* (2017) 33 (21):3461–7. doi: 10.1093/bioinformatics/btx441

17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Gene Ontol Consort Nat Genet* (2000) 25(1):25–9. doi: 10.1038/75556

18. Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, et al. The gene ontology resource: enriching a GOld mine. *Nucleic Acids Res* (2021) 49 (D1):D325–D34. doi: 10.1093/nar/gkaa1113

19. Toth R, Schiffmann H, Hube-Magg C, Buscheck F, Hoflmayer D, Weidemann S, et al. Random forest-based modelling to detect biomarkers for prostate cancer progression. *Clin Epigenetics* (2019) 11(1):148. doi: 10.1186/s13148-019-0736-8

20. Lind AP, Anderson PC. Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. *PloS One* (2019) 14(7):e0219774. doi: 10.1371/journal.pone.0219774

21. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf* (2003) 4:2. doi: 10.1186/1471-2105-4-2

22. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* (2019) 47(W1):W556–W60. doi: 10.1093/nar/gkz430

23. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* (2012) 2(5):401–4. doi: 10.1158/2159-8290.CD-12-0095

24. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signaling* (2013) 6(269):11. doi: 10.1126/scisignal.2004088

25. Li W, Song P, Zhao M, Gao L, Xie J, You C. BOP1 used as a novel prognostic marker and correlated with tumor microenvironment in pan-cancer. *J Oncol* (2021) 2021:3603030. doi: 10.1155/2021/3603030

26. Chen G, Sun J, Xie M, Yu S, Tang Q, Chen L. PLAU promotes cell proliferation and epithelial-mesenchymal transition in head and neck squamous cell carcinoma. *Front Genet* (2021) 12. doi: 10.3389/fgene.2021.651882

27. Ju H-Q, Lin J-F, Tian T, Xie D, Xu R-H. NADPH homeostasis in cancer: functions, mechanisms and therapeutic implications. *Signal Transd Target Ther* (2020) 5(1):231. doi: 10.1038/s41392-020-00326-0

28. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: What, why, and how? *BMJ* (2009) 338:b375. doi: 10.1136/bmj.b375

29. Manjang K, Tripathi S, Yli-Harja O, Dehmer M, Glazko G, Emmert-Streib F. Prognostic gene expression signatures of breast cancer are lacking a sensible biological meaning. *Sci Rep* (2021) 11(1):156. doi: 10.1038/s41598-020-79375-y

30. Inoue S, Yoshida T, Nishino T, Goto M, Aoyama M, Kawakita N, et al. Biomarkers predicting the response to chemotherapy and the prognosis in patients with esophageal squamous cell carcinoma. *Gen Thorac Cardiovasc Surg* (2021) 69 (3):525–33. doi: 10.1007/s11748-021-01586-5

31. Wei X, Zhou X, Zhao Y, He Y, Weng Z, Xu C. A 14-gene gemcitabine resistance gene signature is significantly associated with the prognosis of pancreatic cancer patients. *Sci Rep* (2021) 11(1):6087. doi: 10.1038/s41598-021-85680-x

32. Yoshida T, Miyoshi T, Seike J-I, Yamai H, Takechi H, Yuasa Y, et al. Gene expression changes in a chemoresistant model with human esophageal cancer xenografts using cDNA microarray. *Anticancer Res* (2009) 29(4):1163–8.

33. Li B, Li J, Wen Xu W, Guan XY, Qin YR, Zhang LY, et al. Suppression of esophageal tumor growth and chemoresistance by directly targeting the PI3K/AKT pathway. *Oncotarget* (2014) 5(22):11576–87. doi: 10.18632/oncotarget.2596

34. Dhanda J, Triantafyllou A, Liloglou T, Kalirai H, Lloyd B, Hanlon R, et al. SERPINE1 and SMA expression at the invasive front predict extracapsular spread and survival in oral squamous cell carcinoma. *Br J Cancer* (2014) 111(11):2114–21. doi: 10.1038/bjc.2014.500

35. Yang K, Zhang S, Zhang D, Tao Q, Zhang T, Liu G, et al. Identification of SERPINE1, PLAU and ACTA1 as biomarkers of head and neck squamous cell carcinoma based on integrated bioinformatics analysis. *Int J Clin Oncol* (2019) 24 (9):1030–41. doi: 10.1007/s10147-019-01435-9

36. Liu B, Su Q, Ma J, Chen C, Wang L, Che F, et al. Prognostic value of eight-gene signature in head and neck squamous carcinoma. *Front Oncol* (2021) 11:657002. doi: 10.3389/fonc.2021.657002

37. Monteiro LS, Diniz-Freitas M, Warnakulasuriya S, Garcia-Caballero T, Forteza-Vila J, Fraga M. Prognostic significance of cyclins A2, B1, D1, and E1 and CCND1 numerical aberrations in oral squamous cell carcinomas. *Anal Cell Pathol (Amst)* (2018) 2018:7253510. doi: 10.1155/2018/7253510

38. Li Z, Chen C, Wang J, Wei M, Liu G, Qin Y, et al. Overexpressed PLAU and its potential prognostic value in head and neck squamous cell carcinoma. *PeerJ* (2021) 9:e10746. doi: 10.7717/peerj.10746

39. Seker F, Cingoz A, Sur-Erdem İ, Erguder N, Erkent A, Uyulur F, et al. Identification of SERPINE1 as a regulator of glioblastoma cell dispersal with transcriptome profiling. *Cancers (Basel)* (2019) 11(11):1651. doi: 10.3390/cancers11111651

40. Pan J-X, Qu F, Wang F-F, Xu J, Mu L-S, Ye L-Y, et al. Aberrant SERPINE1 DNA methylation is involved in carboplatin induced epithelial-mesenchymal transition in epithelial ovarian cancer. *Arch Gynecol Obstet* (2017) 296(6):1145–52. doi: 10.1007/s00404-017-4547-x

41. Xu JH, Wang Y, Xu D. CKS2 promotes tumor progression and metastasis and is an independent predictor of poor prognosis in epithelial ovarian cancer. *Eur Rev Med Pharmacol Sci* (2019) 23(8):3225–34. doi: 10.26355/eurrev_201904_17681

42. Li L, Zhu Z, Zhao Y, Zhang Q, Wu X, Miao B, et al. FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of

gastric adenocarcinoma revealed by microarray and bioinformatics. *Sci Rep* (2019) 9(1):7827. doi: 10.1038/s41598-019-43924-x

43. Ai C, Zhang J, Lian S, Ma J, Győrffy B, Qian Z, et al. FOXM1 functions collaboratively with PLAU to promote gastric cancer progression. *J Cancer* (2020) 11(4):788–94. doi: 10.7150/jca.37323

44. Zhang Q, Lei L, Jing D. Knockdown of SERPINE1 reverses resistance of triple–negative breast cancer to paclitaxel *via* suppression of VEGFA. *Oncol Rep* (2020) 44(5):1875–84. doi: 10.3892/or.2020.7770

45. Urban P, Vuaroqueaux V, Labuhn M, Delorenzi M, Wirapati P, Wight E, et al. Increased expression of urokinase-type plasminogen activator mRNA determines adverse prognosis in ErbB2-positive primary breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol* (2006) 24(26):4245–53. doi: 10.1200/JCO.2005.05.1912

46. Chen Z, Chen X, Xie R, Huang M, Dong W, Han J, et al. DANCR promotes metastasis and proliferation in bladder cancer cells by enhancing IL-11-STAT3 signaling and CCND1 expression. *Mol Ther J Am Soc Gene Ther* (2019) 27(2):326–41. doi: 10.1016/j.ymthe.2018.12.015

47. Chen R, Feng C, Xu Y. Cyclin-dependent kinase-associated protein Cks2 is associated with bladder cancer progression. *J Int Med Res* (2011) 39(2):533–40. doi: 10.1177/147323001103900222

48. Fang L, Che Y, Zhang C, Huang J, Lei Y, Lu Z, et al. PLAU directs conversion of fibroblasts to inflammatory cancer-associated fibroblasts, promoting esophageal squamous cell carcinoma progression *via* uPAR/Akt/NF-κB/IL8 pathway. *Cell Death Discov* (2021) 7(1):32. doi: 10.1038/s41420-021-00410-6

49. Qi J, Yu Y, Akilli Öztürk Ö, Holland JD, Besser D, Fritzmann J, et al. New wnt/β-catenin target genes promote experimental metastasis and migration of colorectal cancer cells through different signals. *Gut* (2016) 65(10):1690–701. doi: 10.1136/gutjnl-2014-307900

50. Chen X, Zhao Y. Block of proliferation 1 promotes cell migration and invasion in human colorectal cancer cells *via* the JNK pathway. *J Clin Lab Anal* (2020) 34(7):e23283. doi: 10.1002/jcla.23283

51. Chung KY, Cheng IK, Ching AK, Chu JH, Lai PB, Wong N. Block of proliferation 1 (BOP1) plays an oncogenic role in hepatocellular carcinoma by promoting epithelial-to-mesenchymal transition. *Hepatol (Baltimore Md)* (2011) 54(1):307–18. doi: 10.1002/hep.24372

52. Gupta R, Bugide S, Wang B, Green MR, Johnson DB, Wajapeyee N. Loss of BOP1 confers resistance to BRAF kinase inhibitors in melanoma by activating MAP kinase pathway. *Proc Natl Acad Sci USA* (2019) 116(10):4583–91. doi: 10.1073/pnas.1821889116

53. Ruan JS, Zhou H, Yang L, Wang L, Jiang ZS, Wang SM. CCNA2 facilitates epithelial-to-mesenchymal transition *via* the integrin αvβ3 signaling in NSCLC. *Int J Clin Exp Pathol* (2017) 10(8):8324–33.

54. Leik CE, Su EJ, Nambi P, Crandall DL, Lawrence DA. Effect of pharmacologic plasminogen activator inhibitor-1 inhibition on cell motility and tumor angiogenesis. *J Thromb Haemost JTH* (2006) 4(12):2710–5. doi: 10.1111/j.1538-7836.2006.02244.x

55. Xu B, Bai Z, Yin J, Zhang Z. Global transcriptomic analysis identifies SERPINE1 as a prognostic biomarker associated with epithelial-to-mesenchymal transition in gastric cancer. *PeerJ* (2019) 7:e7091. doi: 10.7717/peerj.7091

56. Kong HJ, Kwon EJ, Kwon OS, Lee H, Choi JY, Kim YJ, et al. Crosstalk between YAP and TGFβ regulates SERPINE1 expression in mesenchymal lung cancer cells. *Int J Oncol* (2021) 58(1):111–21. doi: 10.3892/ijo.2020.5153

57. Danø K, Behrendt N, Høyer-Hansen G, Johnsen M, Lund LR, Ploug M, et al. Plasminogen activation and cancer. *Thromb Haemost* (2005) 93(4):676–81. doi: 10.1160/TH05-01-0054

58. Pestov DG, Strezoska Z, Lau LF. Evidence of p53-dependent cross-talk between ribosome biogenesis and the cell cycle: effects of nucleolar protein Bop1 on G(1)/S transition. *Mol Cell Biol* (2001) 21(13):4246–55. doi: 10.1128/MCB.21.13.4246-4255.2001

59. Arsic N, Bendris N, Peter M, Begon-Pescia C, Rebouissou C, Gadéa G, et al. A novel function for cyclin A2: control of cell invasion *via* RhoA signaling. *J Cell Biol* (2012) 196(1):147–62. doi: 10.1083/jcb.201102085

60. Sternlicht MD, Dunning AM, Moore DH, Pharoah PD, Ginzinger DG, Chin K, et al. Prognostic value of PAI1 in invasive breast cancer: Evidence that tumor-specific factors are more important than genetic variation in regulating PAI1 expression. *Cancer Epidemiol Biomarkers Prev Publ Am Assoc Cancer Res Cosponsored by Am Soc Prev Oncol* (2006) 15(11):2107–14. doi: 10.1158/1055-9965.EPI-06-0351

61. Meng S, Tripathy D, Shete S, Ashfaq R, Saboorian H, Haley B, et al. uPAR and HER-2 gene status in individual breast cancer cells from blood and tissues. *Proc Natl Acad Sci* (2006) 103(46):17361–5. doi: 10.1073/pnas.0608113103

62. Helenius MA, Saramäki OR, Linja MJ, Tammela TLJ, Visakorpi T. Amplification of urokinase gene in prostate cancer. *Cancer Res* (2001) 61(14):5340–4.

63. Lips EH, van Eijk R, de Graaf EJR, Oosting J, de Miranda NFCC, Karsten T, et al. Integrating chromosomal aberrations and gene expression profiles to dissect rectal tumorigenesis. *BMC Cancer* (2008) 8(1):314. doi: 10.1186/1471-2407-8-314

64. Micke P, Edlund K, Holmberg L, Kultima HG, Mansouri L, Ekman S, et al. Gene copy number aberrations are associated with survival in histologic subgroups of non-small cell lung cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer* (2011) 6(11):1833–40. doi: 10.1097/JTO.0b013e3182295917

65. Jin DH, Park SE, Lee J, Kim KM, Kim S, Kim DH, et al. Copy number gains at 8q24 and 20q11-q13 in gastric cancer are more common in intestinal-type than diffuse-type. *PloS One* (2015) 10(9):e0137657. doi: 10.1371/journal.pone.0137657

66. Gao F, Li C, Zhao X, Xie J, Fang G, Li Y. CKS2 modulates cell-cycle progression of tongue squamous cell carcinoma cells partly *via* modulating the cellular distribution of DUTPase. *J Oral Pathol Med Off Publ Int Assoc Oral Pathol Am Acad Oral Pathol* (2021) 50(2):175–82. doi: 10.1111/jop.13116

67. Zhang J, Stevens MF, Bradshaw TD. Temozolomide: mechanisms of action, repair and resistance. *Curr Mol Pharmacol* (2012) 5(1):102–14. doi: 10.2174/1874467211205010102

68. Shriwas O, Arya R, Mohanty S, Mohapatra P, Kumar S, Rath R, et al. RRBP1 rewires cisplatin resistance in oral squamous cell carcinoma by regulating hippo pathway. *Br J Cancer* (2021) 124:2004–16. doi: 10.1101/2020.03.18.998070

69. Zhao C, Liu Z. MicroRNA 617 targeting SERPINE1 inhibited the progression of oral squamous cell carcinoma. *Mol Cell Biol* (2021) 41(6): e0056520. doi: 10.1128/MCB.00565-20

70. Jevrić M, Matić IZ, Krivokuća A, Đorđić Crnogorac M, Besu I, Damjanović A, et al. Association of uPA and PAI-1 tumor levels and 4G/5G variants of PAI-1 gene with disease outcome in luminal HER2-negative node-negative breast cancer patients treated with adjuvant endocrine therapy. *BMC Cancer* (2019) 19(1):71. doi: 10.1186/s12885-018-5255-z

71. Li S, Wu H, Huang X, Jian Y, Kong L, Xu H, et al. BOP1 confers chemoresistance of triple-negative breast cancer by promoting CBP-mediated β-catenin acetylation. *J Pathol* (2021) 254:265–78. doi: 10.1002/path.5676

72. Gao T, Han Y, Yu L, Ao S, Li Z, Ji J. CCNA2 is a prognostic biomarker for ER+ breast cancer and tamoxifen resistance. *PloS One* (2014) 9(3):e91771. doi: 10.1371/journal.pone.0091771

73. Jonsson M, Fjeldbo CS, Holm R, Stokke T, Kristensen GB, Lyng H. Mitochondrial function of CKS2 oncoprotein links oxidative phosphorylation with cell division in chemoradioresistant cervical cancer. *Neoplasia (New York NY)* (2019) 21(4):353–62. doi: 10.1016/j.neo.2019.01.002

74. Bandiera E, Todeschini P, Romani C, Zanotti L, Erba E, Colmegna B, et al. The HIV-protease inhibitor saquinavir reduces proliferation, invasion and clonogenicity in cervical cancer cell lines. *Oncol Lett* (2016) 12(4):2493–500. doi: 10.3892/ol.2016.5008

75. Liu R, Zhang L, Yang J, Zhang X, Mikkelsen R, Song S, et al. HIV Protease inhibitors sensitize human head and neck squamous carcinoma cells to radiation by activating endoplasmic reticulum stress. *PloS One* (2015) 10(5):e0125928. doi: 10.1371/journal.pone.0125928

76. Koltai T. Nelfinavir and other protease inhibitors in cancer: mechanisms involved in anticancer activity. *F1000Research* (2015) 4:9. doi: 10.12688/f1000research.5827.2

77. Trezza A, Cicaloni V, Pettini F, Spiga O. Chapter 2 - potential roles of protease inhibitors in anticancer therapy. In: SP Gupta, editor. *Cancer-leading proteases*. Cambridge, Massachusetts: Academic Press (2020). p. 13–49.

78. Yang P, Li ZY, Li HQ. Potential roles of protease inhibitors in cancer progression. *Asian Pac J Cancer Prev APJCP* (2015) 16(18):8047–52. doi: 10.7314/APJCP.2015.16.18.8047

79. Sapio RT, Nezdyur AN, Krevetski M, Anikin L, Manna VJ, Minkovsky N, et al. Inhibition of post-transcriptional steps in ribosome biogenesis confers cytoprotection against chemotherapeutic agents in a p53-dependent manner. *Sci Rep* (2017) 7(1):9041. doi: 10.1038/s41598-017-09002-w

80. Pecoraro A, Pagano M, Russo G, Russo A. Ribosome biogenesis and cancer: Overview on ribosomal proteins. *Int J Mol Sci* (2021) 22(11):5496. doi: 10.3390/ijms22115496

81. Mahmood N, Mihalcioiu C, Rabbani SA. Multifaceted role of the urokinase-type plasminogen activator (uPA) and its receptor (uPAR): Diagnostic, prognostic, and therapeutic applications. *Front Oncol* (2018) 8(24). doi: 10.3389/fonc.2018.00024

82. Asanuma K, Wakabayashi H, Okamoto T, Asanuma Y, Akita N, Yoshikawa T, et al. The thrombin inhibitor, argatroban, inhibits breast cancer metastasis to bone. *Breast Cancer* (2013) 20(3):241–6. doi: 10.1007/s12282-012-0334-5

83. Asanuma K, Wakabayashi H, Hayashi T, Okuyama N, Seto M, Matsumine A, et al. Thrombin inhibitor, argatroban, prevents tumor cell migration and bone metastasis. *Oncology* (2004) 67(2):166–73. doi: 10.1159/000081004

84. Hua Y, Tang L, Keep RF, Schallert T, Fewel ME, Muraszko KM, et al. The role of thrombin in gliomas. *J Thromb Haemost JTH* (2005) 3(9):1917–23. doi: 10.1111/j.1538-7836.2005.01446.x

85. Kuriakose S, Muleme H, Onyilagha C, Okeke E, Uzonna JE. Diminazene aceturate (Berenil) modulates LPS induced pro-inflammatory cytokine production by inhibiting phosphorylation of MAPKs and STAT proteins. *Innate Immun* (2014) 20(7):760–73. doi: 10.1177/1753425913507488

86. Smith J, Stewart BJ, Glaysher S, Peregrin K, Knight LA, Weber DJ, et al. The effect of pentamidine on melanoma *ex vivo*. *Anti-cancer Drugs* (2010) 21(2):181–5. doi: 10.1097/CAD.0b013e3283340cee

87. Jung HJ, Suh SI, Suh MH, Baek WK, Park JW. Pentamidine reduces expression of hypoxia-inducible factor-1alpha in DU145 and MDA-MB-231 cancer cells. *Cancer Lett* (2011) 303(1):39–46. doi: 10.1016/j.canlet.2011.01.008

88. Chow TY, Alaoui-Jamali MA, Yeh C, Yuen L, Griller D. The DNA double-stranded break repair protein endo-exonuclease as a therapeutic target for cancer. *Mol Cancer Ther* (2004) 3(8):911–9. doi: 10.1158/1535-7163.911.3.8

89. Pathak MK, Dhawan D, Lindner DJ, Borden EC, Farver C, Yi T. Pentamidine is an inhibitor of PRL phosphatases with anticancer activity. *Mol Cancer Ther* (2002) 1(14):1255–64.

90. Hart J, Spencer B, McDermott CM, Chess-Williams R, Sellers D, Christie D, et al. A pilot retrospective analysis of alpha-blockers on recurrence in men with localised prostate cancer treated with radiotherapy. *Sci Rep* (2020) 10(1):8191. doi: 10.1038/s41598-020-65238-z

91. Oh MS, Guzner A, Wainwright DA, Mohindra NA, Chae YK, Behdad A, et al. The impact of beta blockers on survival outcomes in patients with non–small-cell lung cancer treated with immune checkpoint inhibitors. *Clin Lung Cancer* (2021) 22(1):e57–62. doi: 10.1016/j.cllc.2020.07.016

92. Madel M-B, Elefteriou F. Mechanisms supporting the use of beta-blockers for the management of breast cancer bone metastasis. *Cancers* (2021) 13(12):2887. doi: 10.3390/cancers13122887

93. del Rincón SV, Widschwendter M, Sun D, Ekholm-Reed S, Tat J, Teixeira LK, et al. Cks overexpression enhances chemotherapeutic efficacy by overriding DNA damage checkpoints. *Oncogene* (2015) 34(15):1961–7. doi: 10.1038/onc.2014.137