Check for updates

*CORRESPONDENCE
Bin-Bin Cui
cbbhrb@163.com
Yan-Long Liu
liuyanlong@hrbmu.edu.cn

†These authors have contributed
equally to this work

Exploring biomarkers for prognosis and neoadjuvant chemosensitivity in rectal cancer: Multi-omics and ctDNA sequencing collaboration

Xiu-Feng Jiang[1†], Bo-Miao Zhang[1†], Fen-Qi Du[1],
Jun-Nan Guo[1], Dan Wang[2], Yi-En Li[1], Shen-Hui Deng[3],
Bin-Bin Cui[1*] and Yan-Long Liu[1*]

[1]Department of Colorectal Surgery, Harbin Medical University Cancer Hospital, Harbin, China,
[2]Department of Neurology, The Second Affiliated Hospital of Qiqihar Medical University, Qiqihar, China,
[3]Department of Anesthesiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, China

Introduction: This study aimed to identified the key genes and sequencing metrics for predicting prognosis and efficacy of neoadjuvant chemotherapy (nCT) in rectal cancer (RC) based on genomic DNA sequencing in samples with different origin and multi-omics association database.

Methods: We collected 16 RC patients and obtained DNA sequencing data from cancer tissues and plasma cell-free DNA before and after nCT. Various gene variations were analyzed, including single nucleotide variants (SNV), copy number variation (CNV), tumor mutation burden (TMB), copy number instability (CNI) and mutant-allele tumor heterogeneity (MATH). We also identified genes by which CNV level can differentiate the response to nCT. The Cancer Genome Atlas database and the Clinical Proteomic Tumor Analysis Consortium database were used to further evaluate the specific role of therapeutic relevant genes and screen out the key genes in multi-omics levels. After the intersection of the screened genes from differential expression analysis, survival analysis and principal components analysis dimensionality reduction cluster analysis, the key genes were finally identified.

Results: The genes CNV level of principal component genes in baseline blood and cancer tissues could significantly distinguish the two groups of patients. The CNV of HSP90AA1, EGFR, SRC, MTOR, etc. were relatively gained in the better group compared with the poor group in baseline blood. The CNI and TMB was significantly different between the two groups. The increased expression of HSP90AA1, EGFR, and SRC was associated with increased sensitivity to multiple chemotherapeutic drugs. The nCT predictive score obtained by therapeutic relevant genes could be a potential prognostic indicator, and the combination with TMB could further refine prognostic prediction for patients. After a series of analysis in multi-omics association database, EGFR and HSP90AA1 with significant differences in multiple aspects

were identified as the key predictive genes related to prognosis and the sensitivity of nCT.

Discussion: This work revealed that effective combined application and analysis in multi-omics data are critical to search for predictive biomarkers. The key genes EGFR and HSP90AA1 could serve as an effective biomarker to predict prognose and neoadjuvant chemosensitivity.

Introduction

The high incidence and poor prognosis of colorectal cancer (CRC) significantly impact the quality of life and economy of patients worldwide. It has also become the third leading cause of cancer death. Rectal cancer (RC) accounts for approximately one-third of all CRC cases (1). Multimodality therapies have been widely used in the clinical treatment of locally advanced patients due to advancements in pathology, imaging, genome sequencing technology, surgical technology, and instruments. Significant progress has been made in the last 30 years (2). Preoperative neoadjuvant radiotherapy and chemotherapy, total mesorectal excision, and postoperative adjuvant chemotherapy are the primary treatment methods that significantly reduce the local recurrence rate and improve the prognosis of patients. However, that strategy comes at the cost of quality of life (3). Despite this, only about 20% of RC patients treated with neoadjuvant chemotherapy (nCT) will have a complete response (CR) or partial response (PR) during preoperative evaluation or surgery (4). Most of the other patients benefited little from nCT after experiencing a series of toxic side effects. So far, the most significant clinical diagnosis and treatment challenge has been determining the best treatment strategy for each patient, enabling the individualized treatment, reducing side effects, and optimizing the quality of life. In addition, no biomarkers have been identified that can accurately predict the benefits of preoperative neoadjuvant therapy.

Recently, the rapid development of high-throughput sequencing technology has enabled next-generation sequencing (NGS) to detect various gene variants in cancer tissues based on DNA and RNA (5). In clinical applications, NGS-based DNA testing has demonstrated significant advantages in identifying clinically treatable genetic variants to guide patient treatment and prognostic-relevant adverse genomic variants before treatment (6). It also has clinical implications for predicting the outcome of chemotherapy. "Liquid Biopsy" has recently emerged as an accessible, convenient, and reproducible technique for real-time monitoring tumor patients by searching for circulating molecular markers in peripheral blood (7). In different types of cancer, biomarkers such as circulating tumor cells, microRNAs, and DNA have been investigated as potential diagnostic and prognostic markers for personalized therapies (8). This non-invasive blood-based test combined with NGS has significant advantages in patients who require preoperative nCT but cannot access fresh tumor tissue before surgery (9). It can investigate the changes and specific characteristics of the genome in patients at both the baseline and post-treatment level of neoadjuvant therapy, including single nucleotide variants (SNV), copy number variation (CNV) and tumor mutation burden (TMB).

Furthermore, The Cancer Genome Atlas (TCGA), a large genomic database, provides RNA-sequencing data from many RC patients before treatment. Numerous studies have demonstrated that genotyping based on RNA signatures in cancer tissues can accurately predict patient prognosis, chemosensitivity, and immunotherapy sensitivity, with high accuracy across multiple independent cohorts (10, 11). However, TCGA does not provide RNA-sequencing data of RC after treatment, and the genomes of untreated and treated tumors may differ significantly, limiting the analysis and prediction accuracy of neoadjuvant chemosensitivity for RC at the RNA level. A strong association between these two assays that target different genomic levels, such as the CNV level, will affect the RNA expression (12, 13). Meanwhile, increasing the copy number of a gene could be a mechanism for increasing protein expression (14). Therefore, effective combined application and analysis of them in multi-omics data are critical in the future search for predictive biomarkers for neoadjuvant chemosensitivity and to develop multimodality and individualized precision treatment.

In this study, we collected cancer tissues, adjacent tissues, and peripheral blood samples from 16 RC patients before and after nCT. Cell-free DNA (cfDNA) was isolated from the peripheral blood of patients, and the targeted gene capture

panel sequencing was performed. The various gene variations in patients were examined, including SNV, CNV, TMB, copy number instability (CNI), and mutant-allele tumor heterogeneity (MATH). We also identified genes from the blood and tissue by which CNV levels can differentiate the response to nCT. The TCGA database was used to further evaluate the specific role of these genes and screen out the genes with significant differences in CNV in normal and cancer tissues. Then, we identified the genes whose expression levels were significantly correlated with CNV levels. The Clinical Proteomic Tumor Analysis Consortium (CPTAC) database was used for screening after the intersection of the screened genes from the above processes, and the key genes for predicting the prognosis and efficacy of nCT were finally obtained.

Methods

Sample collection

All patients received preoperative nCT, with capecitabine (950–1000 mg/m^2) administered twice daily by oral gavage for 14 days and oxaliplatin (130 mg/m^2) administered intravenously on the first day. Peripheral blood samples before and after nCT and surgically resected tumor tissue were collected from all patients. Following a quality control assessment, 16 patient samples met the criteria and were subjected to further analysis. There were seven patients in the CR group, one in the PR group, seven in the stable disease (SD) group, and one in the progressive disease (PD) group. The study was approved by the Ethics Committee of the Harbin Cancer Hospital Medical University, and all patients signed an informed consent document.

CfDNA extraction and sequencing

Centrifugation at 1600 × g for 10 min separated the peripheral blood lymphocytes and plasma. The supernatant plasma was then transferred to a new 2 mL centrifuge tube and centrifuged at 16,000 × g for 10 min. MagMAXTM cfDNA isolation kit (Life Technologies, California, USA) was used to extract cfDNA in the plasma. Tiangen whole blood DNA kit (Tiangen, Beijing, PRC) was used to extract DNA from peripheral blood lymphocytes according to the manufacturer's instructions. The DNA concentration was determined using either the Qubit dsDNA HS Assay kit or the Qubit dsDNA BR Assay kit (Life Technologies, California, USA).

Genomic DNA was sheared into 150–200 bp fragments with Covaris M220 Focused-ultrasonicatorTM Instrument (Covaris, Massachusetts, USA). Fragmented DNA and cfDNA libraries were constructed by KAPA HTP Library Preparation Kit (Illumina platforms) (KAPA Biosystems, Massachusetts, USA)

as per the manufacturer's instruction. A designed Genescope panel of 1086 genes (Genecast, Beijing, China) was used to capture the DNA libraries that included significant tumor-related genes. The captured samples were paired on an end sequencing Illumina HiSeq X-Ten.

Basic analysis

Following quality control, clean data were aligned to the hg19 human genome using BWA 0.7.17 [15], and duplications were masked by Picard (v2.23.0) [16]. Variants were called using VarScan (v2.4.2) [17] and annotated by ANNOVAR software [18]. Somatic mutations were eliminated using the following steps: (i) Exclude the mutations annotated as synonymous SNVs or located in intergenic or intronic regions; (ii) Exclude mutations annotated with allele frequency ≥ 0.002 in the Exome Aggregation Consortium (ExAC) database [19] and the Genome Aggregation Database [20]; (iii) Exclude mutations with strand bias, support reads < 5, and allele frequency < 0.05 in the tumor sample and allele frequency < 0.01 in the plasma sample. CNVs were called *via* cnvkit (v0.9.2) software using paired mode [21].

We used the principal components analysis (PCA) algorithm to perform dimensionality reduction and cluster analysis based on treatment efficacy information to screen for the potential genes that can predict nCT efficacy. The R package Complex heatmap [22] was performed to draw the landscape of genomic alterations in samples from different sources (patient's baseline blood, post-chemotherapy blood, and CR tissue). The chromosomal locations of the genes identified by the above analyses were visualized by the R package "RCircos" [23]. Furthermore, we also analyzed the differences in CNI, MATH, and TMB in baseline blood. CNI, TMB, and MATH differences were compared using the Mann–Whitney U test.

CNI score calculation

The CNI score is a general measure of chromosomal instability (CIN) and is directly related to the regional chromosomal DNA ploidy [24, 25]. To assess the extent of CIN, we quantified it using the CNI score. After GC content correction and normalization of target region length, read counts were transformed into log2 ratios [26]. The log2 ratios were then converted into Z-score using Gaussian transformations versus a baseline group. The regions with Z-scores greater than the 95th percentile plus twice the absolute standard deviation were defined as unstable regions. The CNI score was calculated by adding the Z-scores of unstable regions [25].

MATH calculation

Somatic mutations with variant allele frequency (VAF) between 2% to 100% were included for MATH analysis. MATH was calculated by the following formula: Median absolute deviation of included somatic mutations/median VAF (27).

TMB calculation

Absolute mutation counts were defined as the number of somatic mutations, and TMB was calculated with the formula: Absolute mutation counts * 1000000/Panel exonic base number (28).

Preprocessing and analysisof TCGA RC samples

We included 165 RC patients in the TCGA database (Data Release 34.0, Release Date: July 27, 2022, https://tcga-data.nci.nih.gov/tcga/). The downloaded information included RNA expression data and CNV information. To search for genes with different CNV levels in normal and cancer tissues, we utilized a chi-square test to evaluate statistical significance.

To evaluate the specific role of therapeutic relevant genes identified by the RC blood samples sequencing results. We performed PCA analysis to extract the main components of those therapeutic relevant genes and then constructed a gene signature in TCGA cohort. Both principal components 1 and 2 were selected as signature scores. A method similar to the gene expression rank index was performed to define the nCT predictive score (nCTPS) of each patient: $nCTPS = \sum PCA1i + \sum PCA2i$ (i is the expression of therapeutic relevant genes). To identify the relative enrichment degree in biological processes of different groups, R packages "GSVA" (29) was used to perform enrichment analysis.

Correlation analysis of genes at CNV and transcriptional levels

To investigate genes associated with CNV and transcript levels, we extracted the RNA expression matrix of the above genes with significantly different CNVs. We divided all samples into four groups: single deletion, normal, single gain, and amplification groups. The R software (version 4.0.5) was then used for statistical analysis. Differences between the two groups were compared using the Wilcoxon test, while comparisons between more than two groups were performed using the Kruskal-Wallis test. In all results, $p < 0.05$ was considered statistically significant.

Identification of potential genes for predicting efficacy of neoadjuvant chemotherapy

After obtaining the genes whose transcriptional levels were associated with CNVs, we intersected these genes with the genes obtained from PCA and cluster analysis that might predict the efficacy of nCT. These genes in the intersection have great potential and value for predicting the efficacy of nCT. Then, to verify the effect of these intersecting genes on chemosensitivity, we performed a drug sensitivity prediction analysis. Gene expression data and chemotherapeutic drug response data were downloaded from CellMiner[TM] (https://discover.nci.nih.gov/cellminer/), these data were from the same batch. We deleted drugs that without FDA-approval or clinical trials, and selected chemotherapy drugs for RC. Then we extracted the genes expression data, and analyzed the correlation between their expression and drug sensitivity.

To further screen and validate the key genes at the protein level, we downloaded the proteomic cohort of RC from the CPTAC database. Meanwhile, a web tool, the University of ALabama at Birmingham CANcer data analysis Portal (UALCAN) (http://ualcan.path.uab.edu/), was chosen to be used, which integrates the proteomic data of all tumor samples from the CPTAC database. We downloaded the immunohistochemical staining images from The Human Protein Atlas project (https://www.proteinatlas.org/). Each sample is represented by 1 mm tissue cores (30). In colon cancer (CC) and RC samples, we analyzed the differential expression of the screened genes in normal and cancer issues, their impact on prognosis, and the correlation between their transcriptional levels and protein levels, respectively. To distinguish the high and low expression groups associated to prognosis, the best cut-off value was estimated by R package "maxstat" (31). Genes with significantly different results in multiple aspects were finally identified as the key genes. Finally, DisNor database (32) (https://disnor.uniroma2.it/) was used to analyze the up- and downstream binding sites and causal interaction of the key genes.

Results

Cluster analysis of SNV and CNV in three types of samples

First, we used the heatmap to present the genes with SNV detected in the baseline blood in order of frequency of occurrence from high to low, which were PTEN, ARID1A, SMARCA4, NPM1, MSH3, SEC16A, AGXT, ACE, KSR2, ERBB2, PTCH1, RB1, FBXW7, APC, DLG5, TP53, MCL1, SETD2, JAK1, NF1, SULT1A1, KDM4D, PIK3CB, SMARCA1, BRAF, CD3EAP,

SERPING1, PDGFRA, COL1A1, HDAC2, EPHX1, MAPK11, and KRAS. The nonsynonymous SNV was the most common mutation type among the genes in the heatmap, occurring almost in every gene, followed by frameshift deletion, non-frameshift deletion, and frameshift insertion (Figure 1A). We found that the SNP level of the principal component genes in baseline blood was less effective for distinguishing the two groups by PCA dimensionality reduction cluster analysis (Figures 1B, C), with an area under the receiver operating characteristic curve (AUC) of only 0.48 (Figure 1D). In addition, after chemotherapy, the SNP level of the principal component genes in blood was less effective in distinguishing the two groups (Figures 1E, F), with an AUC of only 0.62 (Figure 1G).
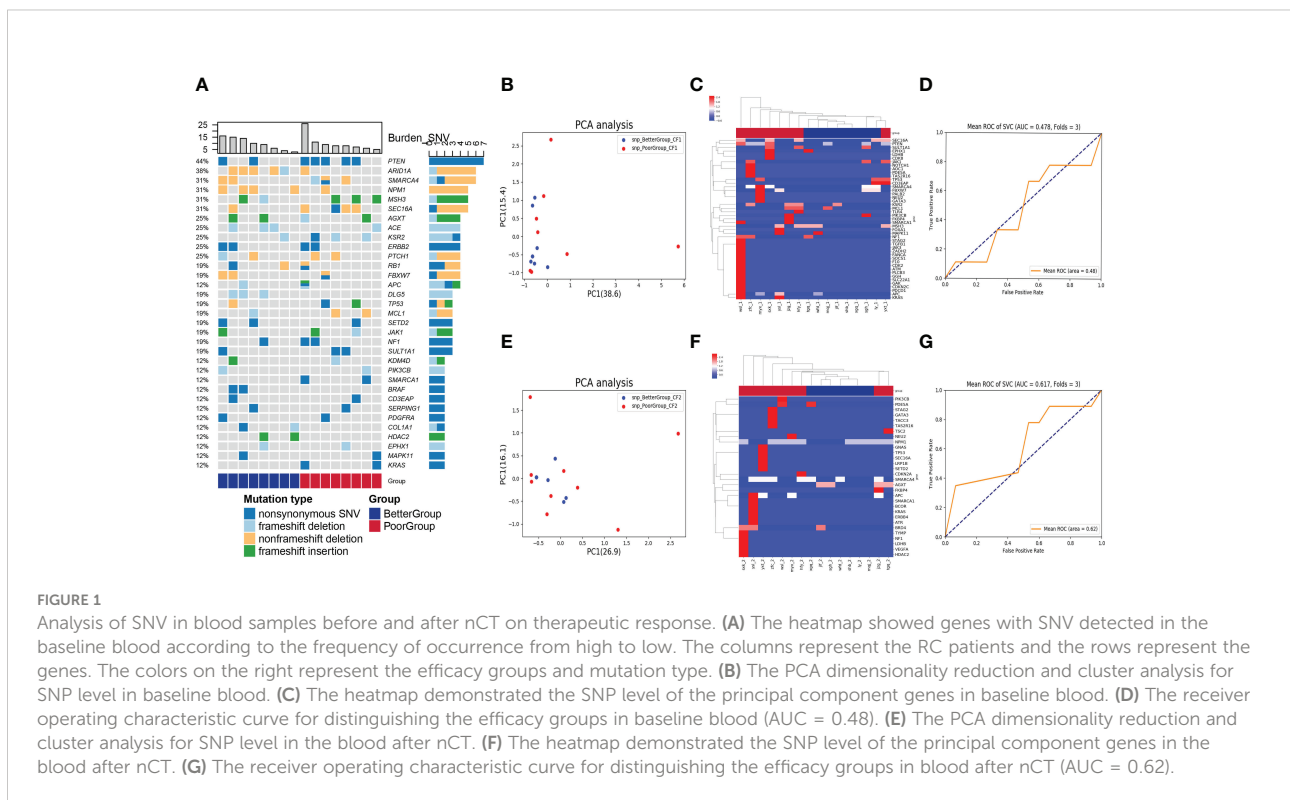
In the CNV analysis, the genes with CNV in descending order of occurrence frequency are: HDAC2, FAM131B, GAPDH, SLC19A1, GGH, REV1, COL1A1, AKR1C3, AREG, HMGB1, HSPA8, CYP2D6, SLC31A1, TGFB1, RNF43, APOA4, KLC3, SPG7, SULT2B1, USP6, ABCC5, AGO2, ARID5B, CAT, CCND2, CDKN1B, CYP24A1, CYP2A6, DHFR, DRD2, GALE, GLP1R, GSTP1, HNF4A, IFNL3, PCK1, PDCD1, TLR2, TPMT, WARS, XRCC1 (Figure 2A). Most of the top 40 genes with CNV occurrence have copy number deletion (32/41), with a small number of genes having copy number gain (13/41). Meanwhile, in the poor group, most patients' genes with CNVs exhibited copy number deletion (Figure 2A). We identified that the CNV level of principal component genes in baseline blood could significantly distinguish the two groups of patients using PCA dimensionality reduction cluster analysis (Figure 2B). The CNVs of DPYP, IL7R, MTRR,
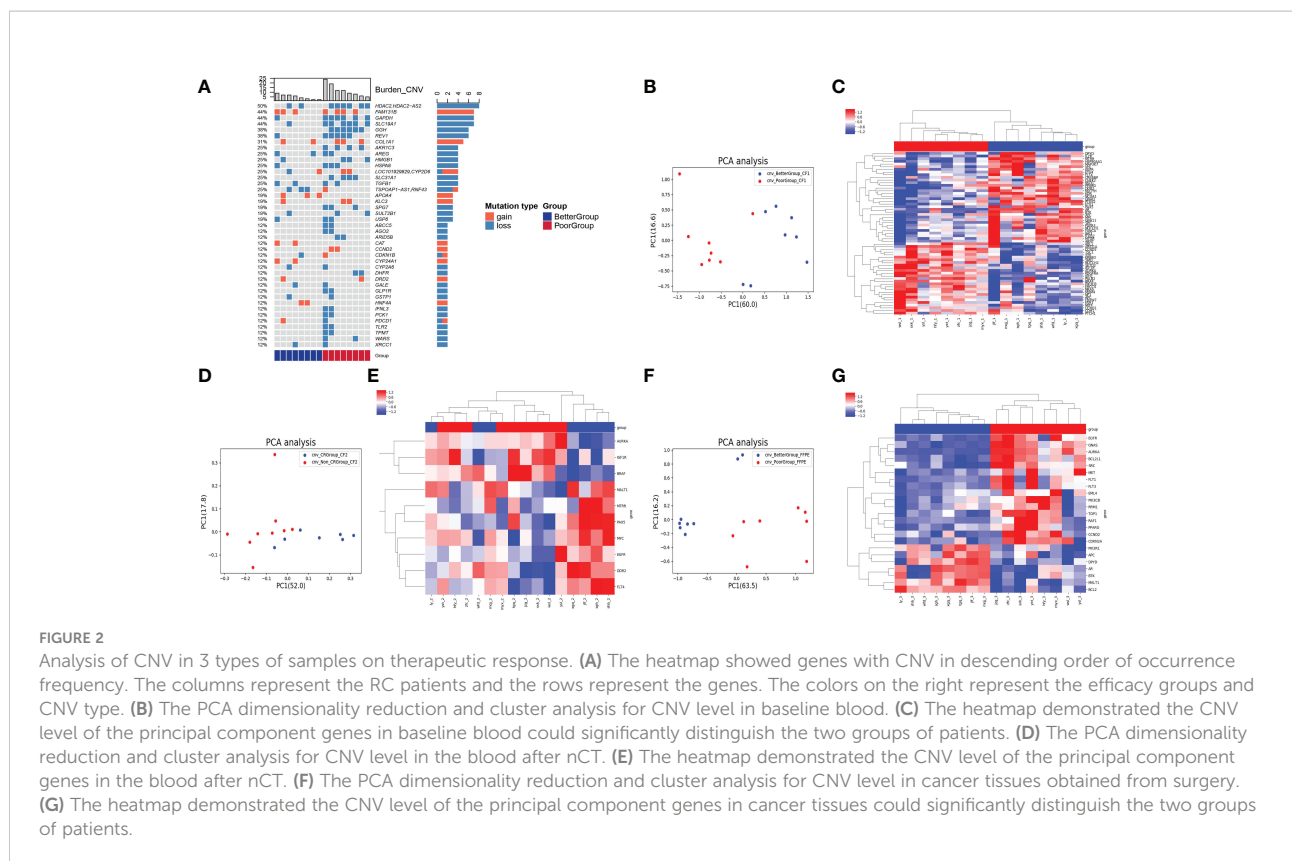
HSP90AA1, MAP3K7, VHL, EGFR, ETV4, FLT4, CREBBP, CHEK2, NSD1, PPARG, DDR2, MUTYH, KDR, NCOA1, GNAQ, PTGS2, ETV1, PAX5, AR, BTK, XPC, MPL, GNA11, SRC, FGGR3, NOTCH1, FANCA, JAK3, ITGB2, mTOR, and ARNT were relatively gained in the better group compared with the poor group. In contrast, the CNVs of ABL1, ARID1A, CCND2, TSC1, SYK, ERBB3, SMO, NOTCH2, SETD2, IGF1R, AURKA, PDGFRA, IDH1, PALB2, BRAF, PIK3CB, ERCC4, NRAS, MDM4, NF1, APC, FBXW7, ESR1, MITF, CCND1, CDH1, and PTCH1 were relatively gained in the poor group compared with the better group (Figure 2C).

However, after clustering by PCA dimensionality reduction in the blood samples after nCT (Figure 2D), the CNV level of the principal component genes could not significantly differentiate the two groups (Figure 2E). The PCA analysis of cancer tissues obtained from surgery (Figure 2F) revealed that the CNV levels of EGFR, GNAS, AURKA, BCL2L1, SRC, MET, FLT1, FLT3, EML4, PIK3CB, RRM1, TOP1, RAF1, PPARG, CCND2, CDKN2A, PIK3R1, APC, DPYD, AR, BTK, MALT1, and BCL2 could significantly differentiate the two groups of patients (Figure 2G).

Analysis of chemotherapy-related loci, CNI, MATH, and TMB in baseline blood

In addition to baseline blood samples, we examined other relevant indicators. The findings revealed that the commonly

FIGURE 1

Analysis of SNV in blood samples before and after nCT on therapeutic response. **(A)** The heatmap showed genes with SNV detected in the baseline blood according to the frequency of occurrence from high to low. The columns represent the RC patients and the rows represent the genes. The colors on the right represent the efficacy groups and mutation type. **(B)** The PCA dimensionality reduction and cluster analysis for SNP level in baseline blood. **(C)** The heatmap demonstrated the SNP level of the principal component genes in baseline blood. **(D)** The receiver operating characteristic curve for distinguishing the efficacy groups in baseline blood (AUC = 0.48). **(E)** The PCA dimensionality reduction and cluster analysis for SNP level in the blood after nCT. **(F)** The heatmap demonstrated the SNP level of the principal component genes in the blood after nCT. **(G)** The receiver operating characteristic curve for distinguishing the efficacy groups in blood after nCT (AUC = 0.62).

FIGURE 2

Analysis of CNV in 3 types of samples on therapeutic response. **(A)** The heatmap showed genes with CNV in descending order of occurrence frequency. The columns represent the RC patients and the rows represent the genes. The colors on the right represent the efficacy groups and CNV type. **(B)** The PCA dimensionality reduction and cluster analysis for CNV level in baseline blood. **(C)** The heatmap demonstrated the CNV level of the principal component genes in baseline blood could significantly distinguish the two groups of patients. **(D)** The PCA dimensionality reduction and cluster analysis for CNV level in the blood after nCT. **(E)** The heatmap demonstrated the CNV level of the principal component genes in the blood after nCT. **(F)** The PCA dimensionality reduction and cluster analysis for CNV level in cancer tissues obtained from surgery. **(G)** The heatmap demonstrated the CNV level of the principal component genes in cancer tissues could significantly distinguish the two groups of patients.

used chemotherapy target genotypes, including RCC1, ERCC2, MTHFR, XPC, and XRCC1, were not significantly different between the two groups of patients (all $p > 0.05$) (Figure 3A). CNI measures genomic instability related to regional chromosomal DNA ploidy (24). The calculation of the CNI value by bioinformatics analysis demonstrated that these values were significantly higher in the better group than in the poor group ($p = 0.0014$) (Figure 3B). The MATH value is a scoring method used to estimate tumor heterogeneity, and its higher value indicates more heterogeneity (33). MATH value calculation illustrates no significant difference between the two groups ($p = 0.23$) (Figure 3C). However, in the case of TMB, it was identified that TMB was significantly higher in the poor group than in the better group ($p = 0.013$) (Figure 3D).

Evaluation of the specific role of therapeutic relevant genes in the TCGA cohort

Given the purpose of our study was to identify the biomarkers and predict therapeutic response before nCT and surgery, we selected the results from the samples available before treatment, that is, the therapeutic relevant genes identified in the baseline blood for subsequent analysis.

To evaluate the specific role of these genes, we used PCA to constructed a gene signature in TCGA cohort, and defined the results as nCTPS. Next, we evaluated the value of the nCTPS in predicting prognosis. After obtaining the best cut-off value through R package "maxstat" (31), we distributed the patients from cohort into high and low nCTPS groups. We used PCA scatterplot to show the distribution of two groups of patients. It can be seen from the figure that the two groups of patients can be also separated by the PCA clustering (Figure 4A). We performed Gene Set Variation Analysis (GSVA) between the two groups. The results showed that enrichment of pathways varied significantly between the two groups, including ErbB signaling pathway, GnRH signaling pathway, mTOR signaling pathway, Wnt signaling pathway, etc. (all q-values < 0.05) (Figure 4B). There was a significant difference in prognosis between the two groups ($p < 0.001$) (Figure 4C). Furthermore, we calculated the MATH and TMB values in TCGA rectal cancer patients. We found that there was no significant difference in MATH and TMB between the high and low score groups (Figures 4D, E). There was a significant difference in the prognosis of patients in the high and low TMB groups (p = 0.009) (Figure 4F). Taking the synergistic effect of the TMB and nCTPS on the prognosis, we performed a stratified prognostic analysis. The results indicated that the nCTPS could be a potential prognostic indicator, and the combination with TMB could further refine prognostic prediction for patients (p = 0.003) (Figure 4G).
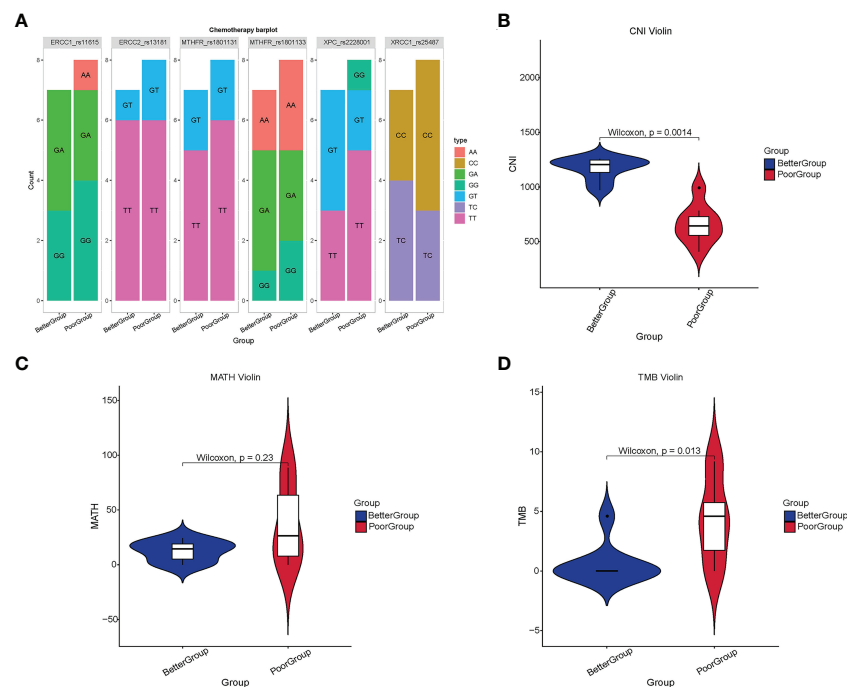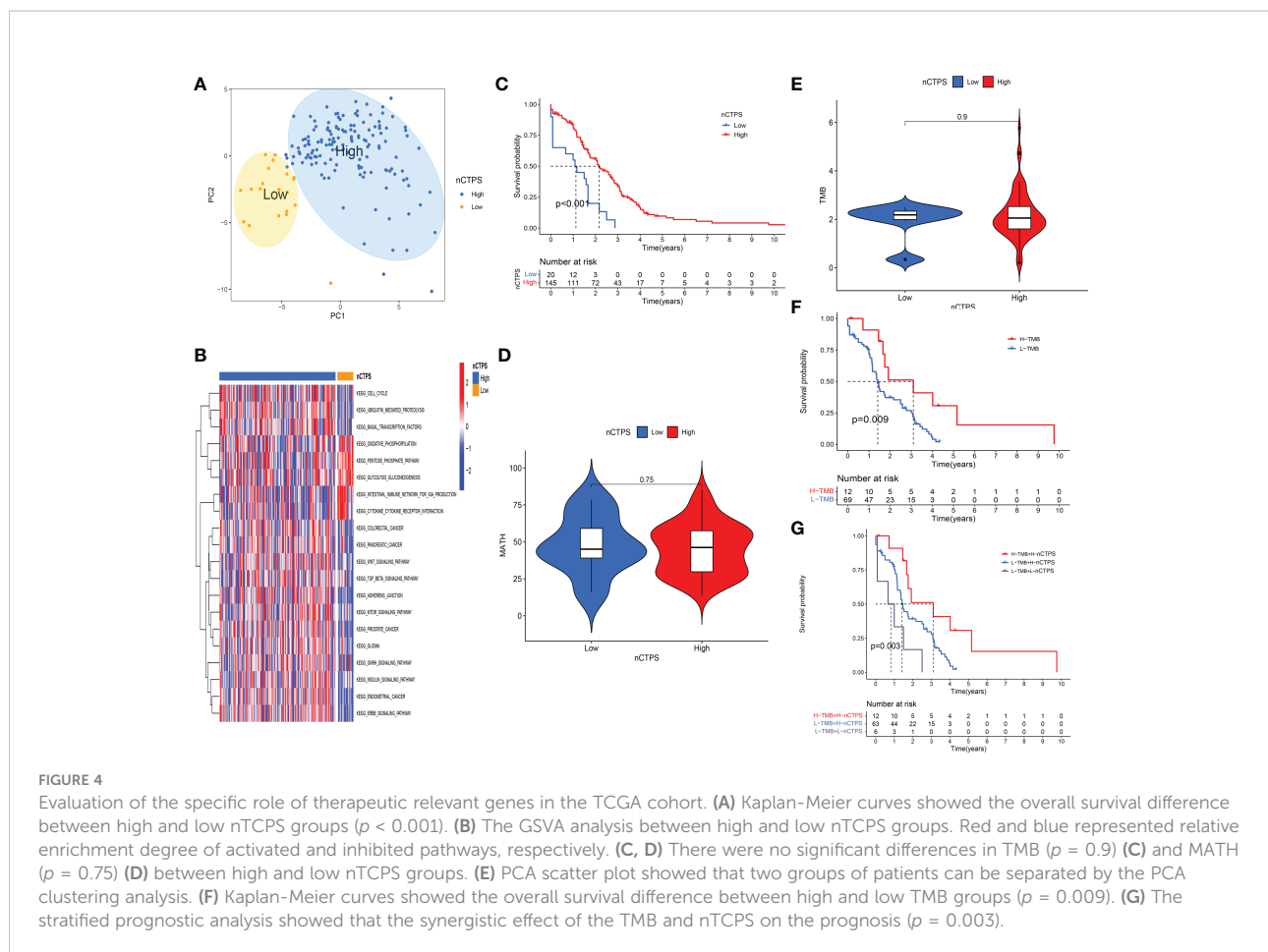
FIGURE 3

Analysis of chemotherapy-related loci, CNI, MATH, and TMB in different therapeutic response groups. **(A)** The commonly used target genotypes of chemotherapy in 16 RC patients. **(B)** The CNI value in the better group was significantly higher than that in the poor group ($p = 0.0014$) **(C)** There was no significant difference in the MATH value between the two groups ($p = 0.23$). **(D)** The TMB in the poor group was significantly higher than that in the better group ($p = 0.013$).

Transcript level combined with CNV to further identify genes associated with chemosensitivity

We first identified 5410 genes in the RC data from TCGA, whose CNVs differ significantly between cancer and normal tissues (all adjust $p < 0.05$). Fifteen of them overlapped with the genes that can differentiate sensitivity to nCT identified using the RC baseline blood sample sequencing results. Figure 5A depicts the chromosomal locations of these genes and CNV alterations, with most of these genes located in Chromosomes 1 and 7. The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of these genes demonstrated that they were significantly enriched in many energy metabolism pathways and classic cancer-related signaling pathways, including regulation of reactive oxygen species biosynthetic process, protein autophosphorylation, EGFR tyrosine kinase inhibitor resistance, ErbB signaling pathway, P53 signaling pathway, and mTOR signaling pathway (Figures 5B, C). The overlapped enriched pathways among GSVA analysis, GO and KEGG enrichment analysis include ErbB signaling pathway, GnRH signaling pathway, mTOR signaling pathway, pancreatic cancer, colorectal cancer, endometrial cancer, adherens junction signaling pathway, etc.

These overlapped signaling pathways are closely related to both tumor progression and drug resistance (34, 35). Therefore, we performed a chemosensitivity analysis on these genes. The findings revealed that the increased expression of genes with CNV levels relatively gained in the poor group was associated with decreased sensitivity to multiple chemotherapeutic drugs, including AURKA and BRAF (Supplementary Figure 1A). However, in the better group, the increased expression of genes with CNV relatively gained was associated with increased sensitivity to multiple chemotherapeutic drugs, including EGFR, SRC, and HSP90AA1 (Supplementary Figure 1B).

Subsequently, we extracted the expression of RNAs with significantly different CNVs in normal and cancer tissues, analyzed the data, and obtained 2643 genes with significant differences in expression levels at different CNV levels. Then, using the sequencing results of RC blood samples, we intersected these genes with the genes that can distinguish the sensitivity to nCT and obtained eight intersecting genes (Figure 5D). The CNV levels of HSP90AA1, EGFR, SRC, and mTOR were higher in the better group compared to the poor group. In contrast, the CNV levels of ARID1A, AURKA, BRAF, and MDM4 were higher in the poor group compared to the better group. The Supplementary Figures 2A–H represents the changes in RNA expression of these eight genes at different CNV levels (all $p < 0.05$).

FIGURE 4

Evaluation of the specific role of therapeutic relevant genes in the TCGA cohort. **(A)** Kaplan-Meier curves showed the overall survival difference between high and low nTCPS groups ($p < 0.001$). **(B)** The GSVA analysis between high and low nTCPS groups. Red and blue represented relative enrichment degree of activated and inhibited pathways, respectively. **(C, D)** There were no significant differences in TMB ($p = 0.9$) **(C)** and MATH ($p = 0.75$) **(D)** between high and low nTCPS groups. **(E)** PCA scatter plot showed that two groups of patients can be separated by the PCA clustering analysis. **(F)** Kaplan-Meier curves showed the overall survival difference between high and low TMB groups ($p = 0.009$). **(G)** The stratified prognostic analysis showed that the synergistic effect of the TMB and nTCPS on the prognosis ($p = 0.003$).
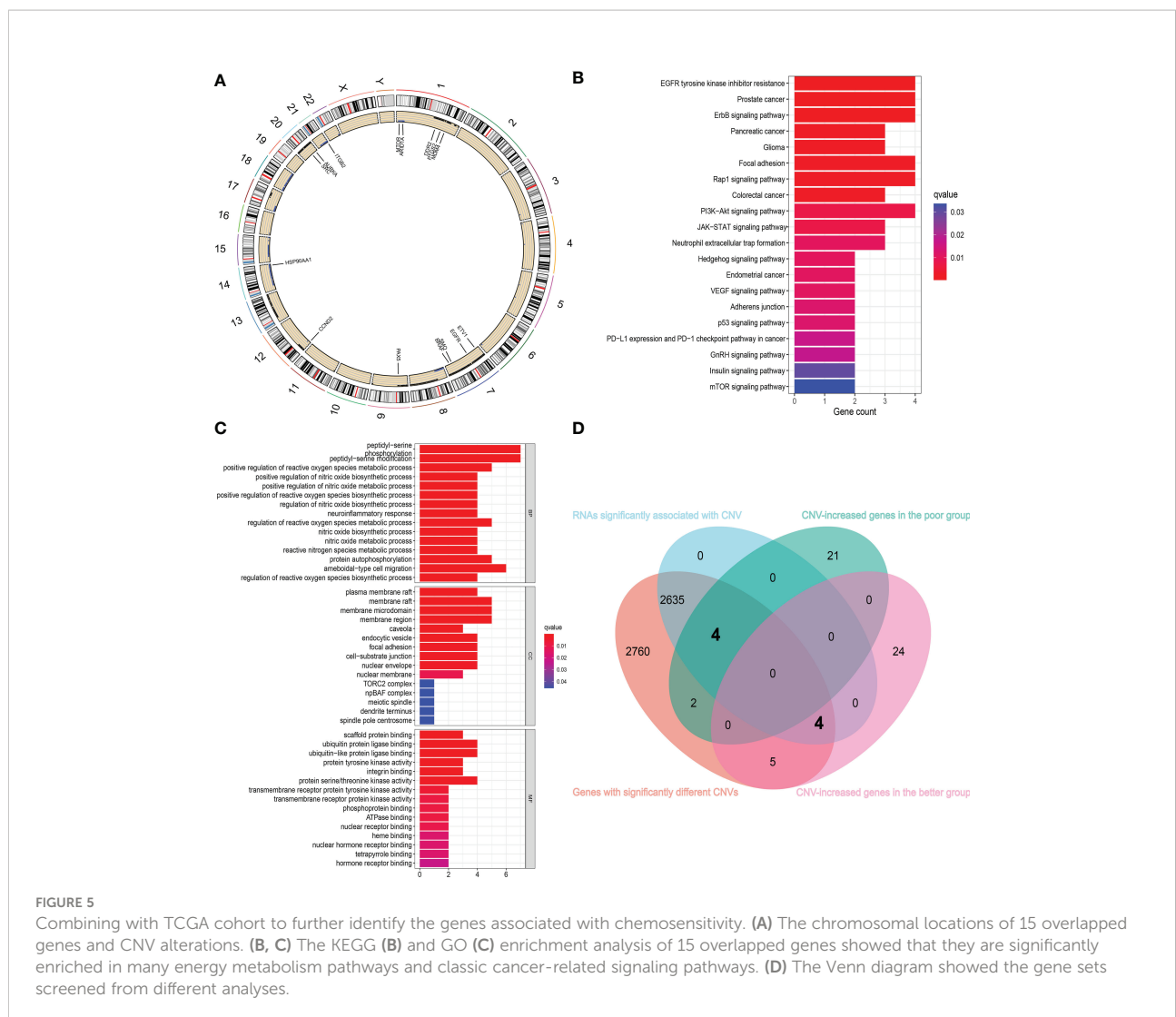
Transcript level combined with protein level to finalize the key genes

A publicly proteomic database was used to analyze expression differences, prognosis, and protein-transcript level correlations for identifying and validating key genes. Because of the scarcity of publicly available RC proteomic data and lack of data on normal tissues, colon cancer (CC) and RC are inseparable from gastrointestinal cancers, with close correlations in various aspects (36). Therefore, we included CC proteomic data in our analysis. At the transcriptional level of CC, EGFR (Figure 6A), HSP90AA1 (Figure 6B), and SRC (Figure 6D) expression levels were significantly different between normal and cancer tissues (all $p < 0.05$). However, there was no significant difference between normal and cancer tissues in MTOR (p = 0.493) (Figure 6C). At the protein level, the expression levels of EGFR (Figure 6E, Supplementary Figures 3A–F), HSP90AA1 (Figure 6F, Supplementary Figures 3G–L), and mTOR (Figure 6G) were significantly different between normal and cancer tissues (all $p < 0.05$). There was no significant difference between normal and cancer tissues in SRC (p = 0.344) (Figure 6H). The expression levels of EGFR (Figure 6I), HSP90AA1 (Figure 6J), and SRC

(Figure 6L) at the transcriptional level of RC were significantly different between normal and cancer tissues (all $p < 0.05$). However, there was no significant difference between normal and cancer tissues in MTOR (p = 0.599) (Figure 6K). Furthermore, the RNA expression of SRC was positively correlated with the protein expression (r = 0.67, $p < 0.001$) (Figure 6P), and the RNA expression of HSP90AA1 was positively correlated with the protein expression, and the statistical significance approached significant (r = 0.35, $p = 0.06$) (Figure 6N). Nevertheless, there was no significant correlations between RNA and protein level in EGFR (Figure 6M) and MTOR (Figure 6O).

In terms of prognosis, at the transcriptional level of RC, the high expression group of EGFR (Figure 7A) and SRC (Figure 7D) had a significantly better prognosis than the low expression group (all $p < 0.05$). The patients with high HSP90AA1 expression seem to have a better prognosis than those with low HSP90AA1 expression, this difference approached statistical significance ($p = 0.051$) (Figure 7B). In contrast to the above results, the high expression group of MTOR had a significantly poor prognosis than the low expression group (p = 0.001) (Figure 7C). However, there was

FIGURE 5

Combining with TCGA cohort to further identify the genes associated with chemosensitivity. **(A)** The chromosomal locations of 15 overlapped genes and CNV alterations. **(B, C)** The KEGG **(B)** and GO **(C)** enrichment analysis of 15 overlapped genes showed that they are significantly enriched in many energy metabolism pathways and classic cancer-related signaling pathways. **(D)** The Venn diagram showed the gene sets screened from different analyses.

no significant difference in prognosis between the high and low expression groups of these four genes at the protein level (all $p > 0.05$) (Figures 7E–7H). The lack of positive results could be attributed to the small sample size. The prognosis of the high expression group of EGFR (Figure 7I), HSP90AA1 (Figure 7J), and mTOR (Figure 7K) at the protein level of CC was significantly better than that of the low expression group (all $p < 0.05$). However, there were no significant difference in survival between high and low SRC expression group (p = 0.16) (Figure 7L).

Altogether, the genes with significant differences in multi-omics and multiple aspects were EGFR and HSP90AA1. Therefore, they were identified as the key predictive genes related to prognosis and the sensitivity of nCT. DisNor database (32) revealed the EGFR and HSP90AA1 up and downstream binding sites and their causal interaction (Figure 8). The up-regulated binding sites of EGFR included PI3K, PI3K3R1, TGFA, SHC1, and SHC3. In contrast, LRIG1

and ERRFI1 were the EGFR down-regulated binding sites. AHSA1 and PTGES3 were the up-regulated, while STIP1 and FNIP1 were the down-regulated binding sites of HSP90AA1, respectively. These findings are critical for understanding the mechanisms and signaling pathways of the key genes involved in drug resistance and identifying potential therapeutic targets.

Discussion

Researchers and clinicians extensively investigate locally advanced RC (LARC) because of its high morbidity and distant metastasis rate (37). Researchers have proposed a new treatment, nCT, to improve the prognosis of such patients. CAO/ARO/AIO-94, CAO/ARO/AIO-04, MRC CR07, and NCIC-CTG C016 all confirmed that nCT could significantly improve the surgical resection rate of LACR, decrease the local recurrence rate, and reduce the occurrence of adverse reactions (38–40). However,
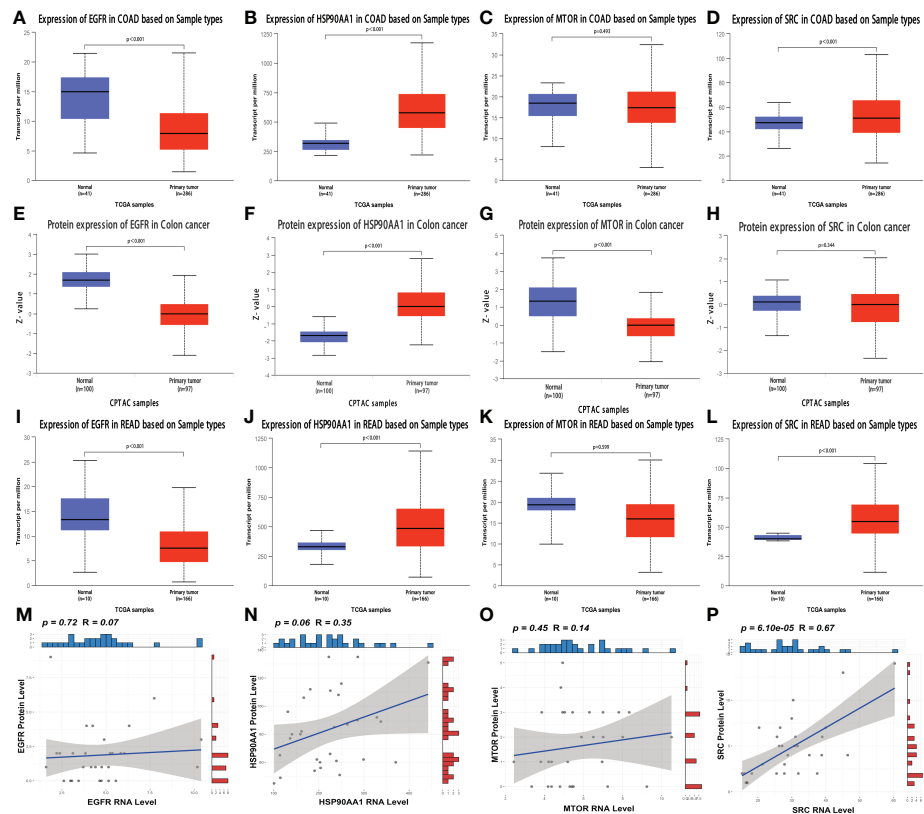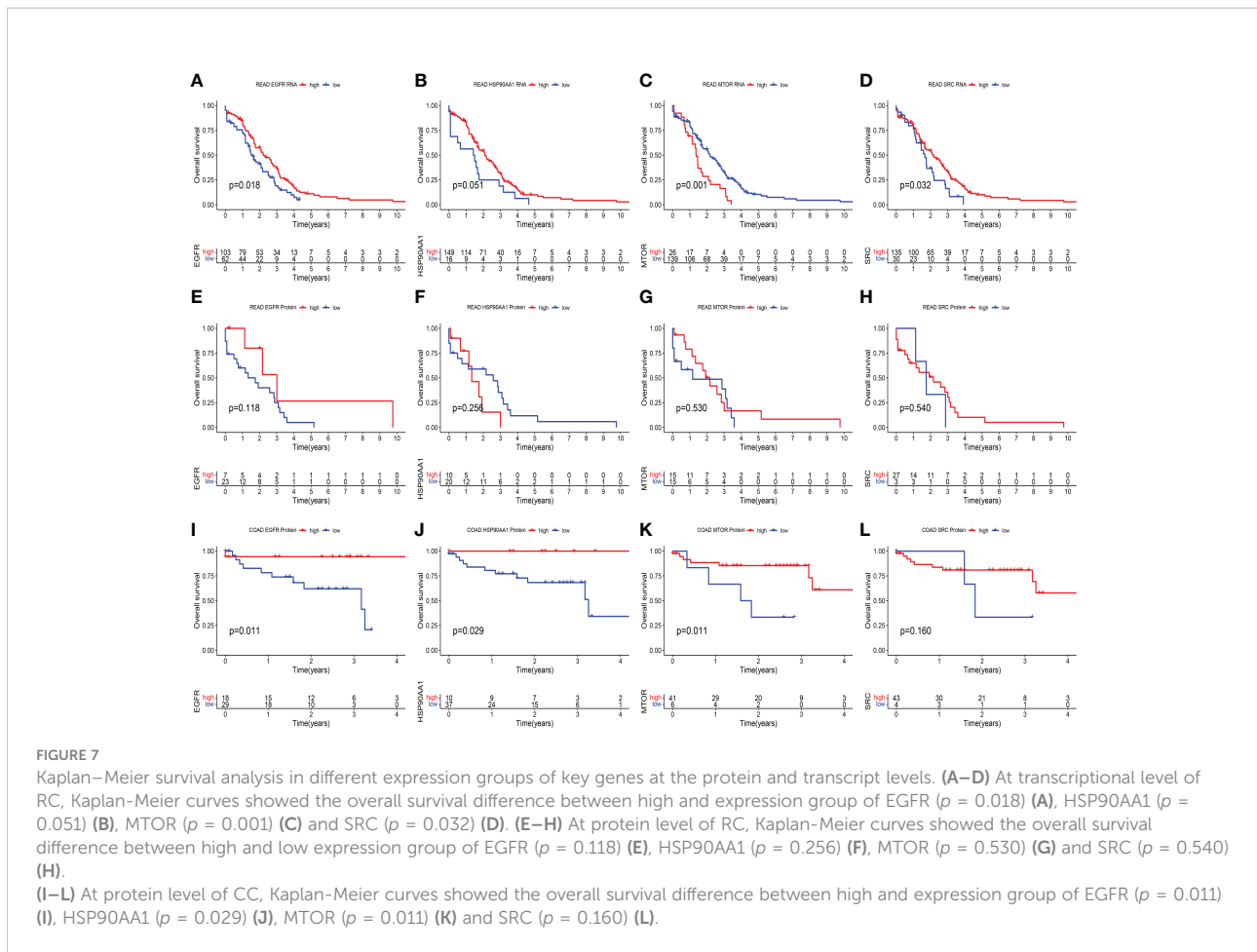
FIGURE 6

Analysis of the expression and correlation of key genes at the protein and transcript levels **(A–D)** The transcriptional expression of EGFR **(A)**, HSP90AA1 **(B)**, MTOR **(C)** and SRC **(D)** in normal and CC tissues. **(E–H)** The protein level of EGFR **(E)**, HSP90AA1 **(F)**, MTOR **(G)** and SRC **(H)** in normal and CC tissues. **(I–L)** The transcriptional expression of EGFR **(I)**, HSP90AA1 **(J)**, MTOR **(K)** and SRC **(L)** in normal and RC tissues. **(M–P)** The correlation analyses between RNA expression and protein level of EGFR **(M)**, HSP90AA1 **(N)**, MTOR **(O)** and SRC **(P)** in RC patients.

numerous studies have found significant differences in patients' ability to respond to nCT. Approximately 50% of patients had a PR, 8–20% achieved pathological CR, and about 20% demonstrated resistance to nCT (41–45).

The response heterogeneity to nCT has resulted in overtreatment, increased cytotoxicity, or economic pressure and may even result in disease progression due to delayed radical surgery. Finding high-performance markers to distinguish which patients can benefit from nCT is thus an important means of promoting the standardization of LACR treatment. Recent studies have identified potential markers, such as DNA mutation, DNA methylation, circulating tumor cells, tumor immune microenvironment, and microRNA (46). However, these markers cannot be widely used due to sensitivity and specificity issues. Therefore, presently there is a need to identify new markers. We identified markers in blood samples that could predict patients' response to nCT and proved that these markers could be used at DNA, transcription and protein levels. This multi-omics-validated marker has high reliability and stability, convenient sampling, and is not

constrained by detection technology. They are highly reliable biomarkers that need further thorough investigations.

At the DNA level, we identified not only the genes associated with nCT response rates but also that CNV, CNI score, and TMB were potential markers, whereas SNP and MATH were not. These findings implied that CNV, CNI, and TMB can be used to predict the patient's ability to respond to nCT (4, 47, 48). In this study, CNI score was used to assess the extent of CIN. Although CIN is ubiquitous in human cancers, its role in tumor evolution is complex and contradictory (49). On the one hand, CIN and complex aneuploidy are associated with resistance to anticancer drugs, such as paclitaxel, in tumor-derived cell lines and clinical settings (50, 51). Conversely, high CIN levels indicate enhanced sensitivity to cytotoxic therapies such as 5-fluorouracil and cisplatin in rectal (52), breast (53) and ovarian cancers (51). Induction of whole chromosome missegregation makes transplanted glioblastoma tumor sensitive to radiotherapy (54). In our analysis, the CNI value was significantly higher in the better group than in the poor group ($p = 0.0014$), which was consistent with previous research. This drastically different effect

FIGURE 7

Kaplan−Meier survival analysis in different expression groups of key genes at the protein and transcript levels. **(A−D)** At transcriptional level of RC, Kaplan-Meier curves showed the overall survival difference between high and expression group of EGFR (*p* = 0.018) **(A)**, HSP90AA1 (*p* = 0.051) **(B)**, MTOR (*p* = 0.001) **(C)** and SRC (*p* = 0.032) **(D)**. **(E−H)** At protein level of RC, Kaplan-Meier curves showed the overall survival difference between high and low expression group of EGFR (*p* = 0.118) **(E)**, HSP90AA1 (*p* = 0.256) **(F)**, MTOR (*p* = 0.530) **(G)** and SRC (*p* = 0.540) **(H)**.
(I−L) At protein level of CC, Kaplan-Meier curves showed the overall survival difference between high and expression group of EGFR (*p* = 0.011) **(I)**, HSP90AA1 (*p* = 0.029) **(J)**, MTOR (*p* = 0.011) **(K)** and SRC (*p* = 0.160) **(L)**.

stems from the complex phenotypes conferred by CIN on cancer cells and the tumor microenvironment (49). CIN can serve as a genomic source of innate immune activation. For example, chromosome segregation errors can directly lead to the activation of immune signaling pathways (55). In addition, CIN can also act as a trigger of tumor immune editing.

FIGURE 8

The up- and downstream binding sites and causal interaction of the key genes' analysis in DisNor.

Chromosomal segregation errors caused by CIN at the early stages of tumorigenesis activate the cGAS-STING pathway, which functions as an innate cellular defense against viral infection (56).

We then perform further analysis at transcript and protein levels. We confirmed that EGFR and HSP90AA1 were the genes with significant differences in multi-omics and multiple aspects. SRC and mTOR did not have this relationship and only differed in a single omics. This demonstrated that EGFR and HSP90AA1 are multi-omics-validated markers that can predict patient response to nCT and require further investigation.

HSP90AA1 is a stress-inducible member of the HSP90 family. It can regulate the tumor-promoting process of many proto-oncogenes, including c-Myc, and is also associated with the malignant tumor phenotype, tumor growth, proliferation, invasion, and chemotherapy resistance (57–59). However, no studies have suggested that HSP90AA1 can be used as a prognostic or predictive marker for nCT. The present study is the first to describe the potential of this molecule as a predictive marker for nCT, which requires further research. EGFR is an important member of the erbB family that significantly regulate cell proliferation, differentiation, division, survival, and cancer development.
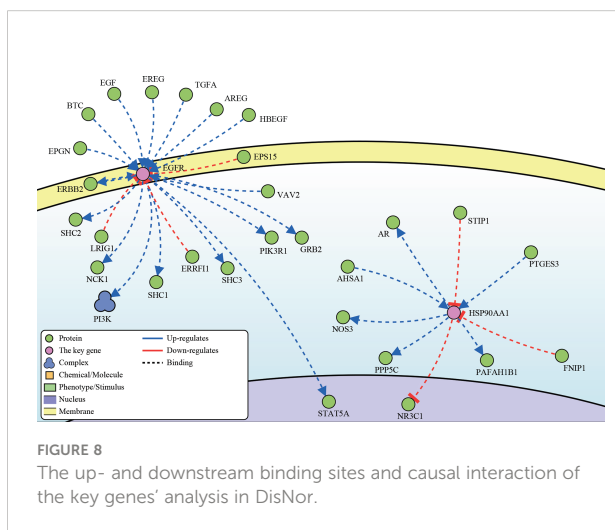
Therefore, EGFR is an important target for targeted therapy (60). Recent studies established that EGFR can be used as a prognostic and predictive molecular marker for nCT in various cancers (61–63). A study about RC demonstrated that circulating EGFR could be used as a potential biomarker for predicting pCR (61). These findings are consistent with our results, which indicated that EGFR could be used as an important target and a sensitive marker to guide RC treatment. SRC and SRC-family protein kinases are proto-oncogenes involved in cell morphology, motility, proliferation, and survival (64). SRC has recently been associated with the prognosis and recurrence of various tumors (65–67). Our findings demonstrated that SRC is not only associated with tumor prognosis but can be used to predict the patient's ability to respond to nCT. mTOR is a serine/threonine kinase regulating various cellular metabolic processes such as protein synthesis and inactivation. mTOR can activate somatic mutations during tumorigenesis, making it an important therapeutic target (68). Some studies have illustrated that components of mTOR-related pathways can be used to predict tumor response to nCT (69, 70). Zhu et al. described in RC that GOLPH3 could predict patient sensitivity to nCT, which plays an important role in mTOR-related pathways (69). However, these studies have not directly demonstrated that mTOR could predict patient sensitivity to nCT. It is worth noting that the present study shows this for the first time.

To pave the way for future studies, we conducted a survival analysis. We described that the above markers could be used as prognostic markers for nCT in CRC, confirming the significance of these markers. We also investigated the signaling pathways of the above markers and found that they can function in multiple signaling pathways. EGFR was enriched in CRC, EGFR tyrosine kinase inhibitor resistance, the FoxO signaling pathway, the Oxytocin signaling pathway, the JAK-STAT signaling pathway, the Pl3K-Akt signaling pathway, and hedgehog signaling pathway. These pathways are of significant value in the occurrence and development of tumors, cancer stem cells, immune response, drug sensitivity, and drug resistance (71–75). Therefore, in-depth research on the markers mentioned above will help explain tumor pathogenesis further and improve patients' prognosis.

However, several limitations remain as follows: Due to the small sample size of proteomic sequencing for RC in TCGA and some missing expression data of individual proteins, fewer analysis can be done and reliable prognostic analysis could not be performed. So, we only use proteomic data to further screen and validate the key genes.

In conclusion, the biomarkers we identified in multi-omics analysis were associated with sensitivity of nCT and prognosis. They require further prospective investigation and are expected to promote nCT standardization and personalization.

Data availability statement

Publicly available datasets were analyzed in this study. These datasets can be found here: TCGA database (http://cancergenome.nih.gov/), DISNOR (https://disnor.uniroma2.it/) (32), The CPTAC database (https://proteomics.cancer.gov/programs/cptac), The UALCAN web resource (http://ualcan.path.uab.edu/). The sequencing data were deposited in the The Genome Sequence Archive for Human, accession number HRA002714 (http://bigd.big.ac.cn/gsa-human).

Ethics statement

The studies involving human participants were reviewed and approved by The Ethical Committees of Harbin Medical University Cancer Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

B-BC and Y-LL designed the study. X-FJ, B-MZ, F-QD and J-NG drafted the manuscript, collected, analyzed, and interpreted the data. DW, Y-EL and S-HD drew the figures. B-BC and Y-LL helped with the final revision of the article. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgment

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/fimmu.2022.1013828/full#supplementary-material

SUPPLEMENTARY FIGURE 1

Correlation analysis in the expression of therapeutic relevant genes with sensitivity of chemotherapy drugs. **(A)** The increased expression of genes with CNV relatively gained in the poor group was associated with decreased sensitivity to multiple chemotherapeutic drugs, including AURKA and BRAF. **(B)** In the better group, increased expression of genes with CNV relatively gained was associated with increased sensitivity to multiple chemotherapeutic drugs, including EGFR, SRC, and HSP90AA1.

SUPPLEMENTARY FIGURE 2

The changes of RNA expression of the 8 genes identified from the Venn diagram at different CNV levels, including HSP90AA1 **(A)**, MDM4 **(B)**, MTOR **(C)**, SRC **(D)**, ARID1A **(E)**, AURKA **(F)**, BRAF **(G)** and EGFR **(H)**.

SUPPLEMENTARY FIGURE 3

Immunohistochemical staining for the key genes EGFR and HSP90AA1 in normal tissues, READ tissues and COAD tissues. (Image credit: Human Protein Atlas, images available from v20.1.proteinatlas.org).

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* (2022) 72(1):7–33. doi: 10.3322/caac.21708

2. Cercek A, Roxburgh CSD, Strombom P, Smith JJ, Temple LKF, Nash GM, et al. Adoption of total neoadjuvant therapy for locally advanced rectal cancer. *JAMA Oncol* (2018) 4(6):e180071. doi: 10.1001/jamaoncol.2018.0071

3. Deng Y, Chi P, Lan P, Wang L, Chen W, Cui L, et al. Neoadjuvant modified FOLFOX6 with or without radiation versus fluorouracil plus radiation for locally advanced rectal cancer: Final results of the Chinese FOWARC trial. *J Clin Oncol* (2019) 37(34):3223–33. doi: 10.1200/JCO.18.02309

4. Frydrych LM, Ulintz P, Bankhead A, Sifuentes C, Greenson J, Maguire L, et al. Rectal cancer sub-clones respond differentially to neoadjuvant therapy. *Neoplasia* (2019) 21(10):1051–62. doi: 10.1016/j.neo.2019.08.004

5. Kato S, Kim KH, Lim HJ, Boichard A, Nikanjam M, Weihe E, et al. Real-world data from a molecular tumor board demonstrates improved outcomes with a precision n-of-One strategy. *Nat Commun* (2020) 11(1):4965. doi: 10.1038/s41467-020-18613-3

6. Ivanov M, Ivanov M, Kasianov A, Rozhavskaya E, Musienko S, Baranova A, et al. Novel bioinformatics quality control metric for next-generation sequencing experiments in the clinical context. *Nucleic Acids Res* (2019) 47(21):e135. doi: 10.1093/nar/gkz775

7. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* (2011) 11(6):426–37. doi: 10.1038/nrc3066

8. Cheng J, Holland-Letz T, Wallwiener M, Surowy H, Cuk K, Schott S, et al. Circulating free DNA integrity and concentration as independent prognostic markers in metastatic breast cancer. *Breast Cancer Res Treat* (2018) 169(1):69–82. doi: 10.1007/s10549-018-4666-5

9. Chen X, Gole J, Gore A, He Q, Lu M , Min J, et al. Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nat Commun* (2020) 11(1): 3475. doi: 10.1038/s41467-020-17316-z

10. Guo JN, Chen D, Deng SH, Huang JR, Song JX, Li XY, et al. Identification and quantification of immune infiltration landscape on therapy and prognosis in left- and right-sided colon cancer. *Cancer Immunol Immunother* (2022) 71(6):1313–30. doi: 10.1007/s00262-021-03076-2

11. Xia T, Guo J, Zhang B, Xue W, Deng S, Liu Y, et al. A novel quantification system combining iTRAQ technology and multi-omics assessment to predict prognosis and immunotherapy efficacy in colon cancer. *Front Bioeng Biotechnol* (2022) 10:862619. doi: 10.3389/fbioe.2022.862619

12. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* (2009) 10:451–81. doi: 10.1146/annurev.genom.9.081307.164217

13. Vollger MR, Guitart X, Dishuck PC, Mercuri L, Harvey WT, Gershman A, et al. Segmental duplications and their variation in a complete human genome. *Science* (2022) 376(6588):eabj6965. doi: 10.1126/science.abj6965

14. Jang S, Su H, Blum FC, Bae S, Choi YH, Kim A, et al. Dynamic expansion and contraction of cagA copy number in helicobacter pylori impact development of gastric disease. *mBio* (2017) 8(1):e01779-16. doi: 10.1128/mBio.01779-16

15. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv e-prints* (2013):1303.3997v2. doi: 10.48550/arXiv.1303.3997

16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009) 25(16):2078–9. doi: 10.1093/bioinformatics/btp352

17. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* (2012) 22(3):568–76. doi: 10.1101/gr.129684.111

18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* (2010) 38(16):e164. doi: 10.1093/nar/gkq603

19. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res* (2017) 45(D1):D840–5. doi: 10.1093/nar/gkw971

20. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* (2020) 581(7809):434–43. doi: 10.1038/s41586-020-2308-7

21. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-wide copy number detection and visualization from targeted DNA sequencing. *PloS Comput Biol* (2016) 12(4):e1004873. doi: 10.1371/journal.pcbi.1004873

22. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* (2016) 32(18):2847–9. doi: 10.1093/bioinformatics/btw313

23. Zhang H, Meltzer P, Davis S. RCircos: an r package for circos 2D track plots. *BMC Bioinf* (2013) 14:244. doi: 10.1186/1471-2105-14-244

24. Chudasama D, Katopodis P, Stone N, Haskell J, Sheridan H, Gardner B, et al. Liquid biopsies in lung cancer: Four emerging technologies and potential clinical applications. *Cancers (Basel)* (2019) 11(3):331. doi: 10.3390/cancers11030331

25. Weiss GJ, Beck J, Braun DP, Bornemann-Kolatzki K, Barilla H, Cubello R, et al. Tumor cell-free DNA copy number instability predicts therapeutic response to immunotherapy. *Clin Cancer Res* (2017) 23(17):5074–81. doi: 10.1158/1078-0432.CCR-17-0231

26. Filloux C, Cedric M, Romain P, Lionel F, Christophe K, Dominique R, et al. An integrative method to normalize RNA-seq data. *BMC Bioinf* (2014) 15:188. doi: 10.1186/1471-2105-15-188

27. Mroz EA, Tward AD, Hammon RJ, Ren Y, Rocco JW. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the cancer genome atlas. *PloS Med* (2015) 12(2):e1001786. doi: 10.1371/journal.pmed.1001786

28. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med* (2017) 9(1):34. doi: 10.1186/s13073-017-0424-2

29. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7

30. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. tissue-based map of the human proteome. *Science* (2015) 347 (6220):1260419. doi: 10.1126/science.1260419

31. Laska E, Meisner M, Wanderling J. A maximally selected test of symmetry about zero. *Stat Med* (2012) 31(26):3178–91. doi: 10.1002/sim.5384

32. Lo Surdo P, Calderone A, Iannuccelli M, Licata L, Peluso D, Castagnoli L, et al. DISNOR: a disease network open resource. *Nucleic Acids Res* (2018) 46(D1): D527–34. doi: 10.1093/nar/gkx876

33. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol* (2013) 49(3):211–5. doi: 10.1016/j.oraloncology.2012.09.007

34. Wang Z. ErbB receptors and cancer. *Methods Mol Biol* (2017) 1652:3–35. doi: 10.1007/978-1-4939-7219-7_1

35. Florio S, Pagnini U, Crispino A, Pacilio C, Crispino L, Giordano A. GnRH and steroids in cancer. *Front Biosci* (2002) 7:d1590–1608. doi: 10.2741/A864

36. Viana Veloso GG, Franco OH, Ruiter R, de Keyser CE, Hofman A, Stricker BC, et al. Baseline dietary glutamic acid intake and the risk of colorectal cancer: The Rotterdam study. *Cancer* (2016) 122(6):899–907. doi: 10.1002/cncr.29862

37. Oronsky B, Reid T, Larson C, Knox SJ. Locally advanced rectal cancer: The past, present, and future. *Semin Oncol* (2020) 47(1):85–92. doi: 10.1053/j.seminoncol.2020.02.001

38. Sauer R, Liersch T, Merkel S, Fietkau R, Hohenberger W, Hess C, et al. Preoperative versus postoperative chemoradiotherapy for locally advanced rectal cancer: results of the German CAO/ARO/AIO-94 randomized phase III trial after a median follow-up of 11 years. *J Clin Oncol* (2012) 30(16):1926–33. doi: 10.1200/jco.2011.40.1836

39. Rödel C, Graeven U, Fietkau R, Hohenberger W, Hothorn T, Arnold D, et al. Oxaliplatin added to fluorouracil-based preoperative chemoradiotherapy and postoperative chemotherapy of locally advanced rectal cancer (the German CAO/ARO/AIO-04 study): final results of the multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol* (2015) 16(8):979–89. doi: 10.1016/s1470-2045(15)00159-x

40. Sebag-Montefiore D, Stephens RJ, Steele R, Monson J, Grieve R, Khanna S, et al. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (MRC CR07 and NCIC-CTG C016): a multicentre, randomised trial. *Lancet* (2009) 373(9666):811–20. doi: 10.1016/s0140-6736(09)60484-0

41. Yang YJ, Cao L, Li ZW, Zhao L, Wu HF, Yue D, et al. Fluorouracil-based neoadjuvant chemoradiotherapy with or without oxaliplatin for treatment of locally advanced rectal cancer: An updated systematic review and meta-analysis. *Oncotarget* (2016) 7(29):45513–24. doi: 10.18632/oncotarget.9995

42. Park IJ, You YN, Agarwal A, Skibber JM, Rodriguez-Bigas MA, Eng C, et al. Neoadjuvant treatment response as an early response indicator for patients with rectal cancer. *J Clin Oncol* (2012) 30(15):1770–6. doi: 10.1200/jco.2011.39.7901

43. Silberfein EJ, Kattepogu KM, Hu CY, Skibber JM, Rodriguez-Bigas MA, Feig B, et al. Long-term survival and recurrence outcomes following surgery for distal rectal cancer. *Ann Surg Oncol* (2010) 17(11):2863–9. doi: 10.1245/s10434-010-1119-8

44. Smith KD, Tan D, Das P, Chang GJ, Kattepogu K, Feig BW, et al. Clinical significance of acellular mucin in rectal adenocarcinoma patients with a pathologic complete response to preoperative chemoradiation. *Ann Surg* (2010) 251(2):261–4. doi: 10.1097/SLA.0b013e3181bdfc27

45. Ryan JE, Warrier SK, Lynch AC, Heriot AG. Assessing pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a systematic review. *Colorectal Dis* (2015) 17(10):849–61. doi: 10.1111/codi.13081

46. Dayde D, Tanaka I, Jain R, Tai MC, Taguchi A. Predictive and prognostic molecular biomarkers for response to neoadjuvant chemoradiation in rectal cancer. *Int J Mol Sci* (2017) 18(3):573. doi: 10.3390/ijms18030573

47. Greenbaum A, Martin DR, Bocklage T, Lee JH, Ness SA, Rajput A. Tumor heterogeneity as a predictor of response to neoadjuvant chemotherapy in locally advanced rectal cancer. *Clin Colorectal Cancer* (2019) 18(2):102–9. doi: 10.1016/j.clcc.2019.02.003

48. Ji D, Zhang D, Zhan T, Jia J, Han W, Li Z, et al. Tumor mutation burden in blood predicts benefit from neoadjuvant chemo/radiotherapy in locally advanced rectal cancer. *Genomics* (2021) 113(1 Pt 2):957–66. doi: 10.1016/j.ygeno.2020.10.029

49. Bakhoum SF, Cantley LC. The multifaceted role of chromosomal instability in cancer and its microenvironment. *Cell* (2018) 174(6):1347–60. doi: 10.1016/j.cell.2018.08.027

50. Bakhoum SF, Danilova OV, Kaur P, Levy NB, Compton DA. Chromosomal instability substantiates poor prognosis in patients with diffuse large b-cell lymphoma. *Clin Cancer Res* (2011) 17(24):7704–11. doi: 10.1158/1078-0432.CCR-11-2049

51. Swanton C, Nicke B, Schuett M, Eklund AC, Ng C, Li Q, et al. Chromosomal instability determines taxane response. *Proc Natl Acad Sci U.S.A.* (2009) 106 (21):8671–6. doi: 10.1073/pnas.0811835106

52. Zaki BI, Suriawinata AA, Eastman AR, Garner KM, Bakhoum SF. Chromosomal instability portends superior response of rectal adenocarcinoma to chemoradiation therapy. *Cancer* (2014) 120(11):1733–42. doi: 10.1002/cncr.28656

53. Jamal-Hanjani M, A'Hern R, Birkbak NJ, Gorman P, Gronroos E, Ngang S, et al. Extreme chromosomal instability forecasts improved outcome in ER-negative breast cancer: a prospective validation cohort study from the TACT trial. *Ann Oncol* (2015) 26(7):1340–6. doi: 10.1093/annonc/mdv178

54. Bakhoum SF, Kabeche L, Wood MD, Laucius CD, Qu D, Laughney AM, et al. Numerical chromosomal instability mediates susceptibility to radiation treatment. *Nat Commun* (2015) 6:5990. doi: 10.1038/ncomms6990

55. Bakhoum SF, Ngo B, Laughney AM, Cavallo JA, Murphy CJ, Ly P, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* (2018) 553(7689):467–72. doi: 10.1038/nature25432

56. Santaguida S, Richardson A, Iyer DR, M'Saad O, Zasadil L, Knouse KA, et al. Chromosome mis-segregation generates cell-Cycle-Arrested cells with complex karyotypes that are eliminated by the immune system. *Dev Cell* (2017) 41638–651 (6):e635. doi: 10.1016/j.devcel.2017.05.022

57. Zuehlke AD, Beebe K, Neckers L, Prince T. Regulation and function of the human HSP90AA1 gene. *Gene* (2015) 570(1):8–16. doi: 10.1016/j.gene.2015.06.018

58. Chu SH, Liu YW, Zhang L, Liu B, Li L, Shi JZ, et al. Regulation of survival and chemoresistance by HSP90AA1 in ovarian cancer SKOV3 cells. *Mol Biol Rep* (2013) 40(1):1–6. doi: 10.1007/s11033-012-1930-3

59. Zhang M, Peng Y, Yang Z, Zhang H, Xu C, Liu L, et al. DAB2IP down-regulates HSP90AA1 to inhibit the malignant biological behaviors of colorectal cancer. *BMC Cancer* (2022) 22(1):561. doi: 10.1186/s12885-022-09596-z

60. Sabbah DA, Hajjo R, Sweidan K. Review on epidermal growth factor receptor (EGFR) structure, signaling pathways, interactions, and recent updates of EGFR inhibitors. *Curr Top Med Chem* (2020) 20(10):815–34. doi: 10.2174/1568026620666200303123102

61. Dayde D, Gunther J, Hirayama Y, Weksberg DC, Boutin A, Parhy G, et al. Identification of blood-based biomarkers for the prediction of the response to neoadjuvant chemoradiation in rectal cancer. *Cancers (Basel)* (2021) 13(14):3642. doi: 10.3390/cancers13143642

62. Beukinga RJ, Wang D, Karrenbeld A, Dijksterhuis WPM, Faber H, Burgerhof JGM, et al. Addition of HER2 and CD44 to (18)F-FDG PET-based clinico-radiomic models enhances prediction of neoadjuvant chemoradiotherapy response in esophageal cancer. *Eur Radiol* (2021) 31(5):3306–14. doi: 10.1007/s00330-020-07439-8

63. Spolverato G, Pucciarelli S, Bertorelle R, De Rossi A, Nitti D. Predictive factors of the response of rectal cancer to neoadjuvant radiochemotherapy. *Cancers (Basel)* (2011) 3(2):2176–94. doi: 10.3390/cancers3022176

64. Roskoski R Jr. Src protein-tyrosine kinase structure and regulation. *Biochem Biophys Res Commun* (2004) 324(4):1155–64. doi: 10.1016/j.bbrc.2004.09.171

65. Hermida-Prado F, Granda-Díaz R, Del-Río-Ibisate N, Villaronga M, Allonca E, Garmendia I, et al. The differential impact of SRC expression on the prognosis of patients with head and neck squamous cell carcinoma. *Cancers (Basel)* (2019) 11(11):1644. doi: 10.3390/cancers11111644

66. Cheng S.J., Kok S.H., Lee J.J., Yen-Ping Kuo M., Cheng S.L., Huang Y.L., et al (2012). Significant association of SRC protein expression with the progression, recurrence, and prognosis of oral squamous cell carcinoma in Taiwan. *Head Neck* 34(9), 1340-1345. doi: 10.1002/hed.21923

67. Zhang L, Teng Y, Zhang Y, Liu J, Xu L, Qu J, et al. C-src expression is predictive of poor prognosis in breast cancer patients with bone metastasis, but not in patients with visceral metastasis. *Apmis* (2012) 120(7):549–57. doi: 10.1111/j.1600-0463.2011.02864.x

68. Murugan AK. mTOR: Role in cancer, metastasis and drug resistance. *Semin Cancer Biol* (2019) 59:92–111. doi: 10.1016/j.semcancer.2019.07.003

69. Zhu K, Zhao Q, Yue J, Shi P, Yan H, Xu X, et al. GOLPH3 overexpression correlates with poor response to neoadjuvant therapy and prognosis in locally advanced rectal cancer. *Oncotarget* (2016) 7(42):68328–38. doi: 10.18632/oncotarget.12008

70. Saeed N, Shridhar R, Hoffe S, Almhanna K, Meredith KL. AKT expression is associated with degree of pathologic response in adenocarcinoma of the esophagus treated with neoadjuvant therapy. *J Gastrointest Oncol* (2016) 7(2):158–65. doi: 10.3978/j.issn.2078-6891.2015.067

71. Talukdar S, Emdad L, Das SK, Fisher PB. EGFR: An essential receptor tyrosine kinase-regulator of cancer stem cells. *Adv Cancer Res* (2020) 147:161–88. doi: 10.1016/bs.acr.2020.04.003

72. Liu Q, Yu S, Zhao W, Qin S, Chu Q, Wu K. EGFR-TKIs resistance *via* EGFR-independent signaling pathways. *Mol Cancer* (2018) 17(1):53. doi: 10.1186/s12943-018-0793-1

73. Farhan M, Wang H, Gaur U, Little PJ, Xu J, Zheng W. FOXO signaling pathways as therapeutic targets in cancer. *Int J Biol Sci* (2017) 13(7):815–27. doi: 10.7150/ijbs.20052

74. Park SY, Lee CJ, Choi JH, Kim JH, Kim JW, Kim JY, et al. The JAK2/STAT3/CCND2 axis promotes colorectal cancer stem cell persistence and radioresistance. *J Exp Clin Cancer Res* (2019) 38(1):399. doi: 10.1186/s13046-019-1405-7

75. Wu C, Zhu X, Liu W, Ruan T, Tao K. Hedgehog signaling pathway in colorectal cancer: function, mechanism, and therapy. *Onco Targets Ther* (2017) 10:3249–59. doi: 10.2147/ott.S139639