A gene mutation-based risk model for prognostic prediction in liver metastases

Bingran Yu[1†], Ning Zhang[1†], Yun Feng[1], Weiqi Xu[1], Ti Zhang[1] and Lu Wang[1*]

Abstract

Background Liver metastasis is the major challenge in the treatment for malignant tumors. Genomic profiling is increasingly used in the diagnosis, treatment and prediction of prognosis in malignancies. In this study, we constructed a gene mutation-based risk model to predict the survival of liver metastases.

Method We identified the gene mutations associated with survival and constructed the risk model in the training cohort including 800 patients with liver metastases from Memorial Sloan-Kettering Cancer Center (MSKCC) dataset. Other 794 patients with liver metastases were collected from 4 cohorts for validation. Furthermore, the analyses of tumor microenvironment (TME) and somatic mutations were performed on 51 patients with breast cancer liver metastases (BCLM) who had both somatic mutation data and RNA-sequencing data.

Results A gene mutation-based risk model involved 10 genes was constructed to divide patients with liver metastases into the high- and low-risk groups. Patients in the low-risk group had a longer survival time compared to those in the high-risk group, which was observed in both training and validation cohorts. The analyses of TME in BCLM showed that the low-risk group exhibited more immune infiltration than the high-risk group. Furthermore, the mutation signatures of the high-risk group were completely different from those of the low-risk group in patients with BCLM.

Conclusions The gene mutation-based risk model constructed in our study exhibited the reliable ability of predicting the prognosis in liver metastases. The difference of TME and somatic mutations among BCLM patients with different risk score can guide the further research and treatment decisions for liver metastases.

Keywords Gene mutation, Risk model, Liver metastasis, Tumor microenvironment

Introduction

Liver metastases are tumors which have spread from primary sites of cancers to liver and are the major cause of treatment failure and mortality in malignant tumors. The most common cancer which metastasizes to the liver is

colorectal cancer, followed by pancreatic cancer, breast cancer, melanoma and lung cancer [1]. Despite variety of treatments including surgery, chemotherapy, targeted therapy and immune checkpoint inhibitors were used in the treatment for liver metastases, poor response and rapid recurrence of tumors were often observed in liver [2]. It is necessary to construct a model which can predict the prognosis of liver metastases and guide the clinical treatment.

Genomic profiling is increasingly used to identify the pathogenic genes, select the targeted treatments and develop prognostic biomarkers [3]. In a phase II non-randomized clinical trial of Hayashi et al. [4], molecularly targeted therapies based on profiling gene

†Bingran Yu and Ning Zhang contributed equally to this work.

*Correspondence:
Lu Wang
wangluzl@fudan.edu.cn
[1] Department of Hepatic Surgery, Shanghai Cancer Center, Shanghai Medical College, Fudan University, No. 270 Dongan Road, Shanghai 200032, People's Republic of China

Yu *et al. BMC Genomics*   (2023) 24:489

Page 2 of 11

expressions and gene alterations by next-generation sequencing (NGS) were applied in the treatment for cancer of unknown primary site and were associated with a favorable survival outcome. The study of Goss et al. [5] used NGS to identify patients with lung squamous cell carcinoma who would derive additional benefit from treatment with afatinib and found ERBB and HER2 mutations could act as predictive markers for afatinib treatment. Furthermore, Long et al. [6] constructed and validated a gene mutation-based gene set to predict the prognosis of patients treated with immune checkpoint therapy.

In the present study, we integrated the cohorts of liver metastases to construct and validate a novel risk model based on gene mutations to predict the prognosis of patients with liver metastases. Additionally, we further analyzed the expression profiles of patients with different risk scores.

Materials and methods

Study population

The mutation data and clinical information of the training, primary liver cancer (PLC) and validation cohorts were obtained from the cBioPortal database (https://www.cbioportal.org). In the training cohort, both mutation data and clinical information were available for 800 patients with liver metastases from Memorial Sloan Kettering Cancer Center (MSKCC) database reported by Zehir et al. [7]. In the validation cohort, both mutation data and clinical information were available for 136 patients with liver metastases from the cohort of Samstein et al. [8], 312 patients with liver metastases from the cohort of Yaeger et al. [9], and 198 patients with liver metastases from the cohort of Pleasance et al. [10]. The PLC cohort, including 138 patients with PLC from MSKCC database reported by Zehir et al. [7], 107 patients with intrahepatic cholangiocarcinoma (ICC) from the cohort of Lowery et al. [11], 114 patients with hepatocellular carcinoma (HCC) from the cohort of Ng et al. [12], 61 patients with combined HCC and ICC from the cohort of Xue et al. [13], and 357 patients with HCC or ICC from The Cancer Genome Altas (TCGA) database, was also used for analysis. The clinical information of each sample enrolled in the analysis were shown in Additional file 1: Fig. S1 and Additional file 2: Table S1-S3.
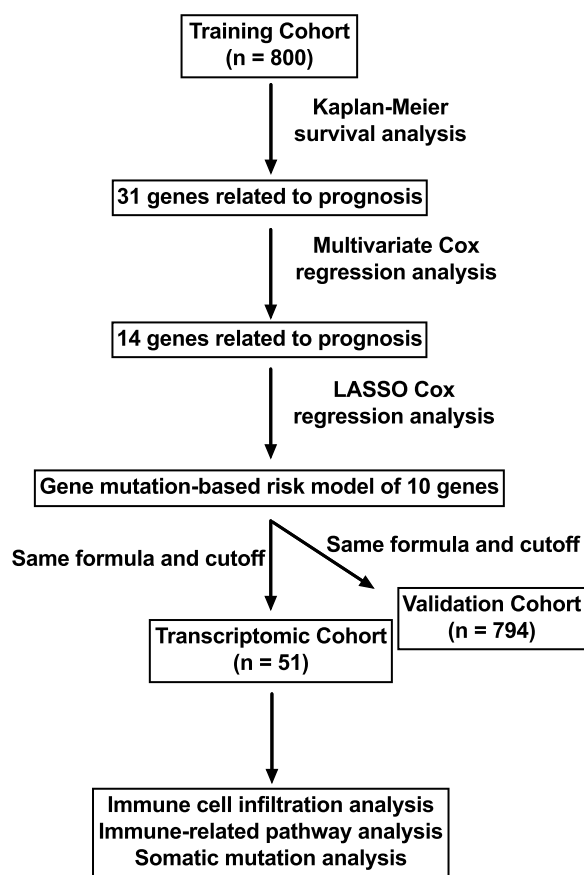
One hundred forty-eight patients with liver metastasis treated in Department of Hepatic Surgery of Fudan University Shanghai Cancer Center (FUSCC) were enrolled in the validation cohort. All patients provided their informed consents. The study was approved by the Institutional Review Board of FUSCC. The clinical information was collected retrospectively and shown in

Additional file 2: Table S2. A customed-designed genetic panel, comprising a hybridization-capture-based assay of 484 genes, was used to identify the mutant genes. The genes used in sequencing were listed in Additional file 2: Table S4. Patient's genomic DNA was extracted from tumor tissues. The libraries were pooled and sequenced using an Illumina HiSeq X TEN platform (Illumina Inc., San Diego, CA, USA). Data were collected using Illumina Real Time Analysis (RTA) and assembled to FASTQ files using Illumina Bcl2Fastq2. Then, the high-quality reads were mapped to the hg19 version of the human reference genome (GRCh37) using BWA aligner with the BWA-MEM algorithm and default parameters. The Genome Analysis ToolKit was used to locally realign the BAM files at intervals with mismatched indels and recalibrate the base quality scores of the reads in the BAM files. Somatic mutations were called from the tissue BAM files using GATK4 Mutect2 with the default parameters. Finally, the variants and annotation results were transferred into Excel spreadsheets for further analyses.

In the cohort of Pleasance et al. [10], 51 patients with breast cancer liver metastases (BCLM) were included into the transcriptomic cohort, whose RNA-sequencing data were acquired from UCSC Xena (https://xenabrowser.net/datapages/). The clinical data of each sample in the transcriptomic cohort were shown in Additional file 1: Fig. S1 and Additional file 2: Table S5.

Study design

Figure 1 and Additional file 1: Fig. S1 summarized the analysis process and cohorts used in the study that included training, validation, PLC and transcriptomic cohorts. First, propensity score matching (PSM) method by "MatchIt" R package was used in the training cohort to balance potentially confounding factors, including sex and cancer types, between the mutant-type and wild-type status of each gene in the MSKCC-IMPACT panel (Additional file 2: Table S6). Then survival data were compared between the mutant-type and wild-type status of each gene using Kaplan–Meier method in univariate analysis. Genes with $P < 0.05$ in univariate analysis were candidates for entry into multivariate Cox regression analysis. Genes with $P < 0.05$ in multivariate analysis and constructed a gene mutation-based risk model. The risk score was calculated according to the formula: Risk score $= \sum C_i X_i$ where C_i is the coefficient of each prognostic gene and X_i is the expression value of each prognostic gene. X_i is 0 when the prognostic gene is wild type, and 1 when the prognostic gene is mutant type. X-tile 3.6.1

Yu *et al. BMC Genomics*   (2023) 24:489

Page 3 of 11

```
┌─────────────────────┐
│  Training Cohort    │
│    (n = 800)        │
└─────────────────────┘
          │ Kaplan-Meier
          │ survival analysis
          ▼
┌─────────────────────────────┐
│ 31 genes related to prognosis│
└─────────────────────────────┘
          │ Multivariate Cox
          │ regression analysis
          ▼
┌─────────────────────────────┐
│ 14 genes related to prognosis│
└─────────────────────────────┘
          │ LASSO Cox
          │ regression analysis
          ▼
┌──────────────────────────────────────┐
│ Gene mutation-based risk model of 10 genes│
└──────────────────────────────────────┘
   Same formula and cutoff    Same formula and cutoff
          │                          ╲
          ▼                           ▼
┌──────────────────────┐   ┌──────────────────────┐
│ Transcriptomic Cohort│   │  Validation Cohort   │
│      (n = 51)        │   │      (n = 794)       │
└──────────────────────┘   └──────────────────────┘
          │
          ▼
┌──────────────────────────────────┐
│  Immune cell infiltration analysis│
│  Immune-related pathway analysis  │
│    Somatic mutation analysis      │
└──────────────────────────────────┘
```

Fig. 1 The workflow of the present study

software was applied to determine the best cutoff for classifying patients into high- and low-risk groups [14]. The same formula and cutoff were used for the validation and transcriptomic cohorts.

Tumor microenvironment analyses
Tumor microenvironment (TME) analyses were compared between the high- and low-risk groups based on the RNA-sequencing data of 51 patients with BCLM in the transcriptomic cohort. The genes used to quantify the enrichment levels of immune cell types and immune-related pathways were acquired from Long et al. [6]. The "GSVA" R package based on single-sample gene set enrichment analysis (ssGSEA) method was used to quantify the enrichment levels of these immune cells and immune-related pathways in each sample [15]. The enrichment scores in ssGSEA analysis were normalized to unity distribution, for which the minimal score is zero and maximal score is one. All immune-related signatures used in the TME analyses were shown in Additional file 2: Table S7 [6].

Drug sensitivity analysis
The "oncoPredict" package was used to assess the drug sensitivity of 198 chemotherapeutic drugs between patients in the high- and low-risk groups.

Functional enrichment analysis
Differential expression genes (DEGs) between the high- and low-risk groups were identified using the "limma" R package with a fold change of 2 and an adjusted *P*-value < 0.05. Functional enrichment analysis of DEGs were performed based on Gene Oncology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) using "clusterProfiler" R package [16].

Somatic mutation analyses
The "maftools" R package was used to analyze the difference of somatic mutations between the high- and low-risk groups in the transcriptomic cohort.

Statistical analyses
Statistical analyses were performed using R 4.2.2 and GraphPad Prism (Version 9.4.1). Cox regression models were applied to assess the independent prognostic value of the risk model. To evaluate the accuracy of the risk model, the receiver operating characteristic (ROC) curves were generated by the "timeROC" package, and the Area Under Curve (AUC) and C-index were compared with other risk factors. The calibration curve was constructed using "rms" R package to explore the predictive accuracy of the risk model. The Wilcoxon test was conducted to compare differences between two groups. *P*-values were two-sided with *P* < 0.05 considered significant.

Results
Construction of the gene mutation-based risk model
We performed PSM analysis and compared the survival outcomes between the mutant-type and wild-type status of each gene in the training cohort. The mutations of 31 genes were found to be associated with survival and the multivariate analysis identified 14 prognostic gene mutations in patients with liver metastasis (Additional file 2: Table S8). The LASSO regression analysis was applied to identified 10 prognostic gene mutations and their coefficients. The risk score was constructed as follow: Risk score $= APC \times (-0.447543265) + B2M \times (1.765754185) + FANCA \times (0.231443411) + FAT1 \times (-0.238096394) + IRS1 \times (1.223490281) + MAP3K13 \times (0.000988332) + NTRK1 \times (0.840546061) + STK11 \times (0.470490681) + TP53 \times (0.198123143) + YES1 \times (0.102246259)$. In the formula, the mutant gene status was coded as 1, and the wild gene status was coded as 0. The optimal

Yu *et al. BMC Genomics*   (2023) 24:489

Page 4 of 11

cutoff value was zero. The patients with risk score ≥ 0 were divided to the high-risk group and those with risk score < 0 were divided to the low-risk group.

Comparison of survival between the high- and low-risk groups

In the training cohort, patients in the low-risk group had a longer overall survival (OS) than those in the high-risk group (26.5 months vs. 12.6 months, $P < 0.001$; Fig. 2a). Subgroup analysis based on sex and cancer types also indicated that the longer OS were observed in the patients with low risk scores compared to those with high risk scores (Fig. 2b-d).

We also compared the OS of PLC cohort to investigate whether the risk model was also applicable to patients with PLC. There was no significant difference in survival between patients in the high- and low-risk groups (33.0 months vs. 37.1 months, $P > 0.05$; Fig. 3a).

The same formula and cutoff were also used in the validation cohort to evaluate the risk model. In the validation cohort, patients in the low-risk group exhibited a better OS than those in the high-risk group (54.5 months vs. 20.7 months, $P < 0.001$; Fig. 3b). Furthermore, subgroup analysis based on different cohorts and cancer types showed that patients in the low-risk group had a

longer survival time than those in the high-risk group (Fig. 3c-d).

Analyses of the predictive value of the gene mutation-based risk model

The AUC of the gene mutation-based risk model in the training cohort was 0.67 at 1 year, 0.64 at 2 years and 0.61 at 3 years (Additional file 1: Fig. S2a); the AUC in the validation cohort was 0.86 at 1 year, 0.73 at 2 years and 0.67 at 3 years (Additional file 1: Fig. S2b). The AUC for different cancer types of training cohorts (Additional file 1: Fig. S2c) and different cohorts of validation cohorts (Additional file 1: Fig. S2d) were also assessed. The calibration curve of the risk model was constructed and showed good agreement between the observations and the predictions in the training (Additional file 1: Fig. S3a) and validation cohorts (Additional file 1: Fig. S3b).

To further analyze the predictive value of the gene mutation-based risk model, the multivariate Cox analysis of sex, age, tumor mutation burden (TMB), primary cancer types and risk model were performed. The risk score and primary cancer types were the independent prognostic factors for patients with liver metastasis in the training cohort (Additional file 1: Fig. S3c), which was also observed in the validation cohort (Additional file 1: Fig. S3d).
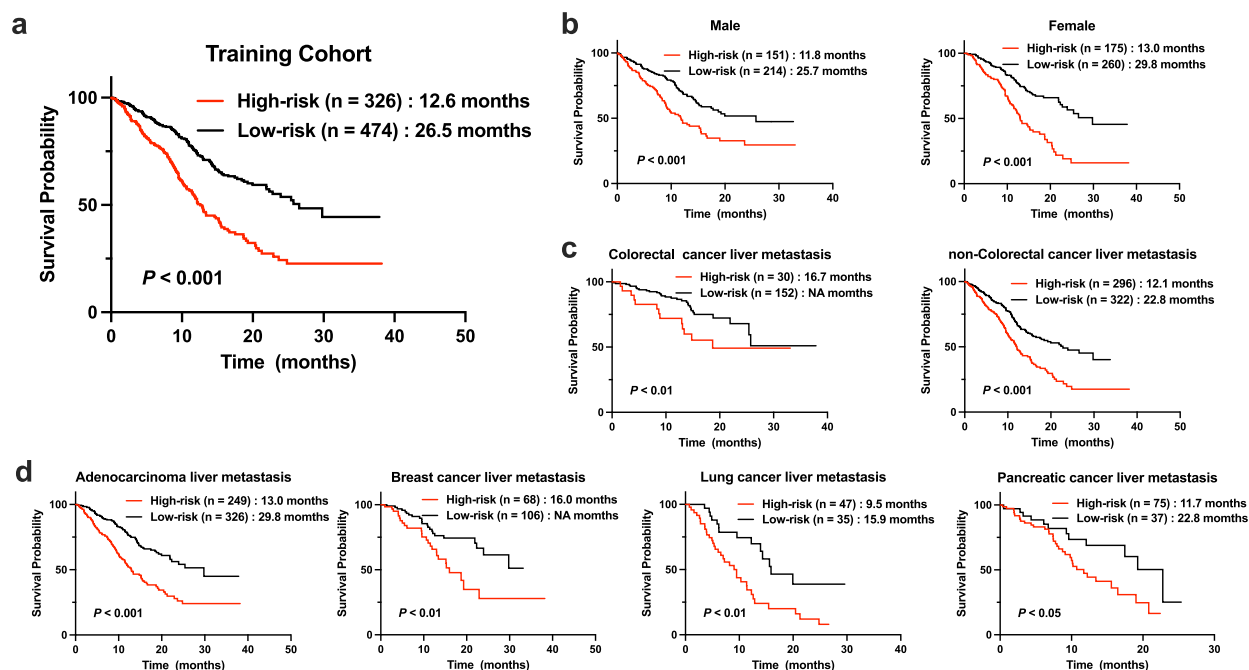
Fig. 2 Survival analysis of gene mutation-based risk model in the training cohort and different subgroups. **a** Survival analysis of gene mutation-based risk model in the training cohort. **b** Survival analysis of gene mutation-based risk model in male and female patients. **c** Survival analysis of gene mutation-based risk model in Colorectal cancer liver metastasis and non-Colorectal cancer liver metastasis. **d** Survival analysis of gene mutation-based risk model in different cancer types (adenocarcinoma liver metastasis, breast cancer liver metastasis, lung cancer liver metastasis and pancreatic cancer liver metastasis)

Yu *et al. BMC Genomics*   (2023) 24:489

Page 5 of 11
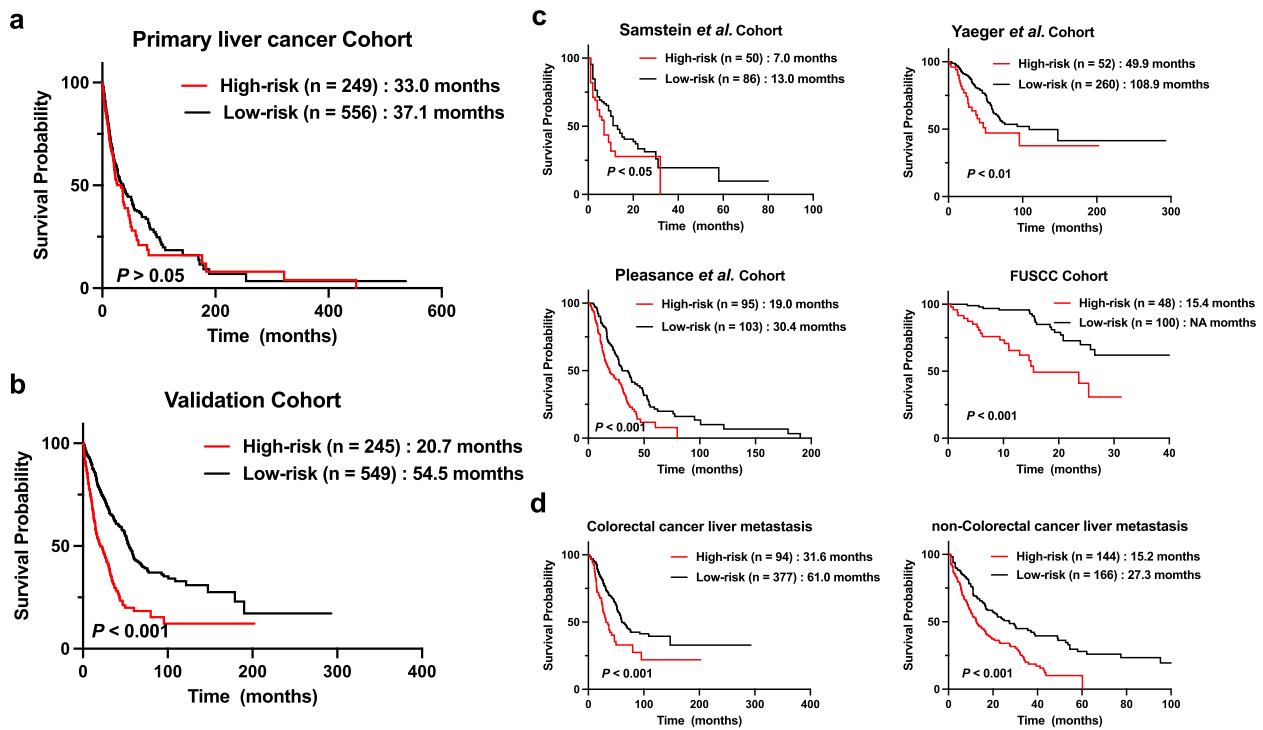
Fig. 3 Survival analysis of gene mutation-based risk model in the primary liver cancer, validation cohorts and different subgroups. **a** Survival analysis of gene mutation-based risk model in the primary liver cancer cohort. **b** Survival analysis of gene mutation-based risk model in the validation cohort. **c** Survival analysis of gene mutation-based risk model in different cohorts of validation cohort. **d** Survival analysis of gene mutation-based risk model in Colorectal cancer liver metastasis and non-Colorectal cancer liver metastasis of validation cohort

We used C-index to compared the performance of the gene mutation-based risk model with 10 genes included in the risk model and found that the risk model exhibited more excellent predictive power than 10 separate genes either in the training cohort or in the validation cohort.

TME analyses between the high- and low-risk groups

We explored the difference of TME based on the RNA-sequencing data of transcriptomic cohort, in which 51 patients with BCLM were classified into the high- and low-risk groups using the same formula and cutoff (Additional file 1: Fig. S1d). A better OS was observed in the low-risk group compared to the high-risk group in the transcriptomic cohort (49.2 months vs. 30.9 months, $P < 0.01$; Fig. 4a). Among

10 genes included in the risk model, *MAP3K13* exhibited a higher expression level in the high-risk group than in the low-risk group (Fig. 4b). We evaluated the enrichment scores of immune cells infiltration and immune-related pathways between the high- and low-risk groups using ssGSEA method. More immune cell infiltrations, including B cells, CD8[+] T cells, Natural Killer (NK) cells, plasmacytoid dendritic cells (pDCs), follicular helper T cells (Tfh) and tumor infiltrating lymphocytes (TILs) were found in the low-risk group (Fig. 4c). Furthermore, the low-risk group also show the higher activity of pathways related to C-C chemokine receptor (CCR), cytolytic activity and human leukocyte antigen (HLA) than the high-risk group (Fig. 4c). The analysis of the expression levels of genes associated

(See figure on next page.)

Fig. 4 Tumor microenvironment analysis between the high- and low-risk groups in patients with breast cancer liver metastases (BCLM). **a** Survival analysis between the high- and low-risk groups in patients with BCLM. **b** mRNA expression profiles analysis of 10 genes included in the risk model between the high- and low-risk groups in patients with BCLM. **c** Comparison of the immune infiltration estimated by the ssGSEA method based on RNA-sequencing data between the high and low-risk groups in patients with BCLM. **d** Comparison of genes associated with chemokines between the high and low-risk groups in patients with BCLM. **e** Drug sensitivity analysis between the high- and low-risk groups in patients with BCLM. **f** Volcano plot of differential expression genes (DEGs) between the high and low-risk groups in patients with BCLM. The high-risk group was the controlled group. **G** GO and KEGG functional enrichment analysis of DEGs between the high and low-risk groups in patients with BCLM
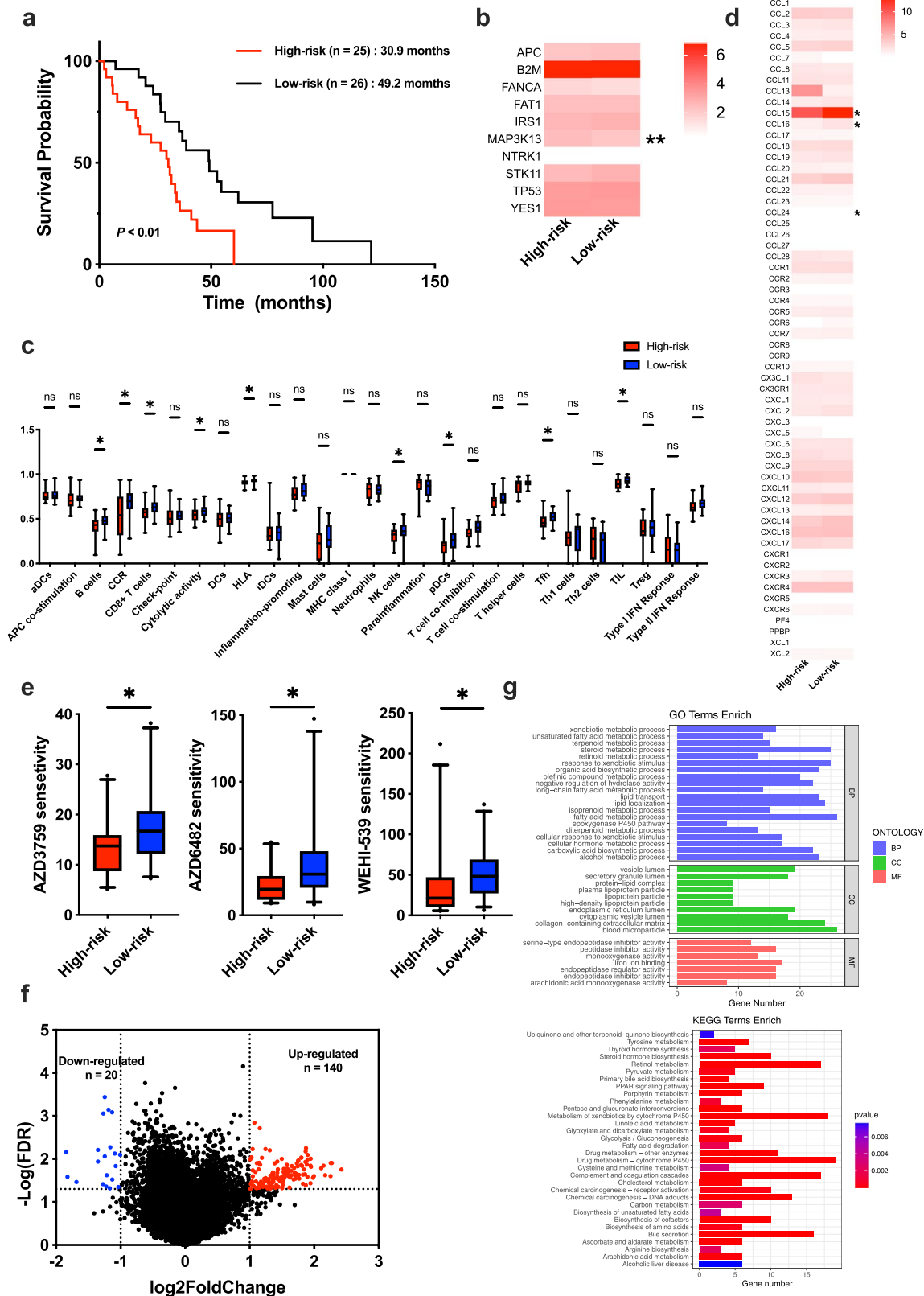
Fig. 4 (See legend on previous page.)

Yu *et al. BMC Genomics*   (2023) 24:489

Page 7 of 11

with chemokines indicated that the high-risk group expressed the higher levels of CCL15, CCL16 and CCL24 than the low-risk group (Fig. 4d). We assessed the drug sensitivity of the high- and low-risk populations to chemotherapeutic drugs and found that the low-risk group had the higher sensitivity to AZD3759, AZD 6482 and WEHI-539 (Fig. 4e).

We performed the analysis of DEGs between the high- and low-risk groups in the transcriptomic cohort. Compared to the high-risk group, a total of 160 DEGs were identified with 140 up-regulated genes and 20 down-regulated genes in the low-risk group (Fig. 4f). The function enrichment analysis of DEGs based on GO and KEGG databases showed that DEGs between the high- and low-risk groups involved in pathways associated with lipid metabolism such as fatty acid metabolism, lipid transport, lipid localization and lipoprotein particle in GO, and PPAR signaling pathway, fatty acid degradation and cholesterol metabolism in KEGG (Fig. 4g, Additional file 2: Table S9).

Somatic mutation analyses between the high- and low-risk groups

Somatic mutation analyses were performed in the transcriptomic cohort. Besides *TP53*, the high-risk group also exhibited higher frequencies of gene mutations in *PTEN* than the low-risk group (Fig. 5a-b). We analyzed the mutation signatures and compared the extracted mutation signatures against the Catalogue of Somatic Mutations in Cancer (COSMIC) by cosine similarity. 10 mutation signatures were identified. Three mutation signatures that associated with defects in DNA-DSB repair by HR were observed in the high-risk group, while other seven mutation signatures were found in the low-risk group (Fig. 5c-d).

Discussion

Genomic profiling has been increasingly used in constructing the risk model to predict either the response to immunotherapy or the prognosis of patients with malignant tumors [6, 17–19]. To the best of our knowledge, the present multicohort study was the first one
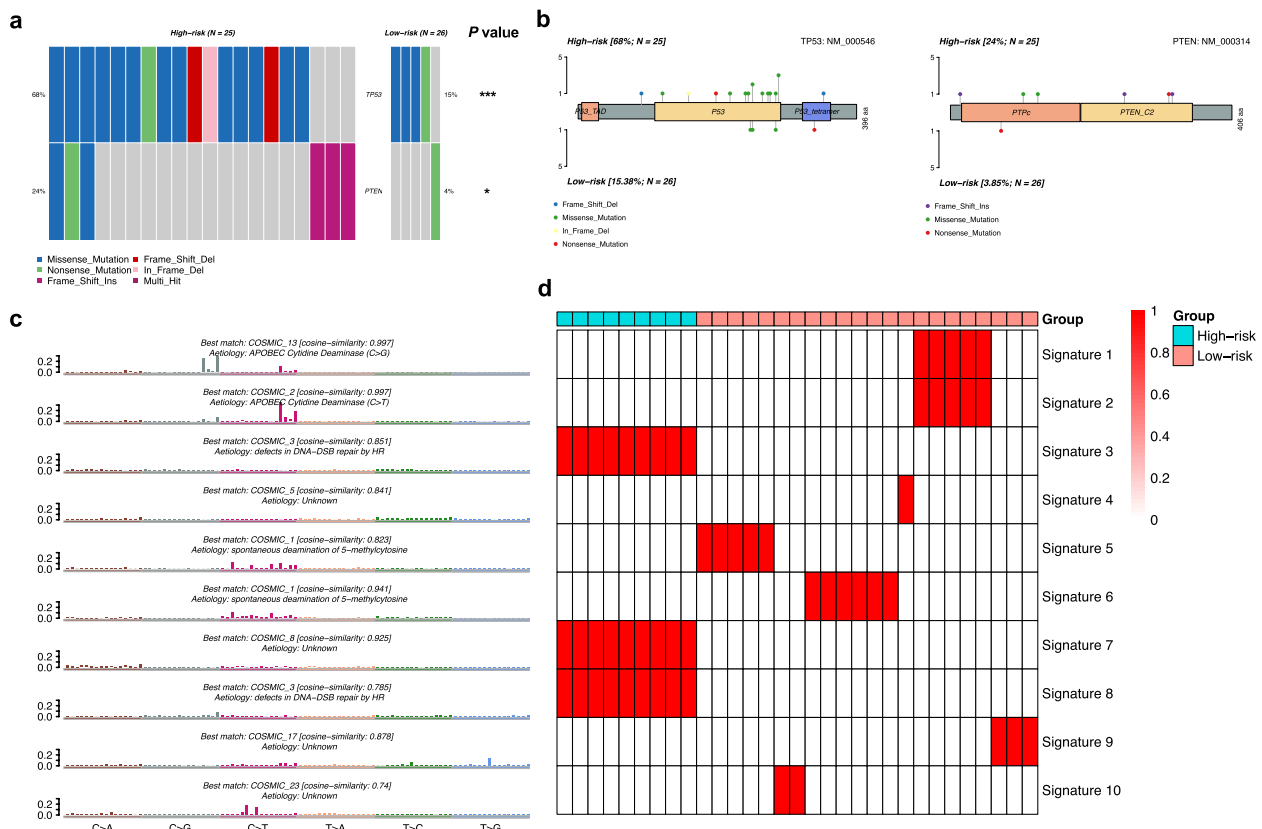
Fig. 5 Somatic mutation analysis between the high and low-risk groups in patients with breast cancer liver metastases (BCLM). **a** Comparison of gene mutations between the high and low-risk groups in patients with BCLM. **b** Lillipop plot of *TP53* and *PTEN* mutation between the high and low-risk groups in patients with BCLM. **c** Mutation signatures of patients with BCLM. **d** Comparison of mutation signatures between the high and low-risk groups in patients with BCLM

Yu *et al. BMC Genomics*   (2023) 24:489

Page 8 of 11

that constructed and validated a prognostic risk model based on gene mutations to predict the survival outcomes of patients with liver metastases. In our study, 10 genes mutations were identified as the prognostic genes mutations. Among them, most of gene mutations have been reported to be associated with the occurrence and progression of cancers. The mutations of *TP53* can impair antitumor activity and confer mutant *p53* protein oncogenic properties [20]. *FAT1* exhibited complexity in modulating tumorigenesis and acted as tumor promoter or suppressor depending on tumor types [21, 22]. Furthermore, tumors without mutations of *APC* were reported to be carry a worse prognosis than single *APC* mutation tumors in patients with colorectal cancers [23]. In our risk model, the coefficient of *FAT1* and *APC* was less than zero, demonstrated the mutations of these two genes suppressed the progression of tumors in patients with liver metastases.

Patients were divided into the high- and low-risk group according to the risk scores and a better OS was observed in the low-risk group compared to the high-risk group. However, it should be mentioned that of patients with colorectal cancer liver metastasis (CRCLM) or other cancer liver metastasis (OCLM), only approximately 20% were divided into the high-risk group. OCLM contained various uncommon cancers which had difference tumor characteristics, thus leading to less patients in the high-risk group. For patients with CRCLM, the reason of the imbalanced distribution could be due to the high frequency of *APC* mutations in CRC and the negative coefficient of *APC* in the risk score [24]. Despite of the different distribution among the different liver metastases, better survival was observed in all low-risk groups. In patients with PLC, there was no significant difference of survival outcomes between the high- and low-risk groups, indicated the risk model wasn't applicable to PLC. Furthermore, the multivariate analysis indicated that the risk score was also the prognostic factor of patients with liver metastases. We also compared predictive accuracy of the risk model with that of 10 individual gene mutations and a higher C-index of risk model was observed. In summary, based on these analyses and results, our gene mutation-based risk model was reliable in the prediction of survival of patients with liver metastases.

We explored the impact of 10 gene mutations included in the risk model on their gene expression levels. In the patients with BCLM, patients in the high-risk group had a significantly higher expression level of *MAP3K13*. High *MAP3K13* expression was reported to be correlated with poor prognosis in breast cancers, which might be one of the reasons that a shorter survival time was observed in the high-risk group of our study [25].

We used the transcriptomic data of a BCLM cohort to explore the difference of TME between the high- and low-risk groups. We found that the low-risk group had a higher infiltration of immune cells, such as a higher level of B cells, CD8$^+$ T cells, NK cells, Tfh and TILs. These immune cells exhibit antitumor effect in the development and progression of tumor [26–28]. It was reported that the depletion of these immune cells was associated with low responses to immunotherapy, thus leading to poor prognosis [29–32]. However, the low-risk group also had more pDCs which impeding T cell-mediated cytotoxicity, and a higher expression level of CCL15, CCL16 and CCL24 that were associated with clinical progression [1, 33]. Therefore, immune cells and chemokines with antitumor or protumor effects were both play a crucial part in the liver metastasis. Furthermore, we found that patients in the low-risk group have a higher response to chemotherapeutic drugs, such as AZD3759 which is a novel epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) and produced promising antitumor effect on non-small cell lung cancer [34]. In addition to AZD3759, AZD6482 (*PI3Kβ* inhibitor) and WEHI-539 (a *BCL2L1* inhibitor) were also reported to have potential antitumor effect. Despite these chemotherapeutic drugs have not been widely used, the results might provide guidance for further studies on treatments of patients with BCLM [35, 36].

In current study, the function enrichment analysis showed that DEGs between the high- and low-risk groups were enriched in pathways associated with lipid metabolism. It was reported that lipid metabolism exhibited vital effect by supporting proliferation, survival, migration, invasion and metastases of cancer cells during tumor progression [37]. The high rates of fatty acid oxidation were associated with the high potential metastases of triple-negative breast cancers [38]. Therefore, our study indicated that fatty acid metabolism play a vital role in prognosis of patients with BCLM in the high- and low-risk groups. Furthermore, *PPAR* modulates the lipid homeostasis in liver and the blockage of *PPARγ* can suppress breast cancer progression [39, 40]. The enrichment of *PPAR* pathway was also observed between the high- and low-risk groups in our study, confirmed the important role of *PPAR* in BCLM.

In the analyses of somatic mutation, higher frequencies of mutations in *TP53* and *PTEN* were observed in the high-risk patients with BCLM. *PTEN* is one of the most frequently mutated human tumor suppressor genes and the breast carcinogenesis is potentially associated with *PTEN* loss of activity owing to *PTEN* mutation [41]. Besides, patients in the high-risk group also had the completely different mutation signatures with those in the

Yu *et al. BMC Genomics*   (2023) 24:489

Page 9 of 11

low-risk group, which might lead to the difference of survival in patients with liver metastases.

This study had several limitations that should be considered. First, this was a study including various liver metastases. Although PSM analysis was performed to balance the bias, some potential factors might still influence the results due to the lack of complete clinical data. Second, despite the risk model exhibited the reliable ability of predicting the survival outcomes in patients with liver metastases, the accuracy was not high enough. Then, biological insight into the basis for the prognostic significance was obtained through analyses of a small ($n=51$) cohort of BCLM samples. Different biological characteristics can be observed among different cancer types. Therefore, analyses of various types of liver metastases were needed to verify the conclusions which were obtained from analyses of BCLM patients. Finally, the in vivo and in vitro functional experiments were needed to verify the difference of TME between the high- and low-risk groups and to investigate the molecular mechanism underlying the influence of each gene on liver metastases.

Conclusion

This study constructed and validated a gene mutation-based risk model to predict survival time for patients with liver metastases. The analyses of TME demonstrated that the patients with BCLM in the low-risk group had more immune cell infiltrations than those in the high-risk group. Furthermore, lipid metabolism played a crucial role in BCLM. Our study also revealed distinct mutation signatures for the high- and low-risk groups. Further studies were needed to verify the predictive ability of our gene mutation-based risk model and explore the biological characteristics of patients with different risk scores.

Abbreviations

MSKCC	Memorial Sloan-Kettering Cancer Center dataset
TME	Tumor microenvironment
BCLM	Breast cancer liver metastases
NGS	Next-generation sequencing
PLC	Primary liver cancer
ICC	Intrahepatic cholangiocarcinoma
HCC	Hepatocellular carcinoma
TCGA	The Cancer Genome Altas
FUSCC	Fudan University Shanghai Cancer Center
PSM	Propensity score matching
ssGSEA	Single-sample gene set enrichment analysis
DEG	Differential expression genes
GO	Gene Oncology
KEGG	Kyoto Encyclopedia of Genes and Genomes
ROC	Receiver operating characteristic
AUC	Area Under Curve
OS	Overall survival
TMB	tumor mutation burden
NK	Natural Killer
pDC	Plasmacytoid dendritic cell
Tfh	Follicular helper T cells
TIL	Tumor infiltrating lymphocyte
CCR	C-C chemokine receptor
HLA	Human leukocyte antigen
COSMIC	Catalogue of Somatic Mutations in Cancer

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09595-9.

Additional file 1: Fig. S1. The summary of patients included in the training (a), validation (b) and primary liver cancer (c) and transcriptomic (d) cohorts. **Fig. S2.** Assessment of the predictive performance of the gene mutation-based risk model. a. Receiver operating characteristic (ROC) curve of the training cohort. b. ROC curve of the validation cohort. c. Area under the ROC curve (AUC) for different cancer types in the training cohort. d. AUC for different cohorts in the validation cohort. **Fig. S3.** Assessment of the predictive accuracy of the gene mutation-based risk model. a. Calibration curve for assessing the predictive accuracy of the gene mutation-based risk model in the training cohort. The gray line represents ideal performance. The red line represents actual performance. b. Calibration curve for assessing the predictive accuracy of the gene mutation-based risk model in the validation cohort. The gray line represents ideal performance. The red line represents actual performance. c. Univariate and multivariate Cox regression analyses of gene mutation-based risk model in the training cohort. d. Univariate and multivariate Cox regression analyses of gene mutation-based risk model in the validation cohort. e. Comparison of C-indexes between the gene mutation-based risk model and 10 gene mutations included in the risk model in the training cohort. f. Comparison of C-indexes between the gene mutation-based risk model and 10 gene mutations included in the risk model in the validation cohort.

Additional file 2: Table S1. The clinical data for each sample used in the training cohort. **Table S2.** The clinical data for each sample used in the validation cohort. **Table S3.** The clinical data for each sample used in the primary liver cancer cohort. **Table S4.** Gene list for FUSCC. **Table S5.** The clinical data for breast cancer liver metastases in the cohort of Pleasance et al. **Table S6.** Gene list for MSKCC. **Table S7.** Immune-related signatures used in the study. **Table S8.** The multivariate Cox regression analysis of gene mutations in patients with liver metastasis. **Table S9.** Functional enrichment analysis of differential expressed genes between the high- and low-risk group in the transcriptomic cohort.

Availability of data and materials

The data and material were derived from cBioportal, and UCSC Xena, which are publicly available at https://www.cbioportal.org and https://xenabrowser.net/datapages/. The datasets of FUSCC cohort are available from the corresponding author on reasonable request.

Yu *et al. BMC Genomics*   (2023) 24:489

Page 10 of 11

Declarations

Ethics approval and consent to participate
The patients in FUSCC cohort included in the study have signed the informed consents. The study was performed in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of Fudan University Shanghai Cancer Center.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Tsilimigras DI, Brodt P, Clavien PA, Muschel RJ, D'Angelica MI, Endo I, Parks RW, Doyle M, de Santibanes E, Pawlik TM. Liver metastases. Nat Rev Dis Primers. 2021;7(1):27.
2. Milette S, Sicklick JK, Lowy AM, Brodt P. Molecular pathways: targeting the microenvironment of liver metastases. Clin Cancer Res. 2017;23(21):6390–9.
3. Siu LL, Conley BA, Boerner S, LoRusso PM. Next-generation sequencing to guide clinical trials. Clin Cancer Res. 2015;21(20):4536–44.
4. Hayashi H, Takiguchi Y, Minami H, Akiyoshi K, Segawa Y, Ueda H, Iwamoto Y, Kondoh C, Matsumoto K, Takahashi S, et al. Site-specific and targeted therapy based on molecular profiling by next-generation sequencing for cancer of unknown primary site: a nonrandomized phase 2 clinical trial. JAMA Oncol. 2020;6(12):1931–8.
5. Goss GD, Felip E, Cobo M, Lu S, Syrigos K, Lee KH, Goker E, Georgoulias V, Li W, Guclu S, et al. Association of ERBB mutations with clinical outcomes of Afatinib- or Erlotinib-treated patients with lung squamous cell carcinoma: secondary analysis of the LUX-lung 8 randomized clinical trial. JAMA Oncol. 2018;4(9):1189–97.
6. Long J, Wang D, Wang A, Chen P, Lin Y, Bian J, Yang X, Zheng M, Zhang H, Zheng Y, et al. A mutation-based gene set predicts survival benefit after immunotherapy across multiple cancers and reveals the immune response landscape. Genome Med. 2022;14(1):20.
7. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med. 2017;23(6):703–13.
8. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, Barron DA, Zehir A, Jordan EJ, Omuro A, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet. 2019;51(2):202–6.
9. Yaeger R, Chatila WK, Lipsyc MD, Hechtman JF, Cercek A, Sanchez-Vega F, Jayakumaran G, Middha S, Zehir A, Donoghue MTA, et al. Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. Cancer Cell. 2018;33(1):125–36. e123.
10. Pleasance E, Titmuss E, Williamson L, Kwan H, Culibrk L, Zhao EY, Dixon K, Fan K, Bowlby R, Jones MR, et al. Pan-cancer analysis of advanced patient tumors reveals interactions between therapy and genomic landscapes. Nat Cancer. 2020;1(4):452–68.
11. Lowery MA, Ptashkin R, Jordan E, Berger MF, Zehir A, Capanu M, Kemeny NE, O'Reilly EM, El-Dika I, Jarnagin WR, et al. Comprehensive molecular profiling of intrahepatic and extrahepatic Cholangiocarcinomas: potential targets for intervention. Clin Cancer Res. 2018;24(17):4154–61.
12. Ng CKY, Dazert E, Boldanova T, Coto-Llerena M, Nuciforo S, Ercan C, Suslov A, Meier MA, Bock T, Schmidt A, et al. Integrative proteogenomic characterization of hepatocellular carcinoma across etiologies and stages. Nat Commun. 2022;13(1):2436.
13. Xue R, Chen L, Zhang C, Fujita M, Li R, Yan SM, Ong CK, Liao X, Gao Q, Sasagawa S, et al. Genomic and Transcriptomic Profiling of Combined Hepatocellular and Intrahepatic Cholangiocarcinoma Reveals Distinct Molecular Subtypes. Cancer Cell. 2019;35(6):932–47. e938.
14. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res. 2004;10(21):7252–9.
15. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.
16. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587–92.
17. Wang Z, Ge Y, Li H, Fei G, Wang S, Wei P. Identification and validation of a genomic mutation signature as a predictor for immunotherapy in NSCLC. Biosci Rep. 2022;42(11):BSR20220892.
18. Liu M, Xia S, Zhang X, Zhang B, Yan L, Yang M, Ren Y, Guo H, Zhao J. Development and validation of a blood-based genomic mutation signature to predict the clinical outcomes of atezolizumab therapy in NSCLC. Lung Cancer. 2022;170:148–55.
19. Deng J, Weng X, Chen W, Zhang J, Ma L, Zhao K. A nomogram and risk classification model predicts prognosis in Chinese esophageal squamous cell carcinoma patients. Transl Cancer Res. 2022;11(9):3128–40.
20. Hu J, Cao J, Topatana W, Juengpanich S, Li S, Zhang B, Shen J, Cai L, Cai X, Chen M. Targeting mutant p53 for cancer therapy: direct and indirect strategies. J Hematol Oncol. 2021;14(1):157.
21. Chen ZG, Saba NF, Teng Y. The diverse functions of FAT1 in cancer progression: good, bad, or ugly? J Exp Clin Cancer Res. 2022;41(1):248.
22. Pastushenko I, Mauri F, Song Y, de Cock F, Meeusen B, Swedlund B, Impens F, Van Haver D, Opitz M, Thery M, et al. Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. Nature. 2021;589(7842):448–55.
23. Schell MJ, Yang M, Teer JK, Lo FY, Madan A, Coppola D, Monteiro AN, Nebozhyn MV, Yue B, Loboda A, et al. A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC. Nat Commun. 2016;7:11743.
24. Lim B, Mun J, Kim J-H, Kim CW, Roh SA, Cho D-H, Kim YS, Kim S-Y, Kim JC. Genome-wide mutation profiles of colorectal tumors and associated liver metastases at the exome and transcriptome levels. Oncotarget. 2015;6(26):22179–90.
25. Han H, Chen Y, Cheng L, Prochownik EV, Li Y. microRNA-206 impairs c-Myc-driven cancer in a synthetic lethal manner by directly inhibiting MAP3K13. Oncotarget. 2016;7(13):16409–19.
26. Tumeh PC, Hellmann MD, Hamid O, Tsai KK, Loo KL, Gubens MA, Rosenblum M, Harview CL, Taube JM, Handley N, et al. Liver metastasis and treatment outcome with anti-PD-1 monoclonal antibody in patients with Melanoma and NSCLC. Cancer Immunol Res. 2017;5(5):417–24.
27. Johansson J, Siarov J, Kiffin R, Molne J, Mattsson J, Naredi P, Olofsson Bagge R, Martner A, Lindner P. Presence of tumor-infiltrating CD8(+) T cells and macrophages correlates to longer overall survival in patients undergoing isolated hepatic perfusion for uveal melanoma liver metastasis. Oncoimmunology. 2020;9(1):1854519.
28. Yang S, Qian L, Li Z, Li Y, Bai J, Zheng B, Chen K, Qiu X, Cai G, Wang S, et al. Integrated Multi-Omics Landscape of Liver Metastases. Gastroenterology 2022;164(3):407–23.
29. Xu Y, Wei Z, Feng M, Zhu D, Mei S, Wu Z, Feng Q, Chang W, Ji M, Liu C, et al. Tumor-infiltrated activated B cells suppress liver metastasis of colorectal cancers. Cell Rep. 2022;40(9):111295.
30. Ducimetiere L, Lucchiari G, Litscher G, Nater M, Heeb L, Nunez NG, Wyss L, Burri D, Vermeer M, Gschwend J, et al. Conventional NK cells and tissue-resident ILC1s join forces to control liver metastasis. Proc Natl Acad Sci U S A. 2021;118(27):e2026271118.
31. Niogret J, Berger H, Rebe C, Mary R, Ballot E, Truntzer C, Thibaudin M, Derangere V, Hibos C, Hampe L, et al. Follicular helper-T cells restore CD8(+)-dependent antitumor immunity and anti-PD-L1/PD-1 efficacy. J Immunother Cancer. 2021;9(6):e002157.
32. Hollern DP, Xu N, Thennavan A, Glodowski C, Garcia-Recio S, Mott KR, He X, Garay JP, Carey-Ewend K, Marron D, et al. B Cells and T follicular helper cells mediate response to checkpoint inhibitors in high mutation burden mouse models of breast cancer. Cell. 2019;179(5):1191–206. e1121.
33. Shen W, Zhang X, Tang J, Zhang Z, Du R, Luo D, Liu X, Xia Y, Li Y, Wang S, et al. CCL16 maintains stem cell-like properties in breast cancer by activating CCR2/GSK3beta/beta-catenin/OCT4 axis. Theranostics. 2021;11(5):2297–317.

Yu *et al. BMC Genomics*   *(2023) 24:489*

Page 11 of 11

34. Ahn MJ, Kim DW, Cho BC, Kim SW, Lee JS, Ahn JS, Kim TM, Lin CC, Kim HR, John T, et al. Activity and safety of AZD3759 in EGFR-mutant non-small-cell lung cancer with CNS metastases (BLOOM): a phase 1, open-label, dose-escalation and dose-expansion study. Lancet Respir Med. 2017;5(11):891–902.

35. Chiou JT, Lee YC, Wang LJ, Chang LS. BCL2 inhibitor ABT-199 and BCL2L1 inhibitor WEHI-539 coordinately promote NOXA-mediated degradation of MCL1 in human leukemia cells. Chem Biol Interact. 2022;361:109978.

36. Xu PF, Yang JA, Liu JH, Yang X, Liao JM, Yuan FE, Liu BH, Chen QX. PI3Kbeta inhibitor AZD6482 exerts antiproliferative activity and induces apoptosis in human glioblastoma cells. Oncol Rep. 2019;41(1):125–32.

37. Bian X, Liu R, Meng Y, Xing D, Xu D, Lu Z. Lipid metabolism and cancer. J Exp Med. 2021;218(1):e20201606.

38. Wright HJ, Hou J, Xu B, Cortez M, Potma EO, Tromberg BJ, Razorenova OV. CDCP1 drives triple-negative breast cancer metastasis through reduction of lipid-droplet abundance and stimulation of fatty acid oxidation. Proc Natl Acad Sci U S A. 2017;114(32):E6556–65.

39. Wang Y, Nakajima T, Gonzalez FJ, Tanaka N. PPARs as metabolic regulators in the liver: lessons from liver-specific PPAR-null mice. Int J Mol Sci. 2020;21(6):2061.

40. Yang PB, Hou PP, Liu FY, Hong WB, Chen HZ, Sun XY, Li P, Zhang Y, Ju CY, Luo LJ, et al. Blocking PPARgamma interaction facilitates Nur77 interdiction of fatty acid uptake and suppresses breast cancer progression. Proc Natl Acad Sci U S A. 2020;117(44):27412–22.

41. Kechagioglou P, Papi RM, Provatopoulou X, Kalogera E, Papadimitriou E, Grigoropoulos P, Nonni A, Zografos G, Kyriakidis DA, Gounaris A. Tumor suppressor PTEN in breast cancer: heterozygosity, mutations and protein expression. Anticancer Res. 2014;34(3):1387–400.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.