ARTICLE    OPEN

Check for updates

# Molecular characteristics of microsatellite stable early-onset colorectal cancer as predictors of prognosis and immunotherapeutic response

Can Lu [1,2,3,11], Xiaopeng Zhang[4,5,11], Josefine Schardey[1,11], Ulrich Wirth[1], Kathrin Heinrich [6], Luca Massiminio[7], Giulia Martina Cavestro[7], Jens Neumann[8,9,10], Alexandr V. Bazhin[1,9,10], Jens Werner[1,9,10] and Florian Kühn[1,9,10 ✉]

The incidence of early-onset colorectal cancer (EO-CRC, in patients younger than 50) is increasing worldwide. The specific gene signatures in EO-CRC patients are largely unknown. Since EO-CRC with microsatellite instability is frequently associated with Lynch syndrome, we aimed to comprehensively characterize the tumor microenvironment (TME) and gene expression profiles of EO-CRC with microsatellite stable (MSS-EO-CRC). Here, we demonstrated that MSS-EO-CRC has a similar pattern of tumor-infiltrating immune cells, immunotherapeutic responses, consensus molecular subtypes, and prognosis as late-onset CRC with MSS (MSS-LO-CRC). 133 differential expressed genes were identified as unique gene signatures of MSS-EO-CRC. Moreover, we established a risk score, which was positively associated with PD-L1 expression and could reflect both the level of tumor-infiltrating immune cells and the prognosis of MSS-EO-CRC patients. Application of this score on the anti-PD-L1 treatment cohort demonstrated that the low-risk score group has significant therapeutic advantages and clinical benefits. In addition, candidate driver genes were identified in the different-sidedness of MSS-EO-CRC patients. Altogether, MSS-EO-CRC exhibits distinct molecular profiles that differ from MSS-LO-CRC even though they have a similar TME characterization and survival pattern. Our risk score appears to be robust enough to predict prognosis and immunotherapeutic response and therefore could help to optimize the treatment of MSS-EO-CRC.

## INTRODUCTION

Colorectal cancer (CRC) is the third most diagnosed carcinoma and the second leading cause of cancer-associated mortality globally[1]. Although the incidence of late-onset CRC (LO-CRC) diagnosed in patients 50 years or older has steadily declined over the last two decades in most Western countries[2], the cases of early-onset CRC (EO-CRC) diagnosed in those younger than 50 years have increased alarmingly worldwide. By the 2030s, it is estimated that EO-CRC will account for one-quarter of rectal cancers and 10 to 12% of colon cancers[2,3]. So far, the underlying causes for the rising trends of EO-CRC are unknown, but early-life exposures, Western-style diet, microbial dysbiosis, and physical inactivity might contribute to the expansion of the EO-CRC population[4,5].

Accumulating studies reported that EO-CRCs tend to have a more advanced TNM (tumor-node-metastasis) stage, higher prevalence of left-sided carcinoma and poorly differentiated tumors, a higher proportion of microsatellite instability-high (MSI-H) status and more germline mutations compared to LO-CRCs[5,6]. Extensive efforts have been made to characterize the somatic mutational profiling of EO-CRCs[7,8], which failed to discover previously unknown alterations to elucidate the pathogenesis of these carcinomas or to guide clinical therapy. However, the unique transcriptional features in EO-CRCs remain elusive. Increasing evidence indicates that most EO-CRCs with MSI-H are associated with Lynch syndrome[9,10]. Regarding the well-known germline mutations of mismatch repair (MMR) genes in Lynch syndrome, we intend to identify the potential molecular mechanism for developing EO-CRCs with microsatellite stable (MSS-EO-CRC).

Primary tumor location is essential in predicting the prognosis and the drug responses for CRC patients[11]. Clinically, right-sided and left-sided CRCs are divided according to proximity to the splenic flexure. Although different sidedness of CRC represents the distinction in the mutational spectrum and molecular expression patterns[12], the biological effects of tumor location on MSS-EO-CRC patients remain unclear.

In the present study, multiple transcriptional profiles were systematically integrated to evaluate the characteristic features of MSS-EO-CRC compared to LO-CRC with MSS (MSS-LO-CRC). Based on the differentially expressed genes in MSS-EO-CRC, we constructed a risk score that significantly correlated with the tumor microenvironment (TME) characterization and showed a promising potential to predict response to anti-programmed death-ligand 1 (PD-L1) immunotherapy. Furthermore, we depicted the genetic variants and endogenous growth factor receptor (EGFR)-related molecules' expression between different sidedness of MSS-EO-CRC.

[1]Department of General, Visceral, and Transplant Surgery, Ludwig-Maximilians-University Munich, 81377 Munich, Germany. [2]Department of Colorectal Surgery and Oncology, Key Laboratory of Cancer Prevention and Intervention (Ministry of Education), The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China. [3]Zhejiang Provincial Clinical Research Center for CANCER & Cancer Center of Zhejiang University, Hangzhou, China. [4]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Gastrointestinal Cancer Center, Peking University Cancer Hospital and Institute, Beijing, China. [5]Institute of Laboratory Medicine, University Hospital of LMU Munich, Munich, Germany. [6]Department of Medicine III, University Hospital, Ludwig-Maximilians-University Munich, 81377 Munich, Germany. [7]Experimental Gastroenterology Laboratory, Gastroenterology and Endoscopy Department, San Raffaele Scientific Institute, Milano, Italy. [8]Institute of Pathology, Medical Faculty, Ludwig-Maximilians-University Munich, 81377 Munich, Germany. [9]German Cancer Consortium (DKTK), Partner Site Munich, 81377 Munich, Germany. [10]Bavarian Cancer Research Center (BZKF), Partner Site Munich, Munich, Germany. [11]These authors contributed equally: Can Lu, Xiaopeng Zhang, Josefine Schardey. ✉email: Florian.Kuehn@med.lmu.de
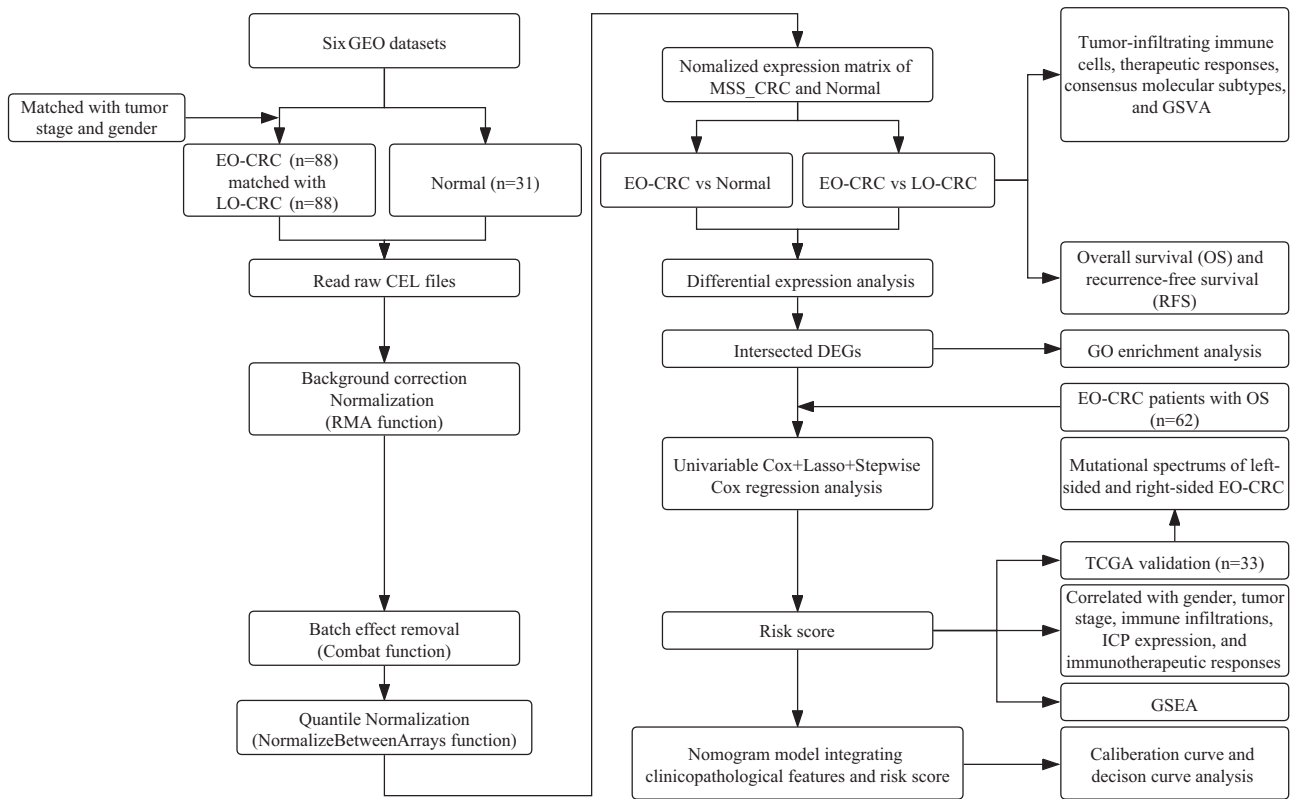
**Fig. 1 Analysis flow diagram of the study.** EO-CRC and LO-CRC patients are from the cohort of CRC with MSS. GEO Gene Expression Omnibus, EO-CRC early-onset colorectal cancer, LO-CRC late-onset colorectal cancer, MSS_CRC colorectal cancer with microsatellite stable, GSVA gene set variation analysis, DEGs differentially expressed genes, GO gene ontology, TCGA The Cancer Genome Atlas, ICP immune checkpoint, GSEA gene set enrichment analysis.

## RESULTS

### Characterization of tumor microenvironment and prognosis of MSS-EO-CRC patients

The study flowchart is depicted in Fig. 1. To exclude the potential effects of confounding variables, we matched MSS-EO-CRC patients with MSS-LO-CRC ones according to gender and tumor stage. Supplementary Table 1 summarizes the clinicopathological characteristics of 88 MSS-EO-CRC patients and 88 MSS-LO-CRC ones. The immune system is widely recognized as a critical factor determining the development and progression of CRC[13,14]. We found that MSS-EO-CRC patients displayed a similar distribution of 22 tumor-infiltrated immune subsets to MSS-LO-CRC ones (Fig. 2a). Meanwhile, no differences between MSS-EO-CRC and MSS-LO-CRC patients were detected in the overall stromal and immune components in the TME (Fig. 2b). We also conducted TIDE and Submap algorithm to predict the treatment response of the population to immunotherapy. These two cohorts have a similar treatment response rate to anti-PD1 or anti-CTLA4 drugs (Fig. 2c, d). Consensus molecular subtypes (CMSs) hold a promising role in deciphering the intrinsic heterogeneity of CRC at the gene expression level[15], playing a crucial role in predicting a patient's prognosis and treatment responses[16]. Our results showed that MSS-EO-CRC patients have a similar composition of CMSs as MSS-LO-CRC ones (Fig. 2e). To identify the potential drugs having different sensitivity in subgroup patients, we predicted that MSS-LO-CRC patients were more sensitive to OSI.906, while MSS-EO-CRC ones were more sensitive to PF.4708671 and Salubrinal (Fig. 2f–h). We found that MSS-EO-CRC patients have a similar survival rate to MSS-LO-CRC in overall survival (OS) and recurrence-free survival (RFS) (Fig. 2i, j).

Furthermore, we matched 33 MSS-EO-CRC with 33 MSS-LO-CRC from TCGA cohort to confirm the above findings. The clinical

characteristics of these patients are listed in Supplementary Table 2. Supplementary Fig. 1 depicted that an independent CRC dataset could get the similar results as Fig. 2, except from the sensitivity difference of some drugs.

### Identification of unique gene signatures in MSS-EO-CRC

To identify the genetic features in MSS-EO-CRC patients, we performed a sequential analysis by comparing the gene expression matrix of MSS-EO-CRC with MSS-LO-CRC and with normal samples. Firstly, we identified 1073 differentially expressed genes (DEGs) in MSS-EO-CRC patients compared to MSS-LO-CRC, including 730 up-regulated and 343 down-regulated genes (Fig. 3a). Functional annotations based on the Gene Set Variation Analysis (GSVA) algorithm showed that these two cohorts displayed significant differences in the enriched hall marker and molecular pathways (Fig. 3b, c). MSS-EO-CRC showed higher enrichment of Wnt beta-catenin signaling, protein secretion, and metabolic activities, whereas MSS-LO-CRC displayed more potent activity in hedgehog signaling. Furthermore, the mTOR signaling pathway, Wnt signaling pathway, and metabolic pathways are markedly enriched in MSS-EO-CRC, and MSS-LO-CRC significantly enriched the extracellular matrix (ECM) receptor interaction pathway. Secondly, 4551 DEGs were obtained from the differential expression analysis between MSS-EO-CRC samples and normal controls (Fig. 3d). Considering the particular entity of EO-CRC in terms of age and carcinoma, we performed the intersection of DEGs between MSS-EO-CRC versus MSS-LO-CRC and MSS-EO-CRC versus normal to identify the genes featured in MSS-EO-CRC patients. In total, 133 DEGs consisting of 102 up-regulated genes and 31 down-regulated genes were identified as the common DEGs (Fig. 3e, f). Furthermore, the
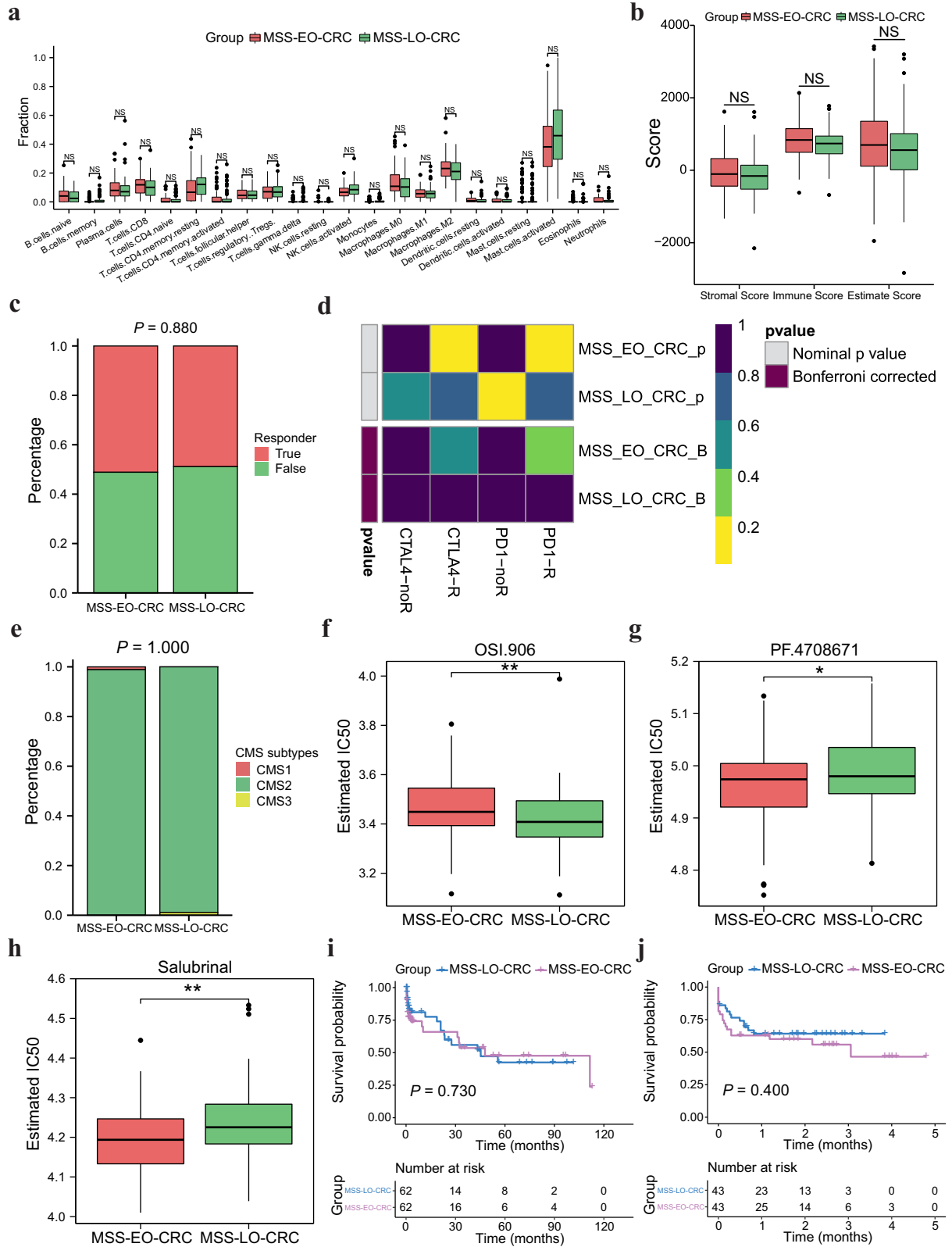
Gene Ontology (GO) enrichment analysis showed that these genes were significantly related to the mitosis activities of chromosomes and DNA (Fig. 3g). Detailed results of this functional enrichment analysis are shown in Supplementary Table 3.

**Development of the risk score for MSS-EO-CRC patients**

Since MSS-EO-CRC has a different transcriptomic landscape than MSS-LO-CRC, it is promising to construct a prognostic model for the subgroup of CRC patients regarding the potential age discrepancies. However, OS information was only available for

**Fig. 2 Characterization of tumor microenvironment and prognosis of early-onset CRC with MSS. a** Comparison of tumor-infiltrating immune cells between MSS-EO-CRC and MSS-LO-CRC. *P*-values were corrected using the Benjamini–Hochberg method. **b** Comparison of the enrichment score between MSS-EO-CRC and MSS-LO-CRC. **c, d** Comparison of Immunotherapeutic responses between MSS-EO-CRC and MSS-LO-CRC using TIDE and SubMap algorithm, respectively. **e** Consensus molecular subtype analysis of MSS-EO-CRC versus MSS-LO-CRC. **f–h** Drug sensitivities comparison between MSS-EO-CRC and MSS-LO-CRC. **i** Overall survival comparison between MSS-EO-CRC and MSS-LO-CRC. **j** Recurrence-free survival comparison between MSS-EO-CRC and MSS-LO-CRC. In the boxplot, the upper and lower boundaries represented the first and third quartiles, respectively, the central line signified the median and the whiskers extended to the most distant data points not considered as outliers (within 1.5 times the interquartile range). Outliers were displayed as points above and below the box-and-whisker diagram. The log-rank test *P* values were shown for each Kaplan–Meier plot. \*\*\**P* < 0.001; \*\**P* < 0.01; \**P* < 0.05. MSS-EO-CRC early-onset colorectal cancer with microsatellite stable, MSS-LO-CRC late-onset colorectal cancer with microsatellite stable, NS no significance.

partial MSS-EO-CRC patients. Thus, MSS-EO-CRC from GEO datasets ($N = 62$) and TCGA cohort ($N = 33$) were separately considered as training and external validation datasets to build the clinical model. The baseline clinical characteristics are summarized in Supplementary Table 4. Twenty-nine genes were identified from the training set using the univariate Cox regression analysis on the genes featured in MSS-EO-CRC (Supplementary Table 5). To refine the parameters incorporated into the model, we subsequently utilized the Least Absolute Shrinkage and Selection Operator (Lasso) Cox regression to select the substantial genes highly predictive of the OS (Fig. 4a). Three genes were identified with the lambda of 0.176 (Fig. 4b). Further, they entered a stepwise Cox regression model using a bidirectional selection strategy. Finally, WASF1 and TNFRSF14 were chosen to construct a prognostic model using a logistic regression algorithm. A risk score for prognosis prediction was determined as follows: risk score = $(16.04519 \times Expr_{WASF1}) + (0.00002 \times Expr_{TNFRSF14})$. We used the time-dependent ROC curves to evaluate the prognostic capacity of this risk score. The area under the curves (AUCs) for 1 year and 3 years OS were 0.70 and 0.74 for the training set (Fig. 4c), 0.83 and 0.87 for the external validation set (Fig. 4d), respectively. A risk score of 42.021 and 60.298 was separately defined as the optimal cut-off value to divide the population of training and validation set into a high-risk group and a low-risk group. Patients in the high-risk groups have significantly worse OSs than the low-risk ones in these two cohorts ($P < 0.05$) (Fig. 4e, f). Moreover, the prognostic capacity of the risk score remained robust in the subgroup analysis stratified by the tumor stage ($P < 0.05$) (Supplementary Fig. 2a, b). In addition, the distribution of gender and tumor stage were similar between high- and low-risk groups of MSS-EO-CRC patients (Supplementary Fig. 2c), which indicated that these two factors have no association with the risk score.

Furthermore, the two genes incorporated into the risk score were prognostic factors for the OS of MSS-EO-CRC patients (Fig. 4g, h), in which the expression level of WASF1 was inversely correlated with the prognosis of patients, whereas patients with higher levels of TNFRSF14 have a better prognosis than lower ones. Compared to MSS-LO-CRC and normal, WASF1 and TNFRSF14 were specifically up-regulated and down-regulated in MSS-EO-CRC, respectively (Fig. 4i). However, no differences were detected between early-stage and advanced MSS-EO-CRC in the expression of these two genes (Supplementary Fig. 2d, e).

## Characterization of the tumor microenvironment and immunotherapeutic responses in high and low-risk score groups

The immune infiltration played a critical role in regulating the development and progression of CRC via conducting pro-tumor or anti-tumor biological effects. We found that the risk score was negatively correlated with the infiltration level of CD8$^+$ T cells, activated memory CD4$^+$ T cells, and activated dendritic cells in the TME of MSS-EO-CRC (Fig. 5a). Since the ICP could significantly alter the function of T lymphocytes, we evaluated the relationship between the risk score and the expression level of seven

ICP-related molecules in MSS-EO-CRC patients. Our study proved that the high-risk group has a markedly higher CD274 (PD-L1) level than the low-risk one (Fig. 5b), suggesting the risk score may correlate with the response to immunotherapy. Thus, we evaluated the capacity of the risk score in predicting the treatment response to anti-PD-L1 antibody Atezolizumab using the IMvigor210 immunotherapy cohort. Patients with high-risk scores had a worse survival rate than patients with low-risk scores (Fig. 5c). The percentage of patients who responded to the anti-PD-L1 drug in the high-risk score group was remarkably lower compared to the low-risk score group (Fig. 5d). However, no difference of neoantigen was detected between high- and low-risk score groups (Fig. 5e).

Targeted therapy has become a promising strategy for advanced-stage cases. Identifying subgroups of patients more sensitive to certain drugs is critical to provide individualized therapy. Our study suggested that the low-risk score group was more sensitive to two tyrosine kinase inhibitors, lapatinib and axitinib, than the high-risk one (Fig. 5f, g). Moreover, we applied the Gene Set Enrichment Analysis (GSEA) analysis on GEO datasets to decipher the molecular mechanism underlying the risk score. As shown in Fig. 5h, the ECM receptor interaction pathway was significantly enriched in the tumors of the low-risk score group.

## Construction and assessment of a predictive nomogram

We performed univariate and multivariate Cox regression analyses on MSS-EO-CRC patients from GEO datasets to assess the risk score, tumor stage, age, and gender as independent prognostic markers. The tumor stage and risk score were identified as the independent prognostic factors for the GEO datasets (tumor stage, HR: 2.05; 95% CI: 1.24–3.38; $P < 0.01$; risk score, HR: 1.32; 95%: 1.06–1.63; $P < 0.05$). We provided the details in Supplementary Table 6. Thus, we integrated the risk score and tumor stage into a nomogram model to maximally increase the predicted probability on 1-year and 3-year OS (Fig. 6a). According to the goodness of fit between the predicted survival probability and actual survival rate on calibration plots, the nomogram has a better prediction on short-term survival (1-year) than long-term survival (3-year) (Fig. 6b). In addition, the nomogram has a higher concordance index (C-index) (0.702, 95% CI: 0.640–0.764) than either tumor stage (0.650, 95% CI: 0.587–0.713) or the risk score (0.649, 95% CI: 0.588–0.710) alone. Decision curve analysis (DCA) demonstrated that the nomogram model has the most significant net benefit for MSS-EO-CRC patients compared to the rest two factors (Fig. 6c, d).

## Mutational spectrums and EGFR-related molecular expression in different sidedness of MSS-EO-CRC patients

Due to the distinctive molecular characteristics of CRC with different sidedness, we intend to explore the potential effects of tumor location on the mutational landscape and genetic expression of MSS-EO-CRC. Based on the somatic mutation data of MSS-EO-CRC patients from TCGA, we predicted the candidate driving genes using the MutSigCV algorithm with a *p*-value less than 0.001. As displayed in Fig. 7a, five genes have been identified as the significantly mutated genes (SMGs) for left-sided
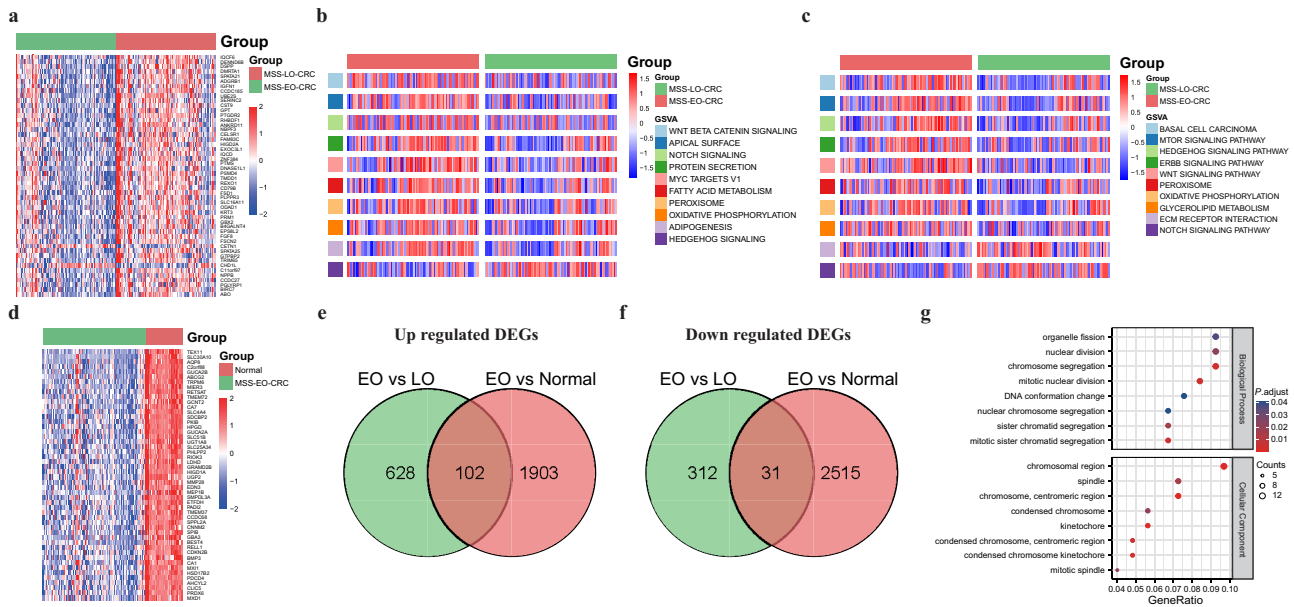
**Fig. 3 Identification of unique gene signatures in early-onset CRC. a** The heatmap of the top 50 DEGs in MSS-EO-CRC compared to MSS-LO-CRC. **b** Heatmap showing the hall markers differences between MSS-EO-CRC and MSS-LO-CRC. **c** Heatmap showing the biological pathway differences between MSS-EO-CRC and MSS-LO-CRC. **d** The heatmap of the top 50 DEGs in MSS-EO-CRC compared to normal samples. **e, f** Up-regulated and down-regulated DEGs, specifically in MSS-EO-CRC, respectively. **g** GO enrichment analysis of unique gene signatures in MSS-EO-CRC. MSS-EO-CRC early-onset colorectal cancer with microsatellite stable, MSS-LO-CRC late-onset colorectal cancer with microsatellite stable, DEGs differentially expressed genes.

MSS-EO-CRC patients, including TP53, FBXW7, KRAS, TGIF1, and CXCL9. Meanwhile, PSD, B2M, HDAC2, and LARP4B might act as the driving genes for the tumorigenesis of right-sided MSS-EO-CRC patients (Fig. 7b).

In addition, multiple studies indicated that right-sided CRC has significantly higher expression of EGFR and its ligands than left-sided ones[12,17,18]. According to the available information on tumor location and MSS status from TCGA, we separately selected 25 MSS-EO-CRC and 285 MSS-LO-CRC patients to evaluate the expression pattern of the above molecules. As is depicted in Fig. 7c, d, MSS-EO-CRC has similar expression changes of AREG with MSS-LO-CRC, whereas EREG has distinct expression characteristics in different sidedness of MSS-LO-CRC but MSS-EO-CRC patients.

## DISCUSSION

The rising trend of EO-CRC will impose an immersive socio-economic burden in a modern-aged society. To reduce the incidence of EO-CRC, the underlying biological mechanism for the tumorigenesis of EO-CRC could provide novel insights to hinder the development of CRC in individuals younger than 50 years of age. Based on tumor stage and gender, we matched MSS-EO-CRC patients with MSS-LO-CRC ones to comprehensively characterize the TME and gene expression patterns of MSS-EO-CRC. Furthermore, a risk score was built to predict the prognosis and immunotherapeutic treatment response of MSS-EO-CRC patients.

We first demonstrated that MSS-EO-CRC patients have a similar composition of tumor-infiltrating immune cells and stromal components with MSS-LO-CRC ones. This finding is in line with data published by Ugai et al., which showed a comparable proportion of nine subsets of T cells, three subtypes of macrophages, and eight subgroups of myeloid cells between MSS-EO-CRC and MSS-LO-CRC patients[19]. Meanwhile, these two CRC cohorts also have identical response rates to ICP inhibitors. Since only CRC patients with MSS were included in our study, most of these subjects are refractory to immune monotherapy, mainly caused by the low levels of tumor-infiltrating lymphocytes and tumor mutation burden[20,21]. According to the classification of

CMSs, our study displayed that CMS2 was the dominant molecular subtype for both MSS-EO-CRC and MSS-LO-CRC patients. In contrast, one clinical study indicated that EO-CRC has a comparable composition of CMS2 with LO-CRC, but also explicitly showed CMS1 was the most common subtype in EO-CRC[22]. This discrepancy is primarily due to the distinct inclusion criteria for MSS status. Furthermore, we proved that MSS-EO-CRC patients have a similar OS and RFS as MSS-LO-CRC, which aligns with multiple studies[23–25]. Therefore, MSS-EO-CRC patients have a similar TME landscape and comparable survival with MSS-LO-CRC ones.

Comprehensively illustrating the enriched pathways can give us more insights into the different potential mechanisms of tumorigenesis between EO-CRC and LO-CRC. EO-CRC patients have more robust Wnt signaling activation than LO-CRC[26]. Consistently, our study depicted that the mTOR signaling pathway and Wnt signaling pathway might play a more significant role in promoting the progression of MSS-EO-CRC than counterparts. It is widely known that Wnt and mTOR pathways play a critical role in promoting the progression of cancer. Therefore, MSS-EO-CRC patients may be more sensitive to Wnt or mTOR-targeting drugs than MSS-LO-CRC ones. In addition, 133 DEGs were identified as the unique gene signatures for the MSS-EO-CRC cohort. The biological enrichment analysis depicted that these genes were involved with the cellular mitosis of cancer cells. Previous studies reported that loss of mitosis regulation could lead to the carcinogenesis via the dysregulated cell cycle and aberrant proliferation[27]. Also, several mitosis-associated molecules are involved with the tumorigenesis and metastasis of cancer[28–31], including CRC. Therefore, the identified gene sets may participate in the development and progression of MSS-EO-CRC via regulating mitosis. These results indicated that MSS-EO-CRC has distinct patterns of molecular mechanism and gene expression compared to MSS-LO-CRC.

Meanwhile, MSS-EO-CRC has the highest WASF1 expression and the lowest TNFRSF14 compared to MSS-LO-CRC and controls. Our study also proved that WASF1 had a detrimental role in MSS-EO-CRC patients, whereas TNFRSF14 seems a protective factor.
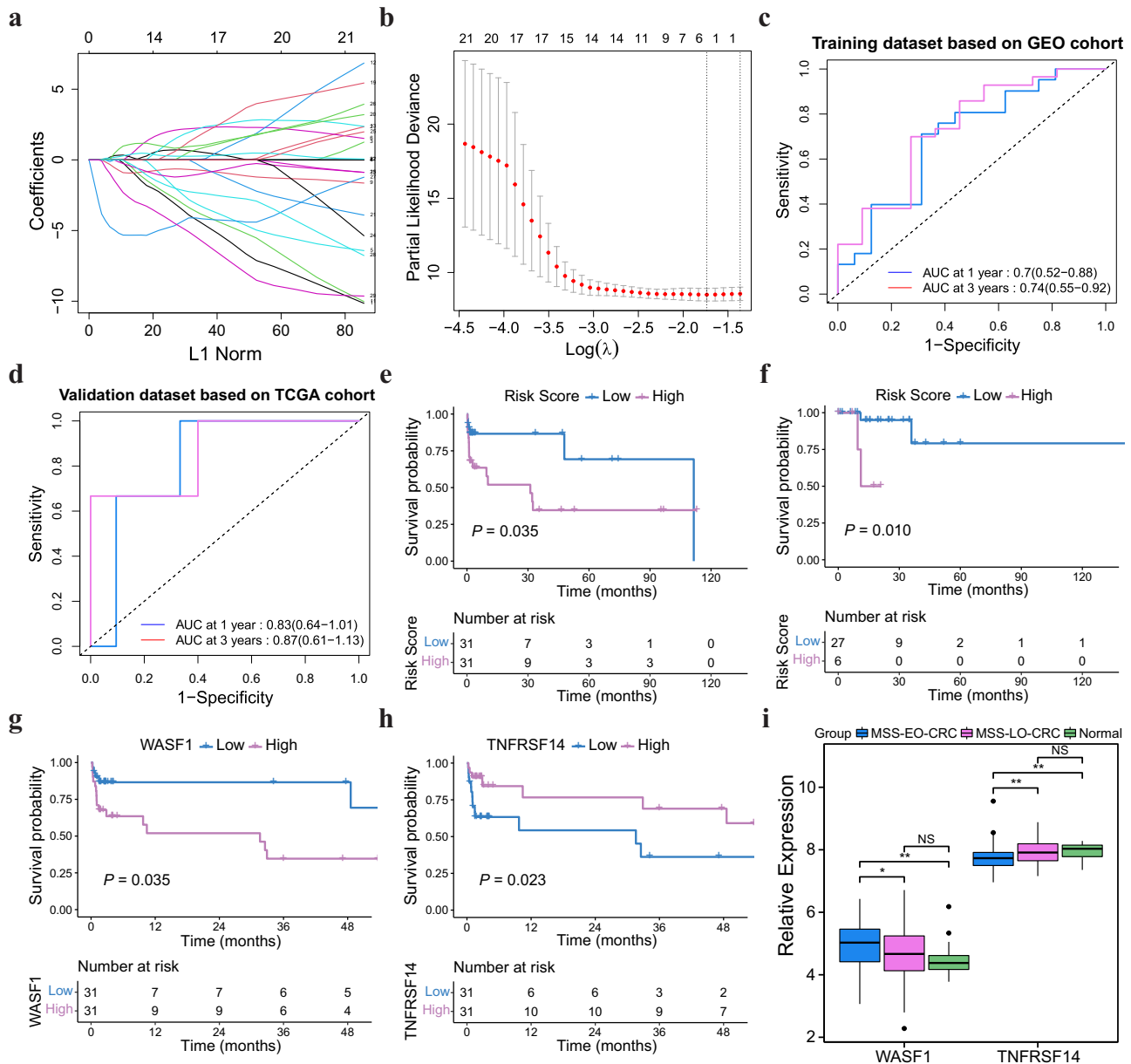
**Fig. 4 Construction of risk score and validation. a** LASSO coefficient of the 29 genes. **b** The 10-fold cross-validation for variable selection in the LASSO model. The dotted line shows the optimal lambda values by using the minimum criteria. **c, d** Time-dependent ROC analyses of the risk score in the training and validation sets, respectively. **e, f** Kaplan–Meier plot of the risk score in the training and validation sets, respectively. **g, h** Kaplan–Meier plot of WASF1 and TNFRSF14 in the training set, respectively. **i** Expression of WASF1 and TNFRSF14 among MSS-EO-CRC, MSS-LO-CRC, and Normal. The upper and lower boundaries represented the first and third quartiles, respectively, while the central line signified the median. The whiskers extended to the most distant data points not considered as outliers (within 1.5 times the interquartile range), and outliers were displayed as points above and below the box-and-whisker diagram. The log-rank test $P$ values were shown for each Kaplan–Meier plot. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$. ROC receiver operating characteristic, AUC area under the curve, MSS-EO-CRC early-onset colorectal cancer with microsatellite stable, MSS-LO-CRC late-onset colorectal cancer with microsatellite stable, NS no significance.

WASF1, also known as WAVE1, could activate the actin-related protein 2/3 complex, causing actin polymerization[32]. Due to the actin cytoskeleton's critical role in mediating cancer cell migration to the blood or lymphatic system[33,34], WASF1 has an essential role in cancer metastasis and invasion. Many studies reported that the down-regulation of WASF1 could significantly inhibit the progression and invasion of prostate cancer and ovarian cancer[35,36] and promote anti-drug-induced apoptosis of leukemia cells[37,38]. It has been reported that WASF1 depletion could decrease the proliferative and invasive ability of epithelial ovarian cancer (EOC) via the PI3K/AKT and p38/MAPK signaling pathways[35]. Also,

elevated expression of WAVE1 is associated with a worse prognosis in EOC[39], which is in line with our findings. However, no available studies depicted the biological function of WASF1 in CRC. TNFRSF14, also known as tumor necrosis factor receptor superfamily 14, encodes the receptor HVEM activating either co-stimulatory or co-inhibitory signaling pathways on immune cells[40,41]. It is expressed in lymphocytes and myeloid lineage cells and highly expressed in endothelial cells and adipocytes[42]. TNFRSF14/BTLA has the similar inhibitory effect with PD-L1/PD-1 to attenuate the activation of T helper cells[43]. Recently, increasing studies indicated the functional activity of TNFRSF14 in
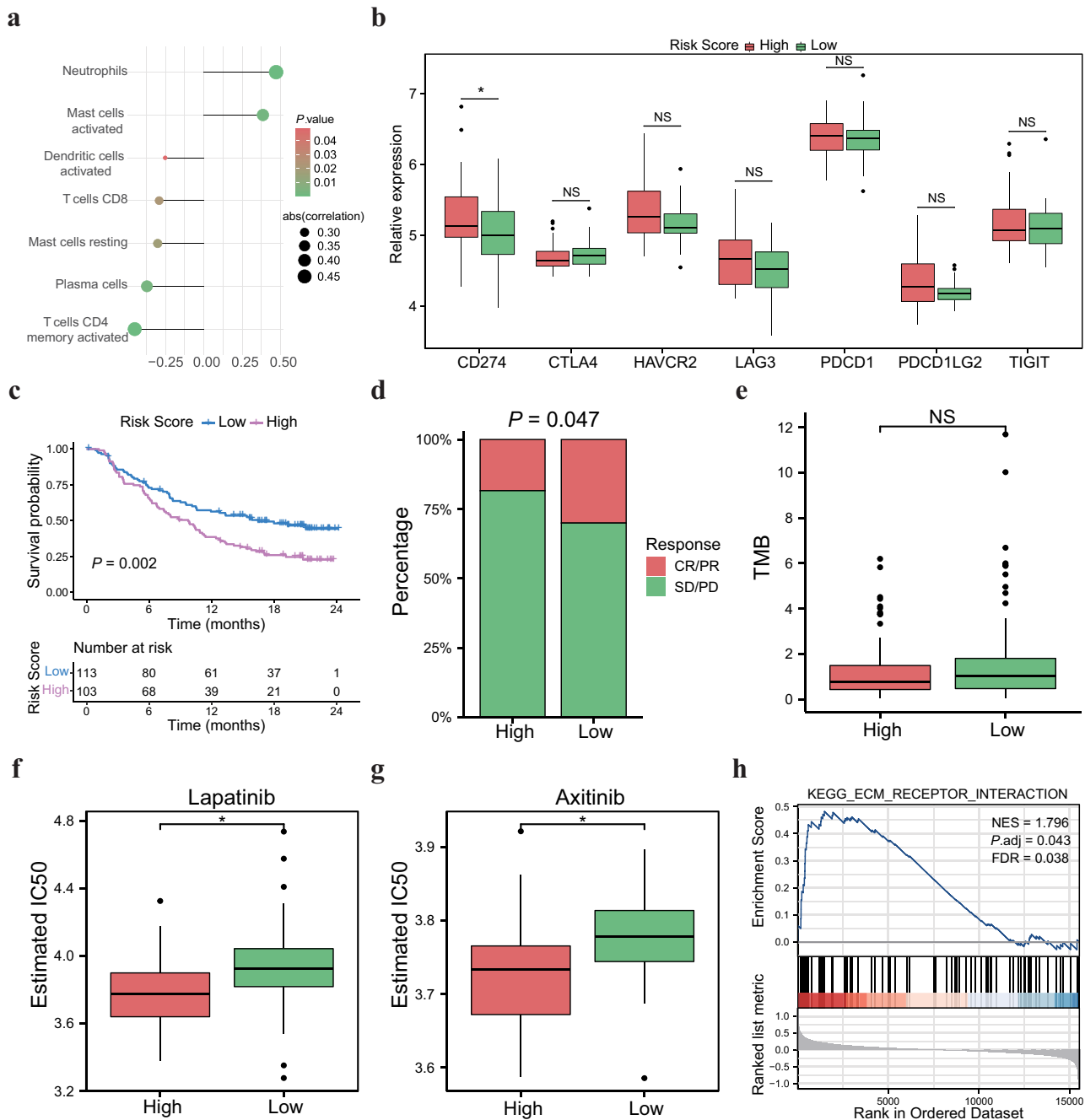
**Fig. 5 Comparison of the tumor microenvironment and immunotherapeutic responses between high and low-risk score groups based on the GEO datasets. a** Correlation analysis between tumor-infiltrating immune cells and the risk score. **b** Expression difference of immune checkpoint between high and low-risk score cohorts. **c** Survival curve showing the low-risk score group had a better prognosis than the high-risk score group. The log-rank test $P$ value were shown on the plot. **d** The proportion of patients with clinical response to anti-PD-L1 treatment in high or low-risk score groups. SD, stable disease; PD, progressive disease; CR, complete response; PR, partial response. **e** Differences in the neoantigen burden between high and low-risk score groups. **f, g** Distribution of estimated IC50 of lapatinib and axitinib between high and low-risk score groups, respectively. **h** GSEA result of the risk score. In the boxplot, the upper and lower boundaries represented the first and third quartiles, respectively, the central line signified the median and the whiskers extended to the most distant data points not considered as outliers (within 1.5 times the interquartile range). Outliers were displayed as points above and below the box-and-whisker diagram. ***$P$ < 0.001; **$P$ < 0.01; *$P$ < 0.05. GSEA gene set expression analysis, NS no significance.

cancer[44–46]. Boice et al. found that it could oppose lymphoma development via the inhibitory cell-cell interactions with BTLA[44]. In bladder cancer, the knockdown of TNFRSF14 significantly enhanced the proliferation of bladder cancer cells through the activation of the Wnt/β-catenin-dependent pathway[46]. Conflicting results have been reported on the prognostic effect of TNFRSF14 on cancer patients; increased expression of TNFRSF14 was correlated with worse OS in chronic lymphocytic leukemia and clear cell renal cell carcinoma[45,47], whereas the opposite correlation was observed in breast cancer and bladder cancer[48,49]. Interestingly, our study indicated that MSS-EO-CRC patients with higher TNFRSF14 expression have better OS than lower ones. Hence, WASF1 and TNFRSF14 have the potential to participate in the development and progression of MSS-EO-CRC.
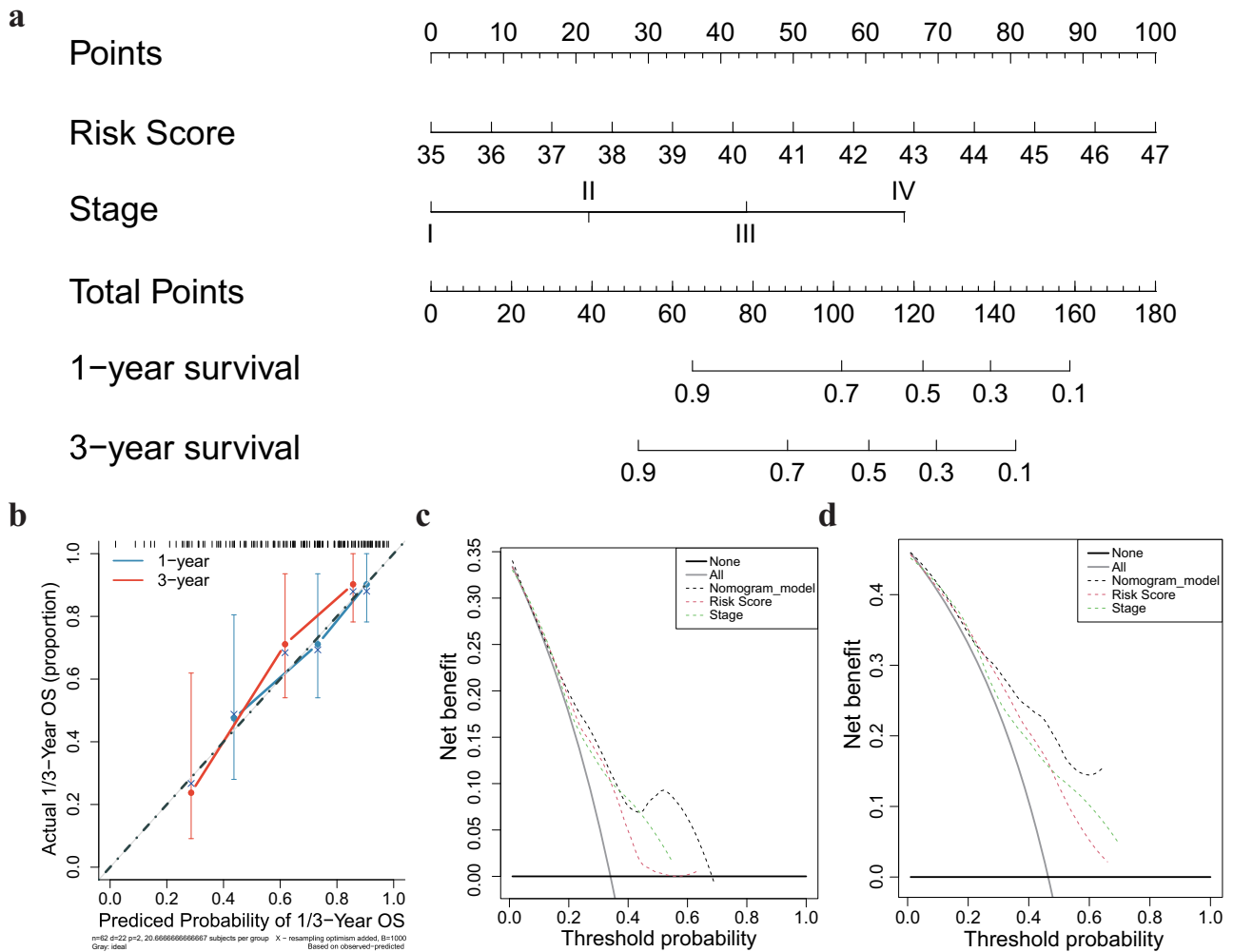
**Fig. 6 Nomogram construction and validation. a** A nomogram predicting survival rate at one and three-year after surgery for MSS-EO-CRC patients. **b** Calibration curves for the nomogram. **c, d** DCA curves depicting the comparison between the nomogram and either risk score or tumor stage for predicting 1- and 3-year overall survival in MSS-EO-CRC, respectively. MSS-EO-CRC early-onset colorectal cancer with microsatellite stable, DCA decision curve analysis.

To our knowledge, this is the first study proposing a prognostic model based on gene expression profiles for MSS-EO-CRC patients. Our risk score was associated with tumor-infiltrated immune cells of MSS-EO-CRC and had a reliable prediction on the prognosis of patients. It is widely recognized that tumors with the infiltration of T cells and PD-L1 expression in the parenchymal are more likely to acquire clinical responses to ICP inhibitors[50]. Here, the risk score has been demonstrated to reflect such features of TME in MSS-EO-CRC. By applying this score to the anti-PD-L1 treatment cohort of metastatic urothelial cancer, we found that the low-risk score group was associated with the immune-inflamed phenotype and higher infiltration of CD8 + T cells, thus, better prognosis and a higher response rate. Besides, the nomogram model was constructed based on the risk score and tumor stage to predict the survival of MSS-EO-CRC patients. We also proved that this model could provide a more reliable prediction than either risk score or tumor stage alone. Consequently, the risk score and the nomogram model could contribute to evaluating the prognosis and immunotherapeutic responses of MSS-EO-CRC patients.

Furthermore, many studies performed the genomic mutational comparison between EO-CRC and LO-CRC patients[7,8,51]. Even so, they failed to depict the mutational landscape in different sidedness of EO-CRC patients. In the present study, we

demonstrated that five genes might act as the driver gene for left-sided MSS-EO-CRC, namely TP53, FBXW7, KRAS, TGIF1, and CXCL9. In contrast, PSD, B2M, HDAC2, and LARP4B might be involved in the development of right-sided MSS-EO-CRC. Several studies consistently pointed out that EO-CRC patients have more frequent TP53 alterations than LO-CRC[7,52,53]. Pilozzi et al. also showed that KRAS mutations have higher rates in EO-CRC than LO-CRC[54]. In addition, we found that two subtypes of CRC have nearly similar expression patterns of EGFR and its ligands except for EREG. These findings indicated that MSS-EO-CRC patients have distinct mutational spectrums in different sidedness.

The primary limitation of our study is the relatively low number of MSS-EO-CRC patients, which made us unable to construct a prognostic model for long-term survival. Meanwhile, a sizeable MSS-EO-CRC cohort is needed to further validate our model's predictive reliability. Due to the unavailability of cell lines or animal models particularly associated with MSS-EO-CRC, we failed to assess the biological function of WASF1 and TNFRSF14 in this subgroup CRC. Although this study included a limited number of EO-CRC subjects with MSS from TCGA, we initially hinted that distinct driver genes might play a significant role in the tumorigenesis of different-sided MSS-EO-CRC. On the other hand, our study comprehensively characterized the molecular and clinical features of MSS-EO-CRC and then proposed a prognostic
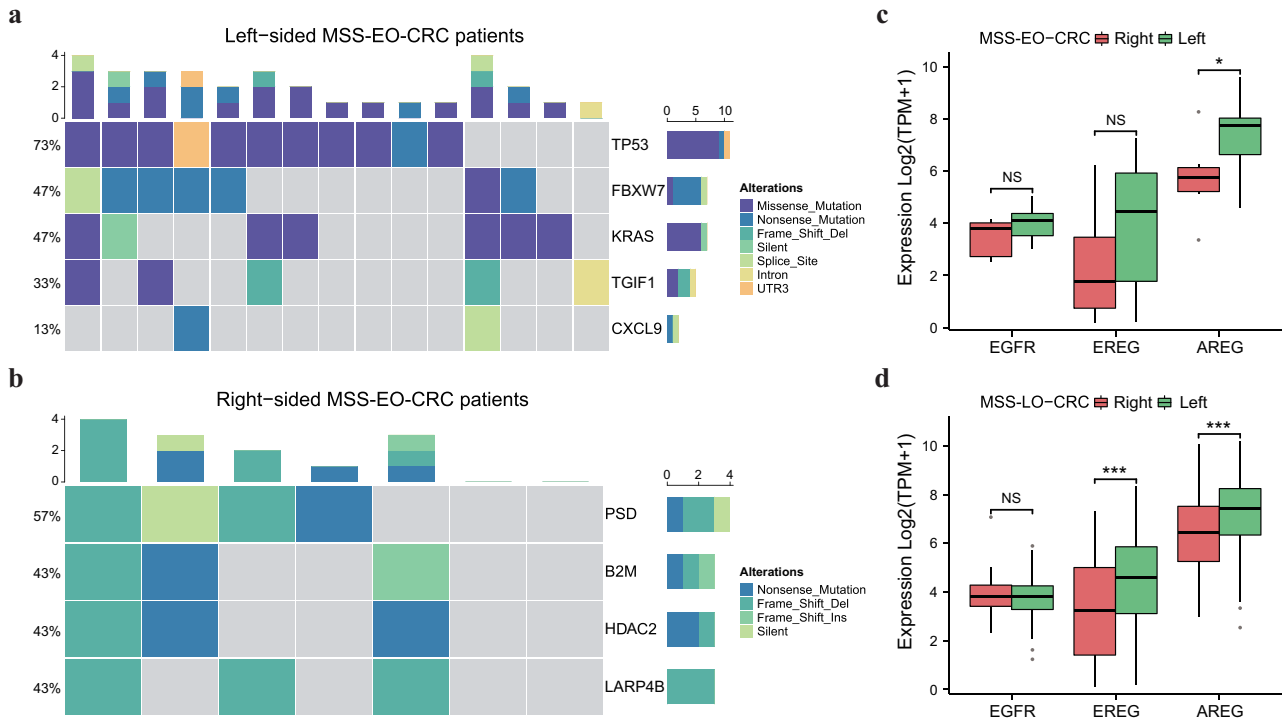
**Fig. 7 Mutational spectrums and EGFR-related molecular expression in different sidedness of early-onset CRC. a, b** Mutational spectrums of left-sided and right-sided MSS-EO-CRC patients, respectively. **c, d** Expression of EGFR-related molecules in MSS-EO-CRC and MSS-LO-CRC, respectively. In the boxplot, the upper and lower boundaries represented the first and third quartiles, respectively, the central line denoted the median and the whiskers extended to the most distant data points not considered as outliers (within 1.5 times the interquartile range). Outliers were displayed as points above and below the box-and-whisker diagram. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$. MSS-EO-CRC early-onset colorectal cancer with microsatellite stable, MSS-LO-CRC late-onset colorectal cancer with microsatellite stable, NS no significance.

model to predict the patients' survival and ICP inhibitors' response.

MSS-EO-CRC has specific gene signatures and different patterns of tumorigenesis from MSS-LO-CRC, whereas they present a similar TME characterization and prognosis. A robust risk score and a nomogram model were established to potentially predict OS and immunotherapeutic responses of MSS-EO-CRC patients, which may contribute to identifying high-risk patients suitable for more intensive therapy.

## MATERIALS AND METHODS

### Data collection and processing

A comprehensive genomic analysis based on available datasets of CRC has been performed. Searching strategy ("colon" or "colorectal" or "rectal") and ("cancer*" or "neoplas*" or "dysplasia") and ("homo sapiens") and ("gse") was conducted on Gene Expression Omnibus (GEO) database to find all suitable CRC datasets. The eligibility criteria of GEO datasets for inclusion in our study were listed in the following (Supplementary Fig. 3): (1) Sequencing data type: transcriptional profiles; (2) Sample type: tissue; (3) Samples size: larger than 20; (4) Clinicopathological information: MSI status, age, and tumor stage. Considering the heterogeneity of GEO datasets across different platforms, a total of six GPL570 platform-based datasets (GSE39582, GSE39084, GSE9348, GSE170999, GSE18088, and GSE75316) were enrolled in this study (Supplementary Fig. 3), among which GSE39582 and GSE9348 contained corresponding normal samples, and GSE39582 and GSE39084 provided survival information. The clinical data of those datasets were downloaded from the corresponding GEO website or published literature, and the details are shown in Supplementary Table 7. Only CRC patients with MSS were recruited to exclude the known genetic effects of inherited cancer syndrome. Nearest

neighbor matching based on tumor stage and gender was performed to match MSS-EO-CRC patients with MSS-LO-CRC ones for genetic and survival analysis in the ratio of 1:1 using the *MatchIt* R package[55]. The standard mean difference evaluated the matching quality before and after matching for each covariate, depicted in Supplementary Fig. 4a–c. 176 CRC patients and 31 normal controls were selected from these six GEO datasets in this study. The robust multichip average algorithm was conducted to uniformly merge the raw CEL files of the above-enrolled subjects for background correction and normalization. The combat function of the *sva* R package and the normalizeBetweenArrays function of the *limma* R package was sequentially applied to remove the batch effects and perform quantile normalization on the merged GEO dataset (Supplementary Fig. 5). The probes were annotated into gene symbols based on the GPL570 annotation files. When multiple probes matched one gene, we regarded the median of these probes as its expression value. In total, 15,620 protein-coding genes were annotated in the merged dataset. Therefore, this final GEO dataset was considered the normalized expression profiles of CRC patients and normal controls.

The Cancer Genome Atlas (TCGA) somatic mutation data were obtained using *TCGAbiolinks* R package[56]. As for gene expression profiles from the TCGA-COAD and READ cohort, the FPKM (fragment per kilobase per million) and counts data, as well as the corresponding clinical data, were downloaded from the Genomic Data Commons (GDC) data portal. Moreover, the immunohistochemistry staining determined the MSI status of these patients according to the expression of mismatch repair proteins. 33 MSS-EO-CRC patients with MSS were selected from the TCGA cohort.

Anti-PD-L1 treatment cohort derived from a multicenter, single-arm clinical trial (IMvigor210) provided the transcriptional profiles and clinical follow-up data of patients with metastatic urothelial

cancer, which was used as the dataset for predicting drug responses for PD-L1 inhibitors[57].

### Immune estimation of the TME
CIBERSORT was utilized to estimate the infiltrating level of 22 immune cells consisting of innate and adaptive immune subsets in the TME[58]. Furthermore, the ESTIMATE algorithm was applied to evaluate the enrichment score of immune and stromal components in the TME, including the immune score, stromal score, and estimate score[59].

### CMS subtypes
The consensus molecular subtype (CMS) of the merged GEO dataset was determined using the single sample predictor implemented in the R package *CMSclassifier*[15].

### Prediction of immunotherapy response
The Tumor Immune Dysfunction and Exclusion (TIDE) score of each sample was calculated to predict drug response to immune checkpoint (ICP) blockade by applying the TIDE algorithm to the expression profiles[60]. Also, the subclass mapping (SubMap) algorithm was utilized to predict the immunotherapy responses by identifying the common subtypes between our expression profiles and one published transcriptional dataset consisting of 47 melanoma patients who received the anti-PD1 or anti-CTLA-4 treatment[61,62].

### Estimation of drug sensitivity
Drug sensitivity was evaluated by the predicted half maximal inhibitory concentration (IC50) based on the analysis of gene expression profiles using the R package *pRRophetic*[63].

### Differential expression analysis
Differential expression analysis was performed to identify the DEGs in MSS-EO-CRC patients versus normal controls and MSS-EO-CRC patients versus MSS-LO-CRC ones of merged GEO datasets via the *limma* R package[64]. Any gene with a P value of <0.05 and |log₂(Fold change)| > 0.2 was considered the DEGs. Furthermore, DEGs consistently changed in the above comparisons were identified as the genes that were specifically dysregulated in MSS-EO-CRC patients.

### Gene set variation analysis
GSVA was conducted to estimate the enrichment scores of signaling pathways and hall marker gene sets using *GSVA* R package[65]. Then, differential analysis was performed to acquire the significantly enriched pathways and hall markers in each patient's cohort. The gene sets were derived from the MSigDB database (https://www.gsea-msigdb.org)[66,67].

### Gene Ontology analysis and Gene set enrichment analysis
GO analysis and GSEA were conducted to determine the potential biological function related to genes or prognostic model using *ClusterProfiler* R package with the *P*-value corrected by Benjamini–Hochberg method[68]. The following parameters were used for GSEA: nPerm = 1000, minGSSize = 10, and maxGSSize = 500. Adjusted *P*-value < 0.05 was regarded as significant.

### Construction of the prognostic model
To predict the OS of MSS-EO-CRC patients, we constructed a prognostic model based on the genes dysregulated in these patients. At first, univariate Cox regression of the above genes was performed to identify the prognostic genes with *p*-values less than 0.2.

Secondly, LASSO Cox regression was performed to reduce dimensionality and select the optimal parameters from the above prognostic genes. We applied ten-fold cross-validation to determine the lambda values and select the best one with the least partial likelihood of deviance. Next, the optimal genes were identified according to the selected lambda.

Thirdly, stepwise Cox regression was conducted to determine the best model choice from the above optimal genes with the bidirectional algorithm and the Akaike information criterion. Each parameter would be assigned a regression coefficient, and a risk score was generated using the following formula:

$$\text{Risk score} = \sum_{n=1}^{\text{Num}} (\text{Expression}_n * \text{RC}_n)$$

Where Num refers to the number of genes, Expression$_n$ represents the expression level of gene$_n$, and RC$_n$ is the regression coefficient of gene$_n$.

Furthermore, univariate and multivariate cox regression were sequentially used to identify the independent prognostic factors with a *p*-value < 0.05 from four variables: age, gender, tumor stage, and risk score. Then, the nomogram model was constructed by integrating the above factors to predict the one-year and three years OS for MSS-EO-CRC patients. Moreover, the calibration curve was conducted to evaluate the goodness-of-fit of the nomogram model. DCA was performed to assess the model's reliability by calculating the clinical net benefit for patients at each threshold probability. Besides, Harrell's C-index were calculated to evaluate the prediction capability of our nomogram model.

*Survival* and *glmnet* R packages were used to perform the Cox regression and LASSO Cox regression analysis, respectively. We applied *Survminer* R package to select the best cut-off point for distinguishing high and low-risk score groups in this study.

### MutSigCV
The MutSigCV (version 1.4.1) algorithm was performed to determine SMGs in specified cohorts of patients[69]. Default settings were used to select the SMGs with a *p*-value < 0.001.

### Statistics and reproducibility
Correlation analysis was performed using the non-parametric Spearman method. The two-sided unpaired Wilcoxon rank-sum test or two-sided Kruskal–Wallis test were conducted to assess the statistical difference of continuous variables. We applied the Benjamini–Hochberg method to correct the *p*-values of multiple testing. The statistical difference among categorical variables was calculated using a chi-squared test. The survival difference between groups was evaluated by a log-rank test in the Kaplan–Meier plot. All analyses were done using R software (version 4.1.0) and MATLAB R2021b. *P*-value < 0.05 was regarded as statistically significant.

### Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY
GEO datasets are publicly available in the National Center for Biotechnology Information Portal (https://www.ncbi.nlm.nih.gov/geo/), including GSE39582, GSE39084, GSE9348, GSE170999, GSE18088, and GSE75316. TCGA datasets enrolled in this study are openly available in the National Cancer Institute GDC Data Portal (https://portal.gdc.cancer.gov/). TCGA data are displayed under the Project IDs "TCGA-COAD" and "TCGA-READ." IMvigor210 dataset is openly available in *IMvigor210CoreBiologies* (http://research-pub.gene.com/IMvigor210CoreBiologies/).

## CODE AVAILABILITY

## REFERENCES

1. Keum, N. & Giovannucci, E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 713–732 (2019).
2. Araghi, M. et al. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *Lancet Gastroenterol. Hepatol.* **4**, 511–518 (2019).
3. Zaborowski, A. M. et al. Characteristics of early-onset vs late-onset colorectal cancer: a review. *JAMA Surg.* **156**, 865–874 (2021).
4. Patel, S. G., Karlitz, J. J., Yen, T., Lieu, C. H. & Boland, C. R. The rising tide of early-onset colorectal cancer: a comprehensive review of epidemiology, clinical features, biology, risk factors, prevention, and early detection. *Lancet Gastroenterol. Hepatol.* **7**, 262–274 (2022).
5. Sinicrope, F. A. Increasing incidence of early-onset colorectal cancer. *N. Engl. J. Med.* **386**, 1547–1558 (2022).
6. Akimoto, N. et al. Rising incidence of early-onset colorectal cancer—a call to action. *Nat. Rev. Clin. Oncol.* **18**, 230–243 (2021).
7. Lieu, C. H. et al. Comprehensive genomic landscapes in early and later onset colorectal cancer. *Clin. Cancer Res.* **25**, 5852–5858 (2019).
8. Cercek, A. et al. A comprehensive comparison of early-onset and average-onset colorectal cancers. *J. Natl Cancer Inst.* **113**, 1683–1692 (2021).
9. Poynter, J. N. et al. Molecular characterization of MSI-H colorectal cancer by MLHI promoter methylation, immunohistochemistry, and mismatch repair germline mutation screening. *Cancer Epidemiol. Biomark. Prev.* **17**, 3208–3215 (2008).
10. Sinicrope, F. A. Lynch syndrome-associated colorectal cancer. *N. Engl. J. Med.* **379**, 764–773 (2018).
11. Gallois, C., Pernot, S., Zaanan, A. & Taieb, J. Colorectal cancer: why does side matter? *Drugs* **78**, 789–798 (2018).
12. Stintzing, S., Tejpar, S., Gibbs, P., Thiebach, L. & Lenz, H. J. Understanding the role of primary tumour localisation in colorectal cancer treatment and outcomes. *Eur. J. Cancer* **84**, 69–80 (2017).
13. Markman, J. L. & Shiao, S. L. Impact of the immune system and immunotherapy in colorectal cancer. *J. Gastrointest. Oncol.* **6**, 208–223 (2015).
14. Fletcher, R. et al. Colorectal cancer prevention: immune modulation taking the stage. *Biochim. Biophys. Acta Rev. Cancer* **1869**, 138–148 (2018).
15. Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
16. Sawayama, H., Miyamoto, Y., Ogawa, K., Yoshida, N. & Baba, H. Investigation of colorectal cancer in accordance with consensus molecular subtype classification. *Ann. Gastroenterol. Surg.* **4**, 528–539 (2020).
17. Missiaglia, E. et al. Distal and proximal colon cancers differ in terms of molecular, pathological, and clinical features. *Ann. Oncol.* **25**, 1995–2001 (2014).
18. Lee, M. S. et al. Association of CpG island methylator phenotype and EREG/AREG methylation and expression in colorectal cancer. *Br. J. Cancer* **114**, 1352–1361 (2016).
19. Ugai, T. et al. Immune cell profiles in the tumor microenvironment of early-onset, intermediate-onset, and later-onset colorectal cancer. *Cancer Immunol. Immunother.* **71**, 933–942 (2022).
20. Ganesh, K. et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 361–375 (2019).
21. Weng, J. et al. Exploring immunotherapy in colorectal cancer. *J. Hematol. Oncol.* **15**, 95 (2022).
22. Willauer, A. N. et al. Clinical and molecular characterization of early-onset colorectal cancer. *Cancer* **125**, 2002–2010 (2019).
23. Lipsyc-Sharf, M. et al. Survival in young-onset metastatic colorectal cancer: findings from cancer and leukemia group B (Alliance)/SWOG 80405. *J. Natl Cancer Inst.* **114**, 427–435 (2022).
24. Jin, Z. et al. Clinicopathological and molecular characteristics of early-onset stage III colon adenocarcinoma: an analysis of the ACCENT database. *J. Natl Cancer Inst.* **113**, 1693–1704 (2021).
25. O'Connell, J. B. et al. Do young colon cancer patients have worse outcomes? *World J. Surg.* **28**, 558–562 (2004).
26. Kirzin, S. et al. Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study. *PLoS ONE* **9**, e103159 (2014).
27. Williams, G. H. & Stoeber, K. The cell cycle and cancer. *Nat. Rev. Mol. Cell Biol.* **226**, 352–364 (2012).
28. Tsoumas, D. et al. ILK expression in colorectal cancer is associated with EMT, cancer stem cell markers and chemoresistance. *Cancer Genom. Proteom.* **15**, 127–141 (2018).
29. Wu, J., Ivanov, A. I., Fisher, P. B. & Fu, Z. Polo-like kinase 1 induces epithelial-to-mesenchymal transition and promotes epithelial cell motility by activating CRAF/ERK signaling. *eLife* **5**, e10734 (2016).
30. Hu, C. et al. ROCK1 promotes migration and invasion of non-small-cell lung cancer cells through the PTEN/PI3K/FAK pathway. *Int. J. Oncol.* **55**, 833–844 (2019).
31. Jing, Z. et al. NCAPD3 enhances Warburg effect through c-myc and E2F1 and promotes the occurrence and progression of colorectal cancer. *J. Exp. Clin. Cancer Res.* **41**, 198 (2022).
32. Takenawa, T. & Suetsugu, S. The WASP-WAVE protein network: connecting the membrane to the cytoskeleton. *Nat. Rev. Mol. Cell Biol.* **8**, 37–48 (2007).
33. Sarmiento, C. et al. WASP family members and formin proteins coordinate regulation of cell protrusions in carcinoma cells. *J. Cell Biol.* **180**, 1245–1260 (2008).
34. Machesky, L. M. Lamellipodia and filopodia in metastasis and invasion. *FEBS Lett.* **582**, 2102–2111 (2008).
35. Zhang, J. et al. WAVE1 gene silencing via RNA interference reduces ovarian cancer cell invasion, migration and proliferation. *Gynecol. Oncol.* **130**, 354–361 (2013).
36. Fernando, H. S., Sanders, A. J., Kynaston, H. G. & Jiang, W. G. WAVE1 is associated with invasiveness and growth of prostate cancer cells. *J. Urol.* **180**, 1515–1521 (2008).
37. Kang, R. et al. WAVE1 regulates Bcl-2 localization and phosphorylation in leukemia cells. *Leukemia* **24**, 177–186 (2010).
38. Zhang, Z. et al. Knockdown of WAVE1 enhances apoptosis of leukemia cells by downregulating autophagy. *Int. J. Oncol.* **48**, 2647–2656 (2016).
39. Zhang, J. et al. High level of WAVE1 expression is associated with tumor aggressiveness and unfavorable prognosis of epithelial ovarian cancer. *Gynecol. Oncol.* **127**, 223–230 (2012).
40. Cai, G. & Freeman, G. J. The CD160, BTLA, LIGHT/HVEM pathway: a bidirectional switch regulating T-cell activation. *Immunol. Rev.* **229**, 244–258 (2009).
41. Murphy, T. L. & Murphy, K. M. Slow down and survive: enigmatic immunoregulation by BTLA and HVEM. *Annu. Rev. Immunol.* **28**, 389–411 (2010).
42. Bassols, J., Moreno, J. M., Ortega, F., Ricart, W. & Fernandez-Real, J. M. Characterization of herpes virus entry mediator as a factor linked to obesity. *Obesity* **18**, 239–246 (2010).
43. Mintz, M. A. et al. The HVEM-BTLA axis restrains T cell help to germinal center B cells and functions as a cell-extrinsic suppressor in lymphomagenesis. *Immunity* **51**, 310–323.e317 (2019).
44. Boice, M. et al. Loss of the HVEM tumor suppressor in lymphoma and restoration by modified CAR-T. *Cells Cell* **167**, 405–418.e413 (2016).
45. Tang, M. et al. High expression of herpes virus entry mediator is associated with poor prognosis in clear cell renal cell carcinoma. *Am. J. Cancer Res.* **9**, 975–987 (2019).
46. Wang, L., Wang, Y., Wang, J., Li, L. & Bi, J. Identification of a prognosis-related risk signature for bladder cancer to predict survival and immune landscapes. *J. Immunol. Res.* **2021**, 3236384 (2021).
47. Sordo-Bahamonde, C. et al. BTLA/HVEM axis induces NK cell immunosuppression and poor outcome in chronic lymphocytic leukemia. *Cancers* **13**, 1766 (2021).
48. Ye, H. & Zhang, N. Identification of the upregulation of MRPL13 as a novel prognostic marker associated with overall survival time and immunotherapy response in breast cancer. *Comput. Math. Methods Med.* **2021**, 1498924 (2021).
49. Zhu, Y. D. & Lu, M. Y. Increased expression of TNFRSF14 indicates good prognosis and inhibits bladder cancer proliferation by promoting apoptosis. *Mol. Med. Rep.* **18**, 3403–3410 (2018).
50. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).
51. Berg, M. et al. Distinct high resolution genome profiles of early onset and late onset colorectal cancer integrated with gene expression data identify candidate susceptibility loci. *Mol. Cancer* **9**, 100 (2010).
52. Perea, J. et al. A clinico-pathological and molecular analysis reveals differences between solitary (early and late-onset) and synchronous rectal cancer. *Sci. Rep.* **11**, 2202 (2021).
53. Berg, M. et al. DNA sequence profiles of the colorectal cancer critical gene set KRAS-BRAF-PIK3CA-PTEN-TP53 related to age at disease onset. *PLoS ONE* **5**, e13978 (2010).
54. Pilozzi, E. et al. Left-sided early-onset vs late-onset colorectal carcinoma: histologic, clinical, and molecular differences. *Am. J. Clin. Pathol.* **143**, 374–384 (2015).
55. Ho, D., Imai, K., King, G. & Stuart, E. A. Matchlt: nonparametric preprocessing for parametric causal inference. *J. Stat. Softw.* **42**, 1–28 (2011).
56. Colaprico, A. et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).

57. Mariathasan, S. et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).

58. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

59. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).

60. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).

61. Hoshida, Y., Brunet, J.-P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Subclass mapping: identifying common subtypes in independent disease data sets. *PLoS ONE* **2**, e1195–e1195 (2007).

62. Roh, W. et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci. Transl. Med.* **9**, eaah3560 (2017).

63. Geeleher, P., Cox, N. & Huang, R. S. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS ONE* **9**, e107468 (2014).

64. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

65. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinforma.* **14**, 7 (2013).

66. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

67. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

68. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A J. Integr. Biol.* **16**, 284–287 (2012).

69. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).

## AUTHOR CONTRIBUTIONS

C.L.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft, writing—editing, visualization. X.P.Z. and J.S.: Software, validation, formal analysis, data curation, writing—review and editing, visualization. U.W., K.H., L.M., G.M.C., and J.N.: Conceptualization, writing—review. A.V.B. and J.W.: Conceptualization, writing—review, resources, project administration. F.K.: Conceptualization, writing—review and editing, resources, supervision, project administration, funding acquisition.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41698-023-00414-8.

**Correspondence** and requests for materials should be addressed to Florian Kühn.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.