RESEARCH

Open Access

Multi-omics cluster defines the subtypes of CRC with distinct prognosis and tumor microenvironment

Yuan Ma[1], Jing Li[1], Xu Zhao[1], Chao Ji[1], Weibin Hu[1], YanFang Ma[1], Fengyi Qu[1], Yuchen Sun[1] and Xiaozhi Zhang[1*]

Abstract

Background Colorectal cancer (CRC) is a complex malignancy characterized by diverse molecular profiles, clinical outcomes, and limited precision in prognostic markers. Addressing these challenges, this study utilized multi-omics data to define consensus molecular subtypes in CRC and elucidate their association with clinical outcomes and underlying biological processes.

Methods Consensus molecular subtypes were obtained by applying ten integrated multi-omics clustering algorithms to analyze TCGA-CRC multi-omics data, including mRNA, lncRNA, miRNA, DNA methylation CpG sites, and somatic mutation data. The association of subtypes with prognoses, enrichment functions, immune status, and genomic alterations were further analyzed. Next, we conducted univariate Cox and Lasso regression analyses to investigate the potential prognostic application of biomarkers associated with multi-omics subtypes derived from weighted gene co-expression network analysis (WGCNA). The function of one of the biomarkers MID2 was validated in CRC cell lines.

Results Two CRC subtypes linked to distinct clinical outcomes were identified in TCGA-CRC cohort and validated with three external datasets. The CS1 subtype exhibited a poor prognosis and was characterized by higher tumor-related Hallmark pathway activity and lower metabolism pathway activity. In addition, the CS1 was predicted to have less immunotherapy responder and exhibited more genomic alteration compared to CS2. Then a prognostic model comprising five genes was established, with patients in the high-risk group showing substantial concordance with the CS1 subtype, and those in the low-risk group with the CS2 subtype. The gene MID2, included in the prognostic model, was found to be correlated with epithelial–mesenchymal transition (EMT) pathway and distinct DNA methylation patterns. Knockdown of MID2 in CRC cells resulted in reduced colony formation, migration, and invasion capacities.

Conclusion The integrative multi-omics subtypes proposed potential biomarkers for CRC and provided valuable knowledge for precision oncology.

Keywords Colorectal cancer, Multi-omics profile, Molecular subtype, Prognostic marker, MID2

*Correspondence:
Xiaozhi Zhang
zhangxiaozhi@xjtu.edu.cn
[1] Department of Radiation Oncology, The First Affiliated Hospital of Xi'an Jiaotong University, Yanta West Road 277, Xi'an 710061, Shaanxi, China

Introduction

Colorectal cancer (CRC) is a major global health concern, which ranks third in morbidity (10.0%) and second in mortality (9.4%) worldwide, with an estimated 1.9 million new cases and 935,000 deaths yearly [1]. There are large differences in survival rate depending on stage of disease at diagnosis [2]. For patients with localized CRC,

Ma *et al. European Journal of Medical Research*   (2024) 29:207

Page 2 of 17

the 5-year survival rate is about 90%. However, approximately 20% of patients already at an advanced stage at the time of diagnosis, the 5-year survival rate drops to 12.5% [2]. Besides, the molecular heterogeneity can result in different outcomes for patients even with similar clinicopathological features [3]. To date, increasing evidence has certified that biomolecules hold great promise in predicting disease prognosis and identifying potential treatment targets. Hence, the molecular subtyping of CRC is urgently needed.

Multi-omics data refer to the amalgamation of transcriptomic, genomic, and epigenetic information that can provide a more comprehensive understanding for cancer heterogeneity. Multi-omics-based classification can help identify the most relevant biomarkers and treatment targets for various types of tumors [4–8]. The initiation and progression of CRC are driven by a series of aggressive gene mutations and epigenetic alterations [9]. By studying the multi-omics, which refers to the analysis of various biological molecules, we can gain a more holistic view of the biological characteristics underlying CRC [10]. An integrative multi-omics study revealed that early-onset CRC have higher tumor mutation burden and different biological and clinical features from late-onset CRC [11]. The CRC Subtyping Consortium proposed classic consensus molecular subtypes (CMS), which have distinct gene expression profiles, genomic alterations, immune infiltrations and therapy responses [12].

Cancer is a complex disease with high heterogeneity, the occurrence of CRC undergoing multiple gene mutations and epigenetic modifications such as DNA methylation [13–15]. DNA methylation and somatic mutation can strongly perturb gene expression [11, 16]. A meta-analysis showed that KRAS, BRAF and p53 mutations were associated with the lymphatic and distant metastases of CRC [17]. A systematic review indicated a 1.49-fold greater risk of colorectal cancer in BRCA1 mutation carriers [18]. Changes in DNA methylation also can serve as biomarkers for the diagnosis, prognosis, and treatment response of CRC [19]. Hypomethylation is observed from early adenomas to metastases, with a linear correlation between demethylation grade and disease stage [20]. LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer and inversely correlated with microsatellite instability (MSI) and CpG island methylator phenotype [21, 22]. The hypermethylation of MGMT, a DNA repair enzyme, is associated with chemotherapy response in metastatic CRC [23, 24].

This study will discuss the use of combinatorial algorithms and multi-omics data in defining the different CRC molecular subtypes and their associated prognostic implications. We identified two distinct subtypes with distinct prognosis and validated with Gene Expression Omnibus (GEO) datasets. Specifically, we also comprehensively depicted the functional annotations, immune status, somatic mutations, copy number variations (CNV), and gene expression patterns of distinct subtypes. We also developed a risk model based on the subtypes related genes and subsequently conducted in vitro experiments to validate the function of the identified gene.

Methods

Data source and preprocessing

Molecular profiles of CRC patients were retrieved from The Cancer Genome Atlas (TCGA) using the "TCGA-biolinks" R package for the multi-omics data analysis. A total of 510 CRC patients with complete RNA-seq profiles, miRNA-seq profiles, the Illumina 27 K and 450 K DNA methylation, somatic mutations, and clinicopathological features were selected for subsequent analysis. And the RNA-seq were converted to the log2 "transcripts per million (TPM)" format for subsequent analysis.

Gene expression profiles of datasets (GSE39582, GSE17538, and GSE41258) with RNA expression data and survival information were downloaded from the GEO database for external validation using the "GEO-query" R package.

Identification of subtypes through integrative multi-omics analysis

To perform clustering with "MOVICS" R package [25], the CRC multi-omics data (mRNA, lncRNA, miRNA, DNA methylation CpG sites, and somatic mutation data) were transformed to features in rows and samples in columns. There are 510 CRC patients with complete multi-omics information. Subsequently, we proceeded by selecting the top 50% of variance factors (mRNA, lncRNA, miRNA, DNA methylation CpG sites) with prognostic value (univariate Cox regression analysis, $p < 0.05$), along with genes with mutation frequencies above 0.1, for further in-depth analyses. The clustering prediction index (CPI) and Gaps-statistics based on above multi-omics data were used to determine the optimal number of subtypes. Subsequently, ten clustering algorithms: SNF, PINSPlus, NEMO, COCA, LRAcluster, ConsensusClustering, IntNMF, CIMLR, MoCluster, and iClusterBayes were used to separate CRC patients into different subtypes. Finally, we used "getConsensusMOIC()" function to identify the final clusters with high robustness based on ten clustering methods.

Evaluation of the activated of signaling pathways

To reveal the function distribution of each subtype, we downloaded 50 tumor-related Hallmark gene sets from Molecular Signatures Database (MSigDB) and 87

Ma *et al. European Journal of Medical Research*   (2024) 29:207

Page 3 of 17

metabolism related gene sets from KEGG [26, 27]. The single sample gene set enrichment analysis (ssGSEA) was used to calculated the enrichment scores for each patient, and pathways with different distribution in two subtypes were visualized in heatmaps. "ClusterProfiler" R package was used to evaluate the most significant different enrichment pathways for each subtype based on different expressed genes (DEGs).

Characterization of genetic alteration on subtypes

We analyzed the mutation landscape through the R package "maftools". Using the "oncoplot", "Oncogenic-Pathways", "somaticInteractions", "mafCompare", and "plotOncodrive" functions of this package, we analyzed the tumor mutation panorama, base conversion and transversion, amino acid mutation hotspot, mutation frequency of mutation alleles, copy number mutation, and mutual exclusion or coexistence mutation across different subtypes. The somatic copy number alteration (SCNA) data were analyzed using GISTIC2.0 algorithm on GenePattern (https://cloud.genepattern.org/).

Immune microenvironment analysis and assessment of response to immunotherapy

We also used ssGSEA and xCell to assess the distribution of immunologic functions, immune cells and stromal cells infiltration in each subtype and visualized with boxplots. And boxplots were also utilized to compare the expression of immunological checkpoint. We used the tumor immune dysfunction and exclusion (TIDE) (http://tide.dfci.harvard.edu) web application to estimate the immunotherapy response of each ESCC patient [28]. GSE78220 and IMvigor210 cohorts were used to verify the predictive value of subtype in response to immunotherapy [29, 30].

External data validation

The "MOVICS" R package was utilized to validate the repeatability of cluster analysis in other CRC cohorts. Firstly, nearest template prediction (NTP) function permits adaptable cross-platform, cross-species, and multiclass predictions, without requiring the optimization of analysis parameters. Then, we compared the prognosis of the predicted cancer subtypes in these validation cohorts.

WGCNA and prognosis model construction

"WGCNA (Weighted Gene Co-Expression Network Analysis)" R package was used to identify co-expressed gene modules correlated with cluster subtypes. We identified a total of 24 co-expression modules through the topological overlap matrix (TOM) calculation, which were marked with different colors by setting a soft threshold power $\beta = 8$, which represented genes that shared

highly similar expression patterns in CRC patients. Then, the most correlated module was identified after assessing the Pearson correlation coefficient between two subtypes and the co-expression modules. Finally, the genes of module with high trait significance were selected for further analysis.

Genes in target module with significant impact on OS were selected for least absolute shrinkage and selection operator (LASSO) regression analysis. Then, genes strongly correlated with prognosis and their coefficients were obtained. The risk score was calculated by multiplying each obtained coefficient by the corresponding gene expression and summing the total values and patients were stratified into low- and high-risk groups based on the median risk score.

Cell culture and plasmid transfection

CRC cell lines SW480 and HCT116 were purchased from Shanghai Institute of Cell Biology and cultured with Dulbecco's modified Eagle's medium (DMEM; Gibco, USA) containing 10% fetal bovine serum (FBS; Hyclone, USA) at 37℃ 5% CO_2. To investigate the impact of MID2 on cellular function, the short-hairpin RNA (shRNA) targeting MID2 and control shRNA were purchased from Sangon Biotech (Shanghai, China). Transfections were carried out according to the Lipofectamine 3000 (Invitrogen, USA) protocol.

Western blot assay

Cells were lysed in RIPA buffer (Beyotime, China), 1 mM PMSF (Beyotime, China), phosphatase inhibitor cocktail (Beyotime, China). The lysates were then denatured in 100℃, separated by 10% SDS-PAGE gel, transferred to PVDF membranes (Millipore, Sigma, USA), and blocked in 5% skimmed milk. The membranes were incubated overnight at 4% with primary antibody: MID2 polyclonal antibody (1:1000, PA5-28457, Thermo Fisher Scientific), GAPDH polyclonal antibody (1:10,000, 10494-1-AP, Proteintech), E-Cadherin polyclonal antibody (1:20,000, 20874-1-AP, Proteintech), N-Cadherin polyclonal antibody (1:2000, 22018-1-AP, Proteintech), Vimentin monoclonal antibody (1:1000, 5741 T, Cell Signaling Technology). Then secondary antibody goat anti-rabbit IgG (1:2000, A0277, Beyotime) was incubated 1 h at room temperature. After washing with TBST solution, the bands were finally visualized using an ECL reagent (Millipore, Sigma, USA).

Colony formation assay

SW480 and HCT116 cells were inoculated in a six-well plate (800/well). After 10 days, the cells were fixed with 4% paraformaldehyde and stained with 1% crystal violet.

Ma *et al. European Journal of Medical Research*   (2024) 29:207

Page 4 of 17

The number of visible colonies was counted to evaluate the colony formation ability of the cells.

Transwell assay

Cell migration and invasion were measured using Boyden chambers in 24-transwell plates (8 µm pores, Corning). 600 µL DMEM medium containing 20% FBS were added to the bottom of plates. Then, 2.5×10^4 cells suspended in a 200 µl serum-free medium were seeded to the upper chambers for migration assay, 5×10^4 cells were seeded in upper chamber pre-coated with 60 µL of Matrigel (BD) for invasion assay. After incubation for 24 h at 37 °C, the membranes were fixed with 4% paraformaldehyde and stained with 1% crystal violet.

Statistical analysis

GraphPad Prism (version 8.0, USA) and R language (version 4.2.1) were used for statistical analysis. Student's t-test and Wilcoxon test were used to compare the discrepancy of continuous data between two groups. Chi-square test and Fisher's exact test were used to compare the distribution of categorical variable. Kaplan–Meier (K-M) method and log-rank test were used to estimate the survival analysis. P value < 0.05 indicated statistical significance.

Results

Overview of multi-omics profiling of two CRC subtypes

The flowchart of this study is presented in Fig. 1. We integrated the multi-omics data of TCGA-CRC, with 510 samples having complete multi-omics and survival data used for subsequent cluster analysis. The optimal number of clusters (k = 2) was determined based on the CPI and gap statistics, the number of clusters that reach the maximum sum of these two statistics is considered optimal (Fig. 2A). In comparison to three, four, or five clusters, two clusters showed superior consistency, as confirmed by the consensus matrix (Fig. 2B, Additional file 1: Fig. S1A, B, and C). The Silhouette value, a clustering quality indicator to assess the effectiveness of clustering for each data point ranging from − 1 to 1, demonstrated that the high silhouette width of the two clusters (0.58 and 0.51), which was closer to 1 than the other clusters, represented the robustness of two subtypes and indicated better clustering performance (Fig. 2C, Additional file 1: Fig. S1E, F, and G). For consensus clustering, ten independent clustering algorithms refer the uniformity of the multi-omics
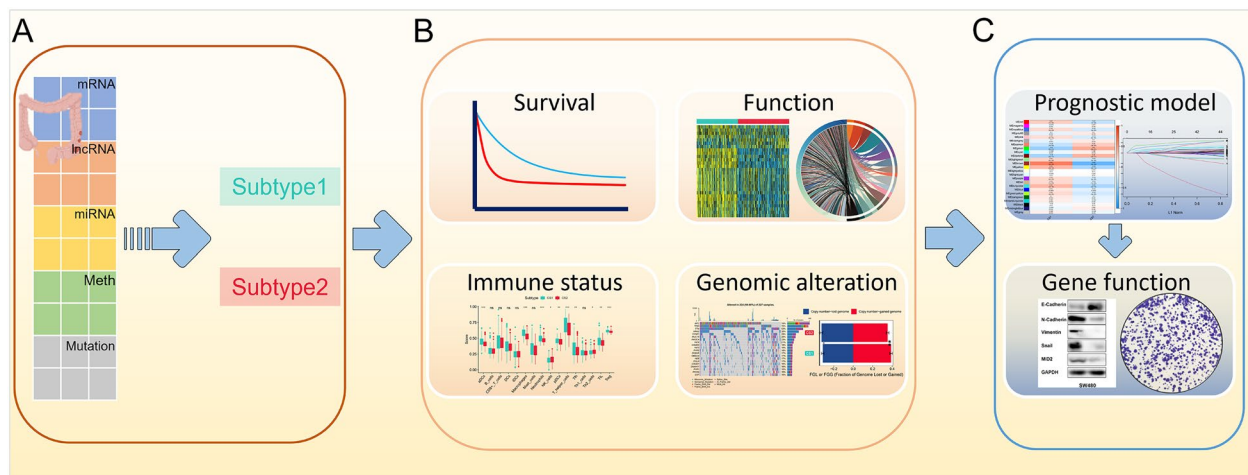
Fig. 1 Schematic diagram of the study design. **A** The mRNA, lncRNA, miRNA, DNA methylation CpG sites, and mutation data from TCGA-CRC were systematically organized into comprehensive multi-omics data, which were utilized to identify two subtypes through integrated clustering algorithms. **B** The association of subtypes with prognoses, enrichment functions, immune status, and genomic alterations were further identified. **C** The risk model, constructed through WGCNA and Cox analysis, exhibited substantial concordance with the prognosis of multi-omics subtypes, and the functionality of the molecular marker MID2 in the risk model was validated

(See figure on next page.)
Fig. 2 Molecular subtypes clustering based on TCGA-CRC multi-omics data. **A** Prediction of optimal cluster number of multi-omics data by cluster prediction index and Gap-statistics. **B** Consensus heatmap for two cluster subtypes based on multi-omics data. **C** The Silhouette value quantify sample similarity based on two cluster subtypes. **D** Clustering of CRC patients via 10 leading-edge clustering methods. **E** Visualization of multi-omics data for mRNA, lncRNA, miRNA, DNA CpG methylation sites and mutant genes. **F** Differential overall survival outcome in two subtypes. **G** Differential disease-free survival in two subtypes, log-rank test
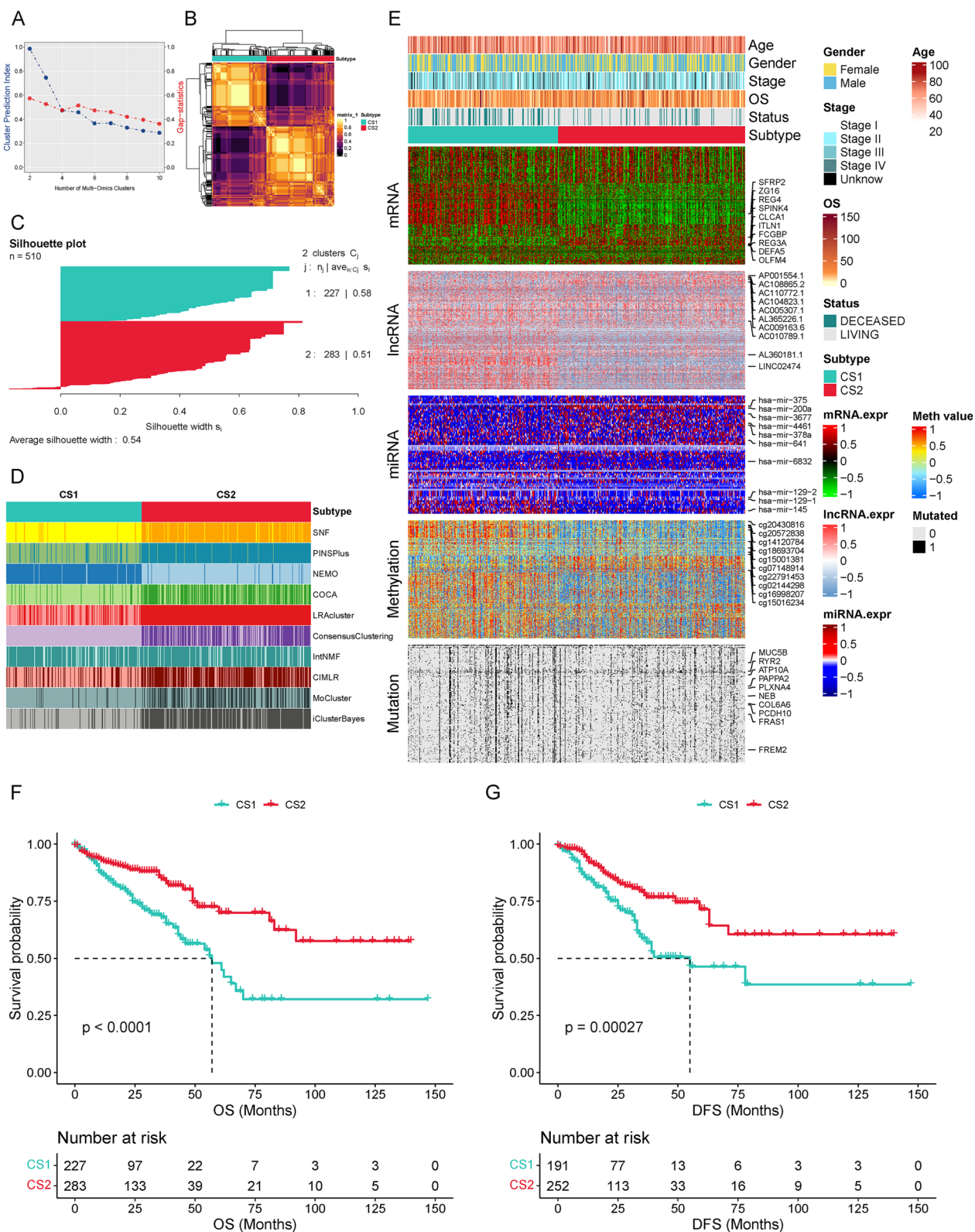
Fig. 2 (See legend on previous page.)

Ma *et al. European Journal of Medical Research*   (2024) 29:207

Page 6 of 17

cluster and we further combined the clustering results via a consensus ensemble approach with "MOVICS" R package (Fig. 2D). The heatmap revealed distinct transcriptomic, genomic, and epigenomic patterns, as well as clinicopathological features of two subtypes (Fig. 2E).

Moreover, the clinical outcomes of patients in the two subtypes were compared. The K-M plots illustrated that the patients in CS1 had worse prognoses than CS2 (OS: $p < 0.0001$; DFS: $p = 0.00027$; Fig. 2F, G). Table 1 shows the demographic features of CRC patients in the TCGA cohort. Notably, CS2 had more N0, T1–2, and I stage, while CS1 was associated with longer tumor dimension, with an average size of 1.38 ± 0.61 cm, compared to CS2 is 1.17 ± 0.50 cm ($p < 0.001$). And multivariable Cox regression analysis indicates that the impact of multi-omics subtypes on survival is independent of other clinicopathological factors (Additional file 2: Fig. S2).

Gene set variation characters of different subtypes

Metabolic reprogramming was known to provide valuable insights into metabolic alterations and the mechanisms of disease progression [26, 31]. In this study, we performed metabolic pathway analysis on each patient and observed that CS2 exhibited more metabolic pathway enrichment compared to CS1. Specifically, key metabolic pathways, such as carbon metabolism, TCA cycle, amino acid metabolism, and fatty acid metabolism were upregulated in CS2, while only Glycosaminoglycan biosynthesis-chondroitin sulfate and heparan sulfate were active in CS1, suggesting potentially differences in metabolic profile and energy utilization pattern between the two subtypes (Fig. 3A). To provide a comprehensive overview of the changes in gene expression, we conducted ssGSEA by employing Hallmark gene sets that represent distinct biological states. Figure 3B shows that CS1 was significantly associated with various malignancy pathways, including epithelial–mesenchymal transition (EMT), angiogenesis, hypoxia, TCF-β, and Notch pathways. In terms of KEGG pathways enrichment, we found that CS1 was associated with cell adhesion and other biological characteristics that are indicative of cancer, such as ECM–receptor interaction, focal adhesion, cell adhesion molecules, PI3K–Akt, Rap1 signaling pathway, and so on. And in line with previous findings, CS2 was also enriched in the progression of cellular metabolism (Fig. 3C, D).

The effect of genetic alteration on subtypes

Gene mutations and copy number alterations are critical events in tumorigenesis and cancer progression. Therefore, we further conducted a deeper analysis on the mutation patterns of different subtypes, and identified genetic alterations that were specifically

Table 1 The clinicopathological parameters of colorectal cancer patients in TCGA

Subtype	CS1 (n = 227)	CS2 (n = 283)	*p*
Age	64.71 ± 13.44	66.09 ± 12.53	0.236
Gender (%)			
Female	107 (47.1)	136 (48.1)	0.906
Male	120 (52.9)	147 (51.9)	
M stage (%)			
M0	160 (70.5)	208 (73.5)	0.654
M1	37 (16.3)	39 (13.8)	
MX	28 (12.3)	31 (11.0)	
Unknown	2 (0.9)	5 (1.8)	
N stage (%)			
N0	114 (50.2)	175 (61.8)	0.005
N1	56 (24.7)	70 (24.7)	
N2	56 (24.7)	38 (13.4)	
NX	1 (0.4)	0 (0.0)	
T stage (%)			
T1	5 (2.2)	14 (4.9)	0.001
T2	24 (10.6)	66 (23.3)	
T3	168 (74.0)	176 (62.2)	
T4	30 (13.2)	26 (9.2)	
Tis	0 (0.0)	1 (0.4)	
Stage (%)			
Stage I	25 (11.0)	66 (23.3)	0.008
Stage II	82 (36.1)	97 (34.3)	
Stage III	74 (32.6)	73 (25.8)	
Stage IV	38 (16.7)	39 (13.8)	
Unknown	8 (3.5)	8 (2.8)	
Cancer type (%)			
Colon adenocarcinoma	137 (60.4)	186 (65.7)	0.247
Colorectal adenocarcinoma	2 (0.9)	0 (0.0)	
Mucinous adenocarcinoma of the colon and rectum	29 (12.8)	28 (9.9)	
Rectal adenocarcinoma	59 (26.0)	69 (24.4)	
Longest dimension (cm)	1.38 ± 0.61	1.17 ± 0.50	0.001
Disease-free status (%)			
Disease free	133 (69.6)	209 (82.9)	0.001
Recurred/progressed	58 (30.4)	43 (17.1)	

associated with each subtype. Waterfall plots revealed that several oncogenes and tumor suppressor genes are mutated in overall cohort. The waterfall plot and comparison forest plot showed that CS1 had relatively more genes mutation than CS2. And CS1 had more TP53, TTN, SYNE1, and FAT3 mutation, while CS2 had more KRAS, RYR2, LRP1B, and SOX9 mutation (Fig. 4A, B). Cross-comparisons showed that CS1 had more TP53, CCDC136, PIK3R3, KCNA6 mutation rate, and CS2 only had more PROKR1 mutation rate (Fig. 4C). In oncogene pathway mutation analysis,
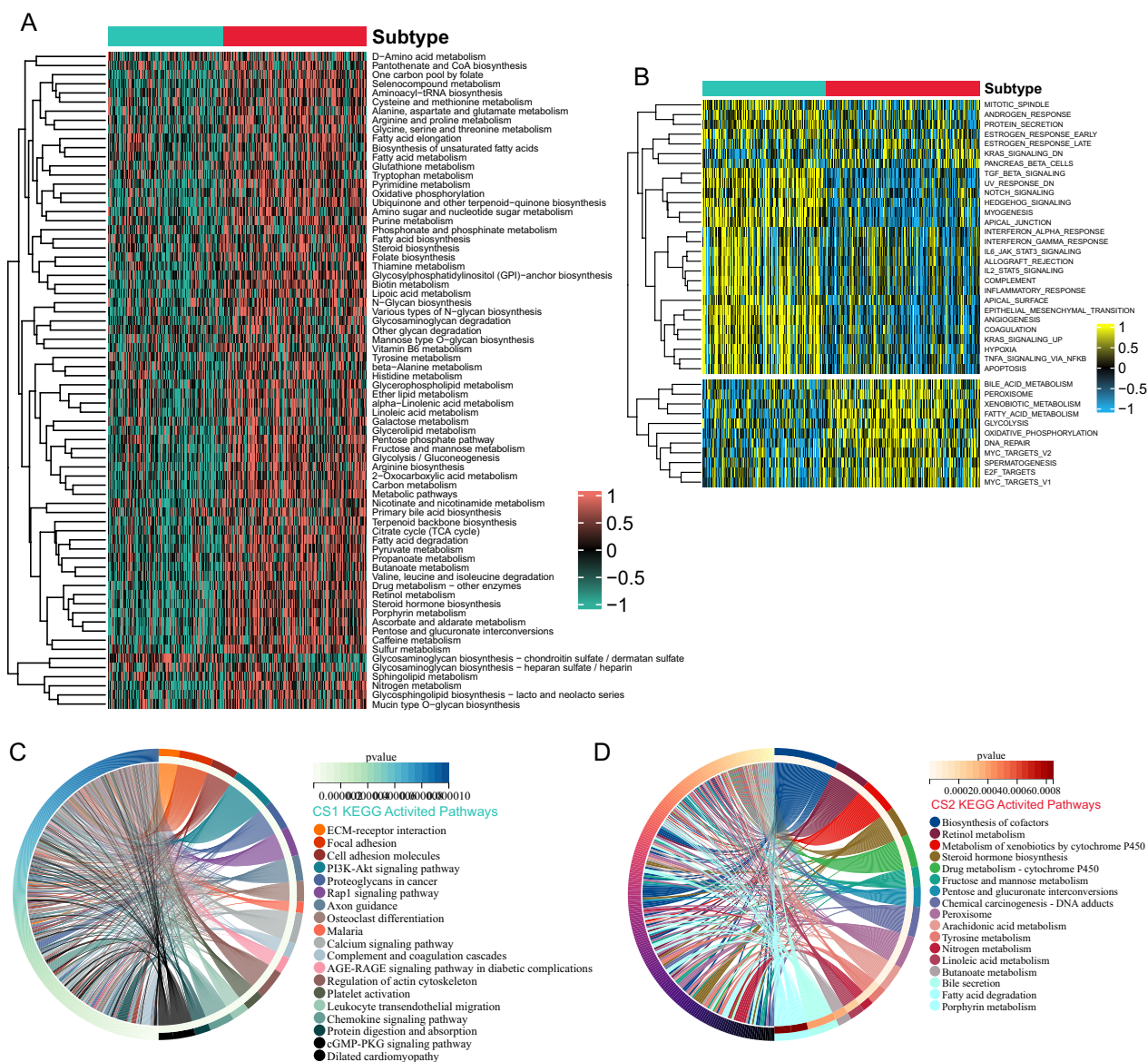
Fig. 3 Differential activity of functional enrichment pathways across two subtypes. **A** Heatmap of differentially activated metabolism signaling pathways. **B** Heatmap of differentially activated Hallmark pathways. **C** The circle plot of CS1 subtype activated KEGG pathways. **D** The circle plot of CS2 subtype activated KEGG pathways

both two subtype have similar gene in these pathways been affected, while samples in CS1 have higher mutation rates in RTK-RAS, WNT, Hippo, and TP53 pathways (Additional file 3: Fig. S3A, B). We also found that TVP23A, GIPC2, NRAS, RPL22, and KRAS were driver genes in CS1, and TMEM60, FAHD2B, ZNF365, and SHC1 were driver genes in CS2 through the plotOncodrive function of the "maftools" R package (Additional file 3: Fig. S3C, D).

Subsequently, we used the compMut and compFGA function of "MOVICS" R package to investigate the difference of tumor mutation burden (TMB) and genome alteration between two subtypes. There was no significant difference in TMB between the two subtypes ($p=0.26$), but CS1 had more copy number variations than CS2 ($p<0.1$, Fig. 4D, E). CS1 had more genomic copy number amplification ($p<0.01$), and CS2 had more copy number lost ($p<0.1$, Fig. 4E). We used GISTIC2.0 algorithm to identify recurrent SCNAs present in different subtypes, Fig. 4F and G depict that CS1 had more frequent copy number gains in chromosome regions 1q, 8q, 13q and 20q, and CS2 had more losses in chromosome

Ma *et al. European Journal of Medical Research*   (2024) 29:207

Page 8 of 17
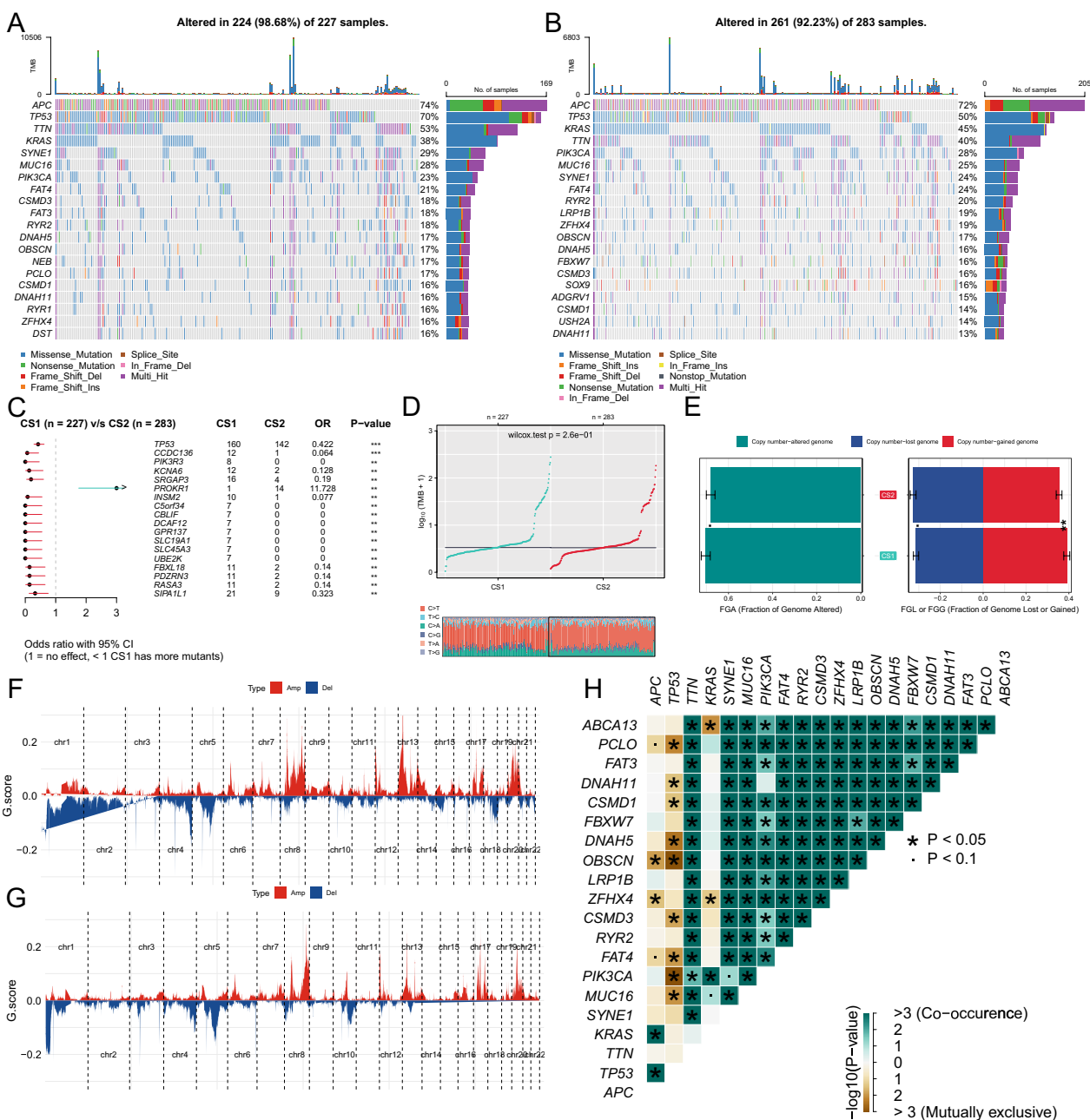
Fig. 4 Diversely genetic alterations among two subtypes. **A** The waterfall plot of the top 20 most frequently mutated genes in CS1. **B** The waterfall plot of the top 20 most frequently mutated genes in CS2. **C** Forest plot of significantly differentially mutated genes between two subtypes. **D** Comparison of tumor mutation burden (TMB) and transitions and transversions (TiTv) among two subtypes. **E** Bar plot of fraction genome altered among two subtypes. **F** The copy number amplifications and deletions among the 22 chromosomes in CS1. **G** The copy number amplifications and deletions among the 22 chromosomes in CS2. **H** The heatmap shows the mutually co-occurring and exclusive mutations of the top 20 frequently mutated genes

regions 1p, 5q, 10q, 15p, and 21p. The co-mutation plot revealed most mutation genes are co-occurrence with others, except APC, TP53, and KRAS (Fig. 4H).

Tumor microenvironment landscape across CRC subtypes

Tumor microenvironment (TME) is a complex system consisting of various cell types, extracellular matrix, and signaling molecules that play critical roles in tumor development, progression, and metastasis. Therefore,

Ma *et al. European Journal of Medical Research*   (2024) 29:207

Page 9 of 17

understanding the intricate crosstalk between tumor cells and the TME is essential to develop efficient cancer treatments. Previous results revealed that CS1 was highly enriched in Hallmark pathways related to immune response, such as interferon-α and interferon-γ response, IL-6/JAK/STAT3 signaling, IL2/STAT5 signaling and TNFA via NF-κB signaling (Fig. 3B). There were also significant differences in cellular composition between the two subtypes. CS1 had relatively higher immune cells, such as dendritic cells (DC), macrophages, neutrophils, Th, Tfh, tumor-infiltrating lymphocytes (TILs), and Treg cells, which can drive and regulate T cell-mediated immune responses and interact with each other, while CS2 had more NK cells known as cytotoxic lymphocytes of the innate immune system (Fig. 5A) [32–34]. In terms of stromal cells, CS1 had more fibroblast, endothelial cell, mesenchymal stem cell (MSC), and pericytes that can establish an inflammatory, immunosuppressive and pro-angiogenic microenvironment, while epithelial and plasma cell are more infiltrated in CS2 (Fig. 5B) [35, 36]. In addition, we utilized ssGSEA to examine the differences in immune function and found that most functions, such as APC co-inhibition and co-stimulation, T cell co-inhibition, inflammation promoting, and type I/II IFN response, are active in CS1 (Fig. 5C). Furthermore, we observed higher expression of immune checkpoints (CD274, CTLA4, IDO1, LAG3, PDCD1, and PDCD1LG2) in CS1 compared to CS2 (Fig. 5D).

Then we employed several methods to predict the potential immunotherapy response of patients in different subtypes. A web platform TIDE integrates large-scale omics data to predict immunotherapy response across various tumor types [28]. Therefore, we used TIDE to generate scores reflecting the likelihood of immunotherapy response based on transcriptomic data for each patient (Fig. 5E). The histogram showed that patients in CS2 exhibit a higher probability of responding to immunotherapy compared to those in CS1 (52% vs 24%, $p = 5.904e{-}10$, Fig. 5F). Subsequently, NTP method was used to predict the multi-omics subtypes of GSE78220 and IMvigor210 cohorts, which provide both immunotherapy response information data and transcriptional data for patients who underwent anti-PD1/PD-L1 therapy [29, 30]. Then we compared the distribution

of immunotherapy responses in the two subtypes of patients. While the Chi-square test did not demonstrate statistical significance, Fig. 5G and H illustrate that patient in CS2 exhibited higher rates of immune complete and partial response (GSE78220: 38% vs. 66%, $p = 0.1124$; IMvigor210: 19% vs. 27%, $p = 0.1192$). These results indicate that patients in CS2 may be better suited for immunotherapy.

Extra validation for molecular subtypes in GEO cohorts

To validate the molecular subtypes identified through our multi-omics analysis, we included external CRC cohorts with transcriptome and follow-up information for further analysis. Three external cohorts (GSE39582, GSE17538, and GSE41258) were downloaded and prepared for validation. Using the "limma" package, we identified the top 200 upregulated genes in CS1 and CS2 subtypes as their marker genes (Additional file 6: Table S1) and applied the NTP method to determine the subtype of each patient in the validation cohorts (Fig. 6A–C). The prognostic predictions for CS1 and CS2 were consistently observed across all three external GEO cohorts, providing robust validation for the molecular subtypes identified in our study (GSE39582: OS $p = 0.019$, DFS $p = 0.011$; GSE17538: OS $p = 0.0024$, DFS $p = 0.0035$; and GSE41258: OS $p = 0.037$, Fig. 6D–H).

Construction of a prognostic model for multi-omics molecular subtype

To enhance the clinical application of the molecular subtype, we aimed to construct a prognostic model for predicting patient outcomes. We utilized the WGCNA to investigate the relationship between gene modules and patient subtypes. Specifically, we constructed gene co-expression networks using the expression profiles of CRC patients, setting β at 8 to ensure scale-free networks (Fig. 7A). We then transformed the adjacency matrix into a topology matrix and used the average-linkage hierarchical clustering method to cluster genes, setting the minimum number of genes in each network module to 30. Next, we merged similar gene modules using the dynamic cut tree method, which resulted in 24 distinct modules (Fig. 7B). This analysis allowed us to identify the gene modules most closely associated with the molecular

(See figure on next page.)

Fig. 5 Comparison of immune status across two subtypes. **A** The boxplot of immune cell infiltrations across two subtypes. **B** The boxplot of stroma cell infiltrations across two subtypes. **C** The boxplot of immune functions across two subtypes. **D** The boxplot of the expression of immune checkpoints across two subtypes. **E** The score of immunotherapy response predicted by the TIDE method. **F** The distribution of immunotherapy responders and non-responders across two subtypes. **G** The distribution of immunotherapy response across the nearest template prediction (NTP) predicted subtypes based on GSE78220 cohort. **H** The distribution of immunotherapy response across the NTP predicted subtypes based on IMvigor210 cohort
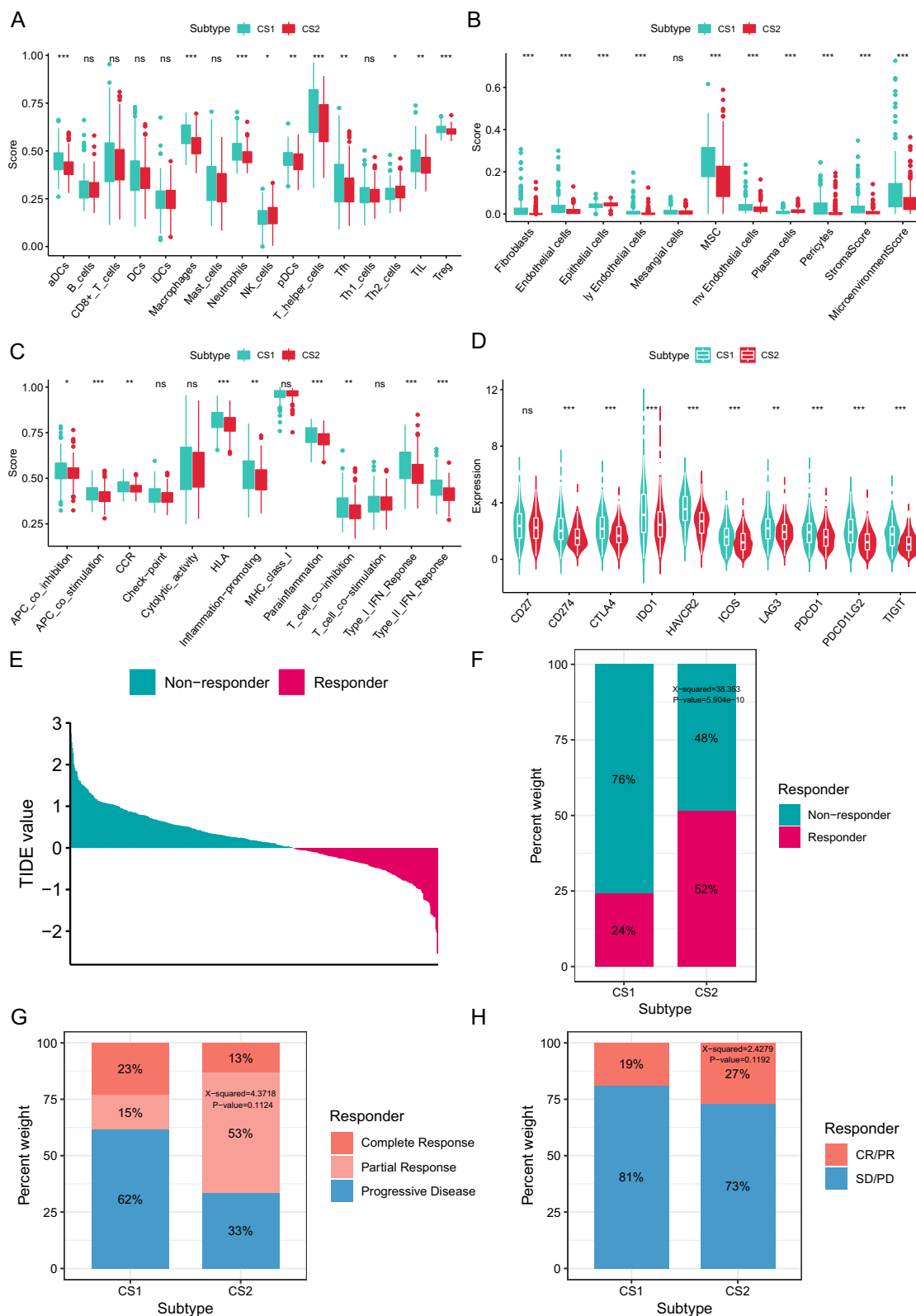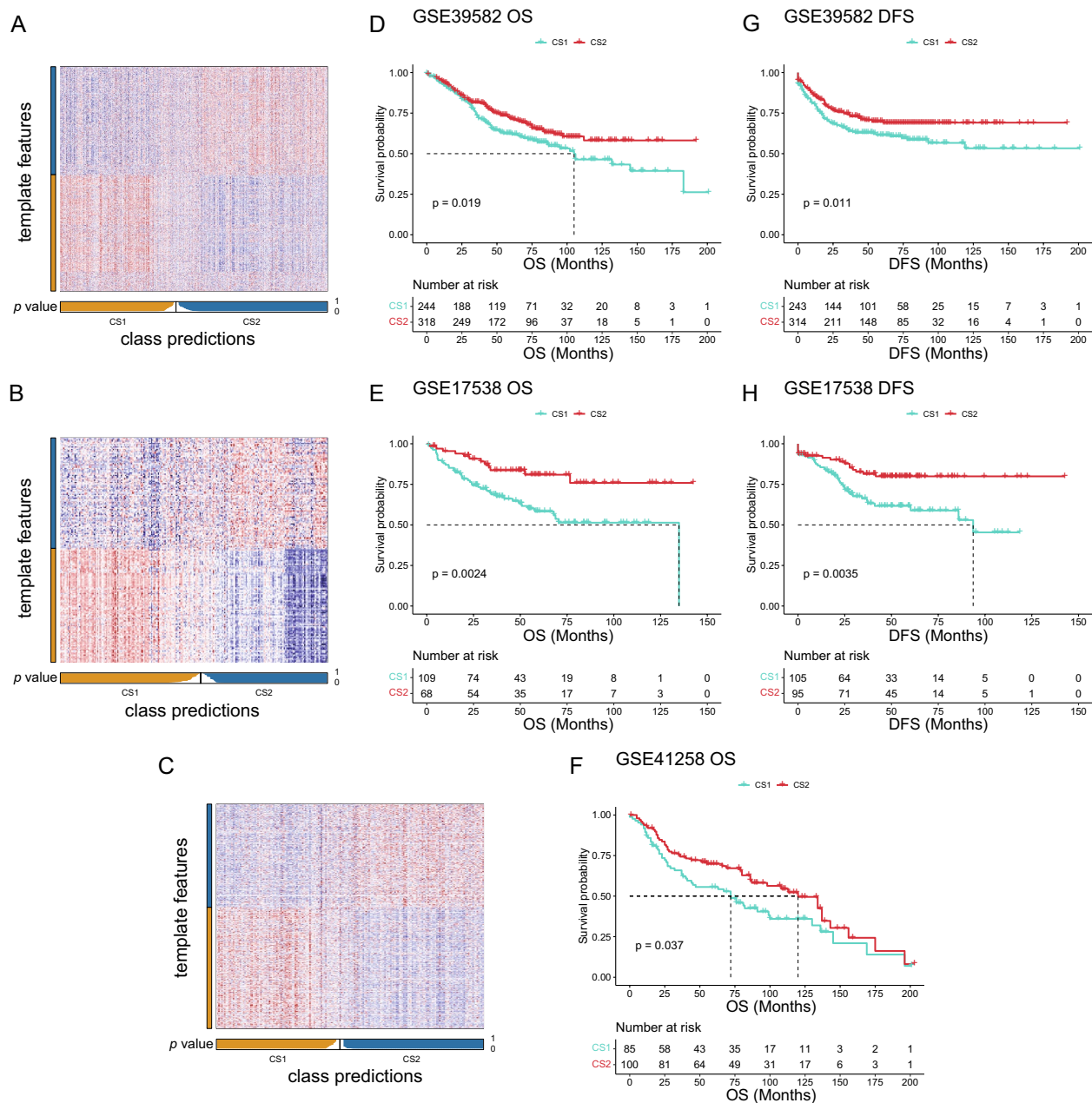
Ma *et al. European Journal of Medical Research*   (2024) 29:207

Page 10 of 17

Fig. 5 (See legend on previous page.)

Fig. 6 Validation of the multi-omics subtypes in external cohorts. **A–C** Heatmap of NTP in GSE39582 (**A**), GSE17538 (**B**), GSE41258 (**C**) cohorts using subtype-specific upregulated biomarkers identified from the TCGA-CRC cohort. **D-F** Differential overall survival outcome in two subtypes in external cohorts. **G-H** Differential disease-free survival in two subtypes in external cohorts

subtype, providing a basis for the development of a robust risk model.

We further calculated the correlations between each module and the subtypes, and found that the brown module had the strongest positive correlation with CS1 and negative correlation with CS2 (Fig. 7C). Next, we performed univariate Cox regression analysis on the 1260 genes in the brown module and identified 44 genes with statistical prognostic significance ($p < 0.01$). To reduce the number of genes and solve multicollinearity problems, we performed LASSO analysis and established a 5-gene prognostic model (Fig. 7D). The final prognostic model is as follows: Risk Score = 0.323 * MID2 + 0.261 * THBS3 + 0.243 * NUMBL + 0.276 * TMEM88 − 0.330 * MRPL37 (Fig. 7E). The CRC patients were then divided into low- and high-risk groups based on the median risk
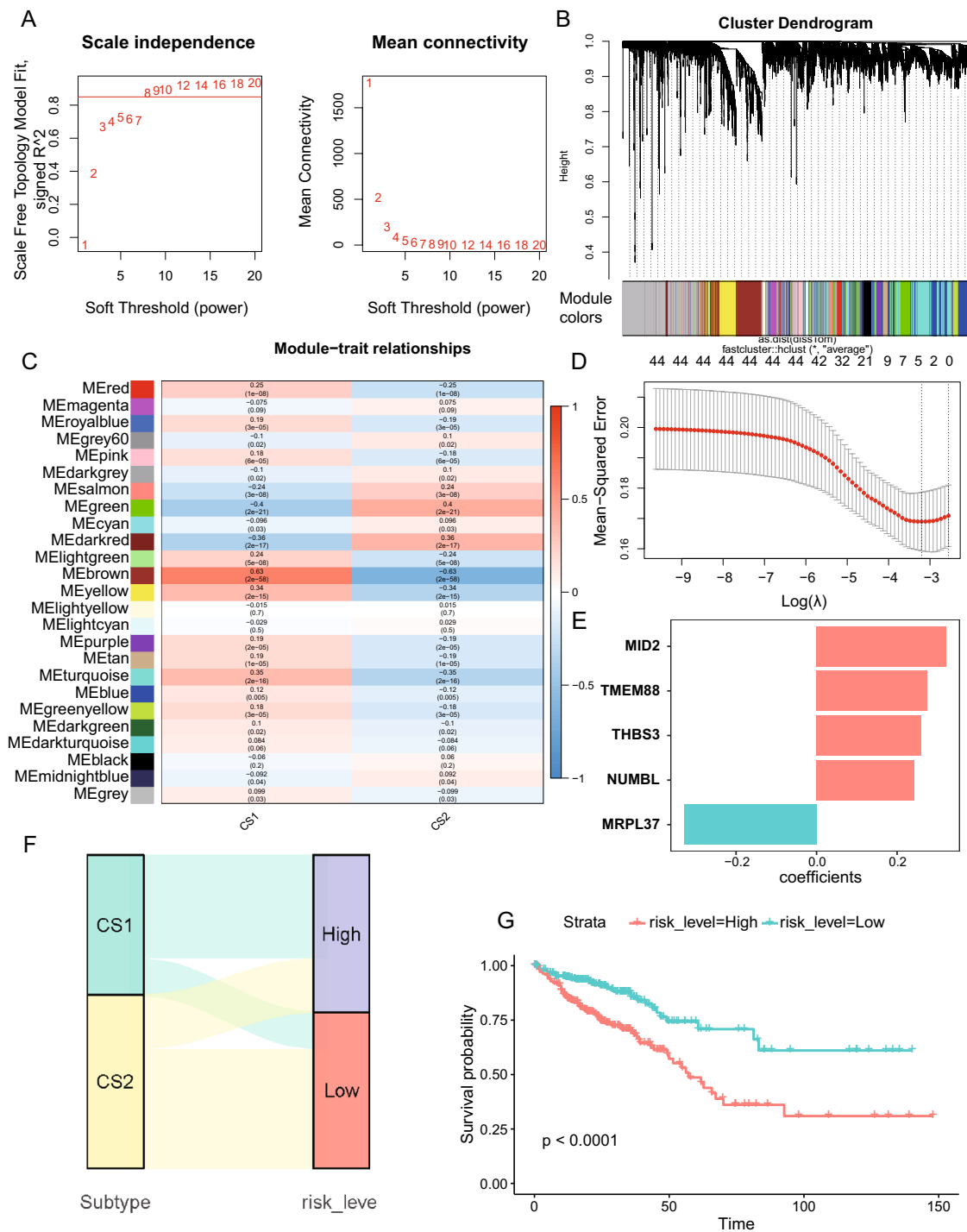
Fig. 7 Construction of prognostic model based on genes associated with subtype. **A** Determine the best soft threshold using network topology analysis. **B** The gene dendrogram and module color of weighted gene co-expression network analysis (WGCNA). **C** Heatmap illustrates the correlation between modules and subtypes. **D** LASSO analysis conducted on 44 genes associated with prognosis. **E** The coefficients of the five genes in the prognostic model. **F** The Sankey plot illustrates the alignment between the subtypes and risk groups. **G** Differential overall survival outcome in high- and low-risk groups based on the prognostic model

score. And the consistency of the risk model and molecular subtypes was confirmed by the Sankey plot, which showed that CS1 was mainly in the high-risk group and CS2 was in the low-risk group (Fig. 7F). Finally, the K-M survival curve demonstrated a significantly lower survival rate in the high-risk group compared to the low-risk group ($p < 0.0001$, Fig. 7G), indicating a significant difference in prognosis between the two groups.

MID2 was involved in EMT function in CRC cells

We observed that MID2, which in previous risk model (Fig. 7E), is a stage-dependent gene, with its expression increasing with CRC stage progression (Fig. 8A). Furthermore, the functional analysis of MID2-related genes using Hallmark gene sets demonstrated a remarkable enrichment in the EMT pathway, with the highest level of significance (Fig. 8B, Additional file 4: Fig. S4). Given our focus on multi-omics research, MID2 displayed a low mutation rate, prompting an investigation into its relationship with DNA methylation. Additional file 5: Fig. S5A indicates that the methylation levels of MID2 CpG sites are not significantly correlated with its expression. Then we explored whether MID2 serves as an upstream regulator of DNA methylation. Additional file 5: Fig. S5B reveals a significant correlation between MID2 and DNA methyltransferase. Additionally, principal component analysis indicates different DNA methylation sites between high and low MID2 expression groups (Additional file 5: Fig. S5C). Previous literatures report the association between DNA methylation in CpG islands and EMT in diseases [37, 38], which implied that the effects of MID2 on cancer progress may be associated with DNA methylation.

Furthermore, we sought to validate the function of MID2 in CRC and confirm the relationship between MID2 and EMT function. Plasmids were used to knockdown the expression of MID2 in SW480 and HCT116 cells. The knockdown efficiency was verified through western blot analysis (Fig. 8C). And the results revealed that the knockdown of MID2 led to upregulation of E-Cadherin and downregulation of N-Cadherin and vimentin (Fig. 8C), which are associated with a mesenchymal phenotype. In addition, we performed the colony formation assay to validate the effect of MID2 on CRC

cells proliferation. Figure 8D shows that CRC cells with sh-MID2 had a significantly reduced colony formation capacity compared to cells with sh-NC. Additionally, because EMT is a critical regulator pathway of tumor invasiveness, we assessed the effect of MID2 on cell migration and invasion using transwell assays. As displayed in Fig. 8E and F, knockdown of MID2 significantly inhibited the migration and invasion ability of CRC cells. These findings suggest that MID2 plays a crucial role in the progression of CRC.

Discussion

CRC is a highly lethal malignancy and ranks as the second leading cause of cancer-related mortality worldwide [1]. Recent advancements in high-throughput biochemical technologies have facilitated the accumulation of multi-omics data, enabling the identification of molecular mechanisms underlying various cancer types. The progression of CRC from colonic epithelium to the development of precancerous polyps, adenomas, and ultimately adenocarcinomas is accompanied by multiple gene mutation, DNA methylation alterations and gene expression changes [39]. Furthermore, it has been observed that early-onset CRC exhibits distinct epigenomic, transcriptomic, proteomic, and metabolic features when compared to late-onset CRC [11, 40]. These findings emphasize the therapeutic and prognostic value of integrating multi-omics data in CRC. In this study, by integrating multiple omics data sets with ten clustering methods, we report two refined and extensively validated CRC subtypes associated with distinct prognosis and molecular characteristics.

In our analysis, CS1 was characterized by poor prognosis and malignant phenotype, such as enriched function of the tumor-related Hallmark pathways, higher genomic alterations, and poor response to immunotherapy. Among the upregulated genes in CS1 (Additional file 6: Table S1), SFRP2 and SFRP4 are Wnt-regulator and can modulate the differentiation of cancer-associated fibroblasts (CAFs), which contribute to the progression and metastasis of the tumor [41–43]. And CAFs can drive FN1, COMP to regulate tumor metastasis and stemness in hepatocellular carcinoma [44, 45]. For CS2 upregulated gene, ZG16 exhibited a sequential decrease from normal

(See figure on next page.)
Fig. 8 Validation the function of MID2 in CRC cell lines. **A** The expression level of MID2 in different stages of CRC patients. **B** GSEA analysis demonstrates enrichment of MID2-related genes in the epithelial–mesenchymal transition (EMT) pathway. **C** SW480 and HCT116 cells were transfected with plasmid carrying sh-NC or sh-MID2. The protein level of MID2 and EMT biomarkers were assessed using western blotting. **D** Colony formation assay to evaluate the effect of MID2 knockdown on the proliferation of CRC cells. **E–F** Representative images and the comparison of cell migration and invasion ability demonstrated by transwell assay between sh-NC and sh-MID2 cells based on SW480 (**E**) and HCT116 (**F**) cell lines. The scale bar represents100 μm. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, and two-tailed unpaired t-test)

Ma *et al. European Journal of Medical Research*   (2024) 29:207
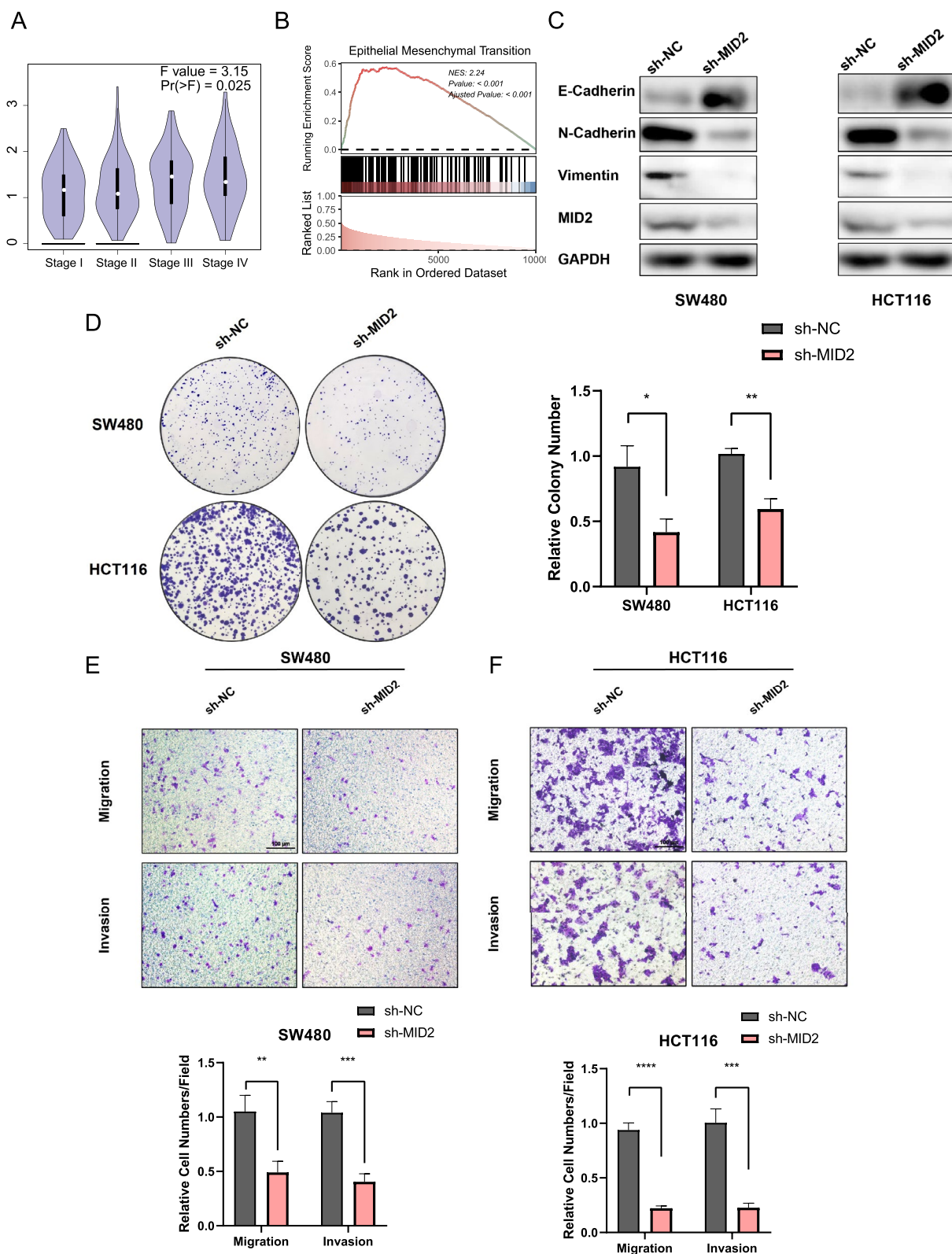
Page 14 of 17

Fig. 8 (See legend on previous page.)

to adenoma, and finally to carcinoma in CRC [46]. Furthermore, ZG16 inhibits PD-L1 expression and promote T cell-mediated immunity [47, 48]. And CS2 upregulated gene ITLN1 serve as tumor suppressor factor in various cancers [49]. Besides, we also found ZG16 and ITLN1 were associated with carbohydrate and fat metabolism [50, 51], which is consistent with CS2 enriched in multiple metabolism pathways (Fig. 3A).

It is well known that somatic mutations are present in the genomes of all cancers [52]. Although there was no difference in TMB across the two subtypes, genomic alteration analyses revealed that CS1 had more mutation genes and CNV than CS2. As shown in Fig. 4H, gene mutations are co-occurrence in CRC, and more synchronization mutation genes are present in the CS1 group (Fig. 2E). And one subtype of triple-negative breast cancer has poor prognosis with high chromosomal instability and highly recurrent CNAs [53]. Since various DNA mutagenic-repair events result in gene mutations and CNV [54, 55], tumors with more genomic alteration are suggest phenotypic malignancy and lead to a poor prognosis.

We analyzed immunotherapy responses across different subtypes based on the TIDE web platform and two external independent immunotherapy cohorts. These results implied that patients in CS2 may derive increased benefits from immunotherapy, providing valuable insights to guide clinical drug application. Since it is not convenient to derive the classification from multi-omics data, utilizing transcriptomic data to stratify patients in clinical applications is recommended. For patients identified in CS2 using the NTP approach or classified as low-risk based on the prognostic model in Fig. 7E, immunotherapy is recommended for the higher response rates.

The WGCNA method uses an unbiased systematic approach to analyze biological problems [56], which allows for the construction of gene co-expression networks to help identify candidate biomarkers or therapeutic targets for various diseases [57]. To facilitate the application of the subtypes to clinical practice, we utilized WGCNA to identify gene modular associated with the multi-omics subtypes and constructed a five-gene prognosis model using Cox and LASSO regression analyses. The Sankey plot demonstrated the coincidence of the risk groups with multi-omics subtypes (Fig. 7F). In addition, we found that the expression of MID2 in the prognosis model increase along with tumor stage progression and was significantly associated with EMT function. MID2, which was firstly identified as a causative gene of the X-linked form of a genetic disorder, is a ubiquitin-conjugating E3 enzyme and can regulate cytokinesis [58, 59]. In breast cancer, MID2 is upregulated and mediates tumor chemoresistance [60, 61]. In this study, in vitro experiments proved that MID2 can mediate the proliferation, migration, and invasion abilities of CRC cells, which consistent with the function of MID2 in neural crest cells and previous in silicon results [62]. These results suggest that MID2 is associated with tumor progression and could be a therapeutic target for CRC.

Our research had some limitations. First, the multi-omics subtypes in this study were based on bioinformatic analysis, which needs more time and research to transform into common medical technology. Second, the metabolomics and proteomics data are also critical to understanding cancer, these data might refine the result of multi-omics data analysis. Third, we preliminary validated the biological function of MID2 in CRC, and its detailed molecular mechanisms on EMT function and DNA methylation require future studies.

Conclusion

In conclusion, our study employed multi-omics data to classify CRC into two subtypes, which were associated with distinct prognoses, enrichment functions, immune microenvironmental characteristics, and altered genomic profiles. Additionally, to facilitate the widespread use of the multi-omics subtypes, a prognostic model based on five genes was constructed that showed strong agreement with the multi-omics subtypes. Besides, through in vitro experiments, we validated the role of MID2, one of the genes in the prognostic model, in promoting EMT function and invasiveness of CRC cells.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40001-024-01805-8.

Additional file 1: Figure S1. (A-C) Consensus heatmap for three (A), four (B), and five (C) subtypes based on multi-omics data. (D-F) The Silhouette value quantify sample similarity based on three (D), four (E), and five (F) cluster subtypes.

Additional file 2: Figure S2. Forest plot for multivariable Cox regression analysis with clinicopathological parameters and multi-omics subtypes

Additional file 3: Figure S3. (A) Bar plot of oncogenic pathways alterations fraction in CS1. (B) Bar plot of oncogenic pathways alterations fraction in CS2. (C) Scatter plot of the variants fraction of onco-drive genes in CS1. (D) Scatter plot of the variants fraction of onco-drive genes in CS2.

Additional file 4: Figure S4. The pathway ranking of MID2-related genes enrichment based on tumor-related Hallmark pathways.

Additional file 5: Figure S5. (A) Heatmap of MID2 expression and methylation levels of MID2-associated CpG sites. (B) Heatmap of MID2 and DNA methyltransferases expression. (C) Principal component analysis of DNA methylation patterns in groups characterized by high and low MID2 expression.

Additional file 6: Table S1. The top 200 upregulated genes in two subtypes according to nearest template prediction (NTP) method.

Ma *et al. European Journal of Medical Research*   (2024) 29:207

Page 16 of 17

Declarations

Ethics approval and consent to participate
This study was approved by the Ethics Committee of The First Affiliated Hospital of Xi'an Jiaotong University (Approval Number: 2017-146).

Consent for publication
All authors have agreed to the publication of this research.

Competing interests
The authors have no conflicts of interest to declare.

References
1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi RE, Corcione F. Worldwide burden of colorectal cancer: a review. Updates Surg. 2016;68:7–11.
3. Punt CJ, Koopman M, Vermeulen L. From tumour heterogeneity to advances in precision treatment of colorectal cancer. Nat Rev Clin Oncol. 2017;14:235–46.
4. Li B, Zhang F, Niu Q, Liu J, Yu Y, Wang P, Zhang S, Zhang H, Wang Z. A molecular classification of gastric cancer associated with distinct clinical outcomes and validated by an XGBoost-based prediction model. Mol Ther Nucleic Acids. 2023;31:224–40.
5. Chen Y, Meng J, Lu X, Li X, Wang C. Clustering analysis revealed the autophagy classification and potential autophagy regulators' sensitivity of pancreatic cancer based on multi-omics data. Cancer Med. 2023;12:733–46.
6. Ruan X, Ye Y, Cheng W, Xu L, Huang M, Chen Y, Zhu J, Lu X, Yan F. Multi-omics integrative analysis of lung adenocarcinoma: an in silico profiling for precise medicine. Front Med (Lausanne). 2022;9: 894338.
7. Guan Y, Yue S, Chen Y, Pan Y, An L, Du H, Liang C. Molecular cluster mining of adrenocortical carcinoma via multi-omics data analysis aids precise clinical therapy. Cells. 2022. https://doi.org/10.3390/cells11233784.
8. Meng J, Lu X, Jin C, Zhou Y, Ge Q, Zhou J, Hao Z, Yan F, Zhang M, Liang C. Integrated multi-omics data reveals the molecular subtypes and guides the androgen receptor signalling inhibitor treatment of prostate cancer. Clin Transl Med. 2021;11: e655.
9. AlMusawi S, Ahmed M, Nateri AS. Understanding cell-cell communication and signaling in the colorectal cancer microenvironment. Clin Transl Med. 2021;11: e308.
10. Sardo E, Napolitano S, Della Corte CM, Ciardiello D, Raucci A, Arrichiello G, Troiani T, Ciardiello F, Martinelli E, Martini G. Multi-omic approaches in colorectal cancer beyond genomic data. J Pers Med. 2022;12(2):128. https://doi.org/10.3390/jpm12020128.
11. Du M, Gu D, Xin J, Peters U, Song M, Cai G, Li S, Ben S, Meng Y, Chu H, Chen L, Wang Q, Zhu L, Fu Z, Zhang Z, Wang M. Integrated multi-omics approach to distinct molecular characterization and classification of early-onset colorectal cancer. Cell Rep Med. 2023;4: 100974.
12. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa EMF, Missiaglia E, Ramay H, Barras D, Homicsko K, Maru D, Manyam GC, Broom B, Boige V, Perez-Villamil B, Laderas T, Salazar R, Gray JW, Hanahan D, Tabernero J, Bernards R, Friend SH, Laurent-Puig P, Medema JP, Sadanandam A, Wessels L, Delorenzi M, Kopetz S, Vermeulen L, Tejpar S. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21:1350–6.
13. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458:719–24.
14. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, Islam SMA, Lopez-Bigas N, Klimczak LJ, McPherson JR, Morganella S, Sabarinathan R, Wheeler DA, Mustonen V, Getz G, Rozen SG, Stratton MR. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.
15. Parmar S, Easwaran H. Genetic and epigenetic dependencies in colorectal cancer development. Gastroenterol Rep (Oxf). 2022;10:goac035.
16. Ghadiri Moghaddam F, Farajnia S, Karbalaei-Mahdi M, Monir L. Epigenetic insights in the diagnosis, prognosis, and treatment selection in CRC, an updated review. Mol Biol Rep. 2022;49:10013–22.
17. Huang D, Sun W, Zhou Y, Li P, Chen F, Chen H, Xia D, Xu E, Lai M, Wu Y, Zhang H. Mutations of key driver genes in colorectal cancer progression and metastasis. Cancer Metastasis Rev. 2018;37:173–87.
18. Oh M, McBride A, Yun S, Bhattacharjee S, Slack M, Martin JR, Jeter J, Abraham I. BRCA1 and BRCA2 gene mutations and colorectal cancer risk: systematic review and meta-analysis. J Natl Cancer Inst. 2018;110:1178–89.
19. Müller D, Győrffy B. DNA methylation-based diagnostic, prognostic, and predictive biomarkers in colorectal cancer. Biochim Biophys Acta Rev Cancer. 2022;1877: 188722.
20. Jung G, Hernández-Illán E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. Nat Rev Gastroenterol Hepatol. 2020;17:111–30.
21. Ogino S, Kawasaki T, Nosho K, Ohnishi M, Suemoto Y, Kirkner GJ, Fuchs CS. LINE-1 hypomethylation is inversely associated with microsatellite instability and CpG island methylator phenotype in colorectal cancer. Int J Cancer. 2008;122:2767–73.
22. Antelo M, Balaguer F, Shia J, Shen Y, Hur K, Moreira L, Cuatrecasas M, Bujanda L, Giraldez MD, Takahashi M, Cabanne A, Barugel ME, Arnold M, Roca EL, Andreu M, Castellvi-Bel S, Llor X, Jover R, Castells A, Boland CR, Goel A. A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. PLoS ONE. 2012;7: e45357.
23. Amatu A, Sartore-Bianchi A, Moutinho C, Belotti A, Bencardino K, Chirico G, Cassingena A, Rusconi F, Esposito A, Nichelatti M, Esteller M, Siena S. Promoter CpG island hypermethylation of the DNA repair enzyme MGMT predicts clinical response to dacarbazine in a phase II study for metastatic colorectal cancer. Clin Cancer Res. 2013;19:2265–72.
24. Calegari MA, Inno A, Monterisi S, Orlandi A, Santini D, Basso M, Cassano A, Martini M, Cenci T, de Pascalis I, Camarda F, Barbaro B, Larocca LM, Gori S, Tonini G, Barone C. A phase 2 study of temozolomide in pretreated metastatic colorectal cancer with MGMT promoter methylation. Br J Cancer. 2017;116:1279–86.
25. Lu X, Meng J, Zhou Y, Jiang L, Yan F. MOVICS: an R package for multi-omics integration and visualization in cancer subtyping. Bioinformatics. 2021;36:5539–41.
26. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25.
27. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45:D353-d361.

28. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, Liu XS. Large-scale public data reuse to model immunotherapy response and resistance. Genome Med. 2020;12:21.

29. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, Seja E, Lomeli S, Kong X, Kelley MC, Sosman JA, Johnson DB, Ribas A, Lo RS. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell. 2016;165:35–44.

30. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, Dawson N, O'Donnell PH, Balmanoukian A, Loriot Y, Srinivas S, Retz MM, Grivas P, Joseph RW, Galsky MD, Fleming MT, Petrylak DP, Perez-Gracia JL, Burris HA, Castellano D, Canil C, Bellmunt J, Bajorin D, Nickles D, Bourgon R, Frampton GM, Cui N, Mariathasan S, Abidoye O, Fine GD, Dreicer R. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet. 2016;387:1909–20.

31. Li Z, Zhang H. Reprogramming of glucose, fatty acid and amino acid metabolism for cancer progression. Cell Mol Life Sci. 2016;73:377–92.

32. Myers JA, Miller JS. Exploring the NK cell platform for cancer immunotherapy. Nat Rev Clin Oncol. 2021;18:85–100.

33. Jung NC, Lee JH, Chung KH, Kwak YS, Lim DS. Dendritic cell-based immunotherapy for solid tumors. Transl Oncol. 2018;11:686–90.

34. Yin X, Chen S, Eisenbarth SC. Dendritic cell regulation of T helper cells. Annu Rev Immunol. 2021;39:759–90.

35. Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. Nat Rev Drug Discov. 2019;18:99–115.

36. Terwoord JD, Beyer AM, Gutterman DD. Endothelial dysfunction as a complication of anti-cancer therapy. Pharmacol Ther. 2022;237: 108116.

37. Wang XC, Song K, Tu B, Sun H, Zhou Y, Xu SS, Lu D, Sha JM, Tao H. New aspects of the epigenetic regulation of EMT related to pulmonary fibrosis. Eur J Pharmacol. 2023;956: 175959.

38. Nowak E, Bednarek I. Aspects of the epigenetic regulation of EMT related to cancer metastasis. Cells. 2021;10:3435.

39. Yang G, Yu XR, Weisenberger DJ, Lu T, Liang G. A multi-omics overview of colorectal cancer to address mechanisms of disease, metastasis, patient disparities and outcomes. Cancers (Basel). 2023;15:2934.

40. Kong C, Liang L, Liu G, Du L, Yang Y, Liu J, Shi D, Li X, Ma Y. Integrated metagenomic and metabolomic analysis reveals distinct gut-micro-biome-derived phenotypes in early-onset colorectal cancer. Gut. 2023;72:1129–42.

41. Kasashima H, Duran A, Martinez-Ordoñez A, Nakanishi Y, Kinoshita H, Linares JF, Reina-Campos M, Kudo Y, L'Hermitte A, Yashiro M, Ohira M, Bao F, Tauriello DVF, Batlle E, Diaz-Meco MT, Moscat J. Stromal SOX2 upregulation promotes tumorigenesis through the generation of a SFRP1/2-expressing cancer-associated fibroblast population. Dev Cell. 2021;56:95-110.e10.

42. Visweswaran M, Keane KN, Arfuso F, Dilley RJ, Newsholme P, Dharmarajan A. The influence of breast tumour-derived factors and wnt antagonism on the transformation of adipose-derived mesenchymal stem cells into tumour-associated fibroblasts. Cancer Microenviron. 2018;11:71–84.

43. Nurmik M, Ullmann P, Rodriguez F, Haan S, Letellier E. In search of definitions: cancer-associated fibroblasts and their markers. Int J Cancer. 2020;146:895–905.

44. Sun L, Wang Y, Wang L, Yao B, Chen T, Li Q, Liu Z, Liu R, Niu Y, Song T, Liu Q, Tu K. Resolvin D1 prevents epithelial-mesenchymal transition and reduces the stemness features of hepatocellular carcinoma by inhibiting paracrine of cancer-associated fibroblast-derived COMP. J Exp Clin Cancer Res. 2019;38:170.

45. Zhang L, Zhang C, Xing Z, Lou C, Fang J, Wang Z, Li M, He H, Bai H. Fibronectin 1 derived from tumor-associated macrophages and fibroblasts promotes metastasis through the JUN pathway in hepatocellular carcinoma. Int Immunopharmacol. 2022;113: 109420.

46. Meng H, Li W, Boardman LA, Wang L. Loss of ZG16 is associated with molecular and clinicopathological phenotypes of colorectal cancer. BMC Cancer. 2018;18:433.

47. Meng H, Yao W, Yin Y, Li Y, Ding Y, Wang L, Zhang M. ZG16 promotes T-cell mediated immunity through direct binding to PD-L1 in colon cancer. Biomark Res. 2022;10:47.

48. Meng H, Ding Y, Liu E, Li W, Wang L. ZG16 regulates PD-L1 expression and promotes local immunity in colon cancer. Transl Oncol. 2021;14: 101003.

49. Kawashima K, Maeda K, Saigo C, Kito Y, Yoshida K, Takeuchi T. Adiponectin and intelectin-1: important adipokine players in obesity-related colorectal carcinogenesis. Int J Mol Sci. 2017. https://doi.org/10.3390/ijms18040866.

50. Javitt G, Kinzel A, Reznik N, Fass D. Conformational switches and redox properties of the colon cancer-associated human lectin ZG16. Febs j. 2021;288:6465–75.

51. Zhao A, Xiao H, Zhu Y, Liu S, Zhang S, Yang Z, Du L, Li X, Niu X, Wang C, Yang Y, Tian Y. Omentin-1: a newly discovered warrior against metabolic related diseases. Expert Opin Ther Targets. 2022;26:275–89.

52. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR, Yates LR, Papaemmanuil E, Beare D, Butler A, Cheverton A, Gamble J, Hinton J, Jia M, Jayakumar A, Jones D, Latimer C, Lau KW, McLaren S, McBride DJ, Menzies A, Mudie L, Raine K, Rad R, Chapman MS, Teague J, Easton D, Langerød A, Lee MT, Shen CY, Tee BT, Huimin BW, Broeks A, Vargas AC, Turashvili G, Martens J, Fatima A, Miron P, Chin SF, Thomas G, Boyault S, Mariani O, Lakhani SR, van de Vijver M, van'tVeer L, Foekens J, Desmedt C, Sotiriou C, Tutt A, Caldas C, Reis-Filho JS, Aparicio SA, Salomon AV, Børresen-Dale AL, Richardson AL, Campbell PJ, Futreal PA, Stratton MR. The landscape of cancer genes and mutational processes in breast cancer. Nature. 2012;486:400–4.

53. Gong Y, Ji P, Yang YS, Xie S, Yu TJ, Xiao Y, Jin ML, Ma D, Guo LW, Pei YC, Chai WJ, Li DQ, Bai F, Bertucci F, Hu X, Jiang YZ, Shao ZM. Metabolic-pathway-based subtyping of triple-negative breast cancer reveals potential therapeutic targets. Cell Metab. 2021;33:51-64.e9.

54. Lee SY, Wang H, Cho HJ, Xi R, Kim TM. The shaping of cancer genomes with the regional impact of mutation processes. Exp Mol Med. 2022;54:1049–60.

55. Pös O, Radvanszky J, Buglyó G, Pös Z, Rusnakova D, Nagy B, Szemes T. DNA copy number variation: main characteristics, evolutionary significance, and pathological aspects. Biomed J. 2021;44:548–59.

56. Zhao W, Langfelder P, Fuller T, Dong J, Li A, Hovarth S. Weighted gene coexpression network analysis: state of the art. J Biopharm Stat. 2010;20:281–300.

57. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:559.

58. Zanchetta ME, Meroni G. Emerging roles of the TRIM E3 ubiquitin ligases MID1 and MID2 in cytokinesis. Front Physiol. 2019;10:274.

59. Li B, Zhou T, Zou Y. Mid1/Mid2 expression in craniofacial development and a literature review of X-linked Opitz syndrome. Mol Genet Genomic Med. 2016;4:95–105.

60. Luo J, Zeng S, Tian C. MORC4 promotes chemoresistance of luminal A/B breast cancer via STAT3-Mediated MID2 Upregulation. Onco Targets Ther. 2020;13:6795–803.

61. Wang L, Wu J, Yuan J, Zhu X, Wu H, Li M. Midline2 is overexpressed and a prognostic indicator in human breast cancer and promotes breast cancer cell proliferation in vitro and in vivo. Front Med. 2016;10:41–51.

62. Qiao Y, Zhou Y, Song C, Zhang X, Zou Y. MID1 and MID2 regulate cell migration and epithelial-mesenchymal transition via modulating Wnt/β-catenin signaling. Ann Transl Med. 2020;8:1021.

Publisher's Note