





## Article

# The Association of R-Loop Binding Proteins Subtypes with CIN Implicates Therapeutic Strategies in Colorectal Cancer

Wenchao Zhao <sup>1,2,3,4,†</sup> , Qian Pei <sup>1,2,3,4,†</sup>, Yongwei Zhu <sup>5,6</sup>, Dongdong Zhan <sup>7</sup>, Guo Mao <sup>8</sup> , Meng Wang <sup>2,3</sup>, Yanfang Qiu <sup>2,3</sup>, Ke Zuo <sup>8</sup>, Haiping Pei <sup>1,2,4</sup>, Lun-Quan Sun <sup>2,3,4,9,10,11</sup>, Ming Wen <sup>2,3,4,9,10,11,\*</sup>  and Rong Tan <sup>2,3,4,9,10,11,\*</sup> 

- <sup>1</sup> General Surgery Department, Xiangya Hospital, Central South University, Changsha 410008, China
  - <sup>2</sup> Xiangya Cancer Center, Xiangya Hospital, Central South University, Changsha 410008, China
  - <sup>3</sup> Key Laboratory of Molecular Radiation Oncology Hunan Province, Changsha 410008, China
  - <sup>4</sup> National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha 410008, China
  - <sup>5</sup> Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha 410008, China
  - <sup>6</sup> Hunan International Scientific and Technological Cooperation Base of Brain Tumor Research, Xiangya Hospital, Central South University, Changsha 410008, China
  - <sup>7</sup> State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China
  - <sup>8</sup> Science and Technology on Parallel and Distributed Processing Laboratory, National University of Defense Technology, Changsha 410073, China
  - <sup>9</sup> Hunan International Science and Technology Collaboration Base of Precision Medicine for Cancer, Changsha 410008, China
  - <sup>10</sup> Center for Molecular Imaging of Central South University, Xiangya Hospital, Changsha 410008, China
  - <sup>11</sup> Hunan Key Laboratory of Aging Biology, Xiangya Hospital, Central South University, Changsha 410008, China
- \* Correspondence: biowming@csu.edu.cn (M.W.); rongtan@csu.edu.cn (R.T.); Tel.: +86-731-84327212 (M.W.); +86-731-84327212 (R.T.)
- † These authors contributed equally to this work.
- ‡ Lead contact.



**Citation:** Zhao, W.; Pei, Q.; Zhu, Y.; Zhan, D.; Mao, G.; Wang, M.; Qiu, Y.; Zuo, K.; Pei, H.; Sun, L.-Q.; et al. The Association of R-Loop Binding Proteins Subtypes with CIN Implicates Therapeutic Strategies in Colorectal Cancer. *Cancers* **2022**, *14*, 5607. <https://doi.org/10.3390/cancers14225607>

Academic Editors: Claudio Luparello and Rita Ferreira

Received: 13 October 2022

Accepted: 8 November 2022

Published: 15 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** R-Loops, finely regulated by R-Loop binding proteins (RLBPs), play pivotal roles in maintaining genomic stability. By integrated proteogenomic analysis, we identified two RLBPs subtypes with distinct prognostic and therapeutic differences in colorectal cancer (CRC). Cluster-I (CI), characterized by high expression of RLBPs, was associated with chromosomal instability (CIN), better prognosis, and sensitivity to drugs targeting genome integrity and EGFR. Cluster-II (CII), characterized by low expression of RLBPs, was associated with mucinous adenocarcinoma, right-sided colon cancer, and poor prognosis. High inflammatory signaling pathway and lymphocyte infiltration enriched in CII, indicating potential application of drugs targeting inflammatory and immune response. Our research might be helpful for the precision treatment of CRC.

**Abstract:** Chromosomal instability (CIN) covers approximately 65 to 70% of colorectal cancer patients and plays an essential role in cancer progression. However, the molecular features and therapeutic strategies related to those patients are still controversial. R-loop binding proteins (RLBPs) exert significant roles in transcription and replication. Here, integrative colorectal cancer proteogenomic analysis identified two RLBPs subtypes correlated with distinct prognoses. Cluster I (CI), represented by high expression of RLBPs, was associated with the CIN phenotype. While Cluster II (CII) with the worst prognosis and low expression of RLBPs was composed of a high percentage of patients with mucinous adenocarcinoma or right-sided colon cancer. The molecular feature analysis revealed that the active RNA processing, ribosome synthesis, and aberrant DNA damage repair were shown in CI, a high inflammatory signaling pathway, and lymphocyte infiltration was enriched in CII. In addition, we revealed 42 tumor-associated RLBPs proteins. The CI with high expression of tumor-associated proteins was sensitive to drugs targeting genome integrity and EGFR in both cell and organoid models. Thus, our study unveils a significant molecular association of the CIN phenotype with

RLBPs, and also provides a powerful resource for further functional exploration of RLBPs in cancer progression and therapeutic application.

**Keywords:** chromosome instability; R-loop binding proteins; molecular subtype; drug sensitivity

## 1. Introduction

Chromosomal instability (CIN) and microsatellites instability (MSI) are two major characteristics of colorectal cancer (CRC) [1,2]. Despite substantial advanced in genomic features for cancer, the clinical significance of those features and potential mechanisms to guide therapeutic strategies have not yet been defined. The MSI in colon cancer is caused by deficiency or dysfunction of the mis-match repair pathway. The early diagnosis of MSI in CRC becomes the therapeutic standard for applications of fluorouracil and immune checkpoint inhibitors, while the molecular mechanisms underlying CIN and related therapeutic sensitivity are still controversial.

R-loop is processed by the transient cross-reaction of RNA with template DNA to form a DNA-RNA hybrid. The structures are prevalent in rDNA loci, telomeric regions, actively transcribed regions, and mitochondrial DNA [3–6]. Extensive physiological and pathological processes, such as immunoglobulin recombination, transcriptional regulation, DNA replication, telomere homeostasis, CpG island promoter methylation, and DNA lagging strand synthesis in activated B cells, have been associated with the formation and mediation of R-loop [5,7]. The R-loop structures cause DNA damage in certain contexts and are regarded as a major source of genomic instability [8]. Thus, homeostatic maintenance of R-loop plays an important role in genome stability [9].

R-loop homeostasis is regulated by a series of R-loop binding proteins (RLBPs). Thus, aberrant expression or dysfunction of R-loop binding proteins impairs genomic integrity and is associated with diverse diseases. Mutations in the DNA/RNA helicase senataxin (SETX) result in neurodegenerative disorders, ataxia with oculomotor apraxia type 2 and amyotrophic lateral sclerosis type 4 (AOA2/ALS4) [10]. The EWS-FIL1 fusion induces R-loop accumulation and perturbs homologous recombination repair, a process that results in Ewing sarcoma and underlies the mechanism of chemosensitivity [11]. Moreover, several factors that are involved in DNA damage repair have been demonstrated to be RLBPs. BRCA1 and the Fanconi anemia (FA) pathway factors FANCA, FANCD2, and TOP1 have been reported to bind R-loop during DNA damage repair [12–15]. Dysfunction of BRCA1 and FA factors is associated with breast cancer and Fanconi anemia, respectively. Due to the significant function of TOP1 in genome maintenance, it is also a desirable chemotherapeutic target. In this regard, advances in elucidating the expression and mechanistic signatures of RLBPs are essential for understanding the regulation of genome stability and cancer, further suggesting desirable therapeutic strategies.

Our study initiated a comprehensive proteomics analysis of 204 RLBPs in three colorectal (CR) proteogenomic databases. We revealed two RLPB clusters with distinct prognoses. The two clusters correlated with CIN phenotypes, anatomical location, and pathological status. Furthermore, the two clusters were distinguished by physiological features and tumor microenvironment (TME). In addition, we also identified 42 tumor-related RLBPs. We revealed that the clusters with high expression of tumor-related RLBPs displayed different drug sensitivity involved in EGFR and genome stability pathways. Thus, the associations of RLBPs with cancer provide a theoretical basis for further functional investigation of R-loop binding proteins in cancer. The RLPB clustering also bridges a gap between CIN and RLBPs, facilitating personalized diagnosis and appropriate selection for therapy in clinical application.

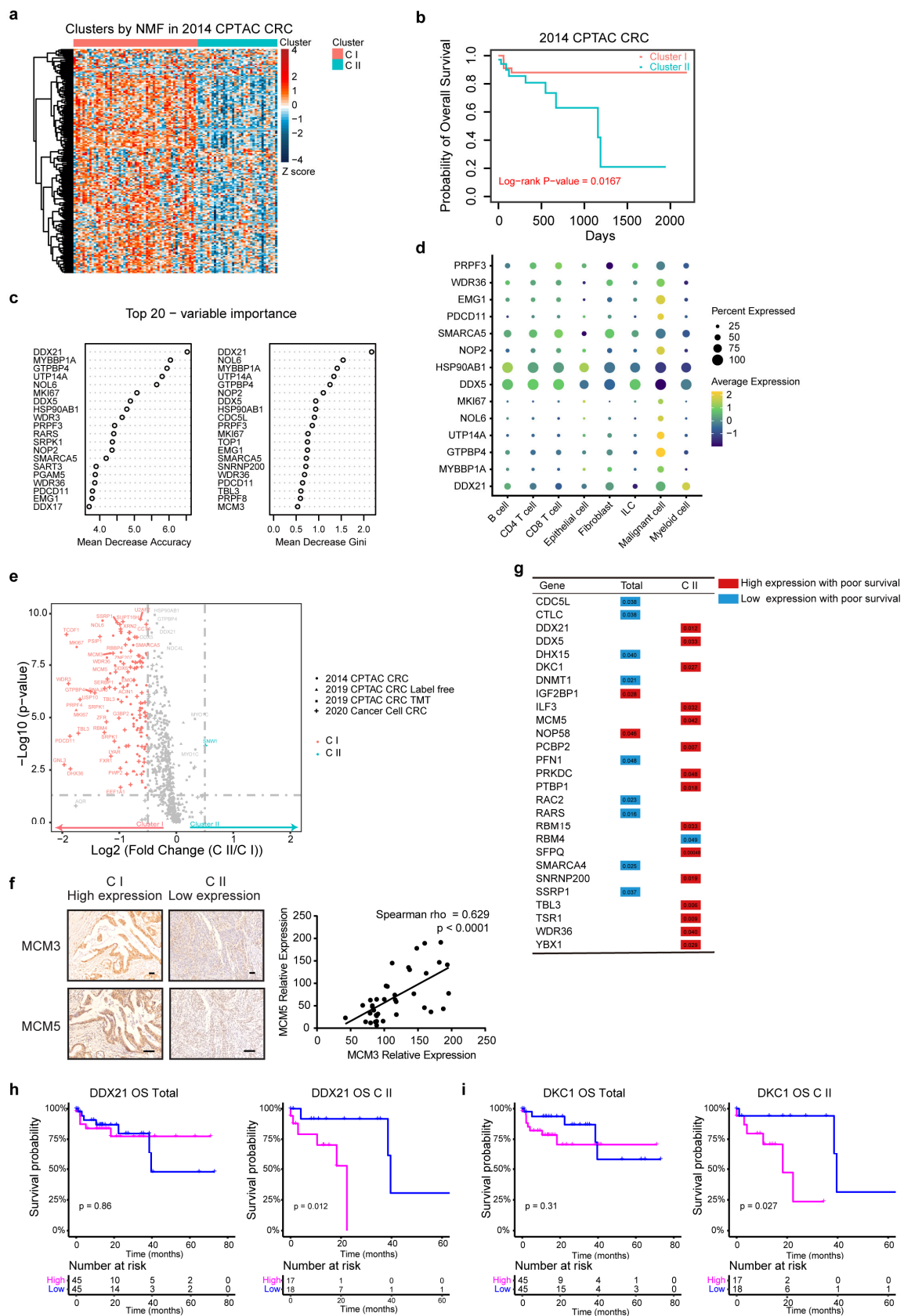
## 2. Results

### 2.1. Proteogenomic Databases for RLBP Analysis

A continuously increasing number of methods have been developed to identify RLBP and establish R-loop-interacting protein databases. The Natalia Gromak group performed affinity purification using the S9.6 antibody to obtain 463 predicted R-loop binding proteins, and the Vivian G. Cheung group employed synthesized hybrids to carry out precipitation to obtain 803 R-loop binding proteins [16,17]. We thus employed a shared 197 overlapping RLBP from the two databases (Supplementary Figure S1a). In addition, some well-established hybrid-binding proteins, such as RNase H1/2, SETX, AQR, and XPG, were lost due to methodological bias [10,18–21]. To define the extensive landscape of RLBP and establish their relevance to colon cancer, we carried out proteogenomic analysis using 204 RLBP, including the 197 detected RLBP and 7 acknowledged RLBP (Supplementary Table S1a), in three colorectal cancer databases and CCLE colorectal cancer cell line databases [22–28]. The 204 RLBP were used for unsupervised clustering analysis and signature identification. In addition, survival differences, clinical relevance, the immune microenvironment, somatic mutation status, genome stability, and drug sensitivity were analyzed.

### 2.2. Clustering of Colorectal Cancers Based on Proteomic Data of RLBP

To further investigate RLBP expression patterns in tumor tissues, we performed non-negative matrix factorization (NMF)-based unsupervised clustering using 204 RLBP in the 2014 CPTAC CRC database [22] (Figures 1a and S1b). The tumor samples can be classified into two subtypes with distinct clinical outcomes. Cluster I (CI) had a better prognosis and Cluster II (CII) had a worse prognosis (Figure 1b and Supplementary Table S1b). To further facilitate stratification, we selected 14 variable important genes (DDX21, MYBBP1A, GTPBP4, UTP14A, NOL6, DDX5, HSP90AB1, PRPF3, NOP2, SMARCA5, WDR36, PDCD11, and EMG1, MKI67) and established a clustering model using the random forest algorithm (Figures 1c and S1c). We also found the 14 variable important genes were enriched in malignant cells in CRC scRNA analyses (Figure 1d). We then used the model to classify 2019 CPTAC or 2020 Cancer Cell CRC databases. We matched the subtypes assigned by the random forest model to NMF subtypes in those two CRC databases and revealed that the accuracy was higher than 77%, indicating the reliability of the clustering method [23,24] (Supplementary Figure S1d,e). Based on the subtypes classified by NMF in 2014 or by clustering model in 2019 and 2020 databases, we noticed that the expressions of the RLBP were consistently high in CI and low in CII (Figures 1e, S1f–i and S2a–c, and Supplementary Table S1c). To further validate the RLBP expression in two clusters, we performed immunohistochemistry (IHC) of MCM3 and MCM5 in colorectal cancer tissues (Figure 1f). The results showed that the two genes exhibited consistent expression patterns in patients from two clusters. Indeed, the IHC scores of MCM3 and MCM5 were highly correlated, further supporting that the RLBP could be markers to distinguish CRC patients. Thus, the CI and CII were subsequently referred to as the RLBP-proficient (CI-pRLBP) cluster and deficient RLBP (CII-dRLBP) cluster, respectively. Among all 204 RLBP, 16 encoding genes showed cluster-dependent correlation with prognosis, as manifested by their significant impact on survival in CII, of note, high expression of 15 genes, including DDX21 and DKC1, led to worse outcomes in CII (Figure 1g–i and Supplementary Table S1d). Thus, RLBP-based clustering revealed the significance of genes associated with tumors, which is frequently underestimated in the observation of total tumor patient cohorts.



**Figure 1.** The consistent expressions of RLBP in two clusters and predictions for prognosis. (a) Heatmap illustrating 204 RLBP expressions in 2014 CPTAC CRC proteomic databases in 90 patients. Annotations indicate tumor clusters CI and CII. (b) Kaplan-Meier curves for overall survival (OS) based

on two clusters in 2014 CPTAC CRC proteomic databases. (c) The top 20 important characteristics indicated by the Mean Decrease Accuracy value and the Mean Decrease Gini value from a random forest model for 2014 CRC databases clustering. (d) Bubble heatmap showing the expression of 14 RLBP in each scRNA-seq cell type. Dot size indicates the percentage of expressed cells, colored by average expression levels. (e) Volcano plot showing 204 RLBP proteomics in four CRC proteomic clusters from three colorectal databases. (Blue and red indicate  $\text{Log}_2$  (fold change (CII/CI))  $> 0.5$  and  $\text{Log}_2$  (fold change (CII/CI))  $< -0.5$  respectively,  $p < 0.05$ ). (f) Representative images of MCM3 and MCM5 immunohistochemistry staining in two clusters of colon cancer tissues (left). Spearman correlation of MCM3 expression with MCM5 expression in colorectal cancer (right). (g) Table showing 204 RLBP that were significantly associated with prognosis in total patients or in CII (Log rank test,  $p < 0.05$ ). (h,i) Kaplan-Meier curves of overall survival (OS) for total patients or CII with DDX21 (h) or DKC1 (i) high and low abundance.

### 2.3. Prognostic Significance of the RLBP-Based Cluster

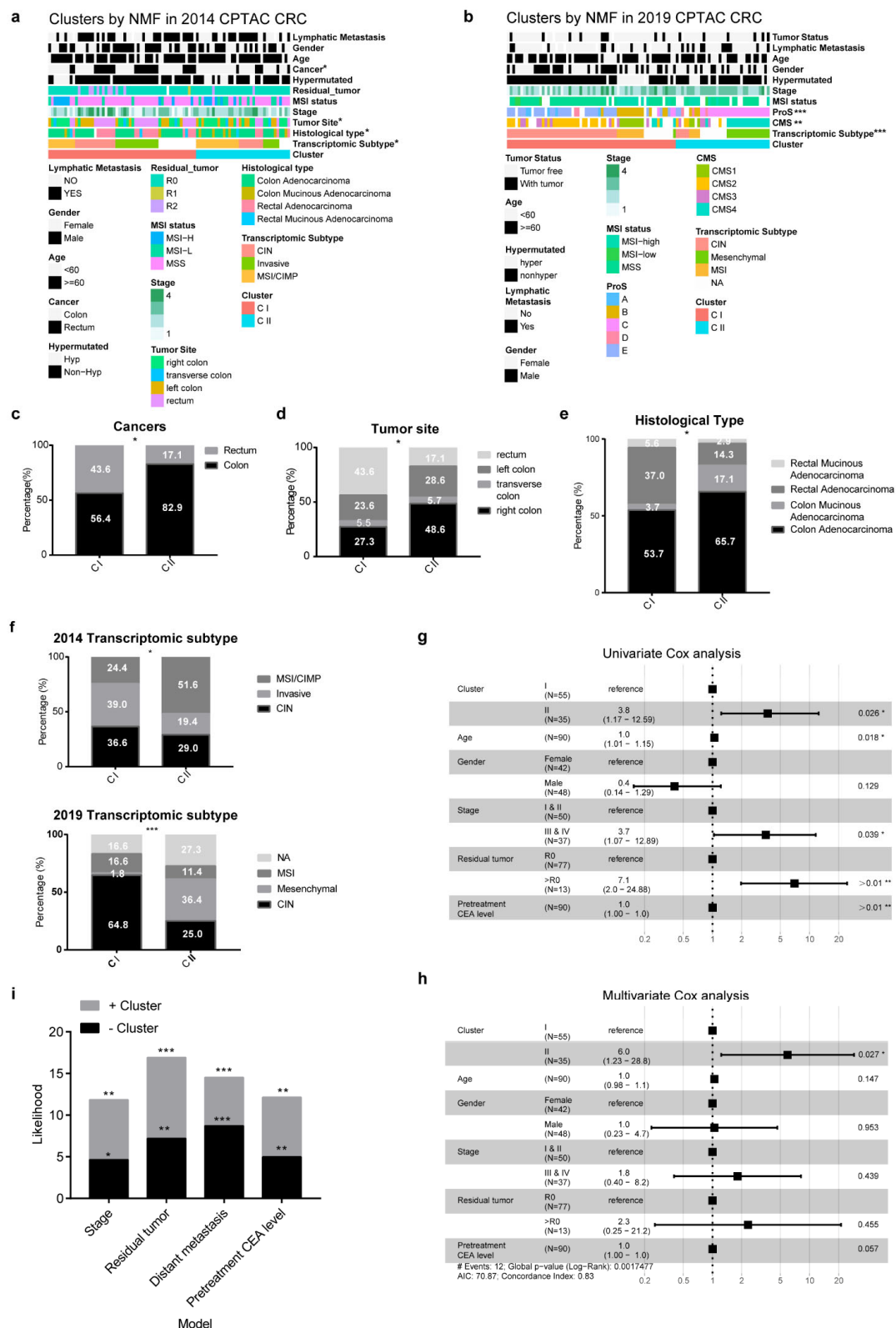
To gain the clinical relevance for RLBP clusters, we analyzed clinical features related to two clusters. The two clusters were highly associated with cancer pathological types, tumor site, histological type, and the transcriptomic subtype in 2014 CPTAC CRC databases, while, we did not observe significant relevance of clusters with MSI status, sex, or lymphatic metastasis [22,23,29] (Figure 2a,b and Supplementary Table S2a,b). The higher percentage of patients with rectum tumors (43.6%) or with left-sided colon cancer (67.2%) were distributed in CI-pRLBP (Figure 2c,d). While the CII-dRLBP contained more patients with right-sided colon cancer (54.3%) or mucinous adenocarcinoma (20%) (Figure 2d,e). Interestingly, we also noticed that CI was highly associated with the CIN phenotype (36.6%) compared to CII (29.0%) in 2014 CPTAC CRC databases. The same result was obtained in the 2019 CPTAC CRC databases, CI (64.8%) was highly enriched in CIN phenotype compared to CII (25.0%) (Figure 2f).

To examine whether RLBP clusters could serve as an independent prognostic factor, we employed univariate and multivariate cox regression models using clinical information, including age, gender, stage, residual tumor, and pretreatment CEA level. The result showed that the cluster was a robust and independent prognostic biomarker for evaluating CRC patient OS (Figure 2g,h; HR = 6.0, 95% CI 1.23–28.1,  $p < 0.05$ ). Moreover, Cox models integrated clusters with stage, residual tumor, distant metastasis, and pretreatment CEA level greatly enhanced prognostic accuracy, implying that RLBP clusters further facilitate prognostic prediction of CRC patients (Figure 2i). The results suggested that RLBP-based clusters reflect the difference clinical features and predict the prognosis of CRC patients.

### 2.4. Different Molecular and Immune Characteristics of the RLBP-Based Cluster

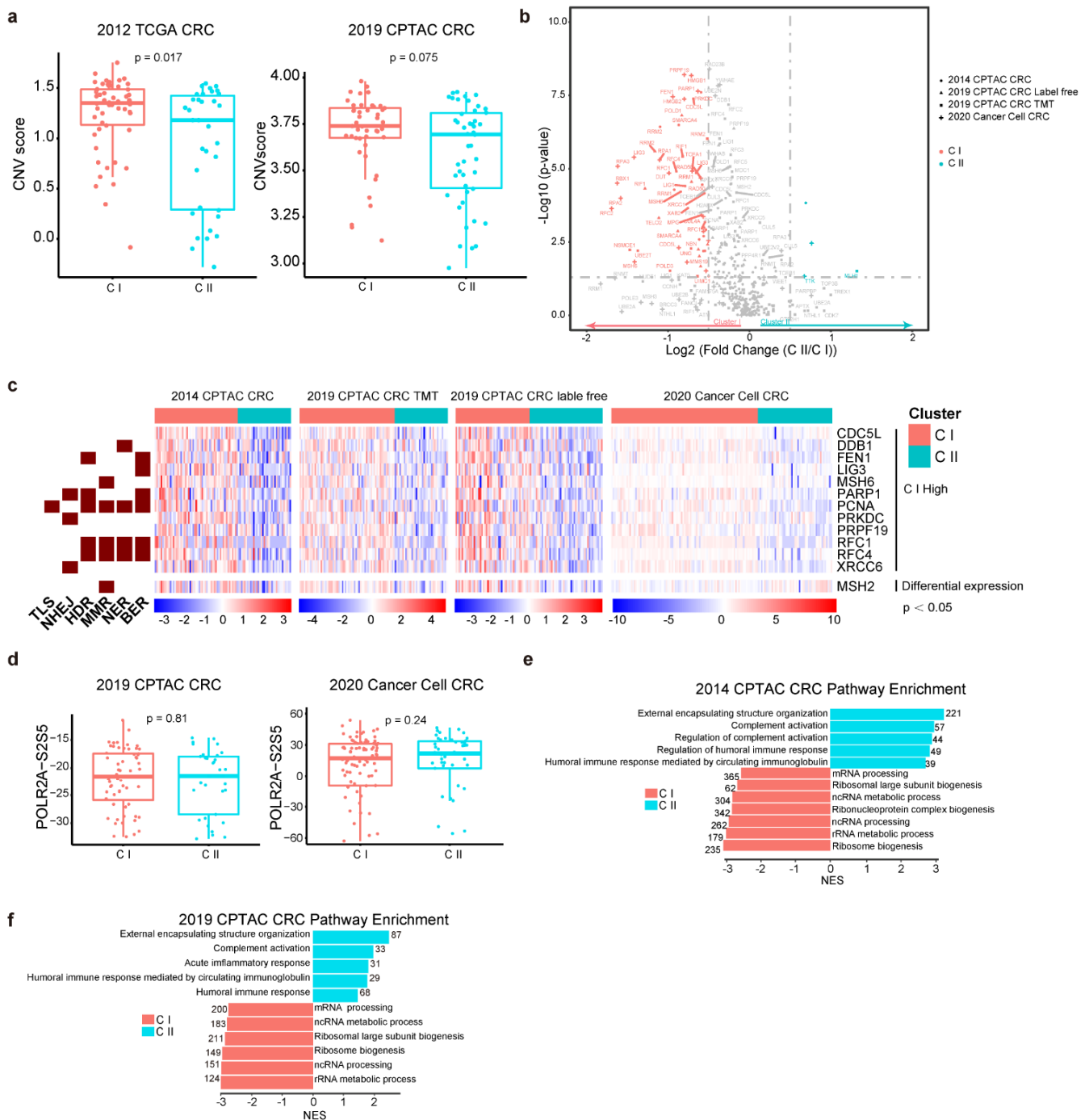
RLBP are essential for genome maintenance; thus, we further compared the mutations between the two subtypes. However, we observed no apparent variations in the indel or point mutation burden between CI and CII in either the TCGA or CPTAC databases (Supplementary Figure S3a,b and Table S2c,d, see Section 4). While, we found the CI cluster showed higher average copy number variation (CNV) scores than CII (Figure 3a). Aberrant expression or dysfunction of DNA damage repair proteins implies inappropriate DNA damage repair, which finally results in genome instability [30]. We analyzed the expression of 276 DNA damage repair proteins in two clusters of three CRC databases. Interestingly, proteins involved in DNA damage repair showed apparently high expression in CI (Figures 3b and S3c–d, and Supplementary Table S2e). Furthermore, 12 DNA damage repair proteins, such as RFC1, RFC4, PAPR1, and PCNA, which engage in one or more specific DNA repair pathways, like translesion DNA synthesis (TLS), nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR) and double-strand breaks (DSBs) repair, showed consistently higher expressions in CI than CII throughout the four proteomic cohorts ( $p < 0.05$ , Wilcoxon signed-rank test). It is notable that the expression of MSH2 showed elevation in CI in both CPTAC databases, while exhibiting the opposite result in the 2020 Cancer cell CRC database (Figure 3c).





**Figure 2.** Clinical features in two clusters and Cox regression analysis of colorectal cancer overall survival. **(a,b)** Clinical relevance in two clusters in 2014 CPTAC CRC **(a)** and 2019 CPTAC CRC **(b)** proteomic databases. Fisher's exact test was used for those variables (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ). **(c–e)** Distributions of cancers **(c)**, tumor site **(d)**, histological type **(e)** in two clusters in 2014 CPTAC CRC databases. **(f)** Distributions of transcriptomic subtype in two clusters in 2014 or 2019 CPTAC CRC databases. **(g,h)** Univariate **(g)** or Multivariate **(h)** Cox regression model analysis, which included

the factors of cluster, age, gender, stage, residual tumor, and pretreatment (i) CoxPH models of stage, residual tumor, distant metastasis, pretreatment CEA level with (Multivariate Cox) or without (Univariate Cox) proteomic Cluster were compared.



**Figure 3.** CI and CII possess different molecular characteristics. (a) Box plot showing CNV score in two clusters in 2012 TCGA CRC and 2019 CPTAC CRC databases. (b) Volcano plot showing expressions of DNA damage proteins in two clusters in three CRC proteomics databases (Blue and red indicate  $\log_2(\text{fold change (CII/CI)}) > 0.5$  and  $\log_2(\text{fold change (CII/CI)}) < -0.5$  respectively,  $p < 0.05$ ). (c) Heatmap showing selected DNA damage protein engaged DNA repair pathways ( $p < 0.05$ ) through four proteomic cohorts. (d) Box plot showing phosphorylation of Polymerase 2 (POLR2A) at S2/5 sites in two clusters in 2019 CPTAC CRC databases (left) and 2020 Cancer cell CRC databases (right) [23,24]. (e,f) Pathway enriched in two clusters in two CRC proteomics databases analyzed by GESA (Two-sided unpaired  $t$  test).

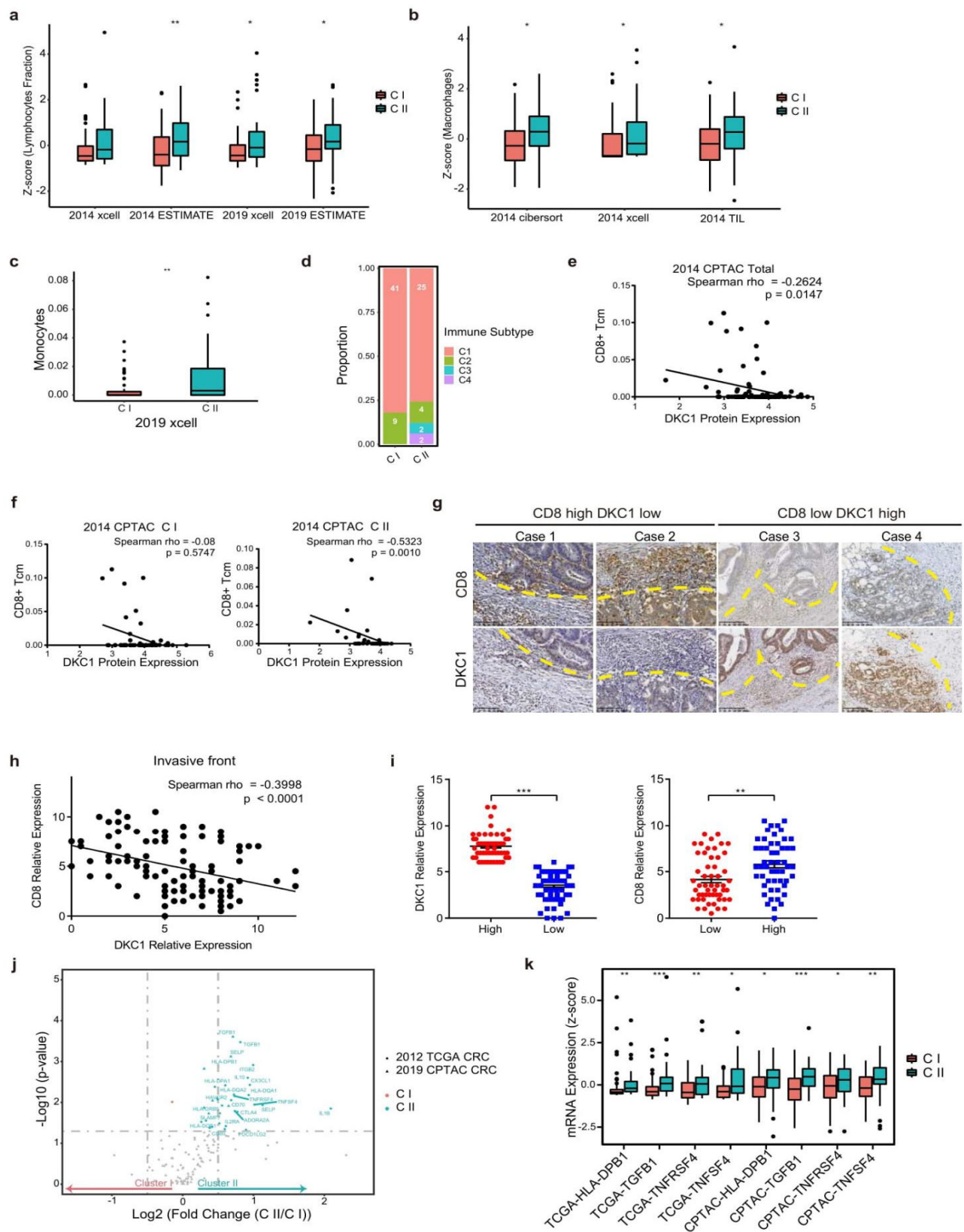
RLBPs play important roles in transcription, translation, and RNA metabolism. However, unexpectedly, the phosphorylations of POLR2A at S2 and S5, which indicate transcriptional activity, exhibited no differences between CI and CII (Figure 3d and Supplementary Table S2f,g). The further Gene Set Enrichment Analysis (GSEA) revealed higher RNA processing activity in CI, while considerable enhancement of inflammatory response in CII in all the three CRC databases (Figures 3e,f and S3e, and Supplementary Table S2h–j). These results suggested that two clusters engaged in different molecular signaling pathways and hold distinct DNA repair capabilities.

### 2.5. The Related Immune Microenvironments in CI and CII

The composition, infiltration, and activity of immune cells constitute the tumor immune microenvironment and confer antitumor activity. To investigate differential immune microenvironments in two subtypes, we employed several methods to profile distinct immune cells signatures mapping to the proteomic clusters [31–34]. Similar as molecular features, the CII showed consistently higher lymphocyte fraction, as evidenced by the immune scores, in both the 2014 and 2019 CPTAC CRC databases (Figure 4a and Supplementary Table S3a–c), indicating extensive immune cell infiltration in the CII subtype compared to the CI subtype of cancer. Furthermore, in spite of different analysis methods, we found that macrophages were highly accumulated in CII in the 2014 CPTAC CRC database (Figure 4b and Supplementary Table S3a,c,d). The immunochemical staining of macrophages in colorectal cancer tissues further emphasized the increased tumor-associated infiltration of macrophages in CII patients (Figure 4b and Supplementary Table S3c) [34]. In the 2019 CPTAC CRC database, we also noticed that the higher percentage of monocytes, which is the precursor of macrophages, is presented in CII (Figure 4c). Six immune subtypes have been previously characterized based on TCGA transcriptome databases [35]. We noticed that CI patients enriched in immune subtypes C1 and C2. While CII patients spanned more board range of immune subtypes, especially accumulating in immune subtypes C3 and C4, which displayed inflammatory and macrophages signatures respectively (Figure 4d).

By analyzing the correlation between the RLBPs and immune cell populations, we found that the expression of DKC1 showed a negative correlation with CD8+ Tcm cells, especially in CII (Figures 4e,f and S4a,b, and Supplementary Table S3e–g). To further validate the correlation of DKC1 with CD8+ T cells, we collected colorectal cancer tissues and immunohistochemically stained the expression of CD8 and DKC1. We measured the correlation of DKC1 expression with CD8+ T cells at tumor sites, stromal sites, and the invasive front (where the tumor invaded the normal lamina propria). The results showed that DKC1 expression was negatively correlated with CD8+ T cells populations at the invasive front (Figure 4g,h), but not at tumor sites and stromal sites (Supplementary Figure S4c). The same results were found in macrophages (Supplementary Figure S4d). DKC1 expression was positively correlated with the macrophages marker, CD68, at the invasive front of colon cancer tumors (Supplementary Figure S4e,f). Meanwhile, we grouped the patients according to DKC1 expression and observed higher expression of CD8 (Figure 4i) or CD68 (Supplementary Figure S5g) in the DKC1-low group, which further demonstrated that RLBPs performed potential regulatory functions in the immune microenvironment. The immune modulators that are secreted by or located in immune cells and nonimmune cells play critical roles in orchestrating the immune environment. We interestingly found a majority of immune modulators were preferentially expressed in CII, further supporting the highly modulatory environment of immune cells in CII in both mRNA and protein levels (Figures 4j and S4h and Supplementary Table S3h,i). In addition, the MHC II component (HLA-DPB1), TGF $\beta$ 1, TNFRSF4, and TNFSF4 showed consistently high expressions in CII in both the 2014 and 2019 CRC databases. However, MHC I components exhibited no consistent differential expression in the two clusters (Figures 4k and S4i). These data indicated that two clusters have different degrees of infiltration by immune cell types and RLBPs could affect immune cells in TME.





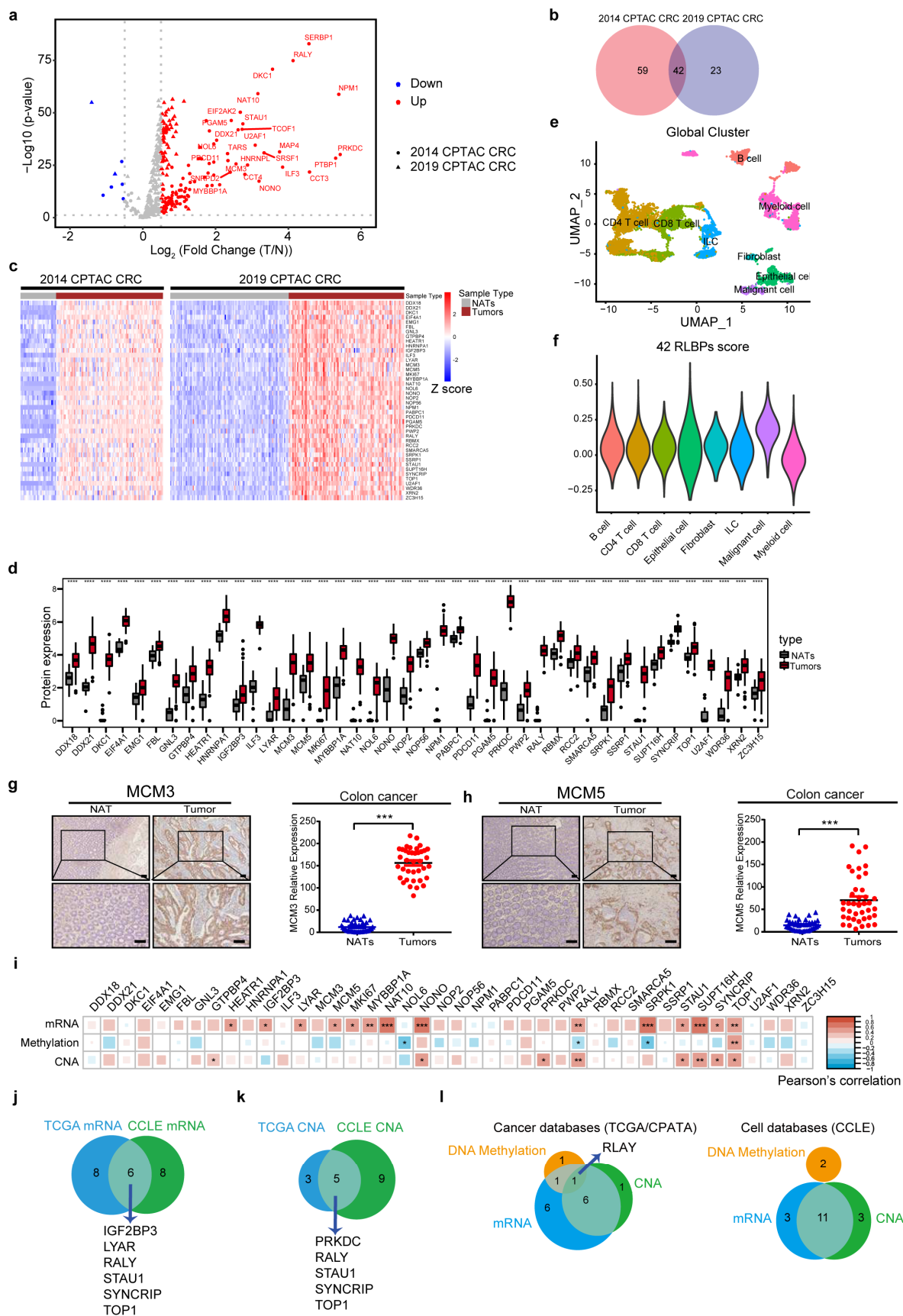
**Figure 4.** Distinct immune microenvironments exist in CI and CII. (a) Comparisons of Lymphocytes Fraction (see Section 4) between two clusters in two CRC databases by different immune fraction analysis, xcell, and ESTIMATE [31,32]. (b) Comparisons of Macrophages between two clusters in 2014 CPTAC CRC databases analyzed by three methods, xcell, CIBERSORT, and immune staining

of CRC databases [31,33,34]. (c) Comparisons of Monocytes between two clusters in 2019 CPTAC CRC Label free databases analyzed by xcell. (d) The distribution of patients belonging to two clusters in four different immune subtypes from Thorsson et al., 2018, Immunity [35]. (e) Scatter diagram showing correlations of DKC1 protein expression with CD8+ Tcm cells in total patients in 2014 CPTAC CRC databases. (f) Scatter diagram showing correlations of DKC1 protein expression with CD8+ Tcm cells in CI or CII cluster in 2014 CPTAC CRC databases. (g) Representative images of DKC1 immunochemistry staining and CD8+ T cells in colorectal cancer tissues. (h) Spearman correlation of DKC1 expression with CD8+ T cells in invasive front (where tumor invaded normal lamina propria) in colorectal cancer. (i) The relative DKC1 and CD8 expression of patients in Immunohistochemical high and low score groups (divided by median expression) (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , error bar indicates  $\pm$ SEM). (j) Volcano plot showing immunomodulators' proteomics in CII versus CI in two CRC proteomics databases. Blue indicates  $\text{Log}_2$  (fold change (CII/CI)  $> 0.5$ ,  $p < 0.05$ ) [35]. (k) The boxplot showing the mRNA expressions of immunomodulators in 2014 and 2019 CRC databases.

## 2.6. A Total of 42 RLBP s Show Consistently High Expression in Tumor Tissues

To obtain the tumor high expression RLBP s, we compared the expressions of 204 RLBP s between tumor tissues and normal adjacent tissues (NATs) in two colorectal databases [22,23]. We obtained 101, and 65 proteins were highly expressed in tumor tissues in the 2014 and 2019 CPTAC colorectal cancer databases, respectively ( $p < 0.05$ ,  $\text{log}_2\text{FC} > 0.5$ ) (Figure 5a). In total, 42 hybrid-binding proteins with consistently high expression were detected in the two proteomic databases and identified as tumor-associated hybrid-binding proteins (Figures 5b–d and S5a and Supplementary Table S4a,b). In addition, MCM3 and IGF2BP3 exhibited consistent upregulation in tumor tissues compared to NATs in six multiple cancer proteomic databases [23,36–40] ( $p < 0.05$ ,  $\text{log}_2\text{FC} > 0.5$ ) (Supplementary Figure S5b and Table S4c). To validate cell types that hold the high expression of those genes, we obtained eight major cell populations referring to fibroblast, epithelial cells, malignant cells, Innate lymphoid cells (ILC), CD4+ T cells, CD8+ T cells B cells, myeloid cells from single-cell RNA-seq (scRNA) datasets [41] (Figure 5e). The 42 RLBP s were highly expressed in a group of malignant cells which was a subgroup from the epithelial cell population, further implying that those proteins may play tumor-promoting roles [41] (Figure 5f). To further verify protein expression, we collected paired colon tumors and NATs and obtained that two of 42-tumor-associated RLBP s, MCM3, and MCM5, exhibited very high expression in tumors compared to the NATs (Figure 5g,h).

High expression of proteins can be affected by copy number alterations (CNA) at the genomic level or by transcriptional/translational regulation. We found high frequency of positive correlations of protein expression with mRNA expression and CNA and a negative correlation with methylation [27,42] (Figures 5i and S5c,d and Supplementary Table S4d,e). We further summarized six genes (IGF2BP3, LYAR, RALY, STAU1, SYNCRIP, and TOP1) that displayed high correlations between mRNA and protein expression in both cancer and cell line databases (Figures 5j and S5e). The expression of 5 genes (PRKDC, RALY, STAU1, SYNCRIP, and TOP1) displayed a strong positive correlation with CNA ( $r > 0.3$ ,  $p < 0.05$ ) (Figure 5k). Notably, seven or eleven genes showed both significantly high gene-level amplifications and mRNA in cancer and cell line databases respectively. Furthermore, the expression of one gene, RALY, showed an intimate correlation with gene amplification, promotor methylation, and mRNA level in the cancer database. Unfortunately, we did not observe a consistent correlation between CpG promotor methylation and protein expression in the cancer and cell line databases (Figure 5l). These findings suggested that the expressions of RLBP s were consistently high in tumor tissues and cells may possess motivation oncogenesis in CRC.



**Figure 5.** The expression and correlation of 42 RLBPs in colorectal tumors. (a) Volcano plot showing 204 RLBPs proteomics in tumors versus normal adjacent tissues (NATs) in two colorectal databases.

2014 CPTAC colorectal cancer (CRC) databases from Bing Zhang et.al., 2014., Nature and 2019 CPTAC colorectal databases from Suhas Vasaikar et.al., 2019, Cell [22,23]. (b) Venn diagram depicting tumor high expressing RLBP in two cancer databases from (a) ( $\log_2$  (fold change)  $> 0.5$ ,  $p < 0.05$ ). (c) The heatmap showing RLBP expression in tumors and NATs in patients from two CRC proteomics databases. (d) The boxplot showing RLBP expression in tumors and NATs in 2014 CPTAC CRC proteomics databases, \*\*\*\*  $p < 0.0001$ . (e) UMAP plots showing major immune and non-immune subsets reproduced by Smart-seq2 scRNA-seq. (f) Violin plot showing the 42 RLBP's score in all immune and non-immune cell types. (g,h) Representative images of immunochemistry staining of MCM3 (e) or MCM5 (f) expressions in colorectal cancer between tumor and NATs. Quantifications of MCM3 and MCM5 expressions in tumors and NATs were shown on right. (i) Pearson correlation of RLBP's proteomic expression with CNA, DNA methylation (promoter CpG clusters) or mRNA from TCGA CRC databases. (j,k) Venn diagram showing genes whose expression was affected by mRNA (j) or CNA (k) in both CCLE databases and TCGA CRC databases. (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ). (l) Venn diagrams showing gene expressions affected by CNA, CpG promotor DNA methylations and mRNA in Cancer Cell Line Encyclopedia (CCLE) databases (DepMap Public) [27] or TCGA CRC databases [42] (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , error bar indicates  $\pm$ SEM).

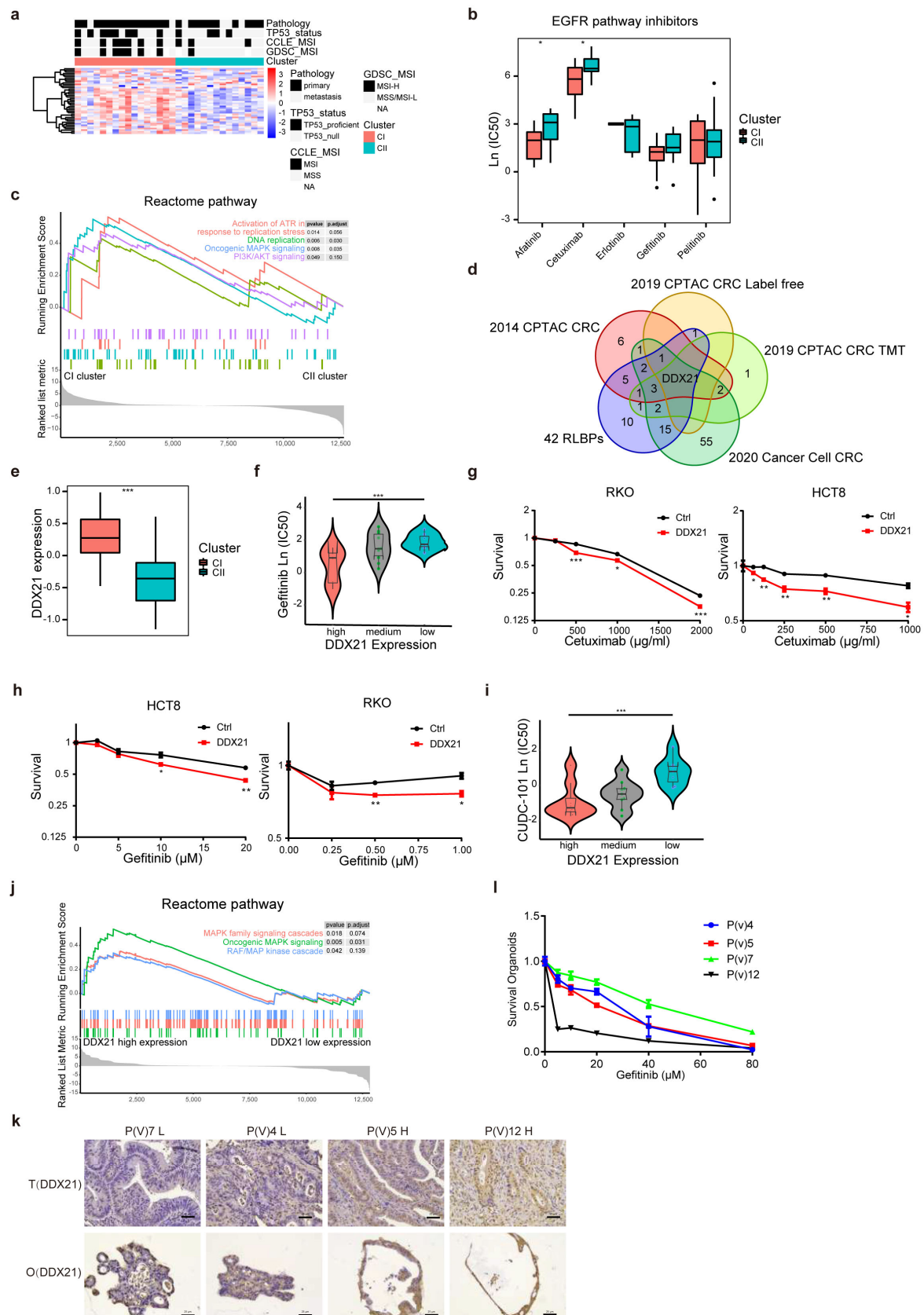
## 2.7. Alternative Therapeutic Drugs Are Selected and Used to Treat RLBP-Expressing Tumors

To further evaluate the therapeutic potential of the 42 tumor-associated genes, we analyzed the CRISPR screening project-CERES scores across the colorectal cancer cell lines [25]. More than 45% of the genes (19/42) had poor therapeutic index, with a score lower than  $-1$  (Supplementary Figure S6a and Table S5a). Moreover, 9 of the 42 tumor-associated genes were identified as core fitness genes across three datasets (Supplementary Figure S6b and Table S5b), implying that those genes are essential for physiological processes and thus have a low therapeutic index [25,27,43–47]. While, the metastatic features showed no apparent correlation with the CERES scores of RLBP (Supplementary Figure S6c,d).

Alternative therapeutic strategies should be considered for tumors with high RLBP expression. Since we observed that 14 genes involved in subtypes classification were enriched in malignant tumor cells (Figure 1d), we divided CRC cancer cells into two clusters and observed the 42 gene expressions were accumulated in CI in cells, which was reminiscent of subtypes in cancer databases (Figures 1a and 6a). We turned to a large-scale drug screening project, the GDSC1 dataset [48]. We interestingly found drugs involved in EGFR and genome stability were highly sensitive in CI compared to CII (Figures 6b and S7a). Pathway enrichment analyses showed CI cell cluster was enriched in oncogenic MAPK signaling and PI3K/AKT pathways, which were two critical downstream pathways for EGFR (Figure 6c). Moreover, DNA replication and DNA damage repair pathways were enriched in the CI cluster (Supplementary Figure S7b).

Furthermore, we revealed that DDX21, which exhibited high expression in tumors, showed predominantly high expression in CI compared to CII across 4 CRC databases (Figure 6d,e,  $p < 10^{-7}$ ). In addition, EGFR pathway inhibitors showed resistance in the median low DDX21 expression group and enhanced resistance in the high-low DDX21 expression group (Figure 6f). This was further validated in cells with high expression of DDX21 (Figures 6g,h and S7c). Consistently, cells with DDX21 were sensitive to the inhibition of CUDU-101, which is a potent multi-targeted inhibitor against HDAC, EGFR, and HER2 pathway (Figure 6i and Supplementary Table S6a). Moreover, cells with DDX21 high expression exhibited higher sensitivity to DNA replication and genome integrity drugs (Supplementary Figure S7d). The pathway enrichment analysis showed MAPK signaling pathway, DNA replication, and DNA damage repair pathway were enriched in DDX21 high expression group (Figures 6j and S7e).





**Figure 6.** The cell and patient model to verify drug sensitivity. **(a)** Heatmap showing 42 RLBP expression and clinical features in CRC cells. **(b)**  $\ln(\text{IC}_{50})$  of EGFR inhibitors in CRC cells' sensitivity between two clusters. **(c)** GSEA Reactome enrichment plot showing enrichment pathways in two clusters



in CRC cells. (d) Venn diagram depicting the overlap genes between tumor high expressed 42 RLBP and cluster significantly expressed RLBP in three CRC proteomic databases ( $p < 10 \times 10^{-7}$ , Wilcoxon rank-sum test). (e) boxplot showing the DDX21 expression in two cell clusters. (f) Violin plot showing distributions of Gefitinib's Ln (IC50) in DDX21 protein high, medium, and low groups (high groups were divided by DDX21 expression higher than 75% cells, low groups were divided by DDX21 expression lower than 25% cells). (g) Viability assays of control or DDX21 overexpressing RKO and HCT8 cells treated with Gefitinib (48 h). (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Error bar indicates  $\pm$ SEM by Student's *t*-test). (h) Viability assays of control or DDX21 overexpressing RKO and HCT8 cells treated with Cetuximab (96 h). (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Error bar indicates  $\pm$ SEM by Student's *t*-test). (i) Violin plot showing distributions of CUDC-101's Ln (IC50) in DDX21 protein high, medium, and low groups (high groups were divided by DDX21 expression higher than 75% cells, low groups were divided by DDX21 expression lower than 25% cells). (j) GSEA Reactome enrichment plot in DDX21 high and low expression groups in CRC cells. (k) Immunohistochemistry staining of DDX21 on PODs and corresponding primary tumors. Scale bar, 20  $\mu$ m. (l) Organoids dose-response to gefitinib, surviving organoids data shown are means  $\pm$  SE from three independent experiments.

Extensive evidences have demonstrated that CR organoids retain molecular and histological features similar to those of the tumor specimens and faithfully predicts the clinical response of patients [49–52]. To further reproduce the responses in clinical patients, we established CR (colorectal) tumor-derived organoids. After histological and pathological validation of the organoids (Supplementary Figure S7f,g), we compared the gefitinib response in organoids with high and low DDX21 expression. As expected, we observed that organoids with higher DDX21 expression were more sensitive to gefitinib (Figure 6k,l). Based on all the evidences above, we concluded that DDX21 could be used as an effective gene to differentiate drug sensitivity in two clusters.

### 3. Discussion

R-loop binding proteins (RLBPs) act extensive essential roles in genome stability. In our study, we profile the RLBP proteomics landscape in colorectal cancer and uncover the associations of the RLBP with CIN phenotype. We reveal two RLBP clusters with distinct prognoses, molecular features, and TME. In addition, we identify 42 tumor-associated RLBP and elucidate the functions of those RLBP on leading differential drug sensitivities.

CIN refers to an abnormal accumulation of gain or loss of entire regions of chromosomes, including intra- and inter-chromosomal aberrations [53]. CIN is considered to promote tumor development by disrupting tumor suppressor genes, increasing the copy number of oncogenes, or altering gene expression [54]. CIN, especially intra-chromosomal aberrations, is the main phenotype of genomic instability in CRC [53]. The diverse dysregulations of genome maintenance, including defects in chromosome segregation, abnormal DNA damage repair, and telomere crisis contribute to CIN. Incorrectly repaired DNA DSBs, the major driver of chromosomal rearrangements, are usually caused by aberrant DNA damage repair systems, such as homologous recombination (HR) and non-homologous end joining (NHEJ) [55,56].

We revealed that CI-pRLBP were more enriched in the CIN subtype which is characterized by high CNVs. More importantly, we observed a high expression of DNA repair proteins in CI. The homeostatic regulations of DNA damage repair pathways play a fundamental role in maintaining genome integrity [57,58]. The high expression of DNA repair proteins in CI ruins the balance of DNA damage repair. The hyperactivated double-strand repair pathways in CI, including non-homologous end-joining (NHEJ) and homologous recombination (HR) pathways, implicate excessive exchanges of chromosomes, which further account for high CNVs. Moreover, DDX21, which is a significant RLBP member, was revealed to engage in the HR pathway directly. The high expression of DDX21 leads to excessive sister chromatid exchanges [59]. PARP1 is required for chromosomal translocations; it displaces the Ku complex and initiates an alternative NHEJ pathway [60]. In addition, the pathway enrichment analysis indicated CI with proficient expression of

RLBPs was accumulated in transcription and DNA replication, cellular processes which are essential for genome stability.

Recent studies showed that CIN is associated with both poor prognosis and drug resistance [61,62]. Inappropriately, DNA damage repair pathways are expected to be synergistically therapeutic strategies for drugs targeting genome stability. This was further validated in the drug sensitivity analysis, as demonstrated by the high percentage of negative correlation of RLPB expression with IC50 values of drugs involved in genome stability (Figure 6). The concept is consistent with our previous findings showing that high expression of RNA helicases, DDX21, renders sensitivity for irinotecan, oxaliplatin, and etoposide [59]. Thus, overall high therapeutic sensitivity in patients would be expected along with better outcomes in CI. Interestingly, we also found CI was more sensitive to genome EGFR inhibitors. EGFR activation is driven by two main pathways: PI3K/AKT/mTOR signal pathway and RAS/RAF/MEK/MAPK signal pathway, which are involved in regulating cell proliferation, survival, cell migration, and angiogenesis [63]. We noticed MAPK and PI3K signaling pathways enriched in CI. The condensed accumulation of MAPK and PI3K genes may attribute to the high growing requirement for CI. Therefore, RLBPs cluster might be regarded as an adequate predictor to evaluate the chemotherapy and targeted therapy. However, the relationship between DNA damage repair and the EGFR signal pathway required further investigations.

CII-dRLBPs contained higher compositions of lymphocytic fractions but exhibited a worse prognosis. Macrophages and progenitor monocytes substantially accumulated in CII. The extensive immunosurveillance factors were preferentially expressed in CII (Figure 4). The comparatively active immune microenvironment may be implied by the lower contribution of DNA damage repair in CII, which leads to an increase in neoantigen expression in tumors. Considering that R-loops and RLBPs play conventional roles in transcription and immune modulation [64], the higher lymphocytic infiltration but lower antigen processing and presentation activity are insufficient to activate the T-cell response and result in a comparatively inactive immune microenvironment. Furthermore, CII could be observed in the consistently high level of TGF $\beta$ 1 which account for tumorigenesis and metastasis and is associated with poor prognosis [35,65]. By analyzing the correlation between gene expression and the immune cell composition, we found that DKC1 showed a negative correlation with macrophages and central memory T (T<sub>cm</sub>) cells, which exhibit long-term memory after activation of immature T-cells by antigens and have the capacity to return to lymph nodes for antigen restimulation. The increased expression of DKC1 with reduced macrophages and CD8<sup>+</sup> T-cells infiltration is consistent with the concept that appropriate immune surveillance is required for tumor control. Previous studies also reported RLBPs are involved in immune microenvironment modulation. The high expression of RNaseH1 leads to R-loop degradation, the process which subsequently reduces cytosolic DNA and inhibits type I IFN-dependent rejection of lymphoma cells [66]. Moreover, METTL3/IGF2BP3 inhibition reduces the stabilization of PD-L1 mRNA, thus enhancing anti-tumor immunity through T-cell activation, exhaustion and infiltration [67]. These results implicate that RLBPs engage in inflammatory signaling pathways and immune cell remodeling.

The locations differences of colorectal tumor are correlated with clinical outcomes. LCRC has a better prognosis than right-sided colon cancer (RCRC) in both chemotherapy and targeted treatments [68]. CI contains higher percentage of left-sided colon (LCRC) adenocarcinoma, while CII contained more patient with right-sided colon cancer or mucinous adenocarcinoma (Figure 2). The location variation on the two clusters supports the survival differences, further suggesting the distinct molecular signature of LCRC and RCRC.

In summary, our study delineates associations of RLBPs with CIN in CRC. We constructed RLBPs-based cluster model to distinguish CRC patients and identify their therapeutic utility in chemotherapy and targeted therapy. With the increasing understanding that RLBPs in CIN phenotypes, drug responses and comprehensive roles in shaping the TME, our results provide some new perspectives on using these proteins as therapeutic strategies and preventing therapeutic failure in future clinical applications.

## 4. Materials and Methods

### 4.1. Differential Protein Analysis

A paired Wilcoxon signed-rank test was employed to detect the difference of protein expression between tumor and paired Para cancerous tissues (Tandem mass tag (TMT) global proteomic analysis of the 100 tumors and 97 paired Para cancerous tissues in 2019 CPTAC CRC, 90 tumors and 30 paired Para cancerous tissues in 2014 CPTAC CRC). Up-regulated or down-regulated proteins in tumors are defined as proteins that are differentially expressed in tumors compared with their matched Para cancerous tissues ( $\log_2$  (fold change (FC)) T/N  $> 0.5$  or  $< -0.5$ ,  $p < 0.05$ ).

### 4.2. Correlations of Protein to mRNA or Copy Number Alteration (CNA) or DNA Methylation in RLBP

A total of 204 RLBP genes in the CRC cell lines or tumor were calculated with pairwise Pearson's correlation coefficients between protein abundance to mRNA or CNA or CpG promotor DNA methylation.

### 4.3. Single-Cell Analyses

The quality control and cell types clustering of CRC Single-Cell Analyses were reproduced from the Smart-seq2 data of CRC single cell databases (GSE146771). The average expression of 42 tumor-associated RLBPs calculated by the Seurat's AddModuleScore function was defined as the 42 RLBPs score.

### 4.4. Survival Analysis

Kaplan-Meier survival curve (log-rank test) was used to compare the overall survival time (OS) of CRC cancer patients with proteomic subtypes, or high and low expression of RLBPs. OS curve was calculated according to the median cutoff.

### 4.5. NMF Clustering for Proteomic of CRC

Non-negative matrix factorization (NMF) and the R-package (CancerSubtypes, Version: 1.14.0, Taosheng Xu, Shenzhen, China) were employed to identify CRC sample clusters. The 204 RLBPs genes were selected for clustering analysis. The preferred clustering result is determined by using the observed cophenetic correlation between the clusters and the silhouette width of the uniform membership matrix. We also employed the random forest predictor implemented in the R package randomForest to assign the molecular subtypes to each sample based on the protein expression matrix.

### 4.6. Random Forest Model Clustering for Proteomic of CRC

We trained the random forest model with the 'train' function in the R caret package. The model parameter was set as: method = 'rf', namely using the 'random forest' algorithm. Performance estimate parameter was set as: trControl = trainControl (method = 'cv', number = 10), namely 10-fold cross-validation (CV), which was implemented as follows: Divide the observations into ten equally sized independent 'folds' (each observation appears in only one fold); Hold out one of these folds (1/Kth of the dataset) to use as a test set; Train a model in the remaining K-1 folds; repeat until each of the folds has been held out once; Performance estimate is the average performance across the K held out folds, namely accuracy of the random forest model. In this study, the final value used for the model was mtry = 2 with the largest value of accuracy (0.9105556).

### 4.7. Pathway Enrichment Analysis

Gene Set Enrichment Analysis was performed using the GSEA software (<https://www.broadinstitute.org/gsea/>; accessed on 15 April 2021). Samples grouped according to the identified subtypes were subjected to GSEA. Molecular Signatures Database (MSigDB) of hallmark gene sets (H), KEGG, Reactome gene sets (C2) and GO gene sets (C5) were used for enrichment analysis,  $p$ -value of 0.05 was used as a cutoff. The enrichment score

(ES) in GSEA was calculated by first ranking the proteins from the most to least significant with respect to the two phenotypes, the entire ranked list was then used to assess how the proteins of each gene set were distributed across the ranked list.

Gene set variation analysis implemented in the GSVA package (version 1.42.0, Robert Castelo, Barcelona, Spain) was used for gene set enrichment analysis. The differences in pathway activities scored per cell line between CI and CII clusters (Figure S6B) were calculated with the LIMMA package (version 3.50.3, Gordon Smyth, Canberra, Australian). The reactome pathway analysis and KEGG analysis between two clusters in CRC cells were performed by the clusterProfiler package (version 4.2.2, Guangchuang Yu, Guangzhou, China).

#### 4.8. Phosphorylated Proteomic Analysis

POLR2A tandem heptapeptide repeats in the C-terminal domain (CTD) can be highly phosphorylated and the phosphorylation occurs mainly at residues 'Ser-2' and 'Ser-5' of the heptapeptide repeat. We calculated the total phosphorylation level of 'Ser-2' and 'Ser-5' in the POLR2A tandem heptapeptide repeats domain of RNA Pol II in each sample, according to the classification of patients. Statistical differences between the two clusters were utilized with Wilcoxon rank-sum test.

#### 4.9. Mutation Burden Evaluation

In order to assess the different mutation burden between the two clusters, we calculated the total number of mutations per trillion bases in tumor tissue, including somatic gene coding errors, base substitutions, gene insertions, and the total number of actual errors. We used the Bioconductor R package Maftools to efficiently and comprehensively analyze somatic variants of CRC patients. Then, the Wilcoxon rank-sum test was used to compare the statistic differences.

#### 4.10. Comparison of DDR Gene Expression between Two Clusters

The expression of 276 DDR genes in Cluster II relative to Cluster I were calculated and  $\log_2(\text{Fold change (C II/C I)}) > 0.5$  or  $< -0.5$ ,  $p < 0.05$  were considered statistically significant.

#### 4.11. Immune Score and Stromal Score

Estimation of stromal and immune cells in malignant tumor tissues was based on the Expression Data (ESTIMATE) algorithm (<https://bioinformatics.mdanderson.org/estimate/>; accessed on 15 April 2021) and Xcell (<https://xcell.ucsf.edu/>; accessed on 15 April 2021). The degree of infiltration of immune and stromal cells in CRC tissues was calculated by ESTIMATE and Xcell algorithms. Statistical differences between the two clusters were utilized with Wilcoxon rank-sum test.

#### 4.12. Comparison of Immune Cell Abundance between Two Clusters

The fraction of immune cell types in a mixed cell population within the leukocyte compartment was estimated using Xcell and CIBERSORT. Indeed, we applied Xcell and CIBERSORT to the 2012 The Cancer Genome Atlas (TCGA) or the 2019 CPTAC CRC RNAseq data, using a set of 64 immune cells (Xcell) or a set of 22 immune cells (CIBERSORT) to determine relative proportions of immune cells in two clusters. Statistical differences between the two clusters were utilized with the Wilcoxon rank-sum test.

#### 4.13. Comparison of Immunomodulators between Two Clusters

From published studies, a 78 immunomodulators (IMs) genes list was obtained [35]. We applied the 2012 TCGA CRC mRNA data to calculate differential expression analysis for IMs between two clusters and then performed a Wilcoxon test for the mean expression level of each gene in the two clusters.

#### 4.14. Dependency Data

Dependency data were taken from three datasets: Achilles CRISPR genome-wide CRISPR-Cas9 KO data [27], and two other published articles [43,45]. Specifically, the CERES score was used to calculate the dependency score for a given 42 tumor-associated genes.

#### 4.15. Association Analysis between RLBP's Protein Expression and Cell Line Drug Response

Cell line drug sensitivity data were downloaded from the Genomics of Drug Sensitivity on the Cancer Project website (GDSC1 and GDSC2). Annotations of compound pathways and targets were also taken from Cancer Cell Line Encyclopedia (CCLE). The correlation between the protein expression of RLBP's individual genes and treatment responses for each drug was determined by Spearman's rank correlation coefficient.

#### 4.16. Cell Viability Measurement

In a 96-well plate, 5000 cells per well were inoculated. The concentration gradient drug treatment with a culture medium was applied on the second day. After being treated with indicated time, cell counting kit-8 reagent (Bimake, Cat#B34304, Houston, TX, USA) and cultural medium were added 1:10 to each well, and plates were incubated at 37 °C for 2 h. Absorbance was measured at 450 nm and normalized against control cells treated with vehicle solution. The experiments were repeated at least three times.

#### 4.17. Immunohistochemistry

IHC staining using previously reported methods. Antibodies were optimized with a pre-determined staining protocol: MCM3 (Abcam, Cat#ab4460, 1:500, Cambridge, UK), MCM5 (Proteintech, Cat#11703-1-AP, 1:100, Rosemont, IL, USA), CD68 (Santa Cruz, Cat#20060, 1:400, Dallas, TX, USA), DKC1 (Novus, Cat#NBP1-85156, 1:100, Littleton, CO, USA), CD8 (Abcam, Cat#ab199016, 1:500) and observed in Leica Microscope.

#### 4.18. Establishment and Culture of PDOs from Colon Cancer

After being washed with iodine tincture and PBS, the biopsies were thoroughly minced into the smallest pieces and digested with 10 mL Gentle Cell Dissociation Reagent (GCDR) (STEMCELL, Cat#07174, Vancouver, BC, Canada) with rotation at 40 rpm for 30 min at 4 °C. Cell pellet was filtered through 70 µm strainer (STEMCELL, Cat#27216) and counted with crypts. Next, 1000 crypts resuspending per dome were plated with Matrigel (Corning, Cat#356231, Corning, NY, USA) at 1:4 into 24 well-pre-warmed cell culture plates. The Matrigel dome was then solidified in a 37 °C incubator for 10 min and cultured with 500 µL IntestiCultTMOGM Human Basal Medium (STEMCELL, Cat#100-0190) mixed with Organoid Supplement (1:1 volume) (STEMCELL, Cat#100-0191), add desired antibiotics 50 µg/mL gentamicin (Thermo Fisher Scientific, Cat#15710064, Waltham, MA, USA) and 100 units penicillin/Streptomycin (Thermo Fisher Scientific, Cat#15140122).

Patient derived organoids (PDOs) were passaged every 5–10 days using TrypLe (Thermo Fisher Scientific, 12604013). In short, 300–600 µL TrypLe were added into the well and thoroughly scrape the Matrigel dome, transferring the organoid mixture to a 15 mL conical tube. Centrifuge the organoid at  $290 \times g$  for 5 min and resuspended in Matrigel and re-seeded at an appropriate ratio. PDOs were cryopreserved with 90% serum + 10% DMSO.

#### 4.19. 3D PDOs Drug Screens

Organoids digestion: first, the organs in the petri dish were digested and incubated at 37 °C for 1–5 min. Observed under the microscope every minute, when the organoids were digested into a single cell mass containing 2–3 cells, the medium was added to terminate the digestion, and a 1200 rpm centrifugation for 5 min, the supernatant was discarded, added to 2 mL 2% Matrigel culture medium, and the cell suspension was prepared. Grafted organoids: allowed the 384-well microplates to polymerize. After being cultured in the incubator for two days, the drug was added according to the drug concentration gradient, and then cultured in the incubator for four days. ATP fluorescence detection: after four days



of drug culture, the drug sensitivity test was carried out. Chemiluminescence detection was carried out by using a chemiluminescence detector. The tumor inhibition rate and IC50 value of the drug were calculated according to the chemiluminescence values of different concentrations of the drug.

#### 4.20. Statistics

The statistical details of all experiments are reported in the form of words, figure legends, and figure diagrams, including statistical analysis, statistical significance, and counting. Data were analyzed with a two-tailed *t*-test and Rstudio software. The correlation between groups was analyzed by Spearman's or Pearson's test, and  $p < 0.05$  was considered to be statistically significant (\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ). Statistical analyses were conducted using GraphPad Prism 7 (Harvey Motulsky, California, USA). For all experiments with error bars, the standard error (SEM) of the average was calculated to represent the change in each experiment. The sample size of each experiment and the number of repetitions of the experiment are included in the figure legends.

### 5. Conclusions

We identified two RLBP subtypes with different therapeutic and prognostic significance in CRC, based on proteogenomic analysis. CI-pRLBPs were associated with CIN, better prognosis, and sensitivity to drugs targeting genome integrity and EGFR.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers14225607/s1>, Figure S1: Proteomics-based molecular subtypes by the random forest model in 2019 and 2020 CRC databases. Figure S2: Proteomics-based clustering of RLBP expression in CRC patients; Figure S3: Landscape of pathway enrichment and DDR genes expression in two clusters; Figure S4: The correlation of RLBP with immunomodulatory factors; Figure S5: The RLBP are affected by CNA, DNA methylation (promoter CpG clusters) or transcriptional regulations; Figure S6: RLBP serve as potential therapeutic targets; Figure S7: The Drug sensitivity analysis in cells with differential RLBP expression; Table S1: The consistent expressions of RLBP in two cluster and predictions for prognosis; Table S2: Clinical signatures and molecular features in CI and CII; Table S3: Immune characteristics of CI and CII; Table S4: 42 RLBP are highly expressed in colorectal tumors; Table S5: Lists of potential therapeutic targets in RLBP; Table S6: The Drug sensitivity analysis in cells with DDX21 expressions.

**Author Contributions:** W.Z. and Q.P. performed the major experiments. R.T. and M.W. (Ming Wen) designed the experiments, performed the data analysis, and provided the project comments. H.P. and Q.P. provided the colon cancer patient sample. Y.Q. and M.W. (Meng Wang) assisted with the experiments and provided technical help. Y.Z., D.Z., G.M. and K.Z. assisted the proteomic data analysis. W.Z., Q.P., L.-Q.S., R.T. and M.W. (Ming Wen) were involved in manuscript preparation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (81974451, 81902888, 81702956, 82030087), the Natural Science Foundation of Hunan Province for Outstanding Young Scholars (2020JJ3062), the Natural Science Foundation of Hunan Province (2020JJ4903, 2020JJ5920), Innovation Foundation of Central South University for Outstanding Young Scholars (2019CX038), Award of Hunan Province for 2021 Hu Xiang Talent (2020RC3061), Youth Research Foundation of Xiangya Hospital, Central South University (2018Q07), and Open Project Program of the State Key Laboratory of Proteomics (SKLP-O202002), the Construction of Innovative Ability of National Clinical Research Center for Geriatric Disorders (2019SK2335), the Strategy-Oriented Special Project of Central South University of China (ZLXD2017003).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Xiangya Hospital, Central South University (protocol code 202103496, approved on 9 March 2021).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patients to publish this paper.

**Data Availability Statement:** The datasets used in this article for analysis are all public. A 204 DNA/RNA Hybrids gene list was obtained from published studies [16,17]. Source proteomic data, mRNA, mutation, and CNA raw data of three colon cancer in this study can be accessed through the Clinical Proteomic Tumor Analysis Consortium (CPTAC) data portal and from the Firehose website (<https://cptac-data-portal.georgetown.edu/study-summary/S045>; accessed on 15 April 2021, <http://gdac.broadinstitute.org>; accessed on 15 April 2021, version 20130523, <http://www.biosino.org/node/project/detail/OEP000729>; accessed on 15 April 2021). Source proteomic data of Clear Cell Renal Cell Carcinoma, Head-and-neck squamous cell carcinoma, Lung adenocarcinoma, and Hepatocellular Carcinoma can be obtained from the CPTAC data portal (<https://cptac-data-portal.georgetown.edu/study-summary/S050>; accessed on 15 April 2021, <https://cptac-data-portal.georgetown.edu/study-summary/S054>; accessed on 15 April 2021, <https://cptac-data-portal.georgetown.edu/study-summary/S056>; accessed on 15 April 2021, <https://cptac-data-portal.georgetown.edu/study-summary/S049>; accessed on 15 April 2021). A 276 DDR gene list was obtained from published studies [30]. The gene fitness scores, proteomic data, mRNA expression, and Copy number alteration of the cell lines are available from the project Score web portal: <https://score.depmap.sanger.ac.uk>; accessed on 15 April 2021. Cancer cell lines' drug responses from this study were downloaded from the GDSC database. CRC scRNA-seq datasets from human samples can be downloaded from (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146771>; accessed on 15 April 2021, GEO accession number GSE146771). The rest of the data available in the article can be found in the supplemental information or from the authors upon request.

**Acknowledgments:** We acknowledge the usage of data from the cancer genome Atlas (TCGA), LinkedOmics, and The Clinical Proteomic Tumor Analysis Consortium (CPTAC). TCGA Data, Resources, and Materials were originally published by the National Cancer Institute. CPTAC data used in this publication were generated by the Clinical Proteomic Tumor Analysis Consortium (NCI/NIH). CCLE data used in this publication were generated by the Cancer Cell Line Encyclopedia. GDSC data used in this publication were generated by the Wellcome Sanger Institute. This work was supported by the National Natural Science Foundation of China (81974451, 81902888, 81702956, 82030087), the Natural Science Foundation of Hunan Province for Outstanding Young Scholars (2020JJ3062), the Natural Science Foundation of Hunan Province (2020JJ4903, 2020JJ5920), Innovation Foundation of Central South University for Outstanding Young Scholars (2019CX038), Award of Hunan Province for 2021 Hu Xiang Talent (2020RC3061) Youth Research Foundation of Xiangya Hospital, Central South University (2018Q07) and Open Project Program of the State Key Laboratory of Proteomics (SKLP-O202002), the Construction of Innovative Ability of National Clinical Research Center for Geriatric Disorders (2019SK2335), the Strategy-Oriented Special Project of Central South University of China (ZLXD2017003) and Beijing Bethune Charitable Foundation.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lengauer, C.; Kinzler, K.W.; Vogelstein, B. Genetic instabilities in human cancers. *Nature* **1998**, *396*, 643–649. [[CrossRef](#)] [[PubMed](#)]
2. Jiricny, J. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 335–346. [[CrossRef](#)] [[PubMed](#)]
3. Chan, Y.A.; Hieter, P.; Stirling, P.C. Mechanisms of genome instability induced by RNA-processing defects. *Trends Genet.* **2014**, *30*, 245–253. [[PubMed](#)]
4. Chen, L.; Chen, J.Y.; Zhang, X.; Gu, Y.; Xiao, R.; Shao, C.; Tang, P.; Qian, H.; Luo, D.; Li, H.; et al. R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol. Cell* **2017**, *68*, 745–757.e5. [[CrossRef](#)]
5. Aguilera, A.; Garcia-Muse, T. R loops: From transcription byproducts to threats to genome stability. *Mol. Cell* **2012**, *46*, 115–124. [[CrossRef](#)]
6. Garcia-Muse, T.; Aguilera, A. R Loops: From Physiological to Pathological Roles. *Cell* **2019**, *179*, 604–618. [[CrossRef](#)]
7. Yu, K.; Chedin, F.; Hsieh, C.L.; Wilson, T.E.; Lieber, M.R. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* **2003**, *4*, 442–451. [[CrossRef](#)]
8. Crossley, M.P.; Bocek, M.; Cimprich, K.A. R-Loops as Cellular Regulators and Genomic Threats. *Mol. Cell* **2019**, *73*, 398–411. [[CrossRef](#)]
9. Richard, P.; Manley, J.L. R Loops and Links to Human Disease. *J. Mol. Biol.* **2017**, *429*, 3168–3180. [[CrossRef](#)]
10. Skourti-Stathaki, K.; Proudfoot, N.J.; Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **2011**, *42*, 794–805. [[CrossRef](#)]
11. Gorthi, A.; Romero, J.C.; Loranc, E.; Cao, L.; Lawrence, L.A.; Goodale, E.; Iniguez, A.B.; Bernard, X.; Masamsetti, V.P.; Roston, S.; et al. EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature* **2018**, *555*, 387–391. [[CrossRef](#)] [[PubMed](#)]

12. Garcia-Rubio, M.L.; Perez-Calero, C.; Barroso, S.I.; Tumini, E.; Herrera-Moyano, E.; Rosado, I.V.; Aguilera, A. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genet.* **2015**, *11*, e1005674. [[CrossRef](#)] [[PubMed](#)]
13. Schwab, R.A.; Nieminuszczy, J.; Shah, F.; Langton, J.; Lopez Martinez, D.; Liang, C.C.; Cohn, M.A.; Gibbons, R.J.; Deans, A.J.; Niedzwiedz, W. The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Mol. Cell* **2015**, *60*, 351–361. [[CrossRef](#)] [[PubMed](#)]
14. Tuduri, S.; Crabbe, L.; Conti, C.; Tourriere, H.; Holtgreve-Grez, H.; Jauch, A.; Pantescio, V.; De Vos, J.; Thomas, A.; Theillet, C.; et al. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* **2009**, *11*, 1315–1324. [[CrossRef](#)]
15. Sordet, O.; Redon, C.E.; Guirouilh-Barbat, J.; Smith, S.; Solier, S.; Douarre, C.; Conti, C.; Nakamura, A.J.; Das, B.B.; Nicolas, E.; et al. Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Rep.* **2009**, *10*, 887–893. [[CrossRef](#)]
16. Wang, I.X.; Grunseich, C.; Fox, J.; Burdick, J.; Zhu, Z.; Ravazian, N.; Hafner, M.; Cheung, V.G. Human proteins that interact with RNA/DNA hybrids. *Genome Res.* **2018**, *28*, 1405–1414. [[CrossRef](#)]
17. Cristini, A.; Groh, M.; Kristiansen, M.S.; Gromak, N. RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep.* **2018**, *23*, 1891–1905. [[CrossRef](#)]
18. Cerritelli, S.M.; Crouch, R.J. Ribonuclease H: The enzymes in eukaryotes. *FEBS J.* **2009**, *276*, 1494–1505. [[CrossRef](#)]
19. Lockhart, A.; Pires, V.B.; Bento, F.; Kellner, V.; Luke-Glaser, S.; Yakoub, G.; Ulrich, H.D.; Luke, B. RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cell Rep.* **2019**, *29*, 2890–2900.e5. [[CrossRef](#)]
20. Sollier, J.; Stork, C.T.; Garcia-Rubio, M.L.; Paulsen, R.D.; Aguilera, A.; Cimprich, K.A. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **2014**, *56*, 777–785. [[CrossRef](#)]
21. Yasuhara, T.; Kato, R.; Hagiwara, Y.; Shiotani, B.; Yamauchi, M.; Nakada, S.; Shibata, A.; Miyagawa, K. Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. *Cell* **2018**, *175*, 558–570.e11. [[CrossRef](#)] [[PubMed](#)]
22. Zhang, B.; Wang, J.; Wang, X.; Zhu, J.; Liu, Q.; Shi, Z.; Chambers, M.C.; Zimmerman, L.J.; Shaddox, K.F.; Kim, S.; et al. Proteogenomic characterization of human colon and rectal cancer. *Nature* **2014**, *513*, 382–387. [[CrossRef](#)] [[PubMed](#)]
23. Vasaikar, S.; Huang, C.; Wang, X.; Petyuk, V.A.; Savage, S.R.; Wen, B.; Dou, Y.; Zhang, Y.; Shi, Z.; Arshad, O.A.; et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **2019**, *177*, 1035–1049.e19. [[CrossRef](#)]
24. Li, C.; Sun, Y.D.; Yu, G.Y.; Cui, J.R.; Lou, Z.; Zhang, H.; Huang, Y.; Bai, C.G.; Deng, L.L.; Liu, P.; et al. Integrated Omics of Metastatic Colorectal Cancer. *Cancer Cell* **2020**, *38*, 734–747.e9. [[CrossRef](#)] [[PubMed](#)]
25. Meyers, R.M.; Bryan, J.G.; McFarland, J.M.; Weir, B.A.; Sizemore, A.E.; Xu, H.; Dharia, N.V.; Montgomery, P.G.; Cowley, G.S.; Pantel, S.; et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **2017**, *49*, 1779–1784. [[CrossRef](#)]
26. Dempste, J.M.; Rossen, J.; Kazachkova, M.; Pan, J.; Kugener, G.; Root, D.E.; Tsherniak, A. Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. *BioRxiv* **2017**, 720243. [[CrossRef](#)]
27. Ghandi, M.; Huang, F.W.; Jane-Valbuena, J.; Kryukov, G.V.; Lo, C.C.; McDonald, E.R., 3rd; Barretina, J.; Gelfand, E.T.; Bielski, C.M.; Li, H.; et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **2019**, *569*, 503–508. [[CrossRef](#)]
28. Nusinow, D.P.; Szpyt, J.; Ghandi, M.; Rose, C.M.; McDonald, E.R., 3rd; Kalocsay, M.; Jane-Valbuena, J.; Gelfand, E.; Schweppe, D.K.; Jedrychowski, M.; et al. Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **2020**, *180*, 387–402.e16. [[CrossRef](#)]
29. Guinney, J.; Dienstmann, R.; Wang, X.; de Reynies, A.; Schlicker, A.; Soneson, C.; Marisa, L.; Roepman, P.; Nyamundanda, G.; Angelino, P.; et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **2015**, *21*, 1350–1356. [[CrossRef](#)]
30. Knijnenburg, T.A.; Wang, L.; Zimmermann, M.T.; Chambwe, N.; Gao, G.F.; Cherniack, A.D.; Fan, H.; Shen, H.; Way, G.P.; Greene, C.S.; et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep.* **2018**, *23*, 239–254.e6. [[CrossRef](#)]
31. Aran, D.; Hu, Z.; Butte, A.J. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **2017**, *18*, 220. [[CrossRef](#)] [[PubMed](#)]
32. Yoshihara, K.; Shahmoradgoli, M.; Martinez, E.; Vegesna, R.; Kim, H.; Torres-Garcia, W.; Trevino, V.; Shen, H.; Laird, P.W.; Levine, D.A.; et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **2013**, *4*, 2612. [[CrossRef](#)] [[PubMed](#)]
33. Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **2015**, *12*, 453–457. [[CrossRef](#)]
34. Saltz, J.; Gupta, R.; Hou, L.; Kurc, T.; Singh, P.; Nguyen, V.; Samaras, D.; Shroyer, K.R.; Zhao, T.; Batiste, R.; et al. Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Rep.* **2018**, *23*, 181–193.e7. [[CrossRef](#)]
35. Thorsson, V.; Gibbs, D.L.; Brown, S.D.; Wolf, D.; Bortone, D.S.; Ou Yang, T.H.; Porta-Pardo, E.; Gao, G.F.; Plaisier, C.L.; Eddy, J.A.; et al. The Immune Landscape of Cancer. *Immunity* **2019**, *51*, 411–412. [[CrossRef](#)] [[PubMed](#)]
36. Huang, C.; Chen, L.; Savage, S.R.; Eguiez, R.V.; Dou, Y.; Li, Y.; da Veiga Leprevost, F.; Jaehnig, E.J.; Lei, J.T.; Wen, B.; et al. Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell* **2021**, *39*, 361–379. [[CrossRef](#)] [[PubMed](#)]

37. Gao, Q.; Zhu, H.; Dong, L.; Shi, W.; Chen, R.; Song, Z.; Huang, C.; Li, J.; Dong, X.; Zhou, Y.; et al. Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* **2019**, *179*, 561–577.e22. [[CrossRef](#)] [[PubMed](#)]
38. Gillette, M.A.; Satpathy, S.; Cao, S.; Dhanasekaran, S.M.; Vasaikar, S.V.; Krug, K.; Petralia, F.; Li, Y.; Liang, W.-W.; Reva, B.; et al. Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. *Cell* **2020**, *182*, 200–225.e35. [[CrossRef](#)]
39. Zhang, H.; Liu, T.; Zhang, Z.; Payne, S.H.; Zhang, B.; McDermott, J.E.; Zhou, J.Y.; Petyuk, V.A.; Chen, L.; Ray, D.; et al. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **2016**, *166*, 755–765. [[CrossRef](#)]
40. Clark, D.J.; Dhanasekaran, S.M.; Petralia, F.; Pan, J.; Song, X.; Hu, Y.; da Veiga Leprevost, F.; Reva, B.; Lih, T.M.; Chang, H.Y.; et al. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* **2019**, *179*, 964–983.e31. [[CrossRef](#)]
41. Zhang, L.; Li, Z.; Skrzypczynska, K.M.; Fang, Q.; Zhang, W.; O'Brien, S.A.; He, Y.; Wang, L.; Zhang, Q.; Kim, A.; et al. Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell* **2020**, *181*, 442–459.e29. [[CrossRef](#)] [[PubMed](#)]
42. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **2012**, *487*, 330–337. [[CrossRef](#)] [[PubMed](#)]
43. Hart, T.; Tong, A.H.Y.; Chan, K.; Van Leeuwen, J.; Seetharaman, A.; Aregger, M.; Chandrashekhar, M.; Hustedt, N.; Seth, S.; Noonan, A.; et al. Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 Genes Genomes Genet.* **2017**, *7*, 2719–2727. [[CrossRef](#)] [[PubMed](#)]
44. Hart, T.; Chandrashekhar, M.; Aregger, M.; Steinhart, Z.; Brown, K.R.; MacLeod, G.; Mis, M.; Zimmermann, M.; Fradet-Turcotte, A.; Sun, S.; et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **2015**, *163*, 1515–1526. [[CrossRef](#)]
45. Behan, F.M.; Iorio, F.; Picco, G.; Goncalves, E.; Beaver, C.M.; Migliardi, G.; Santos, R.; Rao, Y.; Sassi, F.; Pinnelli, M.; et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **2019**, *568*, 511–516. [[CrossRef](#)]
46. Dempster, J.M.; Boyle, I.; Vazquez, F.; Root, D.; Boehm, J.S.; Hahn, W.C.; Tsherniak, A.; McFarland, J.M. Chronos: A CRISPR cell population dynamics model. *BioRxiv* **2021**. [[CrossRef](#)]
47. Chang, L.; Ruiz, P.; Ito, T.; Sellers, W.R. Targeting pan-essential genes in cancer: Challenges and opportunities. *Cancer Cell* **2021**, *39*, 466–479. [[CrossRef](#)]
48. Iorio, F.; Knijnenburg, T.A.; Vis, D.J.; Bignell, G.R.; Menden, M.P.; Schubert, M.; Aben, N.; Goncalves, E.; Barthorpe, S.; Lightfoot, H.; et al. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **2016**, *166*, 740–754. [[CrossRef](#)]
49. Yao, Y.; Xu, X.; Yang, L.; Zhu, J.; Wan, J.; Shen, L.; Xia, F.; Fu, G.; Deng, Y.; Pan, M.; et al. Patient-Derived Organoids Predict Chemoradiation Responses of Locally Advanced Rectal Cancer. *Cell Stem Cell* **2019**, *26*, 17–26.e6. [[CrossRef](#)]
50. Ooft, S.N.; Weeber, F.; Dijkstra, K.K.; McLean, C.M.; Kaing, S.; Werkhoven, E.V.; Schipper, L.; Hoes, L.; Vis, D.J.; Haar, J.V.D.; et al. Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. *Sci. Transl. Med.* **2019**, *11*, eaay2574. [[CrossRef](#)]
51. Ganesh, K.; Wu, C.; O'Rourke, K.P.; Szeglin, B.C.; Zheng, Y.; Sauve, C.G.; Adileh, M.; Wasserman, I.; Marco, M.R.; Kim, A.S.; et al. A rectal cancer organoid platform to study individual responses to chemoradiation. *Nat. Med.* **2019**, *25*, 1607–1614. [[CrossRef](#)] [[PubMed](#)]
52. Kopper, O.; de Witte, C.J.; Lohmussaar, K.; Valle-Inclan, J.E.; Hami, N.; Kester, L.; Balgobind, A.V.; Korving, J.; Proost, N.; Begthel, H.; et al. An organoid platform for ovarian cancer captures intra- and interpatient heterogeneity. *Nat. Med.* **2019**, *25*, 838–849. [[CrossRef](#)] [[PubMed](#)]
53. Mirzaei, G.; Petreaca, R.C. Distribution of copy number variations and rearrangement endpoints in human cancers with a review of literature. *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **2022**, *824*, 111773. [[CrossRef](#)] [[PubMed](#)]
54. Sansregret, L.; Vanhaesebroeck, B.; Swanton, C. Determinants and clinical implications of chromosomal instability in cancer. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 139–150. [[CrossRef](#)]
55. Dahiya, R.; Hu, Q.; Ly, P. Mechanistic origins of diverse genome rearrangements in cancer. *Semin. Cell Dev. Biol.* **2022**, *123*, 100–109. [[CrossRef](#)]
56. Mehta, A.; Haber, J.E. Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016428. [[CrossRef](#)]
57. Li, X.; Baek, G.; Ramanand, S.G.; Sharp, A.; Gao, Y.; Yuan, W.; Welti, J.; Rodrigues, D.N.; Dolling, D.; Figueiredo, I.; et al. BRD4 Promotes DNA Repair and Mediates the Formation of TMPRSS2-ERG Gene Rearrangements in Prostate Cancer. *Cell Rep.* **2018**, *22*, 796–808. [[CrossRef](#)]
58. Day, T.A.; Layer, J.V.; Cleary, J.P.; Guha, S.; Stevenson, K.E.; Tivey, T.; Kim, S.; Schinzel, A.C.; Izzo, F.; Doench, J.; et al. PARP3 is a promoter of chromosomal rearrangements and limits G4 DNA. *Nat. Commun.* **2017**, *8*, 15110. [[CrossRef](#)]
59. Xie, J.; Wen, M.; Zhang, J.; Wang, Z.; Wang, M.; Qiu, Y.; Zhao, W.; Zhu, F.; Yao, M.; Rong, Z.; et al. The Roles of RNA Helicases in DNA Damage Repair and Tumorigenesis Reveal Precision Therapeutic Strategies. *Cancer Res.* **2022**, *82*, 872–884. [[CrossRef](#)]
60. Wray, J.; Williamson, E.A.; Singh, S.B.; Wu, Y.; Cogle, C.R.; Weinstock, D.M.; Zhang, Y.; Lee, S.H.; Zhou, D.; Shao, L.; et al. PARP1 is required for chromosomal translocations. *Blood* **2013**, *121*, 4359–4365. [[CrossRef](#)]
61. Burrell, R.A.; McClelland, S.E.; Endesfelder, D.; Groth, P.; Weller, M.C.; Shaikh, N.; Domingo, E.; Kanu, N.; Dewhurst, S.M.; Gronroos, E.; et al. Replication stress links structural and numerical cancer chromosomal instability. *Nature* **2013**, *494*, 492–496. [[CrossRef](#)] [[PubMed](#)]



- 
62. McGranahan, N.; Burrell, R.A.; Endesfelder, D.; Novelli, M.R.; Swanton, C. Cancer chromosomal instability: Therapeutic and diagnostic challenges. *EMBO Rep.* **2012**, *13*, 528–538. [[CrossRef](#)] [[PubMed](#)]
  63. Normanno, N.; Tejpar, S.; Morgillo, F.; De Luca, A.; Van Cutsem, E.; Ciardiello, F. Implications for KRAS status and EGFR-targeted therapies in metastatic CRC. *Nat. Rev. Clin. Oncol.* **2009**, *6*, 519–527. [[CrossRef](#)] [[PubMed](#)]
  64. Sarkar, K.; Han, S.S.; Wen, K.K.; Ochs, H.D.; Dupre, L.; Seidman, M.M.; Vyas, Y.M. R-loops cause genomic instability in T helper lymphocytes from patients with Wiskott-Aldrich syndrome. *J. Allergy Clin. Immunol.* **2018**, *142*, 219–234. [[CrossRef](#)] [[PubMed](#)]
  65. Massague, J. TGFbeta in Cancer. *Cell* **2008**, *134*, 215–230. [[CrossRef](#)]
  66. Shen, Y.J.; Le Bert, N.; Chitre, A.A.; Koo, C.X.; Nga, X.H.; Ho, S.S.; Khatoo, M.; Tan, N.Y.; Ishii, K.J.; Gasser, S. Genome-derived cytosolic DNA mediates type I interferon-dependent rejection of B cell lymphoma cells. *Cell Rep.* **2015**, *11*, 460–473. [[CrossRef](#)]
  67. Wan, W.; Ao, X.; Chen, Q.; Yu, Y.; Ao, L.; Xing, W.; Guo, W.; Wu, X.; Pu, C.; Hu, X.; et al. METTL3/IGF2BP3 axis inhibits tumor immune surveillance by upregulating N(6)-methyladenosine modification of PD-L1 mRNA in breast cancer. *Mol. Cancer* **2022**, *21*, 60. [[CrossRef](#)]
  68. 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* **2017**, *84*, 69–80. [[CrossRef](#)]