

Article

Identification of Potential Biomarkers and Biological Pathways for Poor Clinical Outcome in Mucinous Colorectal Adenocarcinoma

Chang Woo Kim ¹, Jae Myung Cha ² and Min Seob Kwak ^{2,*}¹ Department of Surgery, Ajou University College of Medicine, Suwon 16499, Korea; kcwgkim@gmail.com² Department of Internal Medicine, Kyung Hee University Hospital at Gangdong, Kyung Hee University College of Medicine, 892 Dongnam-ro, Gandong-gu, Seoul 05278, Korea; 4528359@gmail.com

* Correspondence: kwac63@khu.ac.kr; Tel.: +82-2-440-6119; Fax: +82-2-440-6295

Simple Summary: Patients with mucinous adenocarcinoma (MAC) have been considered to have a faster disease progression than patients with traditional adenocarcinoma (TAC) in colorectal cancer (CRC). However, to date, the roles of MAC in long-term survival remain controversial due to a small sample size and the nature of its relatively rare occurrence, although it potentially represents entities with different aggressiveness and prognoses. Here, using large-scale population data, we found that the patients with the MAC subtype had a significantly worse overall survival rate and a tendency of worse disease-specific survival rate in stage II compared with the patients with the TAC subtype. Furthermore, key gene signatures were identified using the established predictive models for the disease-specific survival of stage II mucinous CRC.

Abstract: Colorectal cancer (CRC) comprises several histological subtypes, but the influences of the histological subtypes on prognosis remains unclear. We sought to evaluate the prognosis of mucinous adenocarcinoma (MAC), compared to that of traditional adenocarcinoma (TAC). This study used the data of patients diagnosed with CRC between 2004 and 2016, as obtained from the Surveillance, Epidemiology, and End Results database. We established a predictive model for disease-specific survival using conditional survival forest, model, non-linear Cox proportional hazards, and neural multi-task logistic regression model and identified the gene signatures for predicting poor prognosis based on the arrayexpress datasets. In total, 9096 (42.1%) patients with MAC and 12,490 (58.9%) patients with TAC were included. Those with the MAC subtype were more likely to have a poorer overall survival rate compared to those with the TAC subtype in stage II CRC ($p = 0.002$). The eight major genes including RPS18, RPL30, NME2, USP33, GAB2, RPS3A, RPS25, and CEP57 were found in the interacting network pathway. MAC was found to have a poorer prognosis compared to TAC, especially in Stage II CRC. In addition, our findings suggest that identifying potential biomarkers and biological pathways can be useful in CRC prognosis.

Keywords: colorectal cancer; mucinous; survival; biomarker; gene



Citation: Kim, C.W.; Cha, J.M.; Kwak, M.S. Identification of Potential Biomarkers and Biological Pathways for Poor Clinical Outcome in Mucinous Colorectal Adenocarcinoma. *Cancers* **2021**, *13*, 3280. <https://doi.org/10.3390/cancers13133280>

Academic Editor: Mark Molloy

Received: 12 June 2021

Accepted: 29 June 2021

Published: 30 June 2021

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



Copyright: © 2021 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/>).

1. Introduction

Traditionally, the TNM staging system has played a fundamental role in the management of patients with colorectal cancer (CRC) as the most powerful and reliable predictor of prognosis [1]. However, clinicians often encounter various clinical outcomes after patients receive standard curative treatment, even among CRC patients with the same stage, which is unexpected and inexplicable. Increasing attention is being focused on the identification of new factors, which may enable a more accurate patient prognostic stratification within each stage in CRC.

Apart from the TNM staging system, histological classification, which is based on histological characteristics and genetic features traits, may influence the clinical features and outcome in CRC; thus, clarifying the effect of the varied histological subtypes will

help clinicians choose the appropriate treatment strategy [2,3]. Mucinous adenocarcinoma (MAC) is defined by the World Health Organization (WHO) as ‘an adenocarcinoma containing more than 50% extracellular mucin within the tumor’ [4], which was found to occur in approximately 10% of all CRCs [5]. Compared with other subtypes, the MAC subtype is characterized by often occurring in a younger age [6], in a more advanced stage [7,8], and at the proximal colon [9,10]. From these findings, patients with MAC may be expected to have a faster disease progression than patients with traditional adenocarcinoma (TAC), which is classified as tubular, villous, or tubulovillous adenocarcinoma.

However, the actual clinical impact on the long-term survival of MAC patients compared that of with TAC patients and the causal link with its tumor biology remains highly controversial, although it potentially represents entities with different aggressiveness and prognoses [11]. To address these issues, we applied the incidence-based mortality method to Surveillance, Epidemiology, and End Results (SEER) data, a national population-based cancer registry [12], to investigate the long-term survival differences between MAC and TAC in CRC, followed by identifying the candidate biomarker genes and their pathways related to prognosis by investigating genomic data.

2. Materials and Methods

2.1. Data Sources

We used data from the National Cancer Institute’s SEER Program database, which was created to collect cancer incidence, prevalence, and survival data from U.S. cancer registries. SEER collects and publishes cancer incidence and survival data from 17 population-based cancer registries (Alaska Native tumor registry, Connecticut, Georgia Center for Cancer Statistics, San Francisco-Oakland, San Jose-Monterey, Greater California, Hawaii, Idaho, Iowa, Kentucky, Los Angeles, Louisiana, Massachusetts, New Mexico, New York, Seattle-Puget Sound, and Utah), covering approximately 34.6% of the U.S. population [13]. The database is broadly representative of the U.S. population. The SEER registry contains information on nine million cancer cases with over 550,000 new cases added to the database each year. It serves as a powerful resource for researchers focused on understanding the natural history of CRC and improving patient treatment [13,14]. This retrospective cohort study was reviewed and approved by the Institutional Review Board of the Kyung Hee University Hospital at Gangdong, Seoul, Republic of Korea (KHNMIC IRB 2020-01-015). The need for informed consent was waived on account of the fact that all of the data used in this study were de-identified.

2.2. Study Population

The SEER registry collects data, including age at diagnosis, sex, race, primary site, histological type, tumor grade, tumor size, tumor depth, and survival. Using the SEER 1975–2016 database (released on 15 April 2019), we analyzed data from all of the patients diagnosed with CRC in the years 2004–2016. We extracted clinical and/or demographic data, including age at diagnosis, sex, race, and tumor information, including location, size, grade, histological type, and American Joint Committee on Cancer 7th TNM stages by using SEER disease codes. Tumor location was determined by using the following codes: cecum (C18.0), appendix (C18.1), ascending colon (C18.2), hepatic flexure (C18.3), transverse colon (C18.4), splenic flexure (C18.5), descending colon (C18.6), sigmoid colon (C18.7), overlapping lesion of colon (C18.8), colon (C18.9), rectosigmoid (C19.9), and rectum (C20.9). The morphology of cancer was categorized according to the third edition of the International Classification of Diseases for Oncology (ICD-O-3) histology and behavior codes: tubular adenocarcinoma, (8211/3); adenocarcinoma in villous adenoma (8261/3), villous adenocarcinoma (8262/3), adenocarcinoma in tubulovillous adenoma (8263/3), and MAC (8480/3). For tumor differentiation grading, we used a four-tier classification (well-differentiated, moderately differentiated, poorly differentiated, undifferentiated), which was proposed by the WHO grading system [15]. Patients were divided into two

groups: the TAC group (patients with tubular adenocarcinoma, villous adenocarcinoma, and tubulovillous adenocarcinoma) and the MAC group (patients with MAC).

2.3. Establishment of Predictive Model

To develop a clinical prediction model for disease-specific survival, we used three predictive models for poor outcomes in CRC patients with MAC based on SEER data. For the tree approach, a conditional survival forest model (CSF) was conducted with 200 decision trees and a maximum depth of 5 to the square root of the total number of features [16]. The non-linear Cox proportional hazards (nCPH) model is frequently used to model survival data or time-to-event data, particularly in the presence of censored survival times [17]. Bent identity was used as an activation function. Furthermore, the learning rate, dropout rate, and number of epochs were set to 0.01, 0.2 and 5000, respectively. Finally, a neural multi-task logistic regression model (N-MTLR), which is based on a multi-task framework allowing the use of neural networks within the original MTLR design, was used [18]. The model was trained over 2000 epochs with a learning rate of 0.0001 using the Adam optimizer and rectified linear unit function. The patients were randomly assigned to a training set (80%) or a test set (20%).

2.4. Gene Signature Evaluation

To identify the different gene signatures for predicting poor outcome using the developed model, two previously published datasets (E-MTAB-863 and E-MTAB-864; ArrayExpress) were retrieved. A total of 37 patients from E-MTAB-863 and 18 patients from E-MTAB-864 were available for analysis after excluding patients with missing variables and other histological types except for MAC. The data were executed in robust multi-array analysis background correction and log₂ transformed, then normalized through quantile normalization.

The linear models for the microarray data package in Bioconductor were utilized to mine statistically significant differentially expressed genes (DEGs) based on the difference in their expression values between MAC and TAC [19]. A *p*-value of <0.05 and |log₂ fold change of | ≥ 1 were used as the cutoff criteria for this analysis. Further analysis, including construction of a human Protein-Protein Interaction (PPI) network, was carried out with the web-based tool NetworkAnalyst (<http://www.networkanalyst.ca>), which supports robust and reliable gene expression analysis [20,21], and pathway activations were selected and matched according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

2.5. Statistical Analysis

The SEER data were obtained using the SEER*Stat software (8.3.6 version; Surveillance Research Program, National Cancer Institute, Bethesda, MD, USA). To account for the potential effect of covariates on each outcome, we created a propensity score combining the covariates of age, sex, race, tumor grade, and tumor size and compared the prognosis of CRC patients with TAC and those with MAC. The cumulative probabilities of survival were assessed according to the Kaplan-Meier life-table analysis. A log-rank test was used to compare the survival rates between the TAC and MAC groups. Demographic differences between the two groups were tested using the Student's *t*-test and Pearson chi-square test. All analyses were performed with Python (version 3.6.9) and R statistical software (version 3.6.0). A two-sided *p* ≥ 0.05 was considered statistically significant.

3. Results

3.1. Construction of the Prediction Model for Disease-Specific Survival

A total of 21,586 patients were included: 9096 (42.1%) patients with MAC and 12,490 (58.9%) patients with TAC in the SEER database. The MAC group had significantly higher proportions of older patients, female sex (except stage III), higher grades, and larger tumor sizes than the TAC group in all TNM stages (Supplementary Table S1). The

propensity score matching eliminated significant differences between the two groups and the distribution of patient characteristics in each group was similar to that of all patients (Table 1).

Table 1. Baseline characteristics of the patients with stage II CRC in SEER database.

KERRYPNX	Before Matching			After Matching		
	MAC n = 4390	TAC n = 2704	p-Value	MAC n = 2618	TAC n = 2618	p-Value
Age at diagnosis, y, mean (SD)	69.8 (14.5)	68.6 (13.9)	<0.001	68.8 (14.7)	68.9 (13.8)	0.877
Sex, n (%)						
Male	2010 (45.8)	1326 (49.0)	<0.001	1267 (48.4)	1263 (48.2)	0.934
Female	2380 (54.2)	1378 (51.0)		1351 (51.6)	1355 (51.8)	
Tumor grade, n (%)						
Well-differentiated	522 (11.9)	341 (12.6)		335 (12.8)	326 (12.5)	
Moderately-differentiated	3144 (71.6)	2105 (77.8)	<0.001	2040 (77.9)	2036 (77.8)	0.151
Poorly-differentiated	615 (14.0)	213 (7.9)		217 (8.3)	211 (8.1)	
Undifferentiated	109 (2.5)	45 (1.7)		26 (1.0)	45 (1.7)	
Primary site, n (%)						
Cecum	1277 (29.1)	915 (33.8)				
Ascending	1219 (27.8)	605 (22.4)				
Hepatic flexure	302 (6.9)	141 (5.2)				
Transverse	531 (12.1)	252 (9.3)	<0.001			
Splenic flexure	155 (3.5)	99 (3.7)				
Descending	217 (4.9)	129 (4.8)				
Sigmoid	592 (13.5)	512 (18.9)				
Rectal	97 (2.2)	51 (1.9)				
Race, n (%)						
Hispanic (All Races)	469 (10.7)	315 (11.6)		304 (11.6)	302 (11.5)	
Non-Hispanic American Indian/Alaska Native	12 (0.3)	10 (0.4)	<0.001	8 (0.3)	7 (0.3)	0.917
Non-Hispanic Asian or Pacific Islander	261 (5.9)	174 (6.4)		173 (6.6)	168 (6.4)	
Non-Hispanic Black	453 (10.3)	364 (13.5)		349 (13.3)	329 (12.6)	
Non-Hispanic White	3195 (72.8)	1841 (68.1)		1784 (68.1)	1812 (69.2)	
Tumor depth, n (%)						
T3	3604 (82.1)	2376 (87.9)	<0.001			
T4	786 (17.9)	328 (12.1)				
Tumor size, mm, mean (SD)	64.5 (42.4)	54.0 (47.6)	<0.001	54.3 (30.0)	53.3 (30.4)	0.211

MAC, mucinous adenocarcinoma; SD, standard deviation; SEER, Surveillance, Epidemiology, and End Results; TAC, traditional adenocarcinoma.

After adjustment, we conducted subgroup analysis by tumor stage to further investigate the disease-specific and overall survival rates. The patients with the MAC subtype were more likely to have a poor overall survival rate compared to patients with the TAC subtype, notably patients with stage II ($p = 0.002$; Figure 1). It did not reach statistical significance for disease-specific survival between the two groups in stage II ($p = 0.161$). In addition, those in the MAC group showed poorer disease-specific survival rates than those in the TAC group with stage III and stage IV ($p < 0.001$ and $p < 0.015$; Figure 1). Similar results were also obtained by Cox proportional hazards analysis (Supplementary Table S2).

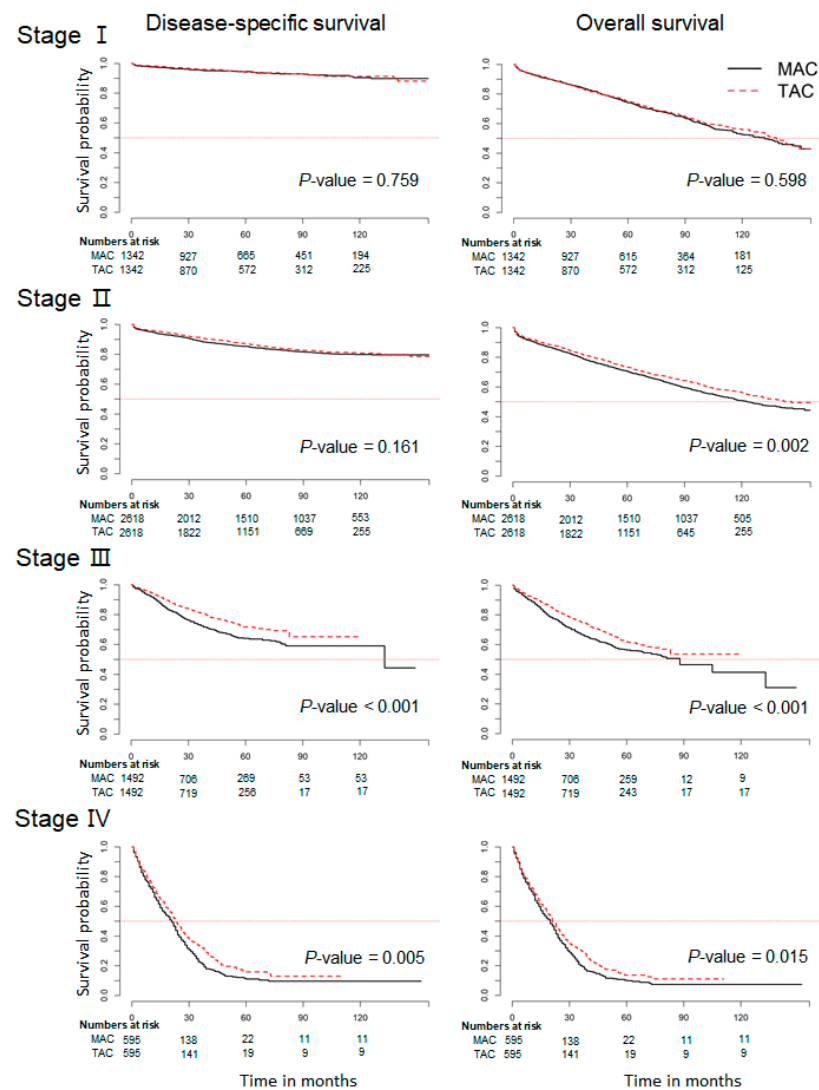


Figure 1. Kaplan–Meier curves of disease-specific and overall survival according to the stage between mucinous (MAC) and traditional colorectal cancer (TAC) in the SEER database.

3.2. Development and Validation of a Risk Stratification Model for MAC with Stage II

Because we found a difference in the not-cause specific survival rate between the MAC and TAC groups in Stage II diseases, a disease-specific survival predictive model in Stage II CRC using CSF, n-CPH, and N-MTLR models were established (Figure 2). Tested on 20% of the SEER dataset using the predictive models, the root mean square errors (RMSEs) were 12.506, 26.714, and 11.529 in CSF, n-CPH, and N-MTLR, respectively. Similarly, the RMSEs were 6.775, 6.447, and 6.230 as a result of the test in the E-MTAB 863 and 864 test sets.

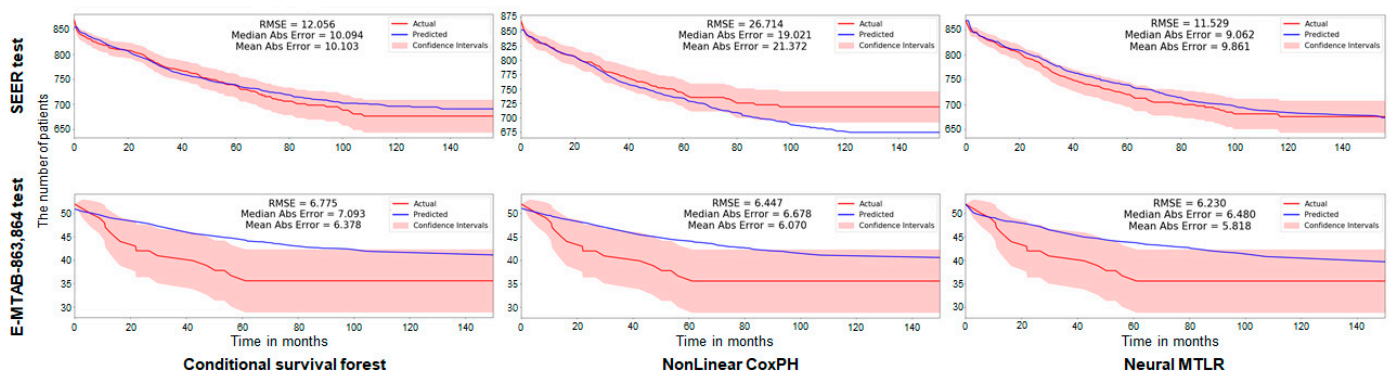


Figure 2. Internal and external validations of three developed models for disease-specific survival in mucinous colorectal cancer. The model shows reasonable discrimination in SEER database and arrayexpress datasets. RMSE, root mean square error.

3.3. Identifying the Differentially Expressed Genes in ArrayExpress Dataset

Of the E-MTAB-863 and E-MTAB-864 datasets, a total of 34 patients, commonly classified into high-risk and low-risk groups from the three established machine learning models (CSF, n-CPH, and N-MTLR), were identified (Table 2). Of 117 DEGs, we identified the 12 DEGs (RPS18, USP33, CENPL, GAB2, RPS3A, RPS25, RPL30, HNMT, NME2, CEP57, ZC3H8, and TRIT1) most highly associated with disease-specific survival in stage II mucinous CRC (Figure 3, Supplementary Figure S1 and Supplementary Table S2).

Table 2. Baseline characteristics of the patients with stage II CRC in array data.

Variables		High Risk <i>n</i> = 18	Low Risk <i>n</i> = 16	<i>p</i> -Value
Prognosis, <i>n</i> (%)	Good prognosis	5 (27.8)	13 (72.2)	0.006
	Poor prognosis	13 (72.2)	3 (16.7)	
Sex, <i>n</i> (%)	male	11 (61.1)	6 (33.3)	0.303
	female	7 (38.9)	10 (55.6)	
Cancer-related death, <i>n</i> (%)	Yes	11 (61.1)	2 (11.1)	0.011
	No	7 (38.9)	14 (77.8)	
Tumor grade, <i>n</i> (%)	Well-differentiated	1 (5.6)	2 (11.1)	0.466
	Moderately-differentiated	12 (66.7)	11 (61.1)	
	Poorly-differentiated	5 (27.8)	2 (11.1)	
	Undifferentiated	0 (0.0)	1 (5.6)	
Lymphovascular invasion, <i>n</i> (%)	Yes	5 (27.8)	2 (11.1)	0.507
	No	11 (61.1)	12 (66.7)	
	N/A	2 (11.1)	2 (11.1)	
T stage, <i>n</i> (%)	3	13 (72.2)	16 (88.9)	0.072
	4	5 (27.8)	0 (0.0)	
Tumor location, <i>n</i> (%)	Caecum	5 (27.8)	6 (33.3)	0.017
	Ascending colon	1 (5.6)	8 (44.4)	
	Hepatic flexure	2 (11.1)	0 (0.0)	
	Transverse colon	5 (27.8)	2 (11.1)	
	Splenic flexure	3 (16.7)	0 (0.0)	
	Descending colon	0 (0.0)	0 (0.0)	
	Sigmoid colon	2 (11.1)	0 (0.0)	

Table 2. Cont.

Variables		High Risk	Low Risk	p-Value
		n = 18	n = 16	
Family history of CRC	Yes	1 (5.6)	0 (0.0)	>0.999
	No	10 (55.6)	6 (33.3)	
	N/A	7 (38.9)	10 (55.6)	
Cancer recurrence	Yes	14 (77.8)	3 (16.7)	0.002
	No	4 (22.2)	13 (72.2)	
Age at diagnosis, y, mean (SD)		78.33 (8.02)	70.13 (5.82)	0.002
Length of follow up, y, mean (SD)		3.14 (2.16)	6.01 (2.74)	0.002
Tumor cell content, (%)		73.89 (14.61)	67.19 (16.22)	0.217
Number of regional LN assessed		14.5 (8.28)	17.5 (8.66)	0.312
Tumor size, cm, mean (SD)		7.02 (2.41)	5.28 (1.77)	0.022

CRC, colorectal cancer; LN, lymph node; SD, standard deviation.

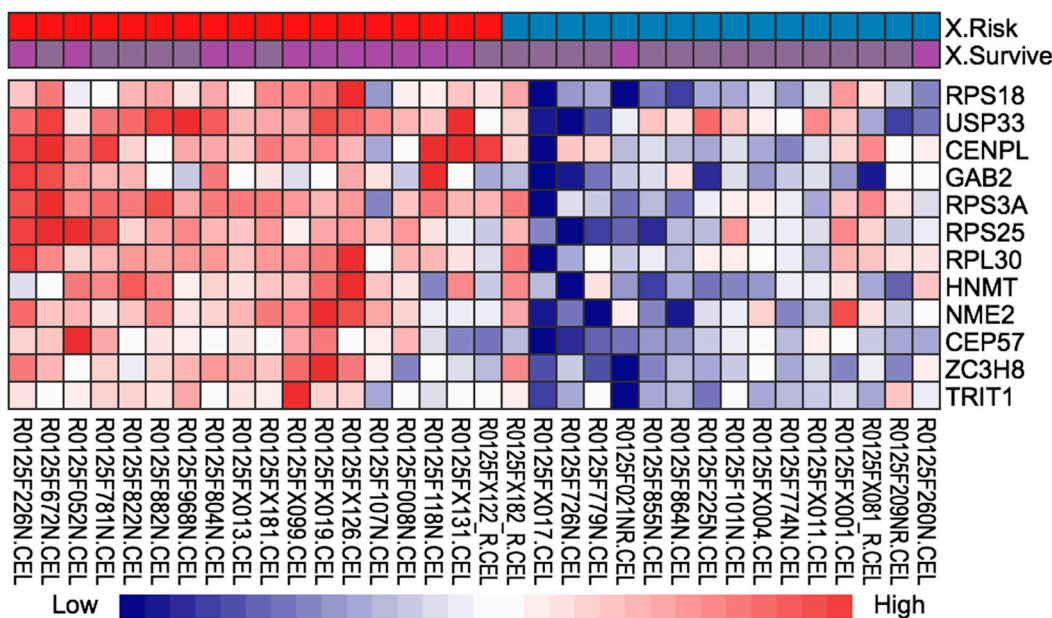


Figure 3. Differentially expressed genes identified from the arrayexpress datasets. Lower horizontal axis marks sample names, and right vertical axis represents gene names. Red represents up-regulated genes and blue represents down-regulated genes.

3.4. KEGG Pathway Analysis and Construction of the PPI Network

The interacting network analysis of responsive genes was further performed to identify the key genes. A total of 117 DEGs were subjected to KEGG pathway analysis (Supplementary Table S3). The genes were significantly enriched in the ‘pathways in cancer’, ‘viral carcinogenesis’, ‘PI3K-Akt signaling pathway’, ‘proteoglycans in cancer’, and ‘Ras signaling pathway’ (Supplementary Table S3). By performing further PPI network analysis, RPS18, RPL30, NME2, USP33, GAB2, RPS3A, RPS25, and CEP57 were indicated to be the common hub genes (Figure 4). Among them, RPS3A constituted the super hub node having the largest degree and highest betweenness (Supplementary Table S4).

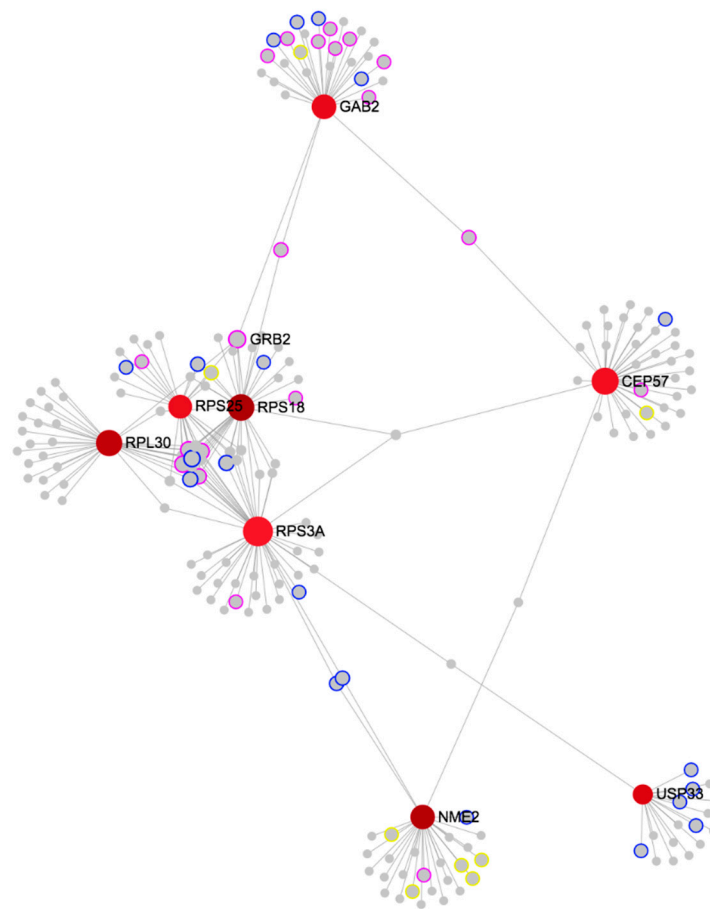


Figure 4. Protein-protein interaction networks of differentially expressed genes associated disease specific survival in mucinous colorectal cancer.

4. Discussion

Using a large-scaled population data, we found that the patients with the MAC subtype had a significantly worse overall survival rate and a tendency of worse disease-specific survival rate in stage II compared with the patients with the TAC subtype. In addition, key gene signatures were identified using the established predictive models for disease-specific survival of stage II mucinous CRC.

Results of the prognosis of MAC compared with TAC are still conflicting. Several studies have reported no correlation of MAC with the oncologic outcomes of CRC patients. Purdie and Piris [22], Xie [23], and Warschkow et al. [24] did not find the presence of MAC to be an independent prognostic factor for overall and disease-specific survival in patients with CRC. Similarly, a study involving 1025 unselected patients from Italy showed that the overall survival of patients with stage II and III colon cancer with MAC was not significantly different from those with non-MAC [25]. One study even reported better overall survival rates in patients with MAC [26]. These discrepancies may reflect heterogeneity in the population of patients, lack of stratification by stage, exclusion of rectal primary tumors in some studies, and a small sample size.

In contrast, some authors reported that the MAC subtype was related with a poor prognosis [5,7,27–29]. Common explanations are that patients with MAC were younger and had more advanced tumor stages compared with those with TAC. Notably, in the stratified analysis according to stage from a large-scale population-based cohort, we found stage II in MAC is the clinically critical timepoint to have more aggressive behavior for survival than other histological subtypes. A meta-analysis showed that MAC patients had worse survival rates than TAC patients, even after the stage at diagnosis was similar in MAC and TAC patients in their meta-analysis [5]. The analysis showed a slightly worse prognosis

for MAC (hazard ratio 1.05, 95% confidence interval 1.02–1.08) even when corrected for the stage at presentation. Another study with a stage-adjusted survival analysis from Germany also showed that the CRC patients with MAC with the same stage had a shorter disease-specific survival length and tended to have a worse disease-specific survival curve with a more advanced stage compared to those with TAC, but failed to achieve a level of statistical significance due to a small sample size [7].

However, even in the studies that showed a correlation of MAC with the prognosis, there are conflicts regarding MAC as an independent prognostic factor for CRC. A large retrospective study showed that MAC was not an independent prognostic factor of CRC in multivariate analysis, although the tumor stage and histological grade were higher in MAC than in TAC [7]. Similarly, another study reported that the MAC subtype was associated with adverse pathologic features, but it was not an independent prognostic factor [29]. These results suggest that other approaches, including biological, molecular, or genetic assessments, are needed to clarify the role of MAC in CRC.

Therefore, we tried to investigate novel insights into the biological differences associated with poor survival in stage II mucinous CRC. Of 12 DEGs for the disease-specific survival of stage II mucinous CRC, eight major genes with a biological relationship within the PPI network (RPS18, RPL30, NME2, USP33, GAB2, RPS3A, RPS25, and CEP57) were identified.

Most of the genes have usually been investigated as prognostic biomarkers in renal cancer, and RPS18, RPL30, NME2, and RPS25 usually have been investigated as unfavorable predictors, while GAB2 has been reported as favorable predictor in renal cancer [30].

RPS18, the molecules within ribosomal proteins, has been demonstrated to be up-regulated in CRC tissues. Its upregulation is a well-known common feature of active proliferation and the proliferation rate of tumor cells [31]. The increased number of ribosomal protein L30 (RPL30) may play a central role in the CRC process and in induced hepatocellular carcinoma [32,33]. The overexpression of NME2 was significantly associated with not only clinical parameters related to tumor progression, invasion, and metastasis but also resistance to 5-FU treatment [34,35]. The low expression of USP33 indicated a high recurrence risk and poor overall prognosis in advanced CRC patients [36]. Grb2-associated binders 2 (GAB2) contributed to tumor growth and angiogenesis through the upregulation of vascular endothelial growth factor expression, and it was also important in BRAF inhibition resistance [37,38]. RPS3A was determined to be a key gene associated with Microsatellite instability (MSI) as an important biological feature of CRC [39]. Although the association between RPS25, CEP57, and CRC has not yet been established, the relationship with other tumors has been reported. RPS25 was shown to be a potential biomarker of lung adenocarcinoma and adult T-cell leukemia [40,41], while CEP57 was associated with prostate cancer [42]. Taken together, with applications to CRC genomic data, the hub genes and networks for disease-specific survival of MAC in stage II revealed potential gene biomarkers. However, this study only represents the first step toward defining the roles of these genes in MAC, because there are only a few studies to date that investigate the roles of these genes. Therefore, additional validation, as well as more experimental research, is still required in order to verify the results of the present study.

Several limitations associated with our study should be noted. First, a lack of histological specificity from the ICD-O-3 codes in the SEER database may not precisely capture all of the patients. This, therefore, could lead to underestimating the true prevalence rate in TAC, rather than in MAC. As a result, due to the relatively higher proportion (42.1%) of MAC in the current study compared to previous reports, the disease-specific survival failed to confirm a statistical significance, despite the trend toward a worse outcome in stage II MAC. Second, its retrospective nature might contain bias despite propensity score-matching. Third is the lack of information on perioperative treatment or molecular tumor characteristics, such as the data on the RAS, BRAF, and MSI status, and serum CEA levels. Fourth, the mass data may reveal some confounders, for which no inter-observer agreement data were available for the assessment of MAC and TAC. In spite of these limitations, our

study suggested three well-established predictive models and gene expression features to understand the MAC subtype based on the external validity and a high degree of power, along with minimizing potential confounding. With this, our study may contribute to a body of evidence that may identify reliable molecular biomarkers to more precisely stratify patients with MAC through the integrative analysis of large-scale clinical and gene expression data.

5. Conclusions

Our results indicate that the MAC subtype yielded worse overall and disease-specific survival compared with TAC subtype in stage II CRC and more advanced stages. The eight genes that were identified bear distinct cancer-specific attributes and, as a group, plays important roles in stage II mucinous CRC. Further studies validating the candidate gene biomarkers would be necessary to clarify the role of MAC in controlled clinical trials as well as in an experimental setting.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/cancers13133280/s1>, Figure S1: Volcano plot of differentially expressed genes between high risk- and low risk- group in mucinous colorectal cancer, Table S1: Baseline characteristics before/after propensity score matching, Table S2: Hazard ratios from Cox proportional hazards model between TAC and MAC, Table S3: Significant genes for high risk prediction from differentially expressed gene in stage II mucinous colon cancer, Table S4: Tissue specific protein-protein interaction in network analysis, Table S5: Key genes selected based on topological parameters like BC and degree.

Author Contributions: Specific author contributions: M.S.K. designed the study. J.M.C. analyzed and interpreted the data and wrote the manuscript. M.S.K. and C.W.K. supervised the project and revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Basic Science Research Program of National Research Foundation of Korea (NRF), which is funded by the Korean Ministry of Science, ICT and Future Planning (grant number: NRF-2019R1C1C1003524).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and was approved by the Institutional Review Board of the Kyung Hee University Hospital at Gangdong, Seoul, Republic of Korea (KHNMIC IRB 2020-01-015).

Informed Consent Statement: The need for informed consent was waived on account of the fact that all of the data used in this study were de-identified.

Data Availability Statement: The data that support the findings of this study are openly available at <https://seer.cancer.gov/> and <http://www.ebi.ac.uk/arrayexpress> (accessed on 30 November 2020).

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

References

1. Edge, S.B.; Byrd, D.R.; Carducci, M.A.; Compton, C.C.; Fritz, A.; Greene, F. *AJCC Cancer Staging Manual*; Springer: New York, NY, USA, 2010; Volume 649.
2. Luo, C.; Cen, S.; Ding, G.; Wu, W. Mucinous colorectal adenocarcinoma: Clinical pathology and treatment options. *Cancer Commun.* **2019**, *39*, 13. [[CrossRef](#)]
3. Dienstmann, R.; Vermeulen, L.; Guinney, J.; Kopetz, S.; Tejpar, S.; Tabernero, J. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* **2017**, *17*, 79–92. [[CrossRef](#)] [[PubMed](#)]
4. Bosman, F.T.; Carneiro, F.; Hruban, R.H.; Theise, N.D. *WHO Classification of Tumours of the Digestive System*; World Health Organization: Geneva, Switzerland, 2010.
5. Verhulst, J.; Ferdinande, L.; Demetter, P.; Ceelen, W. Mucinous subtype as prognostic factor in colorectal cancer: A systematic review and meta-analysis. *J. Clin. Pathol.* **2012**, *65*, 381–388. [[CrossRef](#)] [[PubMed](#)]
6. Park, J.S.; Huh, J.W.; Park, Y.A.; Cho, Y.B.; Yun, S.H.; Kim, H.C.; Lee, W.Y.; Chun, H.K. Prognostic comparison between mucinous and nonmucinous adenocarcinoma in colorectal cancer. *Medicine* **2015**, *94*, e658. [[CrossRef](#)] [[PubMed](#)]
7. Nitsche, U.; Zimmermann, A.; Späth, C.; Müller, T.; Maak, M.; Schuster, T.; Slotta-Huspenina, J.; Kaser, S.A.; Michalski, C.W.; Janssen, K.P.; et al. Mucinous and signet-ring cell colorectal cancers differ from classical adenocarcinomas in tumor biology and prognosis. *Ann. Surg.* **2013**, *258*, 775–782. [[CrossRef](#)] [[PubMed](#)]

8. Yamaguchi, T.; Taniguchi, H.; Fujita, S.; Sekine, S.; Yamamoto, S.; Akasu, T.; Kushima, R.; Tani, T.; Moriya, Y.; Shimoda, T. Clinicopathological characteristics and prognostic factors of advanced colorectal mucinous adenocarcinoma. *Histopathology* **2012**, *61*, 162–169. [[CrossRef](#)]
9. Yoshioka, Y.; Togashi, Y.; Chikugo, T.; Kogita, A.; Taguri, M.; Terashima, M.; Mizukami, T.; Hayashi, H.; Sakai, K.; de Velasco, M.A.; et al. Clinicopathological and genetic differences between low-grade and high-grade colorectal mucinous adenocarcinomas. *Cancer* **2015**, *121*, 4359–4368. [[CrossRef](#)] [[PubMed](#)]
10. Ishihara, S.; Watanabe, T.; Akahane, T.; Shimada, R.; Horiuchi, A.; Shibuya, H.; Hayama, T.; Yamada, H.; Nozawa, K.; Matsuda, K.; et al. Tumor location is a prognostic factor in poorly differentiated adenocarcinoma, mucinous adenocarcinoma, and signet-ring cell carcinoma of the colon. *Int. J. Colorectal Dis.* **2012**, *27*, 371–379. [[CrossRef](#)]
11. Zhang, R.; Zhao, J.; Xu, J.; Chen, Y. Long-term outcomes and prognostic factors of young patients with mucinous and signet-ring cell colorectal cancer. *Arch. Med. Sci.* **2020**, *16*, 359–365. [[CrossRef](#)]
12. Chu, K.C.; Miller, B.A.; Feuer, E.J.; Hankey, B.F. A method for partitioning cancer mortality trends by factors associated with diagnosis: An application to female breast cancer. *J. Clin. Epidemiol.* **1994**, *47*, 1451–1461. [[CrossRef](#)]
13. Daly, M.C.; Paquette, I.M. Surveillance, epidemiology, and end results (SEER) and SEER-medicare databases: Use in clinical research for improving colorectal cancer outcomes. *Clin. Colon Rectal Surg.* **2019**, *32*, 61–68. [[CrossRef](#)]
14. Weiss, J.M.; Pfau, P.R.; O'Connor, E.S.; King, J.; LoConte, N.; Kennedy, G.; Smith, M.A. Mortality by stage for right- versus left-sided colon cancer: Analysis of surveillance, epidemiology, and end results—Medicare data. *J. Clin. Oncol.* **2011**, *29*, 4401–4409. [[CrossRef](#)]
15. Nagtegaal, I.D.; Odze, R.D.; Klimstra, D.; Paradis, V.; Rugge, M.; Schirmacher, P.; Washington, K.M.; Carneiro, F.; Cree, I.A. The 2019 WHO classification of tumours of the digestive system. *Histopathology* **2020**, *76*, 182–188. [[CrossRef](#)] [[PubMed](#)]
16. Wright, M.N.; Dankowski, T.; Ziegler, A. Unbiased split variable selection for random survival forests using maximally selected rank statistics. *Stat. Med.* **2017**, *36*, 1272–1284. [[CrossRef](#)] [[PubMed](#)]
17. Katzman, J.L.; Shaham, U.; Cloninger, A.; Bates, J.; Jiang, T.; Kluger, Y. DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **2018**, *18*, 24. [[CrossRef](#)]
18. Fotso, S.J.a.p.a. Deep neural networks for survival analysis based on a multi-task framework. *arXiv* **2018**, arXiv:1801.05512.
19. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*, 1. [[CrossRef](#)]
20. Xia, J.; Gill, E.E.; Hancock, R.E. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat. Protoc.* **2015**, *10*, 823–844. [[CrossRef](#)]
21. Xia, J.; Benner, M.J.; Hancock, R.E. NetworkAnalyst—Integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res.* **2014**, *42*, W167–W174. [[CrossRef](#)]
22. Purdie, C.A.; Piris, J. Histopathological grade, mucinous differentiation and DNA ploidy in relation to prognosis in colorectal carcinoma. *Histopathology* **2000**, *36*, 121–126. [[CrossRef](#)]
23. Xie, L.; Villeneuve, P.J.; Shaw, A. Survival of patients diagnosed with either colorectal mucinous or non-mucinous adenocarcinoma: A population-based study in Canada. *Int. J. Oncol.* **2009**, *34*, 1109–1115. [[CrossRef](#)]
24. Warschkow, R.; Tarantino, I.; Huttner, F.J.; Schmied, B.M.; Guller, U.; Diener, M.K.; Ulrich, A. Predictive value of mucinous histology in colon cancer: A population-based, propensity score matched analysis. *Br. J. Cancer* **2016**, *114*, 1027–1032. [[CrossRef](#)]
25. Catalano, V.; Loupakis, F.; Graziano, F.; Bisonni, R.; Torresi, U.; Vincenzi, B.; Mari, D.; Giordani, P.; Alessandrini, P.; Salvatore, L.; et al. Prognosis of mucinous histology for patients with radically resected stage II and III colon cancer. *Ann. Oncol.* **2012**, *23*, 135–141. [[CrossRef](#)]
26. Hogan, J.; Burke, J.P.; Samaha, G.; Condon, E.; Waldron, D.; Faul, P.; Coffey, J.C. Overall survival is improved in mucinous adenocarcinoma of the colon. *Int. J. Colorectal Dis.* **2014**, *29*, 563–569. [[CrossRef](#)] [[PubMed](#)]
27. Kanemitsu, Y.; Kato, T.; Hirai, T.; Yasui, K.; Morimoto, T.; Shimizu, Y.; Kodera, Y.; Yamamura, Y. Survival after curative resection for mucinous adenocarcinoma of the colorectum. *Dis. Colon Rectum* **2003**, *46*, 160–167. [[CrossRef](#)] [[PubMed](#)]
28. Kang, H.; O'Connell, J.B.; Maggard, M.A.; Sack, J.; Ko, C.Y. A 10-year outcomes evaluation of mucinous and signet-ring cell carcinoma of the colon and rectum. *Dis. Colon Rectum* **2005**, *48*, 1161–1168. [[CrossRef](#)]
29. Hosseini, S.; Bananzadeh, A.M.; Salek, R.; Zare-Bandamiri, M.; Kermani, A.T.; Mohammadianpanah, M. Prognostic significance of mucinous histologic subtype on oncologic outcomes in patients with colorectal cancer. *Ann. Coloproctol.* **2017**, *33*, 57–63. [[CrossRef](#)] [[PubMed](#)]
30. Uhlén, M.; Fagerberg, L.; Hallström, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419. [[CrossRef](#)]
31. Takemasa, I.; Kittaka, N.; Hitora, T.; Watanabe, M.; Matsuo, E.; Mizushima, T.; Ikeda, M.; Yamamoto, H.; Sekimoto, M.; Nishimura, O.; et al. Potential biological insights revealed by an integrated assessment of proteomic and transcriptomic data in human colorectal cancer. *Int. J. Oncol.* **2012**, *40*, 551–559. [[CrossRef](#)]
32. Jiang, W.; Li, X.; Rao, S.; Wang, L.; Du, L.; Li, C.; Wu, C.; Wang, H.; Wang, Y.; Yang, B. Constructing disease-specific gene networks using pair-wise relevance metric: Application to colon cancer identifies interleukin 8, desmin and enolase 1 as the central elements. *BMC Syst. Biol.* **2008**, *2*, 72. [[CrossRef](#)]

33. Uemura, M.; Nouse, K.; Kobayashi, Y.; Tanaka, H.; Nakamura, S.; Higashi, T.; Ono, T.; Nakayama, E.; Hanafusa, T.; Shiratori, Y. Identification of the antigens predominantly reacted with serum from patients with hepatocellular carcinoma. *Cancer* **2003**, *97*, 2474–2479. [[CrossRef](#)]
34. Kirana, C.; Peng, L.; Miller, R.; Keating, J.P.; Glenn, C.; Shi, H.; Jordan, T.W.; Maddern, G.J.; Stubbs, R.S. Combination of laser microdissection, 2D-DIGE and MALDI-TOF MS to identify protein biomarkers to predict colorectal cancer spread. *Clin. Proteom.* **2019**, *16*, 3. [[CrossRef](#)]
35. Wen, S.; Wang, X.; Wang, Y.; Shen, J.; Pu, J.; Liang, H.; Chen, C.; Liu, L.; Dai, P. Nucleoside diphosphate kinase 2 confers acquired 5-fluorouracil resistance in colorectal cancer cells. *Artif. Cells Nanomed. Biotechnol.* **2018**, *46*, 896–905. [[CrossRef](#)]
36. Liu, H.; Zhang, Q.; Li, K.; Gong, Z.; Liu, Z.; Xu, Y.; Swaney, M.H.; Xiao, K.; Chen, Y. Prognostic significance of USP33 in advanced colorectal cancer patients: New insights into β -arrestin-dependent ERK signaling. *Oncotarget* **2016**, *7*, 81223–81240. [[CrossRef](#)]
37. Ding, C.; Luo, J.; Fan, X.; Li, L.; Li, S.; Wen, K.; Feng, J.; Wu, G. Elevated Gab2 induces tumor growth and angiogenesis in colorectal cancer through upregulating VEGF levels. *J. Exp. Clin. Cancer Res.* **2017**, *36*, 56. [[CrossRef](#)] [[PubMed](#)]
38. Herr, R.; Halbach, S.; Heizmann, M.; Busch, H.; Boerries, M.; Brummer, T. BRAF inhibition upregulates a variety of receptor tyrosine kinases and their downstream effector Gab2 in colorectal cancer cell lines. *Oncogene* **2018**, *37*, 1576–1593. [[CrossRef](#)] [[PubMed](#)]
39. Yu, C.; Hong, H.; Zhang, S.; Zong, Y.; Ma, J.; Lu, A.; Sun, J.; Zheng, M. Identification of key genes and pathways involved in microsatellite instability in colorectal cancer. *Mol. Med. Rep.* **2019**, *19*, 2065–2076. [[CrossRef](#)] [[PubMed](#)]
40. Hsu, C.H.; Hsu, C.W.; Hsueh, C.; Wang, C.L.; Wu, Y.C.; Wu, C.C.; Liu, C.C.; Yu, J.S.; Chang, Y.S.; Yu, C.J. Identification and characterization of potential biomarkers by quantitative tissue proteomics of primary lung adenocarcinoma. *Mol. Cell. Proteom.* **2016**, *15*, 2396–2410. [[CrossRef](#)] [[PubMed](#)]
41. Terol, M.; Gazon, H.; Lemasson, I.; Duc-Dodon, M.; Barbeau, B.; Césaire, R.; Mesnard, J.M.; Péloponèse, J.M., Jr. HBZ-mediated shift of JunD from growth suppressor to tumor promoter in leukemic cells by inhibition of ribosomal protein S25 expression. *Leukemia* **2017**, *31*, 2235–2243. [[CrossRef](#)]
42. Paulo, P.; Maia, S.; Pinto, C.; Monteiro, A.; Peixoto, A.; Teixeira, M.R. Targeted next generation sequencing identifies functionally deleterious germline mutations in novel genes in early-onset/familial prostate cancer. *PLoS Genet.* **2018**, *14*, e1007355. [[CrossRef](#)] [[PubMed](#)]