Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Tao Liang Date

Impact of uncommon genomic alterations on outcomes in metastatic colorectal cancer patients

By

Tao Liang
Master of Science in Public Health

Biostatistics and Bioinformatics

_____

Zhengjia (Nelson) Chen, PhD
(Thesis Advisor)

_____

Xiangqin Cui, PhD
(Reader)

Impact of uncommon genomic alterations on outcomes in metastatic colorectal cancer patients

By

Tao Liang

Bachelor of Science
Tianjin Normal University
2016

Thesis Committee Chair: Zhengjia (Nelson) Chen, PhD
Reader: Xiangqin Cui, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics and Bioinformatics
2019

Abstract

Impact of uncommon genomic alterations on outcomes in metastatic colorectal cancer patients

By Tao Liang

Background: Metastatic colorectal cancer(mCRC) remains to be deadly and there are limited actionable genes to develop targeted therapies. Accumulating studies have shown that patients with specific mutations, such as KRAS, are more likely to be insensitive to anti-epidermal growth factor receptor (anti-EGFR) targeted therapies than patients with wild-type genes. However, there are still many patients with wild-type KRAS but nonresponding to anti-EGFR therapies. This suggests that there are other genetic determinants play roles in this process.

Methods and Materials: Records of 161 mCRC patients who underwent molecular profiling for mutations status of 30 genes were reviewed, including 116 patients with records of surgery date. Genes that are related to primary tumor side are determined by Chi-square test or Fisher Exact test. Univariate survival analysis with Cox model was used to determine genes that are related to poor survival outcome, and multivariate Cox model was used to adjust potential confounders. Overall survival and progression-free survival were assessed separately.

Results: TET2 (P=0.0280), FAM123B (P=0.0011), PTEN (P=0.0244), and BCOR (P=0.0212) were associated with the primary tumor side. After adjusting for patient characteristics, BRCA1/2 (OR, 6.98; HR, 1.05 to 46.45; P=0.0444), FLT3 (OR, 7.55; HR, 1.436 to 39.74; P=0.0170), SOX9 (OR, 10.23; HR, 1.61 to 65.24; P=0.0139) and IRS2 (OR, 31.63; HR, 4.55 to 219.92; P=0.0005) are associated with worse overall survival, CDK8 (OR, 3.122; HR, 1.337 to 7.291; P=0.0044) is associated with worse progression-free survival. More interestingly, we found that females (OR, 10.74; HR, 3.080 to 37.43; P< .0001) bearing mutated CDK8 have worse progression-free survival outcome than males (OR, 1.24; HR, 0.31 to 4.94; P=0.7621) bearing mutated CDK8.

Conclusion: Overall, we found that TET2, FAM123B, PTEN and BCOR are associated with primary side of tumor which is an indicator of worse survival outcome in colorectal cancer. Mutated BRCA1/2, FLT3, SOX9, IRS3 and CDK8 are plausible prognostic predictors for mCRC survival. However, these results need to be confirmed by investigation on separate cohorts.

Impact of uncommon genomic alterations on outcomes in metastatic colorectal cancer patients

By

Tao Liang

Bachelor of Science
Tianjin Normal University
2016

Thesis Committee Chair: Zhengjia (Nelson) Chen, PhD
Reader: Xiangqin Cui, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics and Bioinformatics
2019

Acknowledgement

I would first like to thank my thesis advisor Prof. Zhangjia (Nelson) Chen of Rollins School of Public Health at Emory University. He gave me a great help in academic field, answer my question patiently, taught me how to perform a proper research and keep steering me in the right direction. I would also like thank my thesis reader Prof. Xiangqin Cui for her great help during the revision process. He gave me detailed revision advice to make my thesis more reasonable. Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study. Thank you.

Contents

Introduction

Colorectal cancer (CRC) is a top contributor to cancer morbidity and mortality. It is the second and third leading cancer type, constituting 9% and 8% of estimated death in male and female, respectively (Siegel, Miller, & Jemal, 2017). The SEER (Surveillance, Epidemiology, and End results) summary staging system determines the stages of colorectal cancer based on the invasion of cancer cells into other part of body: In situ, Local, Regional and Distant. The 5-year survival rate of distant stage (i.e metastasis CRC or mCRC), which accounts for 21% of all CRC patients, is only 14%, in contrast to 90% and 71% for patients diagnosed as localized and regional stages, respectively(Edwards et al., 2010).

The overall CRC incidence and mortality rate in the US have been decreasing since 1980, which is attributed to improvements of treatments and screening (Edwards et al., 2010). Panitumumab and Cetuximab are two approved anti-epidermal growth factor receptor (anti-EGFR) treatments for mCRC patients. EGFR is a transmembrane protein and receptor of extracellular growth signal which cause cell division. Overexpression of EGFR is associated with development of various tumors and frequently observed in cancer cells. Anti-EGFR therapy prevents cancer cell growing by using EGFR inhibitors (a monoclonal antibody) to block overexpressed EGFR. However, accumulating studies have shown that outcomes of anti-EGFR treatments are heterogeneous due to mutation status of specific genes. A study showed that mCRC patients without mutated KRAS in exon 2 benefit from Panitumumab-FOLFOX4 treatment, whereas patients whose tumor bear mutated KRAS in exon 2 codon 12 and 13 do not benefit from Panitumumab-FOLFOX4 treatment (Douillard et al., 2013). As with Panitumumab, mCRC patients

harboring specific KRAS mutations are unlikely to benefit from Cetuximab (Karapetis et al., 2008). Noticeably, 40% to 60% mCRC patients with wild-type KRAS are also insensitive to anti-EGFR treatment (Linardou et al., 2008). This suggests that there might be other genetic determinants related to poor clinical outcome. In 2012, the Cancer Genome Atlas project identified frequently mutated genes and involved pathway alterations from 257 CRC tumor samples (Cancer Genome Atlas, 2012). Many investigations have been focusing on biological mechanism and classification of molecular subtypes of CRC in the past decade (Ron Firestein et al., 2008; Guinney et al., 2015; Morris et al., 2008), but limited studies have focused on identifying clinically meaningful genes directly from cohort of CRC patients.

The overall goal of our study is to identify potentially actionable genes associated with poorer survival outcome among distant-stage CRC patients. This can be divided into two objectives: i) Detecting candidate genes that are associated with worse clinical outcome indicators (e.g. tumor stage and location) ii) Detecting candidate genes that directly related to poor survival outcome. To reach the goal, we examined associations between genetic mutations and multiple pathological features (e.g. tumor location) in 161 metastasis colorectal patients. Univariate and multivariate survival analysis have been performed, in which events are defined as all-cause death (overall survival) or relapse of tumor after surgery (progression-free survival). For mutations associated with worse survival outcomes given from univariate survival analysis, multivariate analysis is conducted to adjust for possible confounders.

Method

2.1 Study Design

More than 500 patients with metastatic gastrointestinal cancers underwent molecular profiling with next-generation sequencing technology to identify genomic mutations at Winship Cancer Institute of Emory University between January 01, 2013 and July 01, 2018. Among patients who underwent molecular profiling, 179 patients were identified with metastatic colorectal cancer. A total of 161 patients identified with clinical information/pathologic confirmation were included in final analysis (Figure 1).

There are 30 mutated genes were identified in 161 patients. Patient demographics and clinical information, including date and age of diagnosis, medical and surgical history, pathologic information including tumor differentiation, primary tumor location and pathologic stage, social history including smoking and alcohol usage were obtained from detailed electronic medical records. All statistical analyses were performed with SAS 9.4. All significance level was set to 0.05. All reported p-value are two-sided and were not adjusted for multiple testing.

2.2 Descriptive Characteristic and Association Study

The distributions of the genotype of the 30 mutations in 161 patients were constructed. For continues variable like age at diagnosis and survival time, means and standard deviations were summarized. For categorical variables like age group, gender, smoking history, tumor differentiation status, tumor stages, tumor location, frequencies

and percentages were counted. Specifically, the age groups at diagnosis were divided into young (< 49), middle (50 - 64), old (> 65) according to CRC incidence rates trend by age.

For each mutation, the association of the occurrence of mutant allele and pathological/patients' characteristics were tested. Investigated patients' characteristics include age group (young, middle and old), gender, smoking history (current/former, never, unknown), tumor differentiation status (well/well to moderate, moderate, and moderate to poor/poor), tumor stage (I - II, and III - IV), primary tumor side (left and right). To examine such association, a 2×2 or 2×3 table displaying the distribution of mutation status and above features was constructed, then chi-square test was applied (or Fisher's exact test if there is any cell in the table has a frequency less than 5). All investigated patients' characteristics are considered as risk factors of occurrence or survival of colorectal cancer. The corresponding p-values for testing against non-association were generated.

2.3 Survival Analysis

The beginning of the study is when these patients were diagnosed with metastatic CRC, and the end of the study was set to July 01, 2018. All patients with mutations genotype data on specific mutations were included in survival analysis for each gene. The primary end point was defined as the time from diagnosis of metastatic CRC to death from any cause. The secondary end point was defined as the time from the time of surgery to recurrence of tumor. All survival time were calculated on month scale.

The whole dataset is right-censored. For overall survival, patients who have been followed through this study were recorded as "alive" or "dead", whereas patients who left

the study for any reason were recorded as "unknown". Among 161 patients included in overall survival analysis, 116 patients with surgery records were included in progression-free survival analysis. For censored patients without recurrence records, survival time was calculated as starting from surgery date to the end of study.

The survival of patients in each mutation group was summarized by Kaplan-Meier curve, and the difference between these groups was tested by log-rank test. The hazard ratio and corresponding 95% confidence intervals were generated from Cox regression model with single covariate of mutation status. Multivariate Cox regression model were used to adjust for potential confounders. The following covariates were included in multivariate Cox regression model: gender (male vs. female), age group (young or middle vs. old), primary site of tumor (left vs right), stage at diagnosis (I – II vs. III – IV), smoking history (current/former or never vs. unknown), number of previous therapy (<3 or 3 – 5 vs. >5), tumor differentiation status (well/well to moderate, moderate, and moderate to poor/poor).

The Cox regression model takes the form of:

$$h_0(t|X_i) = h_0 exp(\beta X_{i1} + \ldots + \beta X_{ip}) = h_0(t) exp(\beta X_i) h_0(t|X_i)$$

It can be divided into two parts: the underlying baseline hazard function, $h_0(t|X_i)$, which describes how the risk of the event changes at the baseline level of covariates(mutation status in our analysis); and the effect parameters, $h_0(t)exp(\beta X_i)$, which describes the change of risk according to covariates (potential confounders). Local Wald test was performed to see if there any significant difference among different levels of covariates. Adjusted hazard ratio was calculated for each risk factor in multivariable model. Local Wald tests were conducted for each variable in the final model and p-value were output.

In fitting the Cox PH models, we assumed independence of censoring times to ensure reliable and unbiased survival estimates. Other assumptions we made about the censored time include: i) the reasons for censoring is non-informative (i.e, not related to our interested medical condition); ii) censoring times are independent from each other so the survival estimates are unbiased and reliable; iii) The hazard functions of different levels of covariates are proportional and always independent of time.

Results

3.1 Patients Characteristics and Mutation-associated Features

Table 1 displays genotype frequencies of the 30 mutations among 161 patients. Two most frequently mutated genes are P53 (79.5%) and APC (82.6%), with 128 and 133 of the study population were identified as mutation bearing respectively. All other mutations are relatively uncommon with a less than 15% mutation frequency. Especially, SMAD2/4, FBXW7, FAM123B, PTEN, FLT3, SOX9, CDK8, and MYC are found to be mutated in more than 10 patients.

Table 2 summarizes demographic and disease characteristics of the total study population and of patients with left/right primary tumor site. Patients distributed evenly under age, gender and tumor differentiation groups, whereas 73% patients are recorded as without smoking history and 93% patients are diagnosed with III – IV staged tumors.

Mutated TET2 (P=0.0280), FAM123B (P=0.0011), PTEN (P=0.0244) and BCOR (P=0.0212) are associated with the primary tumor site. TET2 (P=0.0045), RUNX1 (P=0.0350) and RB1 (P=0.0136) are most frequently found in moderate to poor/poorly

differentiated tumors. SOX9 (P=0.0144) is associated with age groups. We did not find genetic alterations that associated with gender, tumor stages and smoking history.

3.2 Overall Survival

Five genes were flagged out with worse overall survival outcome: BRCA1/2, FLT3, SOX9, CDK8 and IRS2. The 3-year overall survival rates for patients with mutated BRCA1/2 tumors is 66.7% versus 97.6% for patients with wild-type BRCA1/2 (log-rank P=0.0023) (hazard ratio for death in the mutated group versus wild-type group, 8.88; 95% confidential interval [CI], 1.63 to 48.55; P=0.0117) (Figure 2A, Table 3). The 3-year overall survival rates for patients with mutated FLT3 tumors and wild-type FLT3 are 97.4% and 83.6% respectively (log-rank P=0.0011) (hazard ratio, 8.17; 95% CI, 1.81 to 36.96; P=0.0064) (Figure 2B, Table 3). The 3-year overall survival rates for patients with mutated SOX9 and wild-type SOX9 are 97.5% and 81.5% respectively (log-rank P=0.0193) (hazard ratio, 6.01; 95% CI, 1.09 to 33.34; P=0.0401) (Figure 2C, Table 3). The 3-year overall survival rates for patients with mutated CDK8 and wild-type CDK8 are 96.5% and 91.7% (log-rank P=0.0245) (hazard ratio, 5.492; 95% CI, 1.04 to 29.14; P=0.0455) (Figure 2D, Table 3). The 3-year overall survival rates for patients with mutated IRS2 and wild-type IRS2 are 97.5% and 71.4% respectively (log-rank P < 0.0001) (hazard ratio, 16.07; 95% CI, 3.52 to 73.34; P < 0.0001) (Figure 2E, Table 3). After adjusting for potential prognostic factors with multivariate Cox regression model, overall survival difference remained significant for FLT3, SOX9, IRS2, RB1(Table 5).

3.3 Progression-free Survival

The progression-free survivals are associated with CDK8 genotype. Among patients with wild-type CDK8, the median progression-free survival was 41 months, with 1-year survival rate of 84.4%. Among patients with mutated CDK8, the median recurrence-free survival was 12 months, with 1-year survival rate of 55.6% (Figure 3, Table 4). The hazard ratio for progression after surgery in CDK8 mutation bearing group as compared with wild-type CDK8 group is 2.98 (95% CI, 1.33 to 6.67; P=0.0079). This recurrence-free survival difference remains significant after adjusting for potential prognostic factors (smoking history, age group, primary tumor site) with multivariate Cox model (hazard ratio, 3.122; 95% CI 1.337 to 7.291; P=0.0044).

Interestingly, we found an interaction of gender and CDK8 mutation. Among male patients, the median progression-free survival was 18 months in mutated CDK8 group and 35 months in wild-type CDK8 group, with 1-year progression-free survival rates of 81.9% and 60.0% respectively (log-rank P=0.5043) (hazard ratio for progression among male patients with mutated CDK8 as compared with patients with wild-type CDK8, 1.49; 95% CI, 0.45 to 4.98; P=0.5136) (Figure 4.A). This difference remains insignificant after controlling for potential confounders specified previously (adjusted hazard ratio, 1.24; 95% CI, 0.31 to 4.94; P=0.7621) However, among female patients, the median progression-free survival was 8.5 months in mutated CDK8 group and 49 months in wild-type CDK8 group, with 1-year progression-free survival rates of 86.7% and 0% respectively (log-rank P < 0.0001) (hazard ratio for progression among female patients with mutated CDK8 as compared with patients with wild-type CDK8, 10.74; 95% CI, 3.080 to 37.43, P=) (Figure 4.B). This difference remains significant after controlling for

potential confounders specified previously (adjusted hazard ratio, 11.4; 95% CI, 2.842 to 45.7; P=0.0006)

3.4 Results Interpretation and Clinical Validity

Our findings show that the mutation status of TET2, FAM123B, PTEN and BCOR genes are more likely diagnosed with right-sided primary tumor among advanced colorectal cancer patients, and we tend to think patients bearing these mutations have poorer survival outcome since the primary side of tumor provides both prognostic and predictive values in colorectal cancer. Venook and colleagues showed that regardless of KRAS mutation status or type of treatment received, patients with left-sided primary tumors display better overall survival and progression-free survival than those with right-sided primary tumors (Venook et al., 2016). Another retrospective study showed that mCRC patients with left-sided primary tumors tends to benefit more from initial treatment with FOLFIRI plus cetuximab than from FOLFIRI alone or FOLFIRI plus bevacizumab (Tejpar et al., 2016). Embryonically, the left side of colon was derived from hindgut, while the right side of colon was derived from midgut. In addition, right-sided tumors have observable symptoms only in advanced cancer patients. Distinctness of right and left-sided tumors imply distinct biological pathways guiding the tumor development in CRC patients. The four genes we identified could potentially play roles in this process, and mutations in these genes might be related to the variable effectiveness of current therapies. Several studies have shown that low-level expression of TET2 or PTEN are associated with worse overall and progression-free survival (Rawluszko-Wieczorek et al., 2015; Sawai et al., 2008). These studies support our results on TET2 and PTEN since

genetic mutation could cause change of gene expression. No previous reports were located on survival outcome among mCRC patients with FAM123B and BCOR mutation.

Our results showed that mutations in BRCA1/2, FLT3, SOX9 and IRS2 are associated with worse overall survival outcome under univariate and multivariate models. Previous studies have shown that the low expression of BRAC1 and overexpression of SOX9 can predict poor overall survival (Grabsch et al., 2006; Lü et al., 2008). Among the few reports on IRS2-related survival analysis in colorectal cancer, Hanyuda reported that CRC patients with low expression of IRS1 and high-level physical activity have superior colorectal cancer-specific survival outcome (Hanyuda et al., 2016). This study indicates that IRS genes might play a role in CRC patients' survival outcome. No previous reports were located on survival outcome among mCRC patients with FLT3 mutation.

CDK8 is particularly interesting not just because it was identified as associated with both poorer overall survival and progression-free survival outcome under univariate Cox model, but also because it was identified with gender-differential effect. Firestein reported that overexpression of CDK8 is associated with higher colon cancer-specific mortality females have higher odds of overexpression of CDK8 (R. Firestein et al., 2010). This study not only supports our result that CDK8 is a prognostic predictor from univariate survival analysis, but also supports our result of the interaction between gender and mutated CDK8 since even if mutation status of CDK8 is the same, different hormonal environment in female and male may predispose CDK8 gene-expression divergence, which leads to difference in progression-free outcomes between genders.

Discussion

4.1 Main Conclusions

In this study, we found that TET2, FAM123B, PTEN, and BCOR are associated with the primary tumor side. Survival analysis showed that BRCA1/2, FLT3, SOX9 and IRS2 are associated with worse overall survival outcome, and CDK8 is associated with worse progression-free survival outcome, after adjusting for patient characteristics. More interestingly, we found that females bearing mutated CDK8 have worse progression-free survival outcome than males bearing mutated CDK8. These findings are supported by previous studies and may suggest clinical significance in patient's response to treatments.

4.2 Limitations

Overall, our study is not perfect. The key deficiencies of our cohort are small sample size, large proportion of censored subjects and low frequency for most genes. These issues not only lower the power of analysis, but also make it hard to validate our results by methods like cross-validation. Accordingly, our results are more suggestive evidence rather than solid conclusion, and we failed to repeat some previous findings in the field with our dataset. For example, mutated KRAS has been reported to related to poor overall survival outcome (De Roock et al., 2010; Phipps et al., 2013) in multiple studies, whereas our analysis didn't detect this relationship, and so for ATM, PIK3CA, NARS, PTEN mutations. Another statistical issue is the lack of a multiple testing control. Since our dataset is highly asymmetric and statistical powers are low as mentioned above, the false negative rate (Type II error rate) would higher than other large-scale studies. To avoid high false negative rate, no correction was applied for multiple testing.

Except for statistical deficiencies above, other potential bias from the dataset may confound our analysis. One issue is the lack of information on mutated alleles. The linkage between poor survival outcome and mutated genes vary given different mutated alleles even though these alleles are on the same gene. Phipps and colleagues conducted a cohort study with 1923 CRC patients containing 593 patients whose tumors carry KRAS mutations, and reported that, compared with wild-type KRAS, the presence of p. G13D (a KRAS mutation) are statistically significantly associated with poorer disease-specific and overall survival outcome, whereas neither of p. G12D nor p. G12V mutations is statistically significantly associated with poorer survival outcome (Phipps et al., 2013). Since our dataset does not contain allele-specific mutation information, it's hard to evaluate or control for allele-specific effect and this could partially explain our failure in detecting the association between KRAS mutation status and survival outcome. Also, instead of mutation status, many studies have focused on the association between gene expression and survival outcome. Although genetic alteration is one of most prevalent reasons of expression alteration, investigations are needed to confirm such linkage for specific genes.

Even with imperfection, we still detect some actionable genes with supportive evidence from literatures and newly detected candidates. Upon further confirmation in larger, separate cohorts, these biomarkers could be effectively used as targets in drug designing or classifier of subtype of mCRC. Additional clarification of effect of specific mutation alleles and treatment type and their interaction are needed to improve the efficiency of treatment among mCRC patients.

Reference

Cancer Genome Atlas, N. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature, 487*(7407), 330-337. doi:10.1038/nature11252

De Roock, W., Claes, B., Bernasconi, D., De Schutter, J., Biesmans, B., Fountzilas, G., . . . Laurent-Puig, P. (2010). Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *The lancet oncology, 11*(8), 753-762.

Douillard, J.-Y., Oliner, K. S., Siena, S., Tabernero, J., Burkes, R., Barugel, M., . . . Jassem, J. (2013). Panitumumab–FOLFOX4 treatment and RAS mutations in colorectal cancer. *New England Journal of Medicine, 369*(11), 1023-1034.

Edwards, B. K., Ward, E., Kohler, B. A., Eheman, C., Zauber, A. G., Anderson, R. N., . . . Ries, L. A. (2010). Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer, 116*(3), 544-573. doi:10.1002/cncr.24760

Firestein, R., Bass, A. J., Kim, S. Y., Dunn, I. F., Silver, S. J., Guney, I., . . . Ogino, S. (2008). CDK8 is a colorectal cancer oncogene that regulates β-catenin activity. *Nature, 455*(7212), 547.

Firestein, R., Shima, K., Nosho, K., Irahara, N., Baba, Y., Bojarski, E., . . . Ogino, S. (2010). CDK8 expression in 470 colorectal cancers in relation to beta-catenin activation, other molecular alterations and patient survival. *Int J Cancer, 126*(12), 2863-2873. doi:10.1002/ijc.24908

Grabsch, H., Dattani, M., Barker, L., Maughan, N., Maude, K., Hansen, O., . . . Mueller, W. (2006). Expression of DNA double-strand break repair proteins ATM and BRCA1 predicts survival in colorectal cancer. *Clin Cancer Res, 12*(5), 1494-1500. doi:10.1158/1078-0432.CCR-05-2105

Guinney, J., Dienstmann, R., Wang, X., de Reynies, A., Schlicker, A., Soneson, C., . . . Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nat Med, 21*(11), 1350-1356. doi:10.1038/nm.3967

Hanyuda, A., Kim, S. A., Martinez-Fernandez, A., Qian, Z. R., Yamauchi, M., Nishihara, R., . . . Ogino, S. (2016). Survival Benefit of Exercise Differs by Tumor IRS1 Expression Status in Colorectal Cancer. *Ann Surg Oncol, 23*(3), 908-917. doi:10.1245/s10434-015-4967-4

Karapetis, C. S., Khambata-Ford, S., Jonker, D. J., O'Callaghan, C. J., Tu, D., Tebbutt, N. C., . . . Zalcberg, J. R. (2008). K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med, 359*(17), 1757-1765. doi:10.1056/NEJMoa0804385

Linardou, H., Dahabreh, I. J., Kanaloupiti, D., Siannis, F., Bafaloukos, D., Kosmidis, P., . . . Murray, S. (2008). Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. *The lancet oncology, 9*(10), 962-972.

Lü, B., Fang, Y., Xu, J., Wang, L., Xu, F., Xu, E., . . . Lai, M. (2008). Analysis of SOX9 expression in colorectal cancer. *American journal of clinical pathology, 130*(6), 897-904.

Morris, E. J., Ji, J.-Y., Yang, F., Di Stefano, L., Herr, A., Moon, N.-S., . . . Dyson, N. J. (2008). E2F1 represses β-catenin transcription and is antagonized by both pRB and CDK8. *Nature, 455*(7212), 552.

Phipps, A. I., Buchanan, D. D., Makar, K. W., Win, A. K., Baron, J. A., Lindor, N. M., . . . Newcomb, P. A. (2013). KRAS-mutation status in relation to colorectal cancer survival: the joint impact of correlated tumour markers. *Br J Cancer, 108*(8), 1757-1764. doi:10.1038/bjc.2013.118

Rawluszko-Wieczorek, A. A., Siera, A., Horbacka, K., Horst, N., Krokowicz, P., & Jagodzinski, P. P. (2015). Clinical significance of DNA methylation mRNA levels of TET family members in colorectal cancer. *J Cancer Res Clin Oncol, 141*(8), 1379-1392. doi:10.1007/s00432-014-1901-2

Sawai, H., Yasuda, A., Ochi, N., Ma, J., Matsuo, Y., Wakasugi, T., . . . Takeyama, H. (2008). Loss of PTEN expression is associated with colorectal cancer liver metastasis and poor patient survival. *BMC gastroenterology, 8*(1), 56.

Siegel, R. L., Miller, K. D., & Jemal, A. (2017). Cancer Statistics, 2017. *CA Cancer J Clin, 67*(1), 7-30. doi:10.3322/caac.21387

Tejpar, S., Stintzing, S., Ciardiello, F., Tabernero, J., Van Cutsem, E., Beier, F., . . . Heinemann, V. (2016). Prognostic and Predictive Relevance of Primary Tumor Location in Patients With RAS Wild-Type Metastatic Colorectal Cancer: Retrospective Analyses of the CRYSTAL and FIRE-3 Trials. *JAMA Oncol*. doi:10.1001/jamaoncol.2016.3797

Venook, A. P., Niedzwiecki, D., Innocenti, F., Fruth, B., Greene, C., O'Neil, B. H., . . . Polite, B. N. (2016). Impact of primary (1º) tumor location on overall survival

(OS) and progression-free survival (PFS) in patients (pts) with metastatic colorectal cancer (mCRC): Analysis of CALGB/SWOG 80405 (Alliance). In: American Society of Clinical Oncology.

Tables and Figures

Genes	Mutated (%)	Not Mutated (%)
P53	128(79.5)	33(20.5)
APC	133(82.6)	28(17.4)
ATM	8(5.0)	153(95.0)
SMAD2/4	19(11.8)	142(88.2)
BRCA1/2	8(5.0)	153(95.0)
MUTYH	5(3.1)	156(96.9)
FBXW7	11(6.9)	150(93.2)
ASXI1	4(2.5)	157(97.5)
TET2	5(3.1)	156(96.9)
DNMT3A	4(2.5)	157(97.5)
RUNX1	4(2.5)	157(97.5)
FAM123B	10(6.25)	151(93.8)
ARID1 a/b	7(4.3)	154(95.7)
RB1	3(1.9)	158(98.1)
RNF43	4(2.5)	157(97.5)
PTEN	13(8.1)	148(91.9)
FLT3	15(9.3)	146(90.7)
ZNF217	6(3.7)	155(96.3)
SOX9	13(8.1)	148(91.9)
BCOR	9(5.6)	152(94.4)
CDK8	13(8.1)	148(91.9)
IRS2	8(5.0)	153(95.0)
SRC amp	5(3.1)	156(96.9)
MYC amp	11(6.9)	150(93.2)
CTNNB1	5(3.1)	156(96.9)
NOTCH1/3	4(2.5)	157(97.5)
BCL2 amp	7(4.3)	154(95.7)
KRAS	88(54.7)	73(45.3)
NRAS	9(5.6)	152(94.4)
PIK3CA	27(16.8)	134(83.2)

Table 1 Mutation frequencies

Variable		Primary Tumor Side		p-value
	All N=161(100)	Left N=113(70)	Right N=48(30)	
Overall Survival Time	36.4±29.1	38.6±30.7	31.2±24.4	< .0001
Progression-Free Survival Time	29.1±24.6	31.1±27.2	25.3±18.1	< .0001
Age at Diagnosis	53.0±12.3	51.8±12.8	56.1±10.7	< .0001
Young (<49)	72(45)	58(51)	14(29)	0.0350
Middle (50 - 64)	57(35)	35(31)	22(46)	
Old (65+)	32(20)	20(18)	12(25)	
Gender				
Male	85(53)	60(53)	25(52)	0.9061
Female	76(47)	53(47)	23(48)	
Smoking History				
Current/Former	42(26)	34(29)	9(19)	0.5067
Never	116(73)	78(70)	38(79)	
Unknown	2(1)	1(1)	1(2)	
Tumor Differentiation Status (Missing = 28)				
Well/well to moderate	17(13)	11(12)	6(14)	0.7873
Moderate	81(61)	56(63)	25(57)	
Moderate to poor/poor	35(26)	22(25)	13(29)	
Tumor Stage				
I – II	12(7)	11(10)	1(2)	0.1107
III – IV	149(93)	102(90)	47(98)	
TET2 Mutation				
Mutated	156(97)	112(99)	44(92)	0.0280
Not mutated	5(3)	1(1)	4(8)	
FAM123B				
Mutated	150(94)	110(98)	40(83)	0.0011
Not mutated	10(6)	2(2)	8(17)	
PTEN				
Mutated	146(92)	106(95)	40(83)	0.0244
Not mutated	13(8)	5(5)	8(17)	
BCOR				
Mutated	152(94)	110(97)	42(87)	0.0212
Not mutated	9(6)	3(3)	6(13)	

Table 2 Patients characteristics and genes related to primary tumor side.

			Overall Survival		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
P53	Non-mutated	33	-		
	Mutated	128	0.545 (0.105 – 2.834)	0.4704	0.4632
APC	Non-mutated	28	-		
	Mutated	133	4399458 (0.000 – .)	0.9936	0.2710
ATM	Non-mutated	153	-		
	Mutated	8	0.000 (0.000 – .)	0.9940	0.5345
SMAD2/4	Non-mutated	141	-		
	Mutated	19	3.125 (0.604 – 16.172)	0.1742	0.1515
BRCA1/2	Non-mutated	153	-		
	Mutated	8	8.881 (1.625 – 48.547)	0.0117	0.0023
MUTYH	Non-mutated	156	-		
	Mutated	5	0.000 (0.000 – .)	0.9957	0.6676
FBXW7	Non-mutated	149	-		
	Mutated	11	0.000 (0.000 – .)	0.9959	0.5137
ASXI1	Non-mutated	157	-		
	Mutated	4	0.000 (0.000 – .)	0.9962	0.7043
TET2	Non-mutated	156	-		
	Mutated	5	0.000 (0.000 – .)	0.9952	0.6314
DNMT3A	Non-mutated	157	-		
	Mutated	4	0.000 (0.000 – .)	0.9956	0.6567
RUNX1	Non-mutated	157	-		
	Mutated	4	0.000 (0.000 – .)	0.9951	0.7495
FAM123B	Non-mutated	150	-		
	Mutated	10	2.885 (0.335 – 24.881)	0.3350	0.3120
ARID1 a/b	Non-mutated	154	-		
	Mutated	7	0.000 (0.000 – .)	0.9955	0.6529
RB1	Non-mutated	158	-		
	Mutated	3	42.911 (4.461 – 412.742)	0.0011	< .0001
RNF43	Non-mutated	156	-		
	Mutated	4	0.000 (0.000 – .)	0.9959	0.7888
PTEN	Non-mutated	146	-		
	Mutated	13	0.000 (0.000 – .)	0.9942	0.5517
FLT3	Non-mutated	146	-		

	Mutated	15	8.172 (1.807 – 36.958)	0.0064	0.0011
ZNF217	Non-mutated Mutated	155 6	- 0.000 (0.000 – .)	0.9945	0.5756
SOX9	Non-mutated Mutated	148 13	- 6.014 (1.085 – 33.342)	0.0401	0.0193
BCOR	Non-mutated Mutated	152 9	- 5.010 (0.578 – 43.411)	0.1435	0.1036
CDK8	Non-mutated Mutated	148 13	- 5.492 (1.035 – 29.144)	0.0455	0.0245
IRS2	Non-mutated Mutated	153 8	- 16.066 (3.520 – 73.336)	0.0003	< .0001
SRC amp	Non-mutated Mutated	156 5	- 0.000 (0.000 – .)	0.9956	0.6607
MYC amp	Non-mutated Mutated	149 11	- 3.326 (0.388 – 28.474)	0.2727	0.2438
CTNNB1	Non-mutated Mutated	156 5	- 0.000 (0.000 – .)	0.9944	0.7153
NOTCH1/3	Non-mutated Mutated	157 4	- 0.000 (0.000 – .)	0.9960	0.6880
BCL2 amp	Non-mutated Mutated	153 7	- 0.000 (0.000 – .)	0.9943	0.5544
KRAS	Non-mutated Mutated	73 88	- 1.037 (0.229 – 4.694)	0.9623	0.9622
NRAS	Non-mutated Mutated	152 9	- 0.000 (0.000 – .)	0.9948	0.6004
PIK3CA	Non-mutated Mutated	134 27	- 3.562 (0.793 – 15.999)	0.0975	0.0766

Table 3 Univariate overall survival analysis

			Progression-Free Survival		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
P53	Non-mutated Mutated	24 92	1.245 (0.611 – 2.534)	0.5463	0.5402
APC	Non-mutated Mutated	22 94	0.604 (0.317 – 1.149)	0.1242	0.1158
ATM	Non-mutated Mutated	111 5	0.476 (0.066 – 3.446)	0.4626	0.4471
SMAD2/4	Non-mutated Mutated	103 12	1.299 (0.557 – 3.030)	0.5443	0.5387
BRCA1/2	Non-mutated Mutated	112 4	0.000 (0.000 – .)	0.9843	0.1065
MUTYH	Non-mutated Mutated	111 5	0.000 (0.000 – .)	0.9834	0.0869
FBXW7	Non-mutated Mutated	107 8	1.134 (0.411 – 3.130)	0.8084	0.8057
ASXI1	Non-mutated Mutated	112 4	1.306 (0.317 – 5.380)	0.7121	0.7086
TET2	Non-mutated Mutated	112 4	0.430 (0.059 – 3.107)	0.4028	0.3814
DNMT3A	Non-mutated Mutated	113 3	1.657 (0.401 – 6.842)	0.4849	0.4753
RUNX1	Non-mutated Mutated	113 3	0.579 (0.080 – 4.193)	0.5883	0.5792
FAM123B	Non-mutated Mutated	107 8	0.446 (0.109 – 1.830)	0.2624	0.2433
ARID1 a/b	Non-mutated Mutated	110 6	1.350 (0.420 – 4.342)	0.6149	0.6103
RNF43	Non-mutated Mutated	113 3	1.228 (0.169 – 8.904)	0.8393	0.8373
PTEN	Non-mutated Mutated	106 9	1.685 (0.664 – 4.273)	0.2721	0.2609
FLT3	Non-mutated Mutated	107 9	1.505 (0.600 – 3.778)	0.3839	0.3745
ZNF217	Non-mutated	111			

	Mutated	5	1.005 (0.313 – 3.228)	0.9933	0.9932
SOX9	Non-mutated Mutated	109 7	0.915 (0.285 – 2.939)	0.8816	0.8802
BCOR	Non-mutated Mutated	108 8	1.468 (0.530 – 4.067)	0.4606	0.4518
CDK8	Non-mutated Mutated	106 10	2.981 (1.332 – 6.671)	0.0079	0.0047
IRS2	Non-mutated Mutated	112 4	1.849 (0.441 – 7.748)	0.4004	0.3872
SRC amp	Non-mutated Mutated	114 2	0.000 (0.000 – .)	0.9845	0.1073
MYC amp	Non-mutated Mutated	111 5	0.614 (0.150 – 2.523)	0.4990	0.4885
CTNNB1	Non-mutated Mutated	112 4	0.542 (0.075 – 3.926)	0.5445	0.5336
NOTCH1/3	Non-mutated Mutated	114 2	0.743 (0.102 – 5.437)	0.7703	0.7652
BCL2 amp	Non-mutated Mutated	112 4	0.622 (0.150 – 2.581)	0.5134	0.5026
KRAS	Non-mutated Mutated	51 65	1.122 (0.671 – 1.878)	0.6600	0.6554
NRAS	Non-mutated Mutated	110 6	1.627 (0.588 – 4.499)	0.3484	0.3373
PIK3CA	Non-mutated Mutated	101 15	1.506 (0.761 – 2.978)	0.2394	0.2301

Table 4 Univariate progression-free survival analysis

Mutation	Survival Type	Hazard Ratio (95% CI)	P value
BRCA1/2	OS	6.98 (1.05 – 46.45)	0.0444
FLT3	OS	7.55 (1.436 – 39.74)	0.0170
SOX9	OS	10.23 (1.61 – 65.24)	0.0139
CDK8	OS	4.71 (0.813 – 27.31)	0.0837
IRS2	OS	31.63 (4.55 – 219.92)	0.0005
CDK8	PFS	3.122 (1.337 – 7.291)	0.0044

Table 5 Multivariate survival analysis adjusting for patient characteristics. Each gene associated with worse survival outcome in univariate survival analysis (See Table 3 and Table 4) was adjusted for patient characteristics by multivariate Cox model. Patient characteristics included in multivariate Cox model are gender, smoking history, tumor stage, age group, and primary tumor side. Adjusted hazard ratios for survival event (OS or PFS) in the mutated group versus non-mutated group were reported in this table. OS indicates overall survival and PFS indicates progression-free survival.

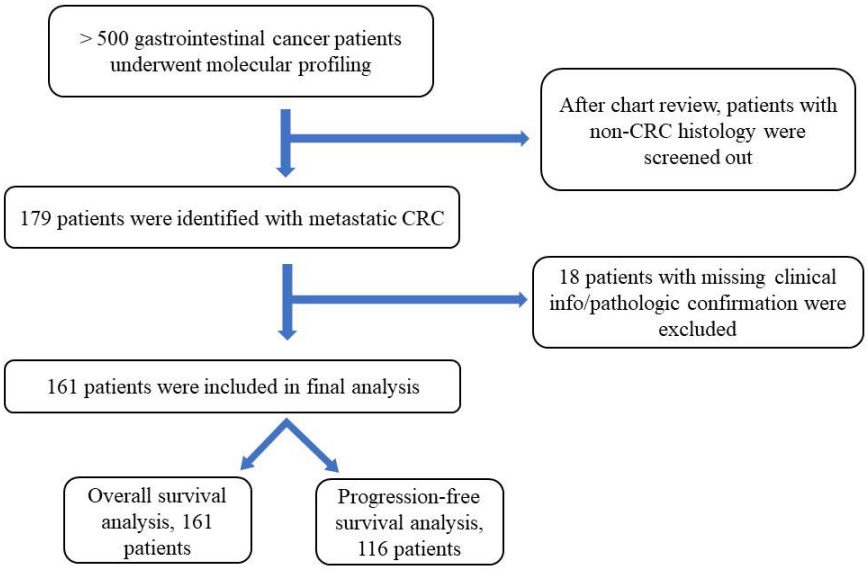

Figure 1 Patient selection flow chart

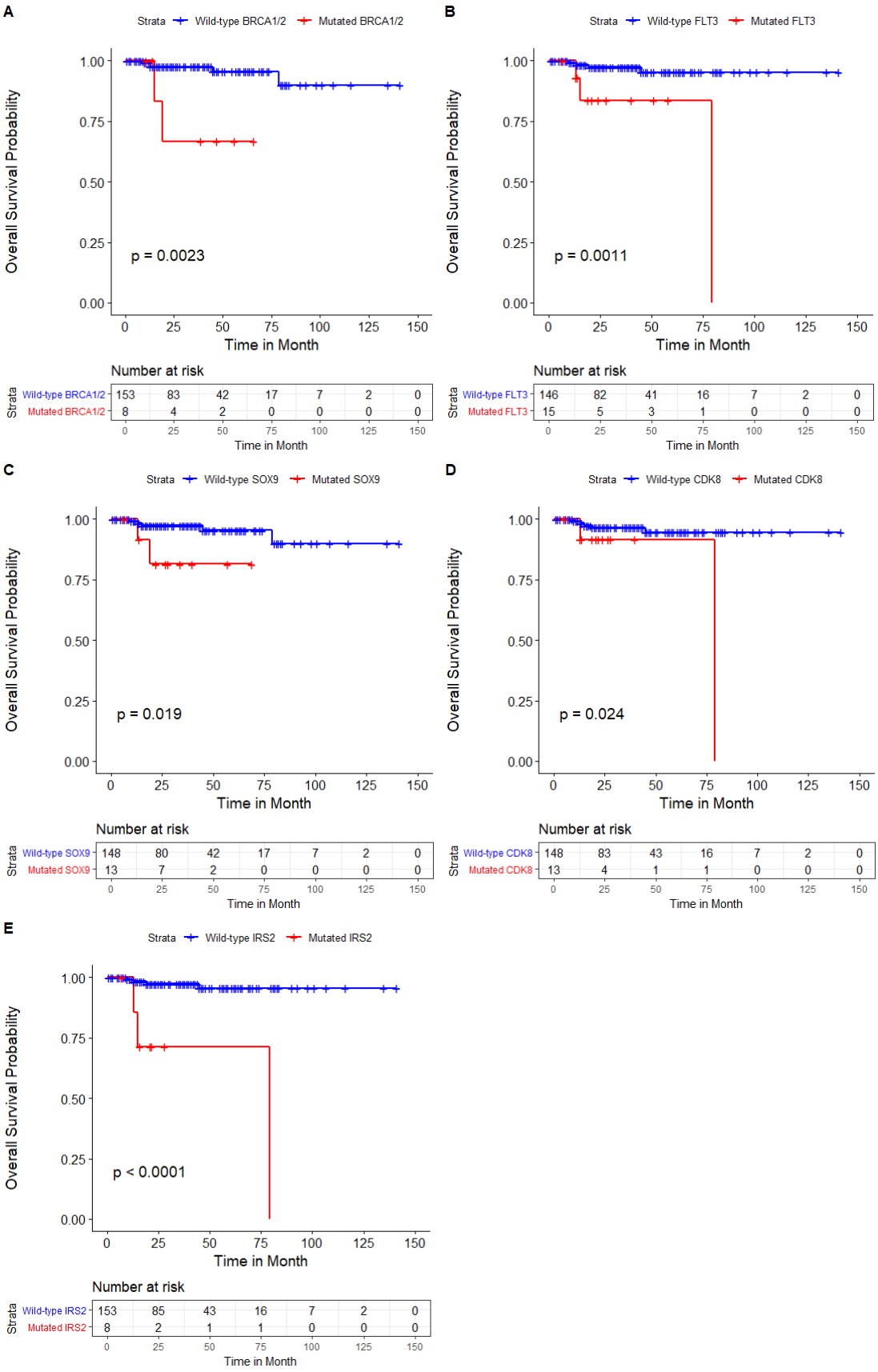
A
Strata Wild-type BRCA1/2 Mutated BRCA1/2
Overall Survival Probability
p = 0.0023
Time in Month
Number at risk
Wild-type BRCA1/2 153 83 42 17 7 2 0
Mutated BRCA1/2 8 4 2 0 0 0 0
B
Strata Wild-type FLT3 Mutated FLT3
Overall Survival Probability
p = 0.0011
Time in Month
Number at risk
Wild-type FLT3 146 82 41 16 7 2 0
Mutated FLT3 15 5 3 1 0 0 0
C
Strata Wild-type SOX9 Mutated SOX9
Overall Survival Probability
p = 0.019
Time in Month
Number at risk
Wild-type SOX9 148 80 42 17 7 2 0
Mutated SOX9 13 7 2 0 0 0 0
D
Strata Wild-type CDK8 Mutated CDK8
Overall Survival Probability
p = 0.024
Time in Month
Number at risk
Wild-type CDK8 148 83 43 16 7 2 0
Mutated CDK8 13 4 1 1 0 0 0
E
Strata Wild-type IRS2 Mutated IRS2
Overall Survival Probability
p < 0.0001
Time in Month
Number at risk
Wild-type IRS2 153 85 43 16 7 2 0
Mutated IRS2 8 2 1 1 0 0 0

Figure 2 Kaplan-Meier plots for overall survival stratified by mutation status. Red and blue lines indicate strata of patients with and without the mutation respectively (A) BRCA1/2; (B) FLT3; (C) SOX9; (D) CDK8; (E) IRS2.

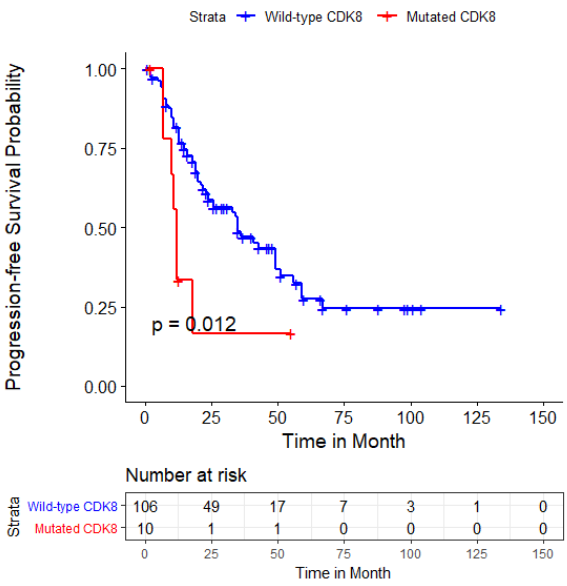

Figure 3. Kaplan-Meier plot for progression-free survival analysis stratified by mutation status of CDK8

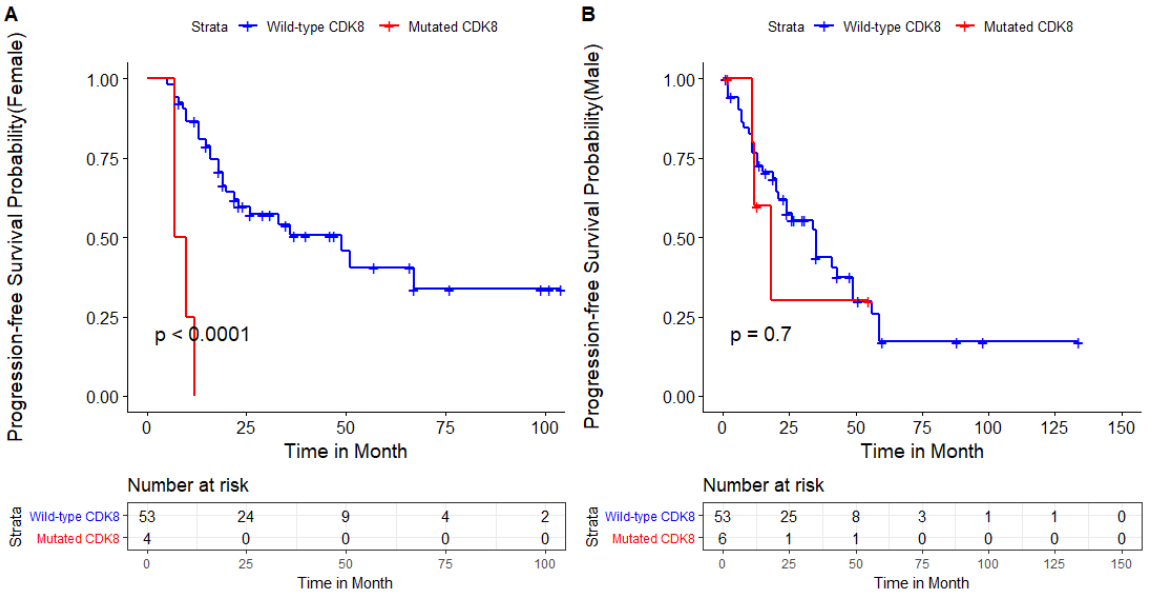

Figure 4. Interaction of gender and mutated CDK8. (A) is female, (B) is male.