Modeling RAS Phenotype in Colorectal Cancer Uncovers Novel Molecular Traits of RAS Dependency and Improves Prediction of Response to Targeted Agents in Patients

Justin Guinney1, Charles Ferté1,2, Jonathan Dry4, Robert McEwen5, Gilles Manceau5, KJ Kao3, Kai-Ming Chang3, Claus Bendtsen4, Kevin Hudson6, Erich Huang1, Brian Dougherty4, Michel Ducreux2, Jean-Charles Soria2, Stephen Friend1, Jonathan Derry1, and Pierre Laurent-Puig5

Abstract

Purpose: KRAS wild-type status is an imperfect predictor of sensitivity to anti-EGF receptor (EGFR) monoclonal antibodies in colorectal cancer, motivating efforts to identify novel molecular aberrations driving RAS. This study aimed to build a quantitative readout of RAS pathway activity to (i) uncover molecular surrogates of RAS activity specific to colorectal cancer, (ii) improve the prediction of cetuximab response in patients, and (iii) suggest new treatment strategies.

Experimental Design: A model of RAS pathway activity was trained in a large colorectal cancer dataset and validated in three independent colorectal cancer patient datasets. Novel molecular traits were inferred from The Cancer Genome Atlas colorectal cancer data. The ability of the RAS model to predict resistance to cetuximab was tested in mouse xenografts and three independent patient cohorts. Drug sensitivity correlations between our model and large cell line compendiums were performed.

Results: The performance of the RAS model was remarkably robust across three validation datasets. (i) Our model confirmed the heterogeneity of the RAS phenotype in KRAS wild-type patients, and suggests novel molecular traits driving its phenotype (e.g., MED12 loss, FBXW7 mutation, MAP2K4 mutation). (ii) It improved the prediction of response and progression-free survival (HR, 2.0; P < 0.01) to cetuximab compared with KRAS mutation (xenograft and patient cohorts). (iii) Our model consistently predicted sensitivity to MAP–ERK kinase (MEK) inhibitors (P < 0.01) in two cell panel screens.

Conclusions: Modeling the RAS phenotype in colorectal cancer allows for the robust interrogation of RAS pathway activity across cell lines, xenografts, and patient cohorts. It demonstrates clinical utility in predicting response to anti-EGFR agents and MEK inhibitors. Clin Cancer Res; 20(1); 265–72. ©2013 AACR.
Modeling the RAS phenotype in colorectal cancer allows for the robust interrogation of RAS molecular dependencies across cell lines, xenografts, and patient cohorts. It demonstrates clinical utility in predicting response to anti-EGF receptor antibodies and MAP–ERK kinase inhibitors.

Translational Relevance

Quality control analysis for outlier detection was performed on all data using principal component analysis. We used the penalized ElasticNet (24) regression model to predict KRAS mutation (codon 12 and 13). Optimal hyperparameters (α and λ in the ElasticNet) were selected using 5-fold cross-validation. Multiple bootstraps of the data (n = 100) were used to assess overfitting. Two-sided, nonparametric methods were used for all statistical evaluations. Multivariate cox regression was used to infer association with progression hazard. To assess the predictive performance of our model, concordance index and proportion of the variance were computed. All statistical analysis was conducted using the R version 3.0 and Bioconductor 2.1.2. All microarray data were processed and normalized using RMA (25). RNA-seq normalization was performed using conditional quantile normalization on the unnormalized count data (26). Comprehensive information of our modeling and statistical approaches can be found in the Supplementary Methods.

Materials and Methods

Patient cohorts

As training set, we used n = 334 fresh-frozen colorectal cancer tissues collected at the Koo Foundation Sun-Yat-Sen Cancer Center (KFSYSCC; Taipei, Taiwan) from 2000–2004 and profiled on the Affymetrix U133 plus 2.0 platform. After RNA and microarray quality control procedures (Supplementary Materials), 322 samples were retained. Taqman real-time PCR was used for detection of mutations in KRAS codon 12 and 13 as previously described (17). Quality control analysis of the microarray data revealed two outliers, which were removed from further analysis. Following the intersection of all samples that had both microarray and KRAS mutation status, 290 samples were available for analysis.

As validation dataset, we used the following publicly available and previously published datasets: Gaedcke and colleagues [ref. 18; n = 65 patients, Gene Expression Omnibus (GEO) id: GSE20842], Kambata-Ford and colleagues [ref. 15; n = 68 patients; GEO id: GSE5851], The Cancer Genome Atlas (TCGA) colorectal cancer dataset [ref. 19; n = 206 patients; https://tcga-data.nci.nih.gov/tcga]. Patient characteristics are described in Supplementary Table S1.

To assess the ability of our model to predict cetuximab response, we used the following datasets: Julien and colleagues [ref. 20; n = 54 mouse xenografts, n = 19 patients; ArrayExpress id: E-MTAB-991], Kambata-Ford and colleagues [ref. 15; n = 68 patients; GEO id: GSE5851], and INSERM (n = 85 patients; GEO id under process). Patient characteristics are described in Supplementary Table S2.

To assess the drug response of MAP–ERK kinase (MEK) inhibition, we use the following datasets: Barretina and colleagues [ref. 21; n = 19 cell lines, http://www.broadinstitute.org/ccle/home], Garnett and colleagues [ref. 22; n = 15 cell lines, http://cancerxgene.org], Jurchott and colleagues [ref. 23; n = 12 cell lines; GEO id: GSE18232], and mouse xenografts (n = 11; ArrayExpress id: E-MEXP-3557).

Bioinformatics and statistical analysis

We first assessed the performance of four recently published gene expression signatures (14, 16, 27, 28) to classify samples with KRAS mutation in several large human colorectal cancer datasets. Notably, these signatures were derived from several disease contexts, and were not specifically optimized for the prediction of RAS within colorectal cancer. We observed that all these signatures performed poorly in distinguishing KRAS mutant from wild-type samples (Supplementary Fig. S1A), underscoring the need of developing a specific RAS colorectal cancer model.

We trained a RAS model on an unpublished gene expression dataset (KFSYSCC; n = 290) derived from primary colon and rectal tumor samples to discriminate KRAS mutation status from wild-type. The output of our predictive model is a continuous value between 0 and 1, which we refer to as the RAS Index Score or RIS, with higher values corresponding to a stronger RAS phenotype. Internal bootstrapping of the data demonstrates robust prediction performance [Supplementary Fig. S1B; median area under the curve (AUC) = 0.81]. The performance of our model was robust across all validation datasets (AUC = 0.78–0.90), regardless of the location of the tumor in the intestine (primary colon, primary rectum, and metastatic colorectal disease) or of the RNA profiling platform (Affymetrix, Agilent, RNAseq; Supplementary Fig. S1C and Supplementary Table S4). The significant genes comprising our model are described in Supplementary Table S3.

To assess the specificity of our model to colorectal cancer, we repeated this evaluation in four lung adenocarcinoma and one endometrial carcinoma datasets. Overall, we observed much lower AUC performance in lung and uterine tumors (AUC = 0.53–0.68) compared with colorectal cancer (Supplementary Table S3), suggesting that our RAS
phenotype is specific to colorectal cancer and not generalizable to other disease types.

Quantitative model uncovers new molecular traits associated with RAS phenotype

We investigated how molecular aberrations contribute to the RAS phenotype using the TCGA colorectal cancer cohort. Expectedly, KRAS, BRAF, and NRAS-mutated samples exhibited a highly significant elevated RIS compared with wild-type samples, \( P < 2 \times 10^{-4} \) (Fig. 1A and Supplementary Fig. S3). Moreover, although we observed that the canonical activating KRAS mutations of codons 12, 13, 61, and 146 had a constitutively high RIS, there were large differences among them. The RIS of KRAS p.A146T and p.G13D samples was significantly lower than samples with a KRAS codon 12 mutation (Fig. 1B). A balanced bootstrap analysis in the KFSYSCC dataset confirmed that the lower RIS of KRAS p.G13D was not an artifact of the lower number of KRAS p.G13D samples relative to codon 12 mutations samples in the training data (Supplementary Fig. S2). Interestingly, the KRAS p.K117N and p.Q22K mutants had a high RIS and may be novel (albeit rare) activating mutations (Fig. 1A). Although BRAF p.V600E showed in aggregate a higher RIS compared with wild-type, it had a significantly lower RIS than KRAS codon 12 mutations \( (P < 5 \times 10^{-5}) \) and NRAS \( (P < 6 \times 10^{-5}) \) (Fig. 1B and Supplementary Fig. S3). We observed no significant contribution of PIK3CA mutations toward RIS variation looking at PIK3CA mutations in aggregate (Fig. 1C) or by PIK3CA exon (Supplementary Fig. S4). Repeating this analysis using colorectal cancer cell lines from the Cancer Cell-Line Encyclopedia (CCLE) dataset confirmed these results in vitro (Supplementary Fig. S5).

The distribution of RIS in samples annotated as RAS wild-type (i.e. KRAS, NRAS, and BRAF wild-type) reveals substantial heterogeneity (Fig. 1A and Supplementary Fig. S3). We hypothesize that this indicates the presence of alternative molecular aberrations that may contribute to the RAS phenotype. To investigate this, we incorporated somatic mutations, amplifications, and homozygous deletions from approximately 400 cancer genes (Sanger Cancer Census; ref. 29) into a feature selection model to identify additional aberrations that could explain RIS. By training multiple bootstrapped models, we were able to observe the frequency an aberration appeared in a model. The most frequently selected genomic aberrations below our false discovery criteria (see Supplementary Methods) are shown in Fig. 2. As a positive control, we observed that KRAS, NRAS, and BRAF are the first, second, and fourth ranked aberrations, respectively. Other aberrations associated with high RIS are NTRK3 mutation, CCND3/TFEB amplification, MED12 loss, and FBXW7 loss. Interestingly, MED12 loss has been recently shown to activate MEK/ERK through activation of TGFBR2 signaling, and may cause resistance to therapies acting upstream to MAPK/ERK (30). Among the aberrations associated with a low RIS is MAP2K4, a mitogen-activating...
kinase implicated in c-jun–NH2–kinase pathway signaling (31), and the tumor suppressor TP53. Although TP53 deletions or mutations are not mutually exclusive with KRAS mutations ($P > 0.5$), we observe a significant, negative association between TP53 aberrations and RIS ($P = 0.0071$, Wilcoxon rank-sum test).

RAS gene expression model improves prediction of response to cetuximab in KRAS wild-type patients

We investigated whether our model was able to predict response to cetuximab in a recently published panel of human colorectal cancer ($n = 19$) and mouse xenografts ($n = 54$; ref. 20). The RAS model separated responders versus nonresponders to cetuximab ($P < 0.05$) better than KRAS and/or BRAF mutation status ($P > 0.1$) both in the human specimens and mouse xenografts (Supplementary Fig. S6).

Next, we applied our model to a 68 metastatic colorectal cancer patient cohort (15) treated with cetuximab monotherapy. Again, our RAS model discriminated patient with disease control (AUC $= 0.75$) better than KRAS status alone (AUC $= 0.70$). Using the median RIS as a threshold, our RAS model separated nonresponders ($P = 3.19e^{-4}$, Fisher exact test) better than KRAS mutation status ($P = 3.3e^{-3}$).
Modeling RAS Phenotype in Colorectal Cancer

( Supplementary Fig. S7). Importantly, the positive predictive value of our model for disease control was higher than the KRAS status alone: 64% (16/25 patients) versus 51% (19/37 patients), respectively.

We next asked whether our RIS score could separate the patient cohort based on progression-free survival (PFS) time. Dichotomizing the patient cohort on the median RIS value, we observed a significant separation of survival time with our model {HR, 2.4; 95% CI, 1.2–4.35; \( P < 0.043 \)} (Supplementary Table S6), treating RIS as a continuous variable. When dichotomized on the median RIS, KRAS mutation status and RIS revealed RIS to be a significant factor {HR, 2.0; 95% CI, 1.03–3.83; \( P = 0.043 \)}; as well as with the KRAS status alone {HR, 2.3; 95% CI, 1.2–4.9; \( P = 0.043 \)}. We performed the same analysis restricting our analysis to the KRAS wild-type patient cohort {HR, 1.56; 95% CI, 0.76–3.2; \( P > 0.05 \)}. A multivariate cox regression analysis including treatment with MEK inhibitor PD-0325901 and selumitinib, \( P < 0.003 \), Spearman rank test; Fig. 4A). The mutation status of KRAS and/or BRAF did not predict MEK response \( (P > 0.2) \).

To corroborate the association between the RAS phenotype and response to MEK inhibition, we applied our model to the 19 colorectal cancer cell lines with drug response data available from CCLE (21) and computed the correlation between RIS and the drug sensitivity values measured for 24 recent anticancer drugs. Interestingly, we observed a significant correlation between RIS and drug sensitivity measurements for the two MEK inhibitors \( (PD-0325901 = 0.85, P = 0.0060), \) the HIF1A drug DMOG \( (R = 0.55, P = 0.0058) \), and the p90 ribosomal S6 kinases inhibitor \( (R = 0.76, P = 0.0058) \).

To expand the drug sensitivity analysis, we tested the RIS association with the cell line drug sensitivity data available from the Sanger dataset (22) with 130 drugs profiled (Fig. 4B). We confirmed the association of the MEK inhibitor AZD6244 (selumitinib) with the RAS phenotype \( (Spearman \text{rho} = 0.629, P = 0.032) \). Moreover, other significant associations included the fibroblast growth factor receptor inhibitor PD-173074 \( (P = 0.0027), \) IGF-IR inhibitor BMS-754807 \( (P = 0.0060), \) the HIF1A drug DMOG \( (R = 0.85, P = 0.0060), \) and the p90 ribosomal S6 kinases inhibitor \( (R = 0.76, P = 0.0058) \).

Figure 4. Volcano plots of Spearman correlation (x-axis) and significance (y-axis) evaluated between drug sensitivity (AUC) and RIS in the (A) CCLE (24 cancer drugs) and (B) Sanger Cell Line drug panel (130 drugs). Significant drug associations \( (P < 0.01) \) are highlighted. C, RIS measured for three colorectal cancer cell lines (HCT116, HT29, and SW480) following treatment with anti-EGFR drug AG1478, and MEK inhibitors PD098059 and U0126. Reduction of RIS is observed in the cell lines treated with MEK \( (P < 0.05) \). As a negative control, the EGFR-treated cell lines exhibited no reduction of RIS. D, xenograft colorectal cancer model showing distribution of RIS pre- (red) and post-treatment with MEK inhibitor selumitinib \( (AZD6244; n = 10), \) RIS reduction is observed in the KRAS-mutant cell line (HCT116) treated with MEK inhibitor, whereas the KRAS wild-type–treated cell line (H9K2) shows a modest increase in RIS.
to a gene expression panel-derived three colorectal cancer cell lines profiled pre- and post MEK inhibition (ref. 23; SW480, HCT116, and HT29; Fig. 4C). We observed that the MEK inhibitor consistently reduced the RIS in all of these cell lines ($P = 0.038$, PD098059; $P = 0.030$, U0126; Fig. 4C). We observed similar findings in a xenograft panel of HCT116 (KRAS p.G13D) and HKH2 (KRAS wild-type) cell lines ($n = 10$) treated with the MEK inhibitor selumitinib (Fig. 4D).

**Discussion**

KRAS mutation status is recognized as a key determinant of resistance to anti-EGFR antibodies in colorectal cancer. The increasing number of targeted agents that impact the EGFR/RAS/MEK signaling pathway and the failure of these therapies to translate seamlessly between disease types provide strong motivation to assess the activity of this pathway in a context-dependent manner, that is lung cancer versus colorectal cancer. Herein, we provide a robust quantitative readout of the RAS phenotype that is specific to colorectal cancer. Using this model, we were able to (i) unravel novel molecular aberrations that contribute to the RAS phenotype (e.g., KRAS-specific amino acid changes, NRAS, and BRAF), (ii) improve the prediction of response and survival to cetuximab treatment in KRAS wild-type patients, and (iii) uncover targeted strategies matched to the patient’s RAS phenotype (e.g., MEK inhibitors).

The specificity of KRAS mutation to predict resistance to anti-EGFR therapy is very high, whereas the wild-type status of KRAS has a low sensitivity anti-EGFR response (2, 4). Our model improves the discrimination of KRAS wild-type patients that are likely to benefit from cetuximab therapy both in terms of PFS and response rate. Particularly, our model demonstrated a significant PFS HR in two independent datasets: HR = 2.0 (1.2–3.3) in the INSERM patient cohort ($P < 0.05$) and HR = 2.3 (1.2–4.9; $P < 0.05$) in the KRAS wild-type patients from the Khambata-Ford dataset. In the latter dataset, we also observed that 64% of the patients with KRAS wild-type and low RIS exhibited partial response or stable disease as compared with only 51% of the patients with KRAS wild-type only. The fact that most of the patients in these two datasets received conventional chemotherapy plus cetuximab may have influenced our models. However, our results are consistent with other studies describing an overlap between RAS and cetuximab gene expression models (32), thus confirming the central role of KRAS in the prediction of response to cetuximab.

Interestingly, our model not only identifies wild-type KRAS patients not suitable for cetuximab, but also uncovers potential alternative treatment strategies. In that regard, we highlight the sensitivity of MEK inhibitors in colorectal cell lines exhibiting a high RAS phenotype. Moreover, we have shown that high RAS phenotype is reversed by MEK inhibition, both in vitro and in xenograft models. Although this information has great potential for changing clinical practice, these warrant confirmation in large prospective cohorts.

Finally, our RAS model has allowed us to probe the relative potency of specific oncogenic mutations contributing to the RAS phenotype. Particularly, we found that KRAS p.G13D-mutated samples had a lower RAS phenotype than p.G12D- or p.G12V-mutant samples. This is consistent with the recent literature describing a different molecular phenotype of the KRAS p.G13D mutation (33, 34) as well as suggesting a possible benefit of cetuximab for patients harboring this specific trait (12, 35). Second, BRAF-mutated samples also exhibited a lower RAS phenotype as compared with KRAS p.G12D- and p.G12V-mutated samples, suggesting a smaller effect in the activation of RAS downstream signaling. This is corroborated in a recent large phase III trial that first-line metastatic patients harboring in BRAF mutations in their tumors could benefit from a combination of chemotherapy plus anti-EGFR therapy (13, 10) but remains a matter of debate for second or additional lines of treatment (36). We did not observe any relationship between specific PIK3CA mutations and the RAS phenotype as measured by RIS, aligning with recent data suggesting an absence of effect of these mutations in predicting cetuximab resistance (37). Further functional studies are warranted to confirm these findings.

In conclusion, we have produced a powerful molecular classifier of the RAS phenotype specific to the colorectal cancer setting. We provide strong evidence by translating efficiently our tools across various platforms (Affymetrix, Agilent, RNAsseq) and datasets (cell line and xenograft panels, and patient cohorts). Importantly, our classifier improves the identification of KRAS wild-type patients likely to benefit from anti-EGFR monoclonal antibodies. Furthermore, our study suggests the application of MEK inhibitors to patients with a high RAS phenotype. These findings have strong potential for practice changes at the bedside, but require prospective validation in patient cohorts.

**Disclosure of Potential Conflicts of Interest**

M. Ducreux has honoraria from speakers’ bureau from Merck Serono and Amgen. P. Laurent-Puig is a consultant/advisory board member of Amgen, Merck Serono, Sanoﬁ, Roche, and Boehringer Ingelheim. No potential conflicts of interest were disclosed by the other authors.

**Authors’ Contributions**

Conception and design: J. Guinney, C. Ferte, J. Dry, C. Bendtsen, K. Hudson, E. Huang, J. Derry

Development of methodology: J. Guinney, C. Ferte, K.J. Kao, P. Laurent-Puig

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): C. Ferte, K.J. Kao, E. Huang, M. Ducreux, J.-C. Soria, P. Laurent-Puig

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Guinney, C. Ferte, J. Dry, R. McElwen, G. Manceau, K.-M. Chang, C. Bendtsen, B. Dougherty, J. Derry, P. Laurent-Puig

Writing, review, and/or revision of the manuscript: J. Guinney, C. Ferte, J. Dry, G. Manceau, C. Bendtsen, K. Hudson, B. Dougherty, M. Ducreux, J.-C. Soria, J. Derry, P. Laurent-Puig

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Guinney, C. Ferte, K.J. Kao, K.-M. Chang

Study supervision: J. Guinney, C. Ferte, S. Friend, J. Derry
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 15, 2013; revised September 27, 2013; accepted October 9, 2013; published OnlineFirst October 29, 2013.

References


