Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Clinical validation of a machine-learning derived signature predictive of outcomes from first-line oxaliplatin-based chemotherapy in advanced colorectal cancer

Single sentence summary: AI-based analysis of colon cancer molecular data results in a platform predictive of OS in TRIBE2 arms (FOLFOX HR=0.629, p=0.04; FOLFOXIRI HR=0.483, p=0.02).

Authors:

Jim Abraham[1], Daniel Magee[1], Chiara Cremolini[2], Carlotta Antoniotti[2], David D. Halbert[1], Joanne Xiu[1], Phillip Stafford[1], Donald A. Berry[3], Matthew Oberley[1], Anthony F. Shields[4], John L. Marshall[5], Mohamed Salem[6], Alfredo Falcone[2], Axel Grothey[7], Michael Hall[8], Alan P. Venook[9], Heinz-Josef Lenz[10], Anthony Helmstetter[1], W. Michael Korn[1,11], David Spetzler[1,11]

Affiliations:

[1]Caris Life Sciences, Phoenix, AZ, USA; [2]Departments of Oncology and Translational Research and New Technologies in Medicine, University Hospital Pisa, Pisa, Italy; [3]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; [4]Department of Oncology, Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA; [5]Ruesch Center for The Cure of Gastrointestinal Cancers, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA; [6]Levine Cancer Institute, Carolinas HealthCare System, Charlotte, NC, USA; [7]Medical Oncology, West Cancer Center, 9745 Wolf River Blvd,

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Germantown, TN, USA; [8]Department of Clinical Genetics, Fox Chase Cancer Center, Philadelphia, PA, USA; [9]Division of Hematology/Oncology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA; [10]University of Southern California, Keck School of Medicine, Norris Comprehensive Cancer Center, Los Angeles, California

[11]**Corresponding Authors:** David Spetzler and W. Michael Korn, Caris Life Sciences, Phoenix, AZ 85040, USA. Phone: 602-464-7537; E-mails: dspetzler@carisls.com, wmkorn@carisls.com.

Running Title: AI Analysis of Molecular Data to Predict FOLFOX Response

Keywords: AI, colon cancer, molecular diagnostics, FOLFOX, FOLFIRI

Disclosure of Potential Conflicts of Interest: JA, DM, DDH, JX, PS, MO, AH, WMK and DS are employees of Caris Life Sciences. DAB and JLM serve on the scientific advisory board of Caris Life Sciences. AFS, MS, AG, MH, and HJL have unpaid consultant/advisory board relationships with Caris Life Sciences.

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Word count, abstract: 250

Word count, translational relevance: 139

Word count, text (from the beginning of Introduction to the end of Discussion): 4927

Reference count: 31

Table count: 3

Figure count: 3

Supplementary material: 5 tables, 3 figures

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Translational Relevance

FOLFOX, FOLFIRI, or FOLFOXIRI chemotherapy in combination with bevacizumab are considered standard 1st line treatment options for patients with metastatic colorectal cancer. To determine if benefit from first-line chemotherapy in colorectal cancer is predictable, we leveraged artificial intelligence algorithms and comprehensive molecular profiling data to identify and validate a predictor of benefit from FOLFOX chemotherapy. A 67 gene signature was predictive of survival in an independent RWE dataset of 412 patients who had received FOLFOX/BV in 1st line and inversely predictive of outcomes in RWE data from 55 patients who had received 1st line FOLFIRI chemotherapy. Blinded analysis of TRIBE2 samples confirmed that FOLFOX*ai* was predictive of OS in both oxaliplatin-containing arms (FOLFOX HR=0.629, p=0.04 and FOLFOXIRI HR=0.483, p=0.02). FOLFOX*ai* was also predictive of benefit from oxaliplatin-containing regimens in advanced esophageal/gastro-esophageal junction cancers as well as pancreatic ductal adenocarcinoma.

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Abstract

Purpose: FOLFOX, FOLFIRI, or FOLFOXIRI chemotherapy with bevacizumab (BV) are considered standard 1st line treatment options for patients with metastatic colorectal cancer (mCRC). We developed and validated a molecular signature predictive of efficacy of oxaliplatin-based chemotherapy combined with BV in patients with mCRC.

Experimental Design: A machine-learning approach was applied and tested on clinical and NGS data from a real-world evidence (RWE) data set and samples from the prospective TRIBE2 study resulting in identification of a molecular signature - FOLFOX*ai*. Algorithm training considered time-to-next-treatment (TTNT). Validation studies used TTNT, PFS and overall survival (OS) as the primary endpoints.

Results: A 67 gene signature was cross-validated in a training cohort (N=105) which demonstrated the ability of FOLFOX*ai* to distinguish FOLFOX-treated mCRC patients with increased benefit (IB) from those with decreased benefit (DB). The signature was predictive of TTNT and OS in an independent RWE dataset of 412 patients who had received FOLFOX/BV in 1st line and inversely predictive of survival in RWE data from 55 patients who had received 1st line FOLFIRI. Blinded analysis of TRIBE2 samples confirmed that FOLFOX*ai* was predictive of OS in both oxaliplatin-containing arms (FOLFOX HR=0.629, p=0.04 and FOLFOXIRI HR=0.483, p=0.02). FOLFOX*ai* was also predictive of treatment benefit from oxaliplatin-containing regimens in advanced esophageal/gastro-esophageal junction cancers (EC/GEJC) as well as pancreatic ductal adenocarcinoma (PDAC).

Conclusions: Application of FOLFOX*ai* could lead to improvements of treatment outcomes for patients with mCRC and other cancers since patients predicted to have

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

less benefit from oxaliplatin-containing regimens might benefit from alternative regimens.

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Introduction

The promise of precision cancer therapy has not yet been fully realized for patients with mCRC. Over the past two decades, conventional chemotherapies (e.g. oxaliplatin and irinotecan) and biologics (including bevacizumab (BV) which targets Vascular Endothelial Growth Factor (VEGF)) and cetuximab and panitumumab (targeting Epidermal Growth Factor-Receptor)) have shown activity in first-line treatment of mCRC. In combination with a fluoropyrimidine, the resulting chemotherapy doublets (FOLFOX and FOLFIRI) have each been found superior to the components[1] and have become standard of care, typically in combination with biologics. However, numerous studies have failed to clearly establish that any of these combination regimens would be superior for any individual patient based on clinical factors. Recently, results of the TRIBE2 phase III study[2] demonstrated that the upfront triple-combination of 5-FU, oxaliplatin, and irinotecan (FOLFOXIRI) with BV followed by the reintroduction of the same regimen after disease progression resulted in improved overall survival compared to the sequential administration of chemotherapy doublets (FOLFOX followed by FOLFIRI), in combination with BV[2]. However, these improved outcomes were achieved at the cost of increased, clinically relevant toxicity, thus limiting broad applicability of this approach.

Since 2008, when the presence of *KRAS* mutations in a tumor were found to preclude benefit from antibodies targeting EGFR[3], it has been anticipated that multi-gene molecular profiling would further refine the ability to personalize treatment of mCRC. However, biomarkers relevant for first-line treatment decisions are limited to KRAS, NRAS, and *BRAF*, as well as microsatellite instability status [4–6]. Efforts to

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

identify biomarkers for chemotherapy in mCRC have been even less fruitful than for the biologics. Genetic polymorphisms in metabolizing enzymes may explain toxicities of fluoropyrimidines and irinotecan but are of limited clinical value. Topoisomerase levels have been shown to be unhelpful when considering irinotecan activity, as have VEGF-A serum levels for bevacizumab[7].

The vast majority of mCRC patients receive FOLFOX-based first-line treatment even though neuropathy almost always limits its use beyond four months. Oxaliplatin has also become a first-line option as part of FOLFOXIRI in mCRC[2] and for other cancers, including FOLFOX in first-line esophageal and gastric cancer[8] and as part of FOLFIRINOX in advanced pancreatic cancer[9]. Given other choices in these diseases, a biomarker predicting the relative efficacy of these regimens would be very helpful. Because oxaliplatin has little activity as monotherapy[1], it is used exclusively in combination with fluoropyrimidines so a biomarker for FOLFOX (as opposed to oxaliplatin alone) would be of pertinent clinical value. The urgent need for predictive biomarkers is highlighted by the fact that a randomized study of 376 patients was conducted that demonstrated that tumor expression of the excision repair cross-complementing-1 gene (ERCC-1) is not a valid predictor of oxaliplatin efficacy in mCRC[7].

Business as usual has not worked in the pursuit of these biomarkers. The routine application of comprehensive molecular profiling, in particular involving next-generation DNA sequencing, has allowed for the creation of increasingly refined molecular portraits of large numbers of tumors from a diverse and representative patient pool[10]. Systematic molecular analyses of colorectal cancers have demonstrated

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

extensive inter- and intra-tumoral heterogeneity and, at the same time, have led to the identification of sub-classes of the disease with different prognostic and therapeutic characeristics[11]. While most of the currently available studies have utilized conventional statistics for disease sub-classification, recent advances in machine learning enable identification of non-intuitive and non-linear patterns and hold the promise of supporting diagnostic and therapeutic decision making with high accuracy.

The availability of large combined clinical and molecular datasets enables development of novel molecular predictors of efficacy of standard treatments. Here, we report the results of a machine learning approach in an attempt to identify a molecular signature predictive of clinical benefit from FOLFOX chemotherapy in previously untreated patients with mCRC. We then sought validation of the putative molecular signature from a large RWE database, a subset of cases from the randomized controlled phase III TRIBE2 study as well as RWE data from patients with advanced EC/GEJC or PDAC who received first-line treatments with oxaliplatin-containing regimens.

Methods

Real-World Evidence (RWE) and TRIBE2 Clinical Trial Cohorts

We utilized an extensive de-identified RWE outcomes dataset collected from the Caris Life Sciences Precision Oncology Alliance registry, and insurance claims data from over 10,000 physicians. The following inclusion criteria were applied for selecting the training cohort: 1.) diagnosis of mCRC, 2.) treatment with FOLFOX-based combination therapy, 3.) completion of at least one full cycle of therapy, 4.) completed next-generation DNA

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

analysis of at least one CRC sample using a 592-gene panel. Patients were excluded if they had prior chemotherapy, including adjuvant therapy. Two separate RWE validation cohorts were generated using the following inclusion criteria: 1.) diagnosis of mCRC, 2.) first-line FOLFOX/BV treatment (FOLFOX/BV cohort) or first-line FOLFIRI-based treatment (FOLFIRI cohort), 3.) completion of at least one full cycle of therapy, 4.) completed next-generation DNA analysis of at least one CRC sample using a 592-gene panel, and 5.) switch to an irinotecan-containing regimen (FOLFOX/BV cohort) or to FOLFOX (FOLFIRI cohort). Inclusion criteria for the FOLFOX/BV cohort were modeled after the TRIBE2 study protocol.

A blinded retrospective-prospective analysis of samples from patients enrolled in the phase III TRIBE2 study, with completed NGS analysis, was performed for further clinical validation. The trial, conducted by the Italian Gruppo Oncologico del Nord-Ovest (GONO), compared the upfront exposure to FOLFOXIRI/BV followed (after maintenance therapy of 5-FU/ BV) by the reintroduction of the same regimen to a preplanned sequential strategy of FOLFOX/BV followed by FOLFIRI/BV after disease progression in the treatment of patients with mCRC. Detailed eligibility criteria and results have been previously reported [2]. All personnel at Caris Life Sciences were blinded to any clinical data associated with these samples. The samples in this trial were subjected to the same quality controls and genomic testing protocols as cases from the RWE training and testing cohorts and outcomes predictions for these cases were returned to GONO for unblinding and assessment of the model's performance. In addition, exploratory analyses were performed in RWE cohorts patients with metastatic PDAC who had received either nab-paclitaxel/gemcitabine or FOLFIRINOX as first-line

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

treatment regimen and patients with metastatic or unresectable esophageal or gastroesophageal adenocarcinoma who had been treated with FOLFOX as first-line treatment regimen. Therapy records for all patients included in the RWE dataset were curated by a board certified medical oncologist prior to inclusion in the study.

Time To Next Treatment (TTNT)

TTNT was defined as the time from first administration of oxaliplatin or 5-fluorouracil following the biopsy or surgical specimen collection to the first administration of irinotecan (indicating a switch to FOLFIRI) or last contact. Patients were algorithmically identified using this method, followed by manual curation by a board-certified medical oncologist to ensure that a FOLFOX regimen had been used appropriately and that the TTNT value was accurate. For algorithm training, a TTNT of 270 days was chosen to define whether a patient benefitted from receiving first-line FOLFOX based on the progression free survival (PFS) noted by Tournigard et al. of approximately 8.5 months[12] and approximately 30 days less than the PFS in the MAVERICC study[7]. Training cases were required to have at least 270 days of follow-up after beginning the FOLFOX regimen if there was no observable switch to FOLFIRI (i.e. short-censored cases). We refer to patients with TTNT < 270 days as having DB to FOLFOX and others as having IB.

Overall Survival (OS)

OS was calculated for all eligible cases which is defined as the time from treatment initiation date to either death for the RWE dataset (from the National Death Index (NDI),

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

National Center for Health Statistics, Centers for Disease Control and Prevention) or last contact in the insurance claims repository. We assumed that any patient without a claim for over 100 days had died, which holds true for over 95% of patients with a recorded death in the NDI. Conversely, patients with a last contact date within 100 days of the most recent refresh of the RWE repository were censored. With regard to the TRIBE2 analysis, OS was defined as time from randomization to death.

Kaplan-Meier Metrics

All listed hazard ratios use the Cox proportional hazards (PH) model and the p-values come from the log-rank statistic. To test whether the signature predicts the same survival benefit for different first-line therapies, we generated a Cox PH model on the combined RWE cohorts for either mCRC, PDAC or EC/GEJC using the model prediction, first-line treatment, and an interaction term between first-line treatment and predicted benefit as covariates. To visualize the effect of the interaction between DB probability and first-line treatment, we used the fitted Cox model to predict relative risk on simulated data using first-line treatment information.

Tumor Samples and Next Generation Sequencing

Tumor-containing formalin fixed paraffin embedded (FFPE) blocks from surgery or biopsy prior to administration of any chemotherapy were used to generate all genomic data used in all analyses as described previously[13]. The genomic data have been deposited in an open-access GitHub repository and are available at https://github.com/carisls/Abraham_2020.

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Algorithm Ensemble and Model Selection

Genomic features obtained via the Caris Life Sciences 592-gene NGS panel were used as feature inputs into an ensemble of over 300 published machine learning algorithms, including random forest, support vector machine, logistic regression, K-nearest neighbor, artificial neural network, naïve Bayes, quadratic discriminant analysis, and Gaussian processes models. Multiple feature selection methods including principal component analysis, mutual information, and variable importance were employed to build models that predict increased benefit (IB) or decreased benefit (DB) to first-line FOLFOX chemotherapy. The final algorithm returns an IB probability obtained from a consensus of model predictions which decreases the dependence on individual fixed decision trees.

Performance of all machine learning models were implemented using the Python package scikit-learn v0.22.1. Biomarker data obtained via NGS for the 105 patients in the training cohort were used as features to classify each patient based on their binary IB or DB status, defined by TTNT $\geq$ 270 days as previously described. We analyzed the performance of models on the training data via 5-fold cross validation, selected due to the relatively small sample size, on metrics including hazard ratio, sensitivity, specificity, positive and negative predictive values, and overall accuracy. Emphasis was placed on correctly identifying the patients with the longest and shortest TTNT due to the obvious clinical implications of misclassifying them. With respect to hazard ratio and the log-rank p-value, predictions on the five hold-out folds were combined and this metric was measured once per iteration in order to avoid possibly misleading values due to low N

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

(21 per fold). Models were trained on both the full biomarker set as well as smaller feature sets after applying various feature selection methods. For example, we used variable importances across folds in tree-based methods like random forests to establish the biomarkers which are most essential to correct classification. The feature set was then trimmed to only include the most important biomarkers, and the same model was retrained and cross-validation metrics were evaluated. The performance improvement, or lack thereof, using the smaller dataset helped reveal whether the culled features were introducing noise that negated our ability to successfully classify the IB and DB cohorts. An exhaustive search using this method where each individual machine learning algorithm was allowed to independently select features and tune hyperparameters yielded five random forest configurations that were able to effectively and consistently separate the IB and DB cohorts without demonstrating statistical differences in mutation frequencies of prognostic variables such as BRAF or the sidedness of the tumor. The algorithms were also evaluated for their ability to predict response in patients that received first-line FOLFIRI in addition to the primary FOLFOX training cohort. Models that exhibited similar hazard ratios in the FOLFOX and FOLFIRI cohorts were discarded as they were likely driven by prognostic factors, and features that enabled such performance were penalized or removed during additional model training. These findings in the training data suggested that our algorithms may be predictive of FOLFOX response and would extrapolate well onto the testing dataset without demonstrating potentially confounding biases. We then focused on establishing a single, aggregated prediction for each patient using the five individual classifiers.

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

In order to address ambiguity or disagreement among models for any given patient, we employed a voting scheme. Without being bound by theory, each model may perform optimally on cases having different characteristics, and in combination the voting scheme accounts for suboptimal or supraoptimal performance of any given model on a subset(s) of cases. In order to achieve a consensus and reduce subtle noise implemented by any individual random forest model, we trained 1,000 instances of each of the five model configurations for a total of 5,000 models. Each of these models were used to vote on the consensus prediction for all patients in the validation cohorts and return a probability of increased benefit using our deliberative analytics (DEAN) framework. The DEAN framework evaluates multiple approaches for reaching a consensus ranging from simple aggregations to sophisticated machine learning models. For FOLFOX, the median of the model probabilities provides the optimal result. To further account for model noise, we introduced a buffer surrounding the 50% IB probability threshold in which cases will be considered a “no call”. We chose this threshold by observing the IB probability range for each patient in this study. A 3% buffer was selected as it is one standard deviation larger than the mean range, so the model will not return a prediction if the IB probability falls within 47-53%. At this time, the locked ensemble predictor was graduated to the testing phase and was applied to the additional RWE and TRIBE2 data. We refer to our patent for full details of the machine learning configurations and inputs[14].

Assessment of Predictive Versus Prognostic Nature of the Model

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

To test whether the signature was merely prognostic, we generated a Cox proportional hazards model on the combined set of RWE patients that received either first-line FOLFOX or first-line FOLFIRI for mCRC and FOLFIRINOX or nab-paclitaxel/gemcitabine for PDAC. The EC/GEJC cohort was excluded from this analysis as only one first-line therapy was included in this work. Three terms were included in the Cox model: first-line treatment, predicted benefit (IB or DB), and an interaction term between first-line treatment and predicted benefit. An additional three term Cox proportional hazards model was fit on the same cohort, with the binary IB/DB prediction replaced with the continuous valued probability of DB from the model. To visualize the effect of the interaction between DB probability and first-line treatment, we used the fitted Cox model to predict relative risk on simulated data where RB probability ranged from 0.01 to 0.99, once with FOLFOX as the first-line treatment and again with FOLFIRI as the first-line treatment for mCRC and similarly FOLFIRINOX or nab-paclitaxel with gemcitabine for PDAC.

Consensus Molecular Subtype (CMS) Classification

A Consensus Molecular Subtype classifier was developed using expression values obtained from de-identified RNASeq data from routine testing at the Caris Life Sciences Laboratory allowing for classifying colorectal cancers into 4 subytpes analagous to Guinney et al[10]. A full 22,948-gene dataset of expression data is produced by the Salmon RNASeq pipeline[15] which provides fast and bias-aware quantification of transcript expression. This pipeline yields discrete TPM (Transcripts Per Million molecules) values for each gene transcript. A classifier was trained against the original

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

CMS datasets as published by Guinney et al.[11] using a classic SVM model as implemented in R. A TCGA dataset of 512 cases was excluded from training. 600 genes were subsequently selected for each of the four CMS subtype classifiers using One vs. All t-test to identify genes uniquely expression in each of the four classes. Cross-validation was performed to optimize the model and finalize SVM parameters. Possible overtraining was evaluated by predicting CMS subtypes from an independent blinded TCGA dataset (512 samples[11]) with an accuracy of 88.3%.

Compliance statement

The study was conducted in accordance with the Declaration of Helsinki and adhered to Good Clinical Practice guidelines. Molecular analyses were performed in a Clinical Laboratory Improvements Amendments (CLIA) approved laboratory. Approval was obtained from the local ethics committees of participating sites, and all TRIBE2 patients provided written informed consent to the study while the RWE analysis was performed on de-identified data. This part of the study was exempt from consent requirement as per review by the Western Institutional Review Board.

Results

Patients

The training cohort consisted of 105 mCRC patients from the RWE dataset who had received first-line FOLFOX-based treatment and who had been profiled by Caris Life Sciences (Fig1). IB and DB cohorts were well balanced in terms of age, gender, tumor location (left, right), mutation status, and biologic agent administered in combination

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

with chemotherapy (Table S1). The first independent validation cohort Included 412 patients (with RWE data on treatments and death dates) treated with FOLFOX/BV and 55 patients who had received FOLFIRI as first-line treatments (Fig 1). Of the FOLFIRI patients, 82% received bevacizumab, 9% panitumumab, and 5% cetuximab, respectively, while 4% (two patients) did not receive a combination with a biologic. In all RWE cohorts, IB, no call, and DB groups were well-balanced in terms of key prognostic features, including age, gender, tumor location (left, right), *KRAS/NRAS/BRAF* mutation status and micro-satellite status (Table 1). Additional RWE datasets included 333 patients with advanced PDAC and EC/GEJC treated in first-line with oxaliplatin-containing regimens (Tables S2-3, Fig 1B). PDAC patients in the nab-paclitaxel/gemcitabine group were significantly older than patients in the FOLFIRINOX group (68.4 vs 59 years; $p<0.0001$), in agreement with current prescribing practices. Comparison of the characteristics of the 296-patient subset from the TRIBE2 trial for whom complete NGS tumor analyses were available (Fig 1), demonstrated that the subset was representative of the entire study population (Table S4, Fig 1C). Twenty-five patient samples in this set of samples did not meet the minimum quality metric (sequencing depth requirement of 300x).

Model Training and Validation in Real-World Evidence Cohorts

The RWE cohort did not include PFS, therefore we used TTNT instead. We compared TTNT and PFS within the TRIBE2 samples in both the FOLFOX (Pearson's $r = 0.98$) and FOLFOXIRI ($r = 0.99$) arms of the trial (Fig S1A) and found them to be highly correlated. Model training was done using TTNT on a patient cohort that included 63

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

patients with IB and 42 DB based on our benefit definition (Methods). Results of 5-fold cross validation demonstrated that the model consistently separated IB from DB cohorts (median HR=0.398 for 100 model cross-validations, 95% CI 0.244 – 0.649, log-rank p<0.001; Fig S2). The final model takes 67 genes into account (Table 2). Among the most relevant features included in the signature were genes involved in mediating WNT signaling (BCL9, CDX2), epithelial-to-mesenchymal transition (INHBA, PRRX1, PBX1, YWHAE), chromatin remodeling (EP300, ARID1A, SMARC4, NSD3), DNA repair (WRN, BRIP1), NOTCH signaling (MAML2) and cell cycle regulation (CNTRL, CCNE1). After locking the algorithm, further validation of the predictive signature (named FOLFOX*ai*) was performed in the FOLFOX/BV cohort (Fig 1).

No call was made if the model output, interpretable as decreased benefit probability, was between 0.47-0.53 (Methods), which was the case in 50 patients (12.1%, Fig 3). There were 242 patients in the IB cohort and 120 in the DB cohort which are well-balanced in terms of known prognostic features with the exception of a higher representation of tumors with indeterminate/high MSI status in patients with predicted DB (Table 1). Kaplan-Meier analysis demonstrates a significant difference in TTNT and overall survival (OS) based on the predicted IB or DB, respectively (median TTNT of 11.5 mo for IB and 8.2 mo for DB, HR = 0.537, 95% CI: 0.428-0.674, log-rank p < 0.0001, Fig 2A; median OS of 42.0 mo for IB and 24.5 mo for DB, HR = 0.466, 95% CI: 0.325-0.670, log-rank p<0.0001, Fig 2B). To analyze specificity of FOLFOX*ai*, we applied the predictor to the FOLFIRI cohort. In contrast to the FOLFOX/BV cohort, the signature prediction resulted in inverted survival curves in the FOLFIRI cohort: patients predicted to have DB from FOLFOX had significantly better outcomes than those with

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

predicted IB and vice versa. The median OS was 18.7 mo for IB and 34.4 mo for DB (HR = 2.631, 95% CI: 1.041-6.649, log-rank p = 0.034; Fig 2C-D). A multivariate Cox PH model was perfomed using all FOLFOX and FOLFIRI patients (Methods). A third-line therapy was present in 25% of FOLFOX and 22% of FOLFIRI patients, primarily in the form of trifluridine and tiparcil (lonsurf) or regorafenib. With respect to time to third treatment, the FOLFOX cohort predicted as IB shows improvement over the predicted DB cohort (HR = 0.429, p < 0.0001) while the FOLFIRI cohort does not show statistical significance (HR = 1.680, p = 0.20). The interaction between the treatment and prediction covariates shows statistical significance so we reject the null hypothesis that FOLFOX*ai* predicts the same OS benefit for both, the FOLFOX and FOLFIRI cohorts (Table S5, Fig 2I-J).

We next asked whether FOLFOX*ai* was predictive of treatment efficacy of oxaliplatin-containing regimens in other disease. To address this possibility, similar analyses were conducted in the PDAC and EC/GEJC cohorts. These analyses demonstrate that the signature was indeed predictive of overall survival in the FOLFIRINOX, with a median OS improvement of 10.1 months in the IB cohort (21.4 mo for IB, 11.3 mo for DB,; HR=0.478, CI: 0.289 – 0.792, log-rank p=0.003) but not the nab-paclitaxel/gemcitabine cohort (median OS 10.8 mo for IB, 9.8 mo for DB; HR=0.958, CI: 0.658 – 1.395, log-rank p=0.823) (Fig S3A-B). Like FOLFOX and FOLFIRI for mCRC, the Cox PH covariate for the interaction between first-line PDAC therapy and FOLFOX*ai* prediction yields p=0.03, so we again reject the null hypothesis that the FOLFOX*ai* prediction provides the same OS benefit for both the FOLFIRINOX and nab-paclitaxel/gemcitabine cohorts (Table S5). Similarly, data from 104 patients with

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

advanced EC/GEJC demonstrate that FOLFOX*ai* is predictive of efficacy of oxaliplatin containing regimens also in this clinical setting (median OS for IB: 14 mo, for DB: 8.9 mo; HR=0.437, CI: 0.250 – 0.763, log-rank p=0.003, Fig S3C). These results demonstrate specificity and potential for broad clinical applicability of FOLFOX*ai*.

Blinded Retrospective-Prospective Analysis of the TRIBE2 Study

Data and samples from 271 patients were available for analysis from the TRIBE2 study. IB vs DB was predicted for 97 vs 36 patients on the FOLFOX/BV arm and 83 vs 20 patients on the FOLFOXIRI arm, respectively. No call was made by the predictive algorithm in 35 (12.9%) patients (Methods). Median PFS1 for patients with IB was 0.9 months longer than for DB (9.6 mo vs 8.7 mo; HR=0.757, 95% CI 0.505 – 1.135, log-rank p=0.18; Fig 2E) and the median OS difference was 6.0 months (24.8 mo vs 18.7 mo; HR=0.629, CI: 0.404 – 0.981, log-rank p=0.04; Fig 2F) in the FOLFOX/BV arm. The differences were also significant in the FOLFOXIRI/BV arm for OS (PFS1: 13.8 mo vs 7.6 mo; HR=0.683, CI: 0.396-1.181 log-rank p=0.17, Fig 2G; OS: 30 mo vs 15.9 mos; HR=0.483, CI: 0.270 – 0.864, log-rank p=0.02, Fig 2H). Thus, this blinded retrospective-prospective analysis of samples from the TRIBE2 trial confirms the signature differentiates IB vs DB in terms of PFS and OS in patients receiving FOLFOX or FOLFOXIRI in combination with BV.

Overlap with Colorectal Cancer Consensus Molecular Subtypes

Investigations into differences in RNA expression profiles in colorectal cancer have revealed four distinct colorectal molecular subtypes (CMS1-4) that are associated with

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

different prognoses and possibly response to chemo- and biologic therapies[11, 16, 17]. To assess whether FOLFOX*ai* merely reproduced this classification, we first validated the Caris WTS-based consensus molecular subtype (CMS) classifier using 2224 WTS profiles available in the Caris database. The classifier assigned a CMS class to the samples analyzed with similar frequency distribution and molecular characteristics as the published, expression array-based classifier[11]. Next, we calculated both, CMS classification as well as the FOLFOX signature in 3744 colorectal cancer cases from the Caris database (Table 3). Cancers with predicted improved benefit from FOLFOX/BV were more likely to be represented in the CMS2 group while cancers classified as CMS1 were more frequently predicted to show decreased benefit from FOLFOX/BV treatment.

Discussion

The quality (depth and durability) of a clinical response following first-line chemotherapy in patients with mCRC usually foreshadows survival. Since oxaliplatin-associated neuropathy develops by the fourth or fifth month of treatment in most patients and tumors acquire chemotherapy resistance over time, the initial therapeutic impact is particularly important. However, reliable molecular predictors of response to chemotherapy are currently unavailable to inform the choice of initial therapy. With this study, we took advantage of an advanced machine learning approach to identify and validate FOLFOX*ai*, a molecular signature predictive of treatment benefit from FOLFOX chemotherapy by analyzing a combined dataset of comprehensive molecular profiling results and clinical outcomes data. The key finding of our studies is that FOLFOX*ai* is

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

predictive of overall survival in patients with mCRC, EC/GEJC, and PDAC who receive oxaliplatin-containing chemotherapy regimens in first-line. To our knowledge, this is the first clinically validated machine-learning powered molecular predictor of chemotherapy efficacy in these diseases with immediate relevance for the initial therapeutic decision-making process.

Molecular landscape studies as well as our own data demonstrate an extensive inter-individual molecular heterogeneity of mCRC and the presence of up to several thousand mutations per case[18, 19]. Thus, it is likely that more than a single mechanism contributes to sensitivity and *de novo* resistance to chemotherapy, a notion that is exemplified by the lack of predictive power for RNA expression level of a single gene (ERCC1), which encodes a base-excision DNA repair enzyme, as demonstrated in a large, randomized phase II study[7]. In contrast, broad molecular characterization of cancers holds the promise of revealing complex systems biology via molecular patterns associated with treatment benefit. Machine-learning algorithms can be instrumental in uncovering such patterns as it has been demonstrated in the context of radiologic imaging[20]. However, machine-learning algorithms as decision support tools for standard treatment decisions in oncology have mostly been limited to cognitive support systems such as IBM Watson[21] or lack sufficient clinical validation[22, 23].

We applied a 2-step clinical validation approach that utilized synthetic study cohorts assembled from RWE data followed by retrospective-prospective analysis of samples from the TRIBE2 study. Inclusion crieteria for the RWE cohorts were modeled after the TRIBE-2 study but maintain some of the variability of treatment schedules observed in daily practice that are not reflected in a randomized clinical trial[24]. The

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

validation through the retrospective-prospective analysis of TRIBE2 samples is complementary to the RWE analysis and represents an accepted validation strategy for predictive biomarkers[25]. The FOLFOXai signature was able to identify patients on both oxaliplatin-based arms of the randomized TRIBE2 trial who would ultimately have increased benefit, with a clinically relevant increase in OS of 6.0 months (FOLFOX/BV arm) or 14.9 months (FOLFIXIRI/BV arm). Moreover, the FOLFOX*ai* predicted treatment benefit from FOLFIRI inversely. Therefore, we propose a clinical decision algorithm that utilizes FOLFOX*ai* to prioritize either FOLFOX or FOLFIRI as first-line chemotherapy in patients who are not candidates for the triple agent FOLFOXIRI regimen and to guide drug discontinuation in case of toxicity in patients receiving FOLFOXIRI as first-line treatment. While these findings suggest incorporation of FOLFOX*ai* in first-line treatment decisions, additional aspects need further consideration. Most importantly, utilization of the signature in the first-line setting requires rapid turn-around times of NGS, signature calculation and reporting, which Caris Life Sciences has achieved with an average turn-around time of 13.5 calendar days. Furthermore, we were not able to prospectively test whether patients predicted to have decreased benefit from FOLFOX will indeed derive greater benefit from FOLFIRI. To address that question formally, a prospective registry study has been implemented.

Our finding that FOLFOX*ai* is also predictive of survival in patients treated with FOLFIRINOX for advanced PDAC highlights the potential extrapolation of this signature to other clinical settings that involve platinum agents and points to potential utility in selecting patients for this more toxic but nonetheless standard regimen compared to nab-paclitaxel/gemcitabine[9, 26]. In concert with our findings in EC/GEJC, these results

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

underscore that FOLFOX*ai* might capture molecular themes relevant for treatment response beyond colorectal cancer, a possibility that is currently under active investigation.

The 67 molecular features included in the signature also provide novel insights into putative biologic mechanisms driving intrinsic resistance to the FOLFOX/bev or FOLFOX combination. Most notably, factors involved in WNT signaling and mediation of epithelial-to-mesenchymal transition (EMT) were among the most important features in the signature which are both well known to confer resistance to chemotherapeutic agents[9]. For example, BCL9 functions as a transcriptional co-activator of the canonical WNT pathway and has been shown to promote a stem-cell like phenotype[27, 28] that was associated with platinum-resistance in NSCLC[29]. High expression of BCL9 was found to be associated with poor outcome in CRC, possibly through mediating neuron-like, multicellular communication properties (in addition to its WNT regulatory function.)[30]

The Consensus Molecular Subtypes have been demonstrated to be prognostic in metastatic CRC and predictive for the efficacy of biologic agents[11, 17]. In the adjuvant setting, the CMS2 subtype has been demonstrated to particularly benefit from FOLFOX chemotherapy[31]. In agreement with this, we observed enrichment of CMS2 in the IB cohort. However, our results suggest that FOLFOX*ai* predicts significant portion of CMS1, 3 and 4 patients to also benefit from oxaliplatin-based treatment with metastatic disease, highlighting its independence form currently established molecular classifiers.

In summary, the totality of this work serves as a validation of FOLFOX*ai*, a molecular signature of efficacy of oxaliplatin-based therapy, which is predictive of overall survival in mCRC. Thus, comprehensive molecular profiling performed at the

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

time of diagnosis of metastatic colorectal cancer (and potentially other cancers) not only delivers key information relevant to targeted and immunotherapies such as mutations in *KRAS*, *NRAS*, and *BRAF*, MSI status, and HER2 amplification, but also provides crucial guidance for the choice of first-line chemotherapy.

Acknowledgements

The authors thank Dr. George Demetri for providing critical review of the manuscript.

This work was supported by Caris Life Sciences.

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

References

1. Rothenberg ML, Oza AM, Bigelow RH, et al: Superiority of oxaliplatin and fluorouracil-leucovorin compared with either therapy alone in patients with progressive colorectal cancer after irinotecan and fluorouracil-leucovorin: Interim results of a phase III trial. Journal of Clinical Oncology 21:2059–2069, 2003

2. Cremolini C, Antoniotti C, Rossini D, et al: Upfront FOLFOXIRI plus bevacizumab and reintroduction after progression versus mFOLFOX6 plus bevacizumab followed by FOLFIRI plus bevacizumab in the treatment of patients with metastatic colorectal cancer (TRIBE2): a multicentre , open-label , phase 3 ,. The Lancet Oncology , 2020

3. Amado RG, Wolf M, Peeters M, et al: Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. Journal of Clinical Oncology 26:1626–1634, 2008

4. Kopetz S, Grothey A, Yaeger R, et al: Encorafenib, binimetinib, and cetuximab in BRAF V600E–mutated colorectal cancer. New England Journal of Medicine 381:1632–1643, 2019

5. Le DT, Uram JN, Wang H, et al: PD-1 Blockade in Tumors with Mismatch-Repair Deficiency [Internet]. New England Journal of Medicine 372:2509–2520, 2015Available from: http://www.nejm.org/doi/10.1056/NEJMoa1500596

6. Andre, T., Shiu, K-K, Kim, TW, Jensen, BV, Jensen, L-H, Punt, CJA, Smith, DM, Garcia-Carbonero, R, Benavides, Gibbs, MP, De La Fouchardiere, C, Rivera, F, Elez, E, Bendell, JC, Le, DT, Yoshino, T, Yang, P, Farooqui, MZH, Marinello, P, Diaz L: Pembrolizumab versus chemotherapy for microsatellite instability-high/mismatch repair deficient metastatic colorectal cancer: The phase 3 KEYNOTE-177 study. J Clin Oncol 38:suppl; abstr LBA4, 2020

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

7. Parikh AR, Lee FC, Yau L, et al: MAVERICC, a randomized, biomarker-stratified, phase II study of mFOLFOX6-Bevacizumab versus FOLFIRI-bevacizumab as first-line chemotherapy in metastatic colorectal cancer. Clinical Cancer Research 25:2988–2995, 2019

8. Al-Batran SE, Hartmann JT, Probst S, et al: Phase III trial in metastatic gastroesophageal adenocarcinoma with fluorouracil, leucovorin plus either oxaliplatin or cisplatin: A study of the Arbeitsgemeinschaft Internistische Onkologie. Journal of Clinical Oncology 26:1435–1442, 2008

9. Conroy T, Desseigne F, Ychou M, et al: FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer [Internet]. N Engl J Med 364:1817–1825, 2011Available from: http://www.ncbi.nlm.nih.gov/pubmed/21561347%5Cnhttp://www.nejm.org/doi/pdf/10.1056/NEJMoa1011923

10. Campbell PJ, Getz G, Korbel JO, et al: Pan-cancer analysis of whole genomes. Nature 578:82–93, 2020

11. Guinney J, Dienstmann R, Wang X, et al: The consensus molecular subtypes of colorectal cancer. Nature Medicine 21:1350–1356, 2015

12. Tournigand C, André T, Achille E, et al: FOLFIRI followed by FOLFOX6 or the reverse sequence in advanced colorectal cancer: A randomized GERCOR study. Journal of Clinical Oncology 22:229–237, 2004

13. Salem ME, Puccini A, Grothey A, et al: Landscape of tumor mutation load, mismatch repair deficiency, and PD-L1 expression in a large patient cohort of gastrointestinal cancers. Molecular Cancer Research 16:805–812, 2018

14. Abraham, J., Spetzler, D., Helmstetter, A., Korn, W.M., Magee D: Next generation molecular profilingWO2020113237, 2020

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

15. Patro R, Duggal G, Love MI, et al: Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods 14:417–419, 2017

16. Stintzing S, Wirapati P, Lenz HJ, et al: Consensus molecular subgroups (CMS) of colorectal cancer (CRC) and first-line efficacy of FOLFIRI plus cetuximab or bevacizumab in the FIRE3 (AIO KRK-0306) trial. Annals of oncology : official journal of the European Society for Medical Oncology 30:1796–1803, 2019

17. Lenz HJ, Ou FS, Venook AP, et al: Impact of consensus molecular subtype on survival in patients with metastatic colorectal cancer: Results from CALGB/SWOG 80405 (Alliance), in Journal of Clinical Oncology. 2019, pp 1876–1885

18. Muzny DM, Bainbridge MN, Chang K, et al: Comprehensive molecular characterization of human colon and rectal cancer [Internet]. Nature 487:330–337, 2012Available from: http://dx.doi.org/10.1038/nature11252

19. Salem ME, Weinberg BA, Xiu J, et al: Comparative molecular analyses of left-sided colon, right-sided colon, and rectal cancers [Internet]. Oncotarget 8, 2017Available from: http://www.oncotarget.com/fulltext/21169

20. Shen L, Margolies LR, Rothstein JH, et al: Deep Learning to Improve Breast Cancer Detection on Screening Mammography. Scientific Reports 9:1–12, 2019

21. Somashekhar SP, Sepúlveda MJ, Puglielli S, et al: Watson for Oncology and breast cancer treatment recommendations: Agreement with an expert multidisciplinary tumor board. Annals of Oncology 29:418–423, 2018

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

22. Kaissis G, Ziegelmayer S, Lohöfer F, et al: A machine learning algorithm predicts molecular subtypes in pancreatic ductal adenocarcinoma with differential response to gemcitabine-based versus FOLFIRINOX chemotherapy. PLoS ONE 14:1–16, 2019

23. Mucaki EJ, Zhao JZL, Lizotte DJ, et al: Predicting responses to platin chemotherapy agents with biochemically-inspired machine learning [Internet]. Signal Transduction and Targeted Therapy 4, 2019Available from: http://dx.doi.org/10.1038/s41392-018-0034-5

24. Stewart M, Norden AD, Dreyer N, et al: An Exploratory Analysis of Real-World End Points for Assessing Outcomes Among Immunotherapy-Treated Patients With Advanced Non–Small-Cell Lung Cancer. JCO Clinical Cancer Informatics 1–15, 2019

25. Parkinson DR, McCormack RT, Keating SM, et al: Evidence of clinical utility:An unmet need in molecular diagnostics for patients with cancer. Clinical Cancer Research 20:1428–1444, 2014

26. von Hoff DD, Ervin T, Arena FP, et al: Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine [Internet]. New England Journal of Medicine 369:1691–1703, 2013Available from: http://www.nejm.org/doi/10.1056/NEJMoa1304369

27. de la Roche M, Worm J, Bienz M: The function of BCL9 in Wnt/β-catenin signaling and colorectal cancer cells. BMC Cancer 8:1–13, 2008

28. Deka J, Wiedemann N, Anderle P, et al: Bcl9/Bcl9l are critical for Wnt-mediated regulation of stem cell traits in colon epithelium and adenocarcinomas. Cancer Research 70:6619–6628, 2010

29. Zhang Y, Zhang Q, Chen H, et al: BCL9 promotes epithelial mesenchymal transition and invasion in cisplatin resistant NSCLC cells via β-catenin pathway [Internet]. Life Sciences 208:284–294, 2018Available from: https://doi.org/10.1016/j.lfs.2018.07.023

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

30. Jiang M, Kang Y, Sewastianik T, et al: BCL9 provides multi-cellular communication properties in colorectal cancer by interacting with paraspeckle proteins [Internet]. Nature Communications 11, 2020Available from: http://dx.doi.org/10.1038/s41467-019-13842-7

31. Song N, Pogue-geile KL, Gavin PG, et al: Clinical Outcome From Oxaliplatin Treatment in Stage II/III Colon Cancer According to Intrinsic Subtypes: Secondary Analysis of NSABP C-07/NRG Oncology Randomized Clinical Trial2:1162–1169, 2016

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Tables

Table 1. Demographics of the cases used in the mCRC RWE testing set.

	First-line FOLFOX							First-line FOLFIRI						
	Actual			Model Prediction				Actual			Model Prediction			
Characteristic	**DB**	**IB**	**p***	**DB**	**No Call**	**IB**	**p***	**DB**	**IB**	**p***	**DB**	**No Call**	**IB**	**p***
Number of patients	179	233		120	50	242		26	29		20	6	29	
Mean Age (Range)	58.0 (31-85)	57.9 (22-86)	0.93	57.5 (33-78)	56.6 (31-80)	58.5 (22-86)	0.44	60.9 (26-85)	55.3 (32-77)	0.09	56.0 (32-77)	62.8 (53-79)	58.4 (26-85)	0.47
Gender (F/M, %)	45/55	42/58	0.73	43/57	48/52	43/57	0.78	27/73	48/52	0.18	35/65	33/67	41/59	0.87
Primary Tumor Site, %[a]	44/38/18	53/29/18	0.13	52/36/12	52/36/12	48/31/21	0.21	35/54/12	62/21/17	0.04	65/30/5	83/17/0	31/45/24	0.04
BRAF, % [b]	91/1/9	97/0/3	0.02	92/0/8	98/2/0	94/0/6	0.03	100/0/0	90/3/7	0.24	90/5/5	83/0/17	100/0/0	0.20
KRAS, % [b]	45/1/54	41/0/59	0.36	42/0/58	36/2/62	45/0/55	0.07	42/0/58	28/0/72	0.39	25/0/75	50/0/50	38/0/62	0.45
MSI, % [c]	92/7/2	91/9/0	0.35	84/13/2	92/6/2	95/5/0	<0.01	96/4/0	97/3/0	1.00	95/5/0	100/0/0	97/3/0	0.85
Second line EGFRi, % [d]	74/26	71/29	0.59	76/24	74/26	70/30	0.46	88/12	72/28	0.25	75/25	83/17	83/17	0.78

[a] left/right/unknown or mix
[b] no pathogenic variant detected/indeterminate/pathogenic variant detected
[c] stable/indeterminate/high
[d] no/yes
* chi-square tests are used for categorical variables and Mann-Whitney U for non-normally distributed age data

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Table 2. List of genomic features used in the algorithm.

ACKR3	BRIP1	EWSR1	IL2	MSI2	PRRX1	TOP1
AKT2	CASP8	EZR	INHBA	MYC	RUNX1T1	TRRAP
AKT3	CCNE1	FAS	KEAP1	NFIB	SBDS	U2AF1
ARFRP1	CDX2	FCRL4	LHFPL6	NFKBIA	SDC4	WRN
ARID1A	CNTRL	FH	LMO1	NSD3	SEPT5	WWTR1
ASXL1	COX6C	FLT1	MAML2	PAX7	SMARCA4	YWHAE
AURKA	CREB1	FLT3	MAP2K4	PBX1	SOX10	ZNF217
BCL9	CRKL	GAS7	MLF1	PCM1	TCL1A	
BIRC3	EP300	GNAS	MN1	PDGFB	TERT	
BRD3	ETV6	HOXA11	MNX1	PER1	TLX3	

Table 3. Predictions of the algorithm by CMS subtype.

CMS Subtype	**DB N (%)**	**No Call N (%)**	**IB N (%)**	**Total N (%)**	**Predicted DB/ No Call/ IB (%)**	**p***	**TCGA Validation Sensitivity (%)**
CMS1	280 (29)	62 (13)	228 (10)	570 (15)	49/11/40	< 0.001	93.4

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

CMS2	158 (16)	135 (29)	842 (37)	1135 (30)	14/12/74	< 0.001	88.6
CMS3	160 (16)	60 (13)	318 (14)	538 (14)	30/11/59	0.11	56.7
CMS4	380 (39)	206 (44)	915 (40)	1501 (40)	25/14/61	0.11	100.0
Total	978	463	2303	3744			88.3

* chi-square test

Figure Legends

Figure 1 - Consort diagram of A) selection of patients on which to train and cross validate the FOLFOXai model, B) Real World Evidence testing data and C) TRIBE2 randomized phase III cases. Outcomes in (A) were last refreshed in September 2018 and outcomes in (B) were last refreshed in July 2020. The FOLFOXai algorithm was locked in December 2018, over nine months before receiving the outcomes associated with (B).

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure 2 - Blinded RWE testing cohorts that received first-line FOLFOX (A-B) or first-line FOLFIRI (C-D), Arms A (FOLFOX, E-F) and B (FOLFOXIRI, G-H) of the TRIBE2 trial, and relative risk of MI FOLFOXai scores between the RWE cohorts (I-J). A) The median TTNT in the Increased Benefit cohort is 3.3 months (40%) longer than the median TTNT in the Decreased Benefit arm (HR = 0.537, 95% CI: 0.428-0.674, log-rank p < 0.001). B) The median OS in the Increased Benefit cohort is 17.5 months (71%) longer than the median OS in the Decreased Benefit arm (HR = 0.466, 95% CI: 0.325-0.670, log-rank p < 0.001). C) The median TTNT in the Increased Benefit cohort is 4.6 months (36%) shorter than the median TTNT in the Decreased Benefit arm (HR = 1.422, 95% CI: 0.797-2.538, log-rank p = 0.233). D) The median OS in the Increased Benefit cohort is 15.7 months (46%) shorter than the median OS in the Decreased Benefit arm (HR = 2.631, 95% CI: 1.041-6.649, log-rank p = 0.032). E) The median PFS in the Increased Benefit cohort is 0.9 months (10%) longer than the median PFS in the Decreased Benefit arm (HR = 0.757, 95% CI: 0.505-1.135, log-rank p = 0.18). F) The median OS in the Increased Benefit cohort is 6.0 months (32%) longer than the median OS in the Decreased Benefit arm (HR = 0.629, 95% CI: 0.403-0.981, log-rank p = 0.04). G) The median PFS in the Increased Benefit cohort is 6.2 months (82%) longer than the median PFS in the Decreased Benefit arm (HR = 0.684, 95% CI: 0.396-1.181, log-rank p = 0.169). H) The median OS in the Increased Benefit cohort is 14.1 months (89%) longer than the median OS in the Decreased Benefit arm (HR = 0.483, 95% CI: 0.270-0.864, log-rank p = 0.02). Relative risk with respect to TTNT (I) and OS (J) as a function of decreased benefit probability on the patients for which we make a call for both the FOLFOX and FOLFIRI RWE cohorts from a Cox proportional hazards model using

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

decreased benefit probability and treatment as the covariates. The intersecting nature of the curves indicates that the DB probability has a different risk effect depending on the treatment received, as evidenced by the reversed signal observed in the RWE FOLFOX (A-B) and RWE FOLFIRI (C-D) cohorts.

Figure 3 - Density of predictions in the RWE testing cohort and the TRIBE2 trial.

Figure S1 - A) Correlation of TTNT and PFS of cases in the TRIBE2 trial. B) Distribution of mCRC RWE TNTs compared with PFS from the FOLFOX and FOLFOXIRI arms of the TRIBE2 trial.

Figure S2 - Cross-validated performance of the training cases. The median TTNT in the Increased Benefit cohort is 3.9 months (52%) longer than the median TTNT in the Decreased Benefit arm (HR = 0.398, 95% CI: 0.244-0.647, log-rank p < 0.001).

Figure S3 - Kaplan-Meier plots of probability of overall survival in patients with advanced PDAC receiving first-line FOLFIRINOX (A) or nab-paclitaxel/gemcitabine (B) or patients with GC/EGJC receiving an oxaliplatin-containing regimen (C). A) The median OS in the Increased Benefit cohort is 10.1 months (90%) longer than the median OS in the Decreased Benefit cohort (HR = 0.478, 95% CI: 0.289-0.792, log-rank p = 0.003). B) The median OS in the Increased Benefit cohort is one month (11%) longer than the median OS in the Decreased Benefit cohort (HR = 0.958, 95% CI: 0.658-1.395, log-rank p = 0.823). C) The median OS in the Increased Benefit cohort is

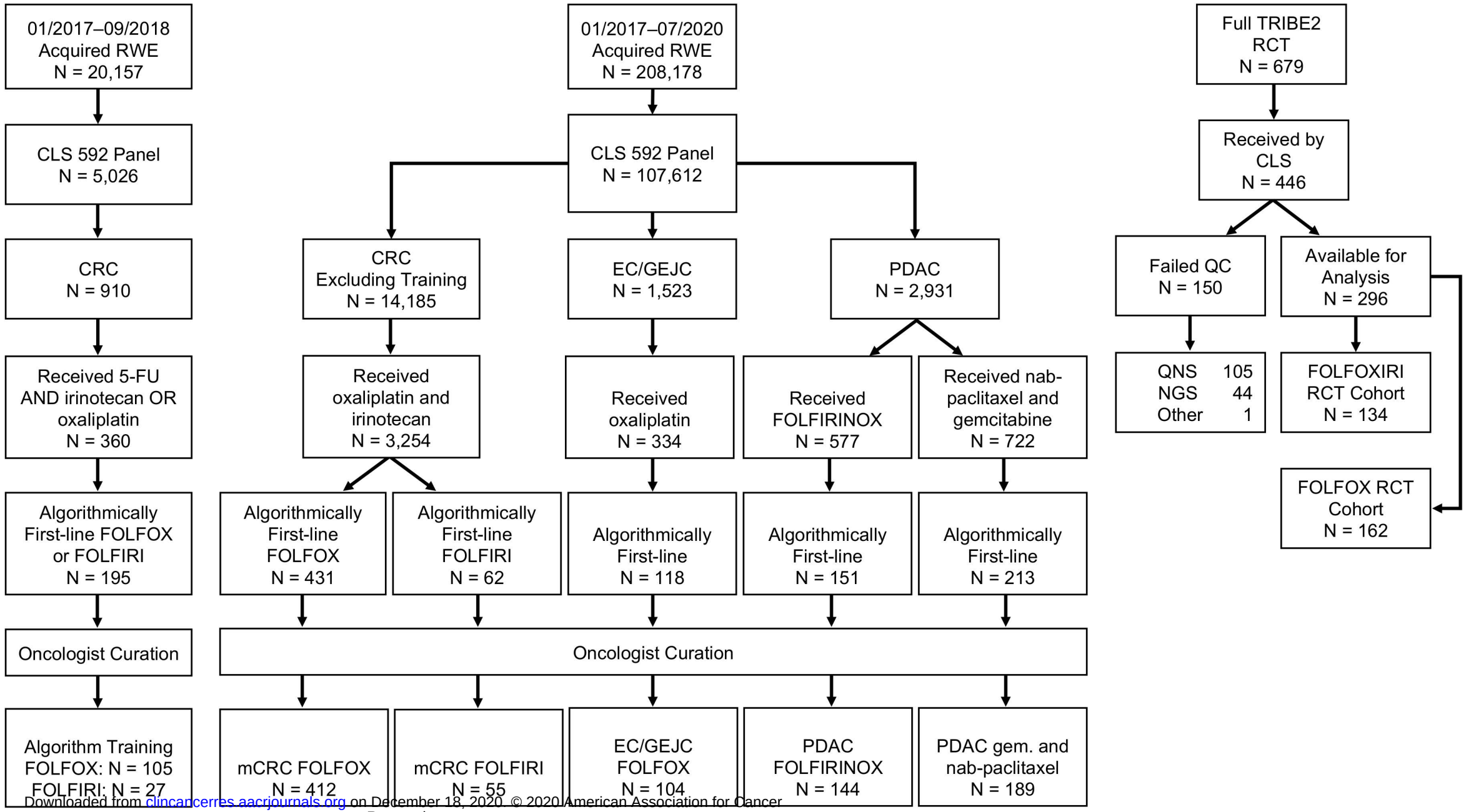
Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

5.1 months (58%) longer than the median OS in the Decreased Benefit cohort (HR = 0.437, 95% CI: 0.250-0.763, log-rank p = 0.003).

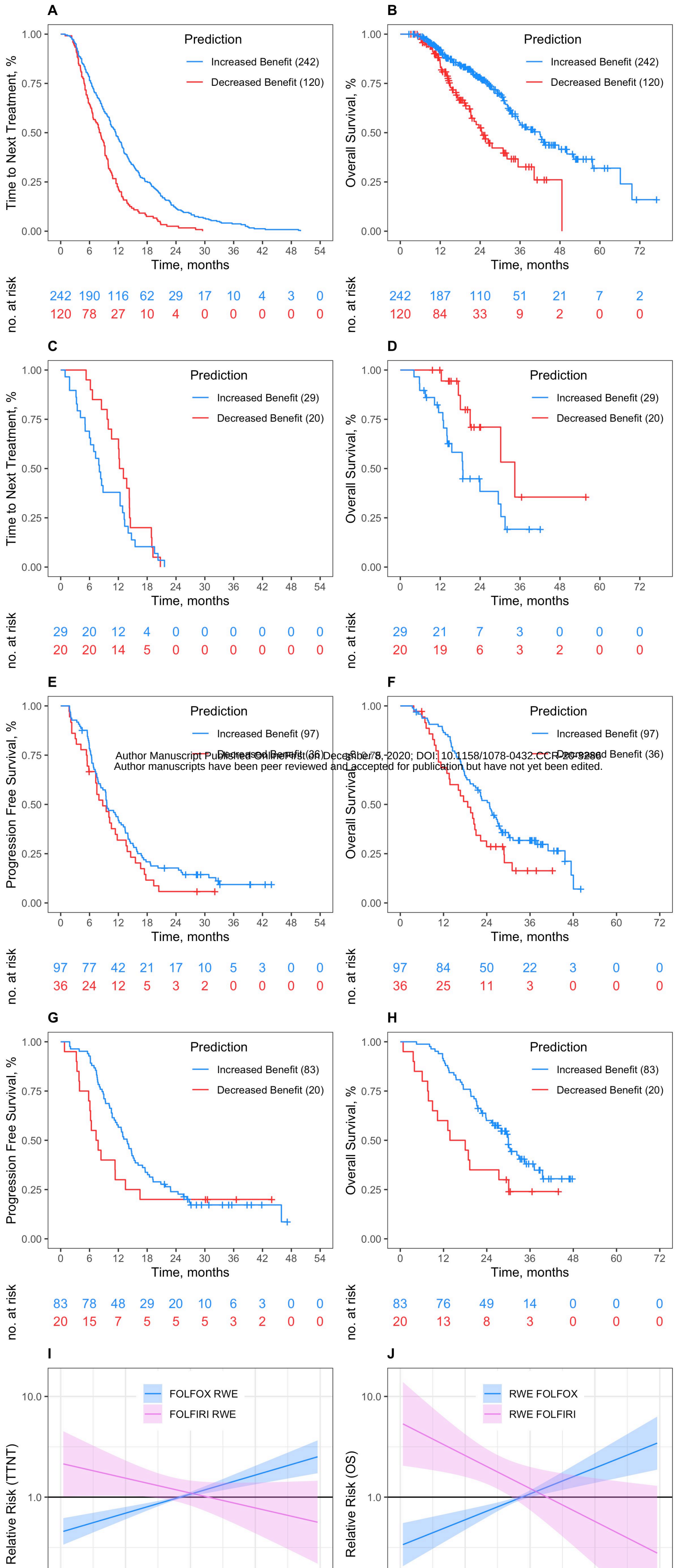
Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

FIGURE 1



Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

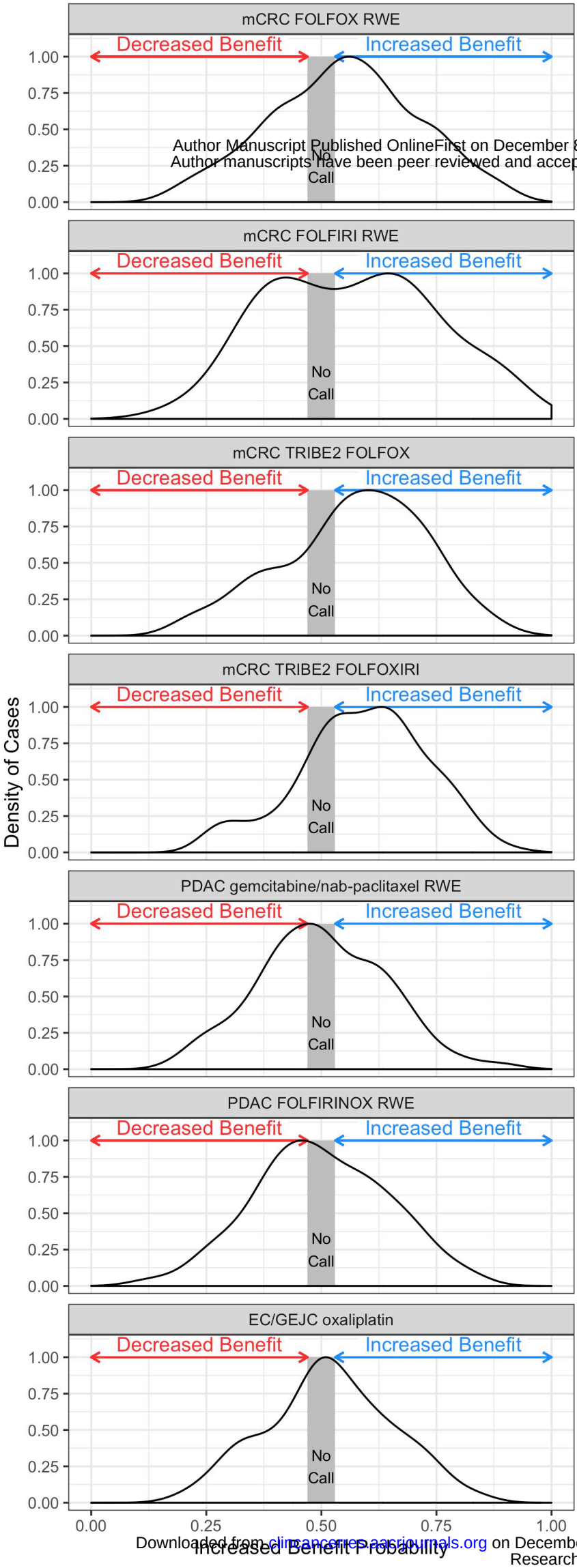
FIGURE 2



Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.

FIGURE 3



Author Manuscript Published OnlineFirst on December 8,
Author manuscripts have been peer reviewed and accept
Downloaded from clincancerres.aacrjournals.org on Decembe
Research.

Author Manuscript Published OnlineFirst on December 8, 2020; DOI: 10.1158/1078-0432.CCR-20-3286
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

American Association for Cancer Research

Clinical Cancer Research

Clinical validation of a machine-learning derived signature predictive of outcomes from first-line oxaliplatin-based chemotherapy in advanced colorectal cancer

Jim P Abraham, Daniel Magee, Chiara Cremolini, et al.

Clin Cancer Res Published OnlineFirst December 8, 2020.

Updated version	Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-20-3286
Supplementary Material	Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2020/12/05/1078-0432.CCR-20-3286.DC1
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/early/2020/12/05/1078-0432.CCR-20-3286. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.

Downloaded from clincancerres.aacrjournals.org on December 18, 2020. © 2020 American Association for Cancer Research.