Predictive Polygenic Score for Outcome after First-Line Oxaliplatin-Based Chemotherapy in Colorectal Cancer Patients Using Supervised Principal Component Analysis

Hanla A. Park[1,2], Dominic Edelmann[3], Federico Canzian[4], Tabitha A. Harrison[5], Xinwei Hua[6,7,8], Qian Shi[9], Allison Silverman[10], Martin Schneider[11], Richard M. Goldberg[12], Steven R. Alberts[13], Michael Hoffmeister[14], Hermann Brenner[14,15,16], Andrew T. Chan[17,18,19], Ulrike Peters[5,20], Polly A. Newcomb[5,20], and Jenny Chang-Claude[1,21]

ABSTRACT

Background: Associations between candidate germline genetic variants and treatment outcome of oxaliplatin, a drug commonly used for patients with colorectal cancer, have been reported but not robustly established. This study aimed to construct polygenic hazard scores (PHSs) as predictive markers for oxaliplatin treatment outcome by using a supervised principal component approach (PCA).

Methods: Genome-wide association analysis for overall survival, including interaction terms (SNP*treatment type) was carried out using two phase III trials, 3,098 resected stage III colon cancer (rCC) patients of NCCTG N0147 and 506 metastatic colorectal cancer (mCRC) patients of NCCTG N9741, separately. SNPs showing interaction with genome-wide significance ($P < 5 \times 10^{-8}$) were selected for PCA to derive a PHS. PHS interaction with treatment was included in Cox regression models to predict outcome. Replication of

prediction models was performed in an independent cohort, DACHS.

Results: The two PHSs based on the first two principal components of selected SNPs (15SNPs for rCC and 13SNPs for mCRC) were used to construct interaction terms with treatment type and included in models adjusted for clinical covariables. However, in the DACHS study, the addition of the two PHS terms to clinical models did not improve the prediction error in either patients with rCC or mCRC. PHS interaction was also not replicated.

Conclusions: The PHSs derived using principal components efficiently combined multiple predictive SNPs for estimating likelihood of benefit from oxaliplatin versus other treatment but could not be replicated.

Impact: These results highlight the potential but also challenges in generating evidence for a predictive polygenic score for oxaliplatin efficacy.

Introduction

Oxaliplatin is a platinum drug often given in combination with other anticancer drugs to treat colorectal cancer. None of the previously reported candidate genetic polymorphism with oxaliplatin treatment outcomes has been robustly validated (1). Applying a machine-learning approach on 592-gene-sequencing data, a recent study derived a 67-gene signature predictive of oxaliplatin-based treatment efficacy in metastatic colorectal cancer (mCRC; ref. 2).

However, no germline genetic variants are currently available as robust predictive markers, i.e., differential association with outcome according to the type of chemotherapy (e.g., with oxaliplatin versus without oxaliplatin) for clinical practice. To efficiently extract the information value of multiple genetic variants showing significant association, we used a supervised principal components approach (3) to construct polygenic hazard scores (PHS) for differential outcomes in colorectal cancer patients receiving oxaliplatin-containing chemotherapy versus others.

[1]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [2]Medical Faculty, University of Heidelberg, Heidelberg, Germany. [3]Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [4]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. [5]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington. [6]Department of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts. [7]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, Massachusetts. [8]Department of Cardiology, Peking University Third Hospital, Peking University, Beijing, China. [9]Department of Quantitative Science, Mayo Clinic, Rochester, Minnesota. [10]Epidemiology Program, Fred Hutchinson Research Cancer Research Center, Seattle, Washington. [11]Department of General, Visceral and Transplantation Surgery, University of Heidelberg, Heidelberg, Germany. [12]West Virginia University Cancer Institute, Morgantown, West Virginia. [13]Medical Oncology, Mayo Clinic, Rochester, Minnesota. [14]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [15]Division of Preventive Oncology, National Center for Tumor Diseases (NCT) and German Cancer Research Center (DKFZ), Heidelberg, Germany. [16]German Cancer Consortium (DKTK),

German Cancer Research Center (DKFZ), Heidelberg, Germany. [17]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts. [18]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. [19]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts. [20]School of Public Health, University of Washington, Seattle, Washington. [21]Cancer Epidemiology Group, University Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

Corresponding Author: Jenny Chang-Claude, Unit of Genetic Epidemiology, Division of Cancer Epidemiology German Cancer Research Center (DKFZ) Im Neuenheimer Feld 581, Heidelberg 69120, Germany. Phone: 4962-2142-2373; Fax: 4962-2142-2203; E-mail: j.chang-claude@dkfz.de

Materials and Methods

Details of all studies and inclusion criteria are found in Supplementary Methods. Two phase III trials served as discovery cohorts, including 3,098 resected stage III colon cancer (rCC) patients of N0147, 96% of patients received oxaliplatin-containing regime (4), and 506 mCRC patients of N9741, 79% of patients received oxaliplatin-containing regime (ref. 5; Supplementary Fig. S1). Both studies were led by the North Central Cancer Treatment Group (NCCTG). NCCTG is now part of the Alliance for Clinical Trials in Oncology. All participants provided written informed consent and the studies were approved by the institutional review board of all participating centers in accordance with Declaration of Helsinki.

All analyses were conducted separately for rCC and mCRC. First, genome-wide association analysis for overall survival (OS) was conducted for each SNP. Cox models included an interaction term of SNP with treatment type (predictive effect). Models were adjusted for age, sex, grade (only for NCCTG N0147), KRAS mutation, and the first three principal components (PC) of all tested SNPs. Manhattan plots were used to illustrate the P values for 4,695,046 imputed/genotyped SNP interactions with type of treatment across chromosomes (Supplementary Fig. S2). Predictive SNPs with genome-wide significance ($P < 5 \times 10^{-8}$) were used for principal component analyses (PCA) to derive PHSs separately for rCC and mCRC. The PHSs (composed of selected PCs), treatment type, and interaction terms of each PHS with treatment type were added to clinical models [including age, sex, grade (only for patients rCC), and KRAS mutation] to predict outcome.

We attempted validation of the prediction models based on the discovery cohort using patients of the population-based DACHS study (6, 7), separately for 549 patients with rCC and 437 patients with mCRC (Supplementary Fig. S1). The study was approved by ethics committees of the Medical Faculty of the University of Heidelberg and the State Medical Boards of Baden-Wuerttemberg and Rhineland-Palatinate. The prediction error curve for OS comparing the clinical models and prediction models was evaluated using the loss function approach (8) in the R package "pec".

Data availability

The data from NCCTG N0147 and NCCTG N9741 analyzed in this study are available from dbGaP (phs001290.v1.p1). The data from DACHS study are available on reasonable application from the corresponding author.

Results

Selected characteristics of patients from participating studies according to the type of chemotherapy are presented in Supplementary Table S1. In the discovery cohorts, GWAS yielded 15 significant SNPs for rCC and 13 SNPs for mCRC. PCA of these predictive SNPs revealed a linear combination of the first two PCs as sufficient for an appropriate presentation of their information value, with corresponding weights given in **Table 1**. The PHSs, that is, first two PCs of the selected SNPs, were used to construct interaction terms with treatment type. In the discovery cohorts, PHS showed significant interaction with treatment,

Table 1. Characteristics of the selected SNPs included into PHS.

Variation rsID	Chr: Position (hg19)	Overlapped Gene	Annotation	Nearest gene	1st eigenvector weight	2nd eigenvector weight
Resected stage III colon cancer						
rs35498763	6:104040991	–	–	SNORA33, R3HDM2P2	2.6×10^{-2}	0.50
rs34110997	6:104043890	–	–	SNORA33, R3HDM2P2	2.6×10^{-2}	0.50
rs35185174	6:104060351	–	–	SNORA33, R3HDM2P2	2.7×10^{-2}	0.50
rs71568154	6:104092576	–	–	SNORA33, R3HDM2P2	2.6×10^{-2}	0.50
rs2893355	7:29804143	–	–	DPY19L2P3, WIPF3	1.0×10^{-2}	-9.06×10^{-3}
rs72733293	9:78627731	PCSK5	Intronic	–	-1.72×10^{-3}	5.82×10^{-3}
rs7286060	22:25154776	PIWIL3	Intronic, noncoding intronic	–	−0.34	1.1×10^{-2}
rs8137801	22:25157455	PIWIL3	Intronic, noncoding intronic	–	−0.34	1.5×10^{-2}
rs11912167	22:25161052	PIWIL3	Intronic, noncoding intronic	–	−0.34	1.5×10^{-2}
rs76939478	22:25161635	PIWIL3	Intronic, noncoding intronic	–	−0.34	1.6×10^{-2}
rs78096049	22:25163992	PIWIL3	Intronic, noncoding intronic	–	−0.34	1.7×10^{-2}
rs112693385	22:25164291	PIWIL3	Intronic, noncoding intronic	–	−0.32	2.1×10^{-2}
rs111319837	22:25164498	PIWIL3	Intronic, noncoding intronic	–	−0.32	2.1×10^{-2}
rs75413996	22:25164604	PIWIL3	Intronic, noncoding intronic	–	−0.32	2.1×10^{-2}
rs61396473	22:25560990	KIAA1671	Intronic, noncoding intronic	–	−0.31	2.2×10^{-2}
mCRC						
rs17406943	1:56918398	–	–	RP4-710M16.2, PPAP2B	8.3×10^{-2}	−0.49
rs56027745	1:56922419	–	–	RP4-710M16.2, PPAP2B	8.3×10^{-2}	−0.49
rs74457219	1:56925869	–	–	RP4-710M16.2, PPAP2B	8.3×10^{-2}	−0.49
rs72662283	1:56926659	–	–	RP4-710M16.2, PPAP2B	8.3×10^{-2}	−0.49
rs115226504	2:2182286	MYT1 L	Intronic, noncoding intronic	–	2.1×10^{-2}	9.0×10^{-3}
rs75318197	2:2186642	MYT1L	Intronic, noncoding intronic	–	2.1×10^{-2}	9.0×10^{-3}
rs112014744	7:46713850	AC011294.3	Noncoding intronic (splice_site)	–	0.37	6.6×10^{-2}
rs10254077	7:46723871	AC011294.3	Noncoding intronic	–	0.38	6.6×10^{-2}
rs10282041	7:46735867	AC011294.3	Intronic, noncoding intronic	–	0.38	6.5×10^{-2}
rs11767153	7:46774786	AC011294.3	Noncoding intronic	–	0.38	5.8×10^{-2}
rs74362872	7:46788622	AC011294.3	Noncoding intronic	–	0.37	6.1×10^{-2}
rs75279668	7:46791193	AC011294.3	Noncoding intronic	–	0.37	6.1×10^{-2}
rs76460164	7:46791195	AC011294.3	Noncoding intronic	–	0.37	6.1×10^{-2}

Abbreviation: PC, principle component.

Table 2. Estimates of clinical model with/without the PHS for overall mortality in patients with resected stage III colon cancer and mCRC in discovery and replication cohorts.

| | Discovery cohort | | | | Replication cohort | | | |
| | Model with clinical covariates | | Model with clinical covariates and PHS | | Model with clinical covariates | | Model with clinical covariates and PHS | |
Output	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
In resected stage III colon patients								
Age (per 5-years)	1.09 (1.05–1.13)	6.1×10^{-6}	1.09 (1.05–1.13)	3.4×10^{-6}	1.15 (1.06–1.25)	1.2×10^{-3}	1.15 (1.06–1.26)	1.0×10^{-3}
Gender (female)	0.83 (0.71–0.96)	2.0×10^{-6}	0.83 (0.71–0.97)	2.0×10^{-2}	1.06 (0.78–1.43)	0.70	1.05 (0.78–1.42)	0.73
Grade (3–4)	1.60 (1.36–1.88)	2.1×10^{-8}	1.63 (1.38–1.92)	5.2×10^{-9}	1.46 (1.07–1.97)	1.5×10^{-2}	1.44 (0.06–1.96)	0.02
KRAS (Wild type)	1.23 (0.92–1.64)	0.16	1.25 (0.93–1.67)	0.14	0.93 (0.63–1.37)	0.72	0.93 (0.63–1.37)	0.72
Treatment (Oxaliplatin-based)	1.05 (0.74–1.48)	0.79	1.13 (0.76–1.68)	0.53	0.64 (0.44–0.91)	1.4×10^{-2}	0.71 (0.48–1.07)	0.10
PHS:								
PC1	–	–	0.50 (0.39–0.64)	1.4×10^{-8}	–	–	0.96 (0.82–1.13)	0.63
PC2	–	–	2.34 (1.78–3.08)	1.3×10^{-9}	–	–	0.99 (0.82–1.18)	0.88
Treatment*PHS:	–	–			–	–		
PC1			2.02 (1.57–2.59)	4.2×10^{-8}			1.22 (0.89–1.69)	0.22
PC2	–	–	0.44 (0.33–0.58)	1.5×10^{-8}	–	–	0.93 (0.65–1.32)	0.68
In mCRC patients								
Age (per 5-years)	1.01 (0.97–1.05)	0.58	1.01 (0.97–1.05)	0.66	1.04 (0.99–1.09)	0.13	1.04 (0.99–1.09)	0.15
Gender (Female)	1.02 (0.85–1.23)	0.82	0.99 (0.82–1.19)	0.91	1.08 (0.88–1.33)	0.44	1.09 (0.89–1.34)	0.41
Treatment (Oxaliplatin-based)	0.70 (0.56–0.88)	2.0×10^{-3}	0.46 (0.36–0.58)	1.8×10^{-10}	0.80 (0.65–1.00)	0.05	0.81 (0.63–1.03)	0.08
PHS:								
PC1	–	–	3.83 (2.58–5.69)	3.1×10^{-11}	–	–	0.96 (0.84–1.10)	0.58
PC2	–	–	0.69 (0.54–0.88)	3.2×10^{-3}	–	–	1.03 (0.90–1.18)	0.69
Treatment*PHS:	–	–			–	–		
PC1			0.24 (0.16–0.37)	2.0×10^{-11}			0.94 (0.64–1.37)	0.74
PC2	–	–	1.84 (1.38–2.46)	4.0×10^{-5}	–	–	1.02 (0.76–1.37)	0.90

Abbreviation: PC, principle component.

as expected, because the SNPs were selected for interactions using the same dataset.

The prediction models (model parameters in **Table 2**) were not validated in DACHS. The prediction error of the 5-year OS remained unchanged after adding the PHS terms to the clinical models (0.173 in resected stage III colon patients; 0.145 in mCRC patients) (**Fig. 1**).

In DACHS, we also failed to replicate the two PHSs and the interaction with treatment in Cox models for OS (**Table 2**). Unadjusted Kaplan–Meier curves also showed no difference in OS for patients according to PHSs (Supplementary Fig. S3).

Discussion

We used an unconventional method to construct predictive PHSs for oxaliplatin treatment outcome for rCC and for mCRC based on genome-wide interaction analysis. However, neither the prediction models nor the PHSs was validated in the independent replication cohort. We cannot rule out that the findings in the discovery cohort were simply due to chance or type I error. The meager discriminative ability of PHS compared to clinical factors may have remained undetected in the replication cohort due to limited power.

The principal components approach may have limitations when applied to highly correlated SNP data. Therefore, we constructed the classical polygenic score using thresholding and pruning for comparison; however, the prediction did not improve.

Limitations of the study should be considered when interpreting the results. Discovery cohorts, NCCTG N0147 and NCCTG N9741, were not originally designed to evaluate the efficacy of oxaliplatin. Patients were defined as having received oxaliplatin-based treat-

ment if they were assigned regimens including oxaliplatin. Otherwise, they were considered to have received the other treatment without oxaliplatin. For that reason, only 4% of the patients from the NCCTG N0147 study did not receive the oxaliplatin-based regime, which limits the statistical power to detect interaction with treatment. The patients who did not receive oxaliplatin showed longer OS compared to the patients who received oxaliplatin (Supplementary Table S1). In addition, we were unable to examine tumor response to oxaliplatin using metrics based on tumor size, like early tumor shrinkage or depth of response, which have been shown to be clinical outcomes in metastatic colorectal cancer. OS is, however, one of the widely used end-points in prospective studies to predict long-term survival.

Strengths of this study include assessing predictive markers (interaction with treatment). Considering that oxaliplatin is not used alone but in association with other chemotherapeutic drugs, prognostic markers from patients who received a combination of oxaliplatin and the other drugs provide limited information on oxaliplatin specifically as a treatment option. Another strength of this study was testing a novel approach to generate PHSs that efficiently combined information from multiple predictive SNPs, thus avoiding the necessity of multiple interaction terms in prediction models for estimating the likelihood of benefit from a particular treatment versus another. Lastly, we accounted for all relevant clinical variables to address the research question on integrating PHS to improve existing clinical models. All analyses were stratified for tumor stage to account for heterogeneity in patient characteristics and corresponding different clinical outcome.

In conclusion, despite efficiently combining multiple SNPs to predict individual benefit from oxaliplatin-containing treatment the predictive PHSs failed to replicate.
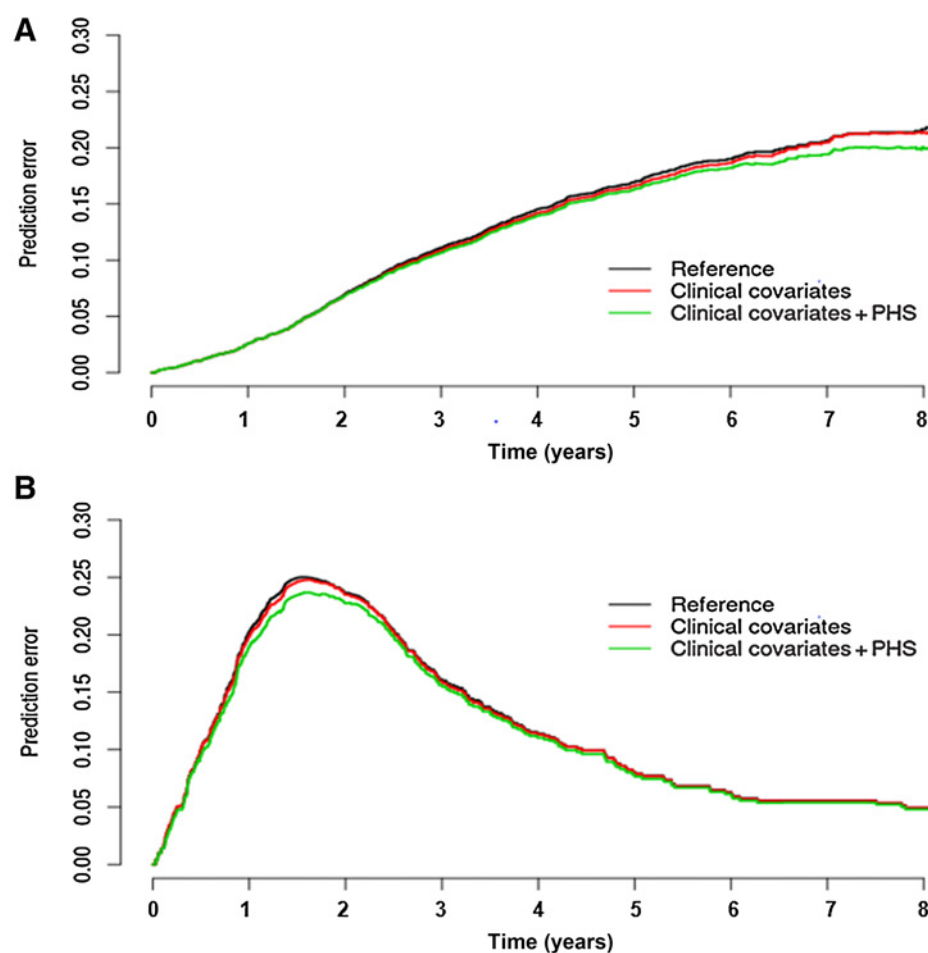
Figure 1.
Prediction error curve for overall survival in resected stage III colon patients (**A**), and in metastatic colorectal cancer patients (**B**) in the replication cohort, DACHS study. Red line represents clinical model, and green line represents prediction model after adding the PHS terms to the clinical models).

Authors' Disclosures

Authors' Contributions

H.A. Park: Formal analysis, validation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **D. Edelmann:** Methodology, writing–review and editing. **F. Canzian:** Writing–review and editing. **T.A. Harrison:** Data curation, writing–review and editing. **X. Hua:** Data curation, writing–review and editing. **Q. Shi:** Data curation, writing–review and editing. **A. Silverman:** Data curation. **M. Schneider:** Writing–review and editing. **R.M. Goldberg:** Data curation, writing–review and editing. **S.R. Alberts:** Data curation, writing–review and editing. **M. Hoffmeister:** Data curation, funding acquisition. **H. Brenner:** Data curation, funding acquisition, writing–review and editing. **A.T. Chan:** Data curation. **U. Peters:** Data curation, writing–review and editing. **P.A. Newcomb:** Data curation, writing–review and editing. **J. Chang-Claude:** Conceptualization, data curation, supervision, funding acquisition, methodology, writing–review and editing.

Acknowledgments

Note

References

1. Park HA, Seibold P, Edelmann D, Benner A, Canzian F, Alwers E, et al. Validation of genetic markers associated with survival in colorectal cancer patients treated with oxaliplatin-based chemotherapy. Cancer Epidemiol Biomarkers Prev 2022; 31:352–61.
2. Abraham JP, Magee D, Cremolini C, Antoniotti C, Halbert DD, Xiu J, et al. Clinical validation of a machine-learning-derived signature predictive of outcomes from first-line oxaliplatin-based chemotherapy in advanced colorectal cancer. Clin Cancer Res 2021;27:1174–83.
3. Bair E, Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. PLoS Biol 2004;2:E108.
4. Alberts SR, Sargent DJ, Nair S, Mahoney MR, Mooney M, Thibodeau SN, et al. Effect of oxaliplatin, fluorouracil, and leucovorin with or without cetuximab on survival among patients with resected stage III colon cancer: a randomized trial. JAMA 2012;307:1383–93.
5. Goldberg RM, Sargent DJ, Morton RF, Fuchs CS, Ramanathan RK, Williamson SK, et al. A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. J Clin Oncol 2004; 22:23–30.
6. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M. Protection from colorectal cancer after colonoscopy: a population-based, case-control study. Ann Intern Med 2011;154:22–30.
7. Brenner H, Chang-Claude J, Jansen L, Knebel P, Stock C, Hoffmeister M. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. Gastroenterology 2014;146:709–17.
8. Gerds TA, Schumacher M. Consistent estimation of the expected Brier score in general survival models with right-censored event times. Biom J 2006;48: 1029–40.