Title: Discovery and features of an alkylating signature in colorectal cancer

Running title: Alkylating mutational signature and colorectal cancer

Carino Gurjao 1,2+, Rong Zhong 3,4+, Koichiro Haruki 3+, Yvonne Y. Li 1,2, Liam F. Spurr 1,2,5, Henry Lee-Six 6, Brendan Reardon 1,2, Tomotaka Ugai 3,7, Xuehong Zhang 8,9, Andrew D. Cherniack 1,2, Mingyang Song 7,8,9,10,11, Eliezer M. Van Allen 1,2, Jeffrey A. Meyerhardt 1, Jonathan A. Nowak 12, Edward L. Giovannucci 7,8,9, Charles S. Fuchs 13, Kana Wu 9#, Shuji Ogino 2,3,7,12#, Marios Giannakis 1,2#

1. Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA

2. Broad Institute of MIT and Harvard, Cambridge, MA, USA

3. Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

4. Department of Epidemiology and Biostatistics and Ministry of Education Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

5. Pritzker School of Medicine, Biological Sciences Division, The University of Chicago, Chicago, IL, USA

6. Wellcome Sanger Institute, Hinxton, UK

7. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

8. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

9. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA

10. Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

11. Division of Gastroenterology, Massachusetts General Hospital, Boston, MA; USA

12. Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Boston, MA, USA

13. Yale Cancer Center, Yale School of Medicine, Smilow Cancer Hospital, New Haven, CT, USA

+: co-first authors

#: co-senior authors

Corresponding author: Marios Giannakis**,** Dana-Farber Cancer Institute, 450 Brookline Ave., Boston, MA 02215-5450. E-mail: Marios_Giannakis@dfci.harvard.edu, Telephone: 617.582.7263

Competing interests: M.G. receives research funding from Bristol-Myers Squibb, Merck, Servier and Janssen. C.S.F. previously served as a consultant for Agios, Bain Capital, Bayer, Celgene, Dicerna, Five Prime Therapeutics, Gilead Sciences, Eli Lilly, Entrinsic Health, Genentech, KEW, Merck, Merrimack Pharmaceuticals, Pfizer Inc, Sanofi, Taiho, and Unum Therapeutics; C.S.F. also serves as a Director for CytomX Therapeutics and owns unexercised stock options for CytomX and Entrinsic Health. J.A.M. received institutional research funding from Boston Biomedical. J.A.M. has also served as an advisor/consultant to Cota Healthcare and served on an NCCN grant panel funded by Taiho Pharmaceutical. E.M.V. has advisory/ consulting roles at Tango Therapeutics, Genome Medical, Invitae, Enara Bio, Janssen and Manifold Bio, has research support from Novartis and BMS; equity at Tango Therapeutics, Genome Medical, Syapse, Enara Bio, Manifold Bio, Microsoft. E.M.V receives travel reimbursement from Roche/Genentech. E.M.V has Institutional patents filed on chromatin mutations and immunotherapy response, and methods for clinical interpretation.

Abstract

Several risk factors have been established for colorectal carcinoma (CRC), yet their direct mutagenic effects in patients' tumours remain to be elucidated. Here, we leveraged whole-exome sequencing data from 900 CRC cases that had occurred in three US-wide prospective studies with extensive dietary and lifestyle information. We found an alkylating signature which was previously undescribed in CRC, and then showed the existence of a similar mutational process in normal colonic crypts. This alkylating signature is associated with high intakes of processed and unprocessed red

meat prior to diagnosis. Additionally, this signature was more abundant in the distal colorectum, predicted to target cancer driver mutations *KRAS* p.G12D, *KRAS* p.G13D and *PIK3CA* p.E5454K, and associated with poor survival. Together, these results link for the first time a colorectal mutational signature to a component of diet, and further implicate the role of red meat in CRC initiation and progression.

Statement of significance

Colorectal cancer (CRC) has several lifestyle risk factors, but the underlying mutations for most have not been observed directly in tumours. Analysis of 900 CRCs with WES and epidemiological annotations revealed an alkylating mutational signature, which associated with red meat consumption, distal tumour location and predicted to target *KRAS* p.G12D/p.G13D.

Introduction

Most tumour mutations are passengers that have no to little functional role in cancer. However, their positional context in the genome may reveal information about the underlying mutational processes[1]. Snapshots of these processes, called mutational signatures, were originally deconvoluted using a Non-Negative Matrix Factorization approach (NMF)[2] on a large collection of whole-genome and whole-exome sequencing (WES) data[3]. Mutational signatures may elucidate the roles of mutagens in cancer, and inform prevention and treatment efforts. Several studies have been conducted to associate mutational signatures with cellular processes or exposures. These include rare cancer predisposition syndromes[4], environmental agents[5], and microbiota[6]. Such association studies have relied on either DNA sequencing datasets or preclinical models, such as organoids. However, although many lifestyle-related factors have been linked to colorectal cancer[7], larger and more comprehensive datasets are needed to enable the discovery of the associated signatures. Consequently, past efforts have not been able to capture the cumulative effect of putative mutagens, such as dietary components, over decades. In particular, red meat consumption has been consistently

linked to the incidence of colorectal cancer[8–10]. The suggested mechanism is mutagenesis through alkylating damage induced by N-nitroso-compounds (NOCs), which are metabolic products of blood haeme iron or meat nitrites/ nitrates[11]. Nevertheless, this mutational damage is yet to be observed directly in patients' tumors.

Results

Active mutational signatures in colorectal tumours and normal colonic crypts

To address this gap, we leveraged a database of incident colorectal carcinoma (CRC) cases that had occurred in three U.S.-wide prospective cohort studies, namely the Nurses' Health Studies I and II (NHS) and the Health Professionals Follow-up Study (HPFS)[12]. Study participants (more than 230,000 women and 50,000 men) repeatedly provided data on diet, lifestyle, and other factors without knowing their future CRC diagnosis, if any. We performed WES on matched primary untreated tumour-normal pairs in 900 CRC patients with adequate tissue materials (**Fig. 1A, Table S1**).

NMF signal separation revealed the existence of seven mutational processes (See ***Methods*** and **Fig. 1B, Fig 1C and Supplemental Figure 1**). We confirmed the robustness of the deconvolution by using another signature assignment program (SigProfiler[3]); we again found seven mutational processes (**Supplemental Figure 2**, left panel) that are highly similar to the ones obtained using the standard NMF approach (**Supplemental Figure 2**, right panel).

To uncover the aetiology of these colorectal signatures (that we name *c*-signatures), we first used a cosine similarity metric (cossim) to compare the deconvoluted signatures to reference COSMIC Single Base Substitution (SBS) signatures[3]. The seven *de novo* signatures displayed the highest similarity with four known mutational processes (**Supplemental Figure 3**), namely *POLE* deficiency (c-POLEa/SBS10a, cossim = 0.95 and c-POLEb/SBS10b, cossim = 0.86), ageing (c-Age/SBS1, cossim = 0.95), deficient mismatch repair (dMMR) (c-dMMRa/SBS15, cossim = 0.90 and c-dMMRb/SBS26,

cossim = 0.90) and exposure to alkylating agents (c-Alkylation/SBS11, cossim = 0.94). c-SBS40 matched the closest to SBS40 (cossim = 0.84), which is a featureless signature with unknown aetiology, and found in most cancers[3].

We substantiated the aetiology of the four mutational processes by integrating clinical, pathology and methylation data (**Fig. 2A**). Tumours harbouring a *POLE* exonuclease domain mutation were significantly enriched in signatures c-POLEa and c-POLEb (p = 2.3×10-5 and p =1.8×10-6 respectively, Mann-Whitney U test). Similarly, patients with orthogonally assessed microsatellite instability (MSI)-high status were significantly enriched in signatures c-dMMRa and c-dMMRb (p <2×10-16 for both, Mann-Whitney U test). Signature c-Age also displayed a significant association with patients' age at diagnosis (p = 1.7×10-5, Mann-Whitney U test). Lastly, we support the aetiology of the alkylating-like signature -not previously described in CRC- by assessing the *MGMT* (O-6-methylguanine-DNA methyltransferase) promoter methylation status in tumours from the NHS/HPFS cohorts. *MGMT* is a central gene in the repair of alkylating lesions. Among the sequenced specimens with available *MGMT* promoter methylation data, we observed that tumours with methylated *MGMT* promoters were enriched in the signature c-Alkylation (p = 6.6×10-3, Mann-Whitney U test) (**Fig. 2A**), further supporting that this signature represents the biological consequence of increased alkylating damage. Of note, SBS18, which is associated with *MUTYH*-associated polyposis[3] (MAP), is absent in the tumour samples we sequenced. We believe this is the case because of low occurrence of *MUTYH* deficiency generally in CRC (less than 1%[13]), as well as further under sampling of patients with germline predisposition mutations as only healthy individuals were enrolled prospectively in NHS/HPFS.

NMF signal separation in The Cancer Genome Atlas (TCGA) colorectal tumours (n=540) revealed the existence of seven signatures (**Supplemental Figure 4** and **Supplemental Figure 5** for SigProfiler results) similar to the ones found in NHS/ HPFS (**Supplemental Figure 6**), thus suggesting the existence of the same underlying mutational processes in all CRC cohorts. Analysis of the TCGA colorectal tumours (**Fig. 2B**) substantiated the same aetiologies for the POLE signatures c-POLEa and c-POLEb (p = 6.2×10-7 and p = 7.3×10-7 respectively, Mann-Whitney U test) as well as dMMR

signatures c-dMMRa and c-dMMRb ($p < 2\times10\text{-}16$ for both, Mann-Whitney U test). We also observed that TCGA tumours with *MGMT* promoter methylation were enriched in signature c-Alkylation ($p = 9.7\times10\text{-}5$, Mann-Whitney U test). Of note, in TCGA, signature c-Alkylation displayed the highest similarity with SBS30 (cossim = 0.812), followed by SBS11. Conversely, SBS30 was the second most similar signature to the c-Alkylation one in the NHS/HPFS cohorts (**Fig. 2C** and **Supplemental Figure 3**). SBS30 resembles SBS11 (cossim of 0.76, **Fig. 2C**) and is attributed to base excision repair deficiency[3], which is also a central pathway in repairing damage from alkylated bases. We nevertheless found no association between germline polymorphisms in *NHTL1* and other genes of the Base Excision Repair (BER) pathway and the alkylating signature in the TCGA specimens (See ***Methods*** and **Supplemental Figure 7**). The presence of SBS30 ahead of SBS11 in the TCGA CRC dataset could instead be attributed to a smaller sample size of CRCs in TCGA compared to NHS/HPFS (See "Under sampling simulations" in ***Methods*** and **Supplemental Figure 8**). The Fanconi Anemia (FA) and Translesion Synthesis (TLS) DNA damage repair pathways, also do not show an association with the alkylating signature (See ***Methods*** and **Supplemental Figure 9A and 9B**).

We also estimated the effect size for the Mann-Whitney U tests by calculating the rank-biserial correlation r_{rb} for each mutational signature and the respective molecular or clinical phenotype shown in **Fig. 2B**. We observed that the effect sizes were similar for the alkylating signatures and the ageing signature ($r_{rb} = 0.14$ and $r_{rb} = 0.16$ respectively), and smaller than the hypermutator dMMR and POLE signatures ($r_{rb} > 0.8$ for dMMR and POLE signatures in both TCGA and NHS/ HPFS).

Interestingly, a previously published survey of mutational signatures in normal colorectal crypts[14] from the European Genome-phenome Archive (EGA) showed the existence of a signature (named SBSC) that we found to be similar to the alkylating one that we observed in NHS/HPFS CRCs (cossim = 0.85). Of note, SBSC matched closely to SBS23 which, similarly to SBS30, also resembles SBS11 (cossim of 0.77, **Fig. 2C**). The hierarchical clustering of the SBSC with the seven signatures deconvoluted from NHS/

HPFS and TCGA confirmed the similarity of EGA SBSC with the alkylating imprints (**Fig. 2C**).

Dietary patterns of alkylation damage

To test whether dietary components contributed to the alkylating signature in CRC, we leveraged prospectively collected repeated measurements of meat, poultry, and fish consumption in grams per day in the NHS and HPFS cohorts. All available red meat variables showed significant positive associations between pre-diagnosis intakes and alkylating damage in CRCs (**Fig. 3A**, overall red meat: $p = 0.017$/ $r_{rb} = 0.14$; unprocessed red meat: $p = 7.8\times10\text{-}3$/ $r_{rb} = 0.16$; and processed red meat $p = 7.3\times10\text{-}3$/ $r_{rb} = 0.16$, Mann-Whitney U test). Other dietary variables (fish and chicken intake, **Fig. 3B**) and lifestyle factors (body-mass index, alcohol consumption, smoking and physical activity in **Supplemental Figure 10**) did not show any significant association with the alkylating signature. In addition, no other CRC mutational process showed a significant association with red meat intake **(Supplemental Figure 11**). Of note, *MGMT* promoter methylation did not differ by red meat consumption (two-sided Mann-Whitney U test p= 0.51, **Supplemental Figure 12**). When adjusted for red meat intake there was no difference in alkylating damage between male and female CRC patients (two-sided Mann Whitney U test $p = 0.27$ for patients with high overall red meat consumption).

Previous studies[9,10] showed a positive association between processed red meat and CRC incidence in the distal colon. Thus, we also investigated how the alkylating damage might differ by tumour location. We found that, compared to the proximal colon, the distal colorectal specimens exhibited higher alkylating damage in tumours ($p = 1.4\times10\text{-}4$ in NHS/HPFS and $p = 1.9\times10\text{-}8$ in TCGA, Mann-Whitney U test) and normal crypts ($p = 0.022$, Mann-Whitney U test) (**Fig. 3B**).

Carcinogenicity of alkylation damage

Mutational processes increase the likelihood of specific driver mutations in certain trinucleotide contexts. To find such driver mutations that associate with the alkylating signature, we devised a simple model (**Fig. 4A**, see ***Methods***) that predicts the relative likelihood of mutational processes to target CRC recurrent drivers in non-MSI-high, non-*POLE*-mutated tumours.

In particular, the alkylating signature appeared to be the dominant one that targets *KRAS* p.G12D (relative likelihood = 1) and p.G13D (relative likelihood = 0.91) (**Fig. 4A**). This is due to p.G12D and p.G13D being in trinucleotide contexts (ACC>ATC and GCC>GTC respectively) mainly targeted by the alkylating signature. *PIK3CA* p.E545K (TCA>TTA) is also predicted to be predominantly targeted by the alkylating signature (relative likelihood = 87%). Supporting this, we showed that CRCs having either *KRAS* p.G12D, *KRAS* p.G13D or *PIK3CA* p.E545K mutant CRCs were enriched with the alkylating signature compared to all other tumours (**Fig. 4B**, p = 0.013, Mann-Whitney U test).

Lastly, we examined patient survival across ordinal alkylating mutational signature quartiles and found that patients whose tumours have high alkylation damage (top quartile) had a worse CRC-specific survival (log-rank test p_{trend} = 0.036, **Fig. 4C**, Supplementary Table S2 and S3). Furthermore, higher alkylating signature contribution was associated with worse CRC-specific survival, in both univariable and multivariable Cox proportional hazards regression analyses (p_{trend} = 0.015 and p_{trend} = 0.036 respectively, **Fig.4D** and **Supplementary Table S3**).

Discussion

Our work demonstrated the presence of a novel alkylating mutational signature, which we deconvoluted directly from WES of colorectal tumours. Interestingly, this signature is highly similar to SBS11, which was originally discovered in patients with prior exposure to temozolomide[1]. Temozolomide is an alkylating agent used as a treatment of brain

gliomas with *MGMT* promoter methylation[1], and induces the same lesions as dietary N-nitroso-compounds (NOCs), and in the same proportions[15,16] (80% of N7-methylguanine and N3-methylguanine, and 10% of O6-methylguanine).

Previous attempts have shown the existence of alkylating lesions in normal colorectal mucosa, notably caused by NOCs[17]. The latter can be formed endogenously after nitrosylation of haeme iron from blood[17,18], but have also been associated with red meat intake in a small cohort of subjects[19]. However, these previous studies were based on limited datasets (small sample sizes and/ or use of laboratory methylating agents) and lack comprehensive sequencing that would enable the discovery of the full mutational spectrum induced by red meat. Crucially, past efforts have focused on normal colorectal tissues and not examined CRC. Our analysis reveals the existence of an alkylating signature in CRC, which is associated with high pre-diagnosis intake of processed and unprocessed red meat.

Earlier work also hypothesized that the distal colon has increased DNA damage from exposure to dietary carcinogens, as a result of faeces storage and water resorption in this portion of the large intestine[20]. This is believed to explain the association observed between distal cancer incidence and red meat consumption[9,10,20]. Consistently, we found an enrichment in tumours and normal crypts in the distal colon and rectum.

In support of the International Agency for Research on Cancer (IARC) Monograph Working Group which classified processed meat as carcinogenic[8], our results provide molecular evidence of this dietary factor's mutagenic impact. In addition, our analyses further implicate unprocessed meat intake and suggest *MGMT* as a factor of susceptibility to red meat induced damage. The existence of a similar alkylating signature in normal colorectal crypts also suggests that mutational changes due to such damage may start to occur early in the path of colorectal carcinogenesis.

Our analysis predicted *KRAS* p.G12D, p.G13D and *PIK3CA* p.E545K to be mainly targeted by the alkylating signature in non-hypermutated CRCs. We showed that there was indeed higher alkylating damage in tumours harbouring these driver mutations.

Independent epidemiological analyses have also shown a positive association between high consumption of red meat products and *KRAS* p.G12D and p.G13D[21,22]. Although the number of mutations due to alkylation damage was lower than other mutational processes, we showed that alkylation might have considerable carcinogenic potential, by targeting driver mutations in KRAS and PIK3CA. We also demonstrated a significantly worse survival for patients with high levels of alkylation signature contribution.

Our study has leveraged a comprehensive dataset with repeated dietary measures over years, without patients knowing their upcoming CRC diagnosis, and WES on a large collection of colorectal tumours. It provides unique evidence supporting the direct impact of dietary behaviours on colorectal carcinogenesis. Moreover, the presence of a similar alkylating signature in normal mucosa advocates for the utility of early dietary interventions and suggests potential precision prevention approaches in *MGMT* methylated pre-malignant tissue. Similarly, the association of the signature with cancer driver mutations -such as *KRAS* and *PIK3CA* ones- may offer future potential therapeutic opportunities. More generally, our study exemplifies the potential role of large-scale molecular epidemiologic studies in elucidating cancer pathogenesis[23] and guiding prevention efforts through lifestyle modifications, such as dietary interventions.

Methods

Study population, specimens and sequencing

We utilised data from three prospective cohort studies in the U.S., the Nurses' Health Study I (NHS1, including 121,701 women aged 30-55 years at enrollment who had been followed since 1976), the Nurses' Health Study II (NHS2, including 116,429 women aged 25-42 years followed since 1989), and the Health Professionals Follow-up Study (HPFS, including 51,529 men aged 40-75 years followed since 1986)[12]. The study participants have been sent questionnaires biennially to update information on lifestyle factors and newly-diagnosed diseases including colorectal carcinoma. The follow-up rate had been more than 90% for each follow-up questionnaire cycle in the three cohort studies. The patients were followed until death or end of follow-up (January 1, 2016 for HPFS; June 1, 2016 for NHS1; June 1, 2015 for NHS2), whichever came first. Study physicians, who were blinded to exposure data, reviewed medical records of 4855 incident CRC cases to confirm the disease diagnosis and to collect data on tumour size, tumour anatomical location, and disease stage. Archival formalin-fixed paraffin-embedded (FFPE) tissue blocks of tumour and normal colon were collected in a subset of CRC. We previously showed that in our cohorts demographic features of cases did not differ appreciably by tissue availability[24]. The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health (Boston, MA), and those of participating registries as required. Written informed consent was obtained from all CRC subjects.

We prioritized relatively more recent CRC cases for sequencing in order to mitigate the potential impact of FFPE artefacts. Given the number of NHS versus HPFS participants (2:1 female/male ratio) we also sequenced relatively more specimens from male patients to obtain more balanced sequencing data. **Supplementary Table S4** shows the clinical and pathological characteristics of the 4855 CRC patients.

Whole exome sequencing was carried as previously described[25]. Briefly, using guide hematoxylin-and-eosin-stained slides, tumour areas were selected to extract tumour-

enriched DNA from tissue sections of tumour FFPE blocks. Normal DNA was extracted from resection margins or other areas free from tumours. DNA specimens underwent hybrid capture with SureSelect v.2 Exome bait (Agilent Technologies), followed by sequencing on Illumina HiSeq 2000 instruments. The obtained average coverage was 85x in tumours and matched adjacent normal colon tissue (see **Supplementary Table S5**).

Dietary variables

Ascertainment of diet was carried out as previously described[9]. To assess dietary intake in each cohort, food frequency questionnaires (FFQs) were initially collected in 1980 for NHS, and in 1986 for HPFS. For the NHS, a 61-item semi-quantitative FFQ was used at baseline[26], which was expanded to approximately 130 food and beverage items in 1984, 1986, and every four years thereafter. For the HPFS cohorts, baseline dietary intake was assessed using a 131-item FFQ that was also used for updates generally every four years subsequently[27]. In particular, unprocessed red meat consumption was evaluated based on forms on the intake of “beef or lamb as main dish,” “pork as main dish,” “hamburger,” and “beef, pork, or lamb as a sandwich or mixed dish.” Processed meat diets included “bacon,” “beef or pork hot dogs,” “salami, bologna, or other processed meat sandwiches,” and “other processed red meats such as sausage, kielbasa, etc.”. Consumption of red meat, chicken, poultry, and fish was evaluated in grams per day. For the remainder of our analysis, we considered the top decile of each variable to determine the “high intake” patients, and considered the rest as “low intake” patients, since only the top decile patients show a substantial difference in overall red meat intake (**Supplemental Figure 13A** and **Supplemental Figure 13B**). Data was based on the most recent pre-diagnosis reported intake for each patient.

MGMT promoter methylation, MSI and *POLE* deficiency status

MGMT promoter methylation analysis in the NHS/HPFS cohorts was carried out using bisulfite conversion and real-time PCR as previously described[28]. MSI status was

evaluated using 10 microsatellite markers (D2S123, D5S346, D17S250, BAT25, BAT26, BAT40, D18S55, D18S56, D18S67, and D18S487) as formerly detailed[12].

POLE deficiency was assessed by sequencing and manual Integrated Genome Viewer (IGV) curation of *POLE* exonuclease domain mutations in hypermutated non-MSI-high tumours (> 400 mutations).

Somatic variants calling

We have utilised the CGA WES Characterisation pipeline (https://github.com/broadinstitute/CGA_Production_Analysis_Pipeline) developed at the Broad Institute of MIT and Harvard to call, filter, and annotate somatic mutations. All analyses were carried out on the human genome build hg19. The pipeline employs the following tools: MuTect[29], ContEst[30], Strelka[31], DeTiN[32], AllelicCapSeg[33], MAFPoNFilter[34], RealignmentFilter, GATK[35], PicardTools. FFPE-specific artefacts are filtered similarly to previous publications[25,36]. Briefly, FFPE artefacts arise from formaldehyde deamination of cytosines resulting in C-to-T transition mutations, which presents itself as an “Orientation bias” (excess of C>T sites in F1R2 read pairs and an excess of G>A in F2R1 read pairs). In the pipeline we used, the “Orientation Bias Filter” tool[37] filters out FFPE-specific artefacts. To further filter spurious SNV calls, we used Burrows-Wheeler Aligner BWA-MEM (http://bio-bwa.sourceforge.net/) to realign sequenced reads associated with the mutations to a set of sequences derived from the human reference assembly. The Panel of Normal was created using normal samples with less than 1% of cross-sample contamination (as evaluated by Contest[30]), and less than 1% of tumour in Normal (as outputted by DeTIN[32]. We illustrate the variant calling pipeline in **Supplemental Figure 14.**

TCGA data analysis

Clinical, methylation, as well as somatic mutation data from the Cancer Genome Atlas (TCGA), were downloaded from the Data Coordination Center (DCC) data portal at https://dcc.icgc.org/releases/current/Projects/COAD-US and https://dcc.icgc.org/releases/current/Projects/READ-US as of March 2020). For

consistency, only WES datasets were used. Altogether, we pooled 540 TCGA patients with somatic mutation data, among which 523 patients also had methylation data.

We evaluate *MGMT* promoter methylation status using the MGMT-STP27 prediction model[38]. In short, two probes (cg12434587 and cg12981137) were used to predict *MGMT* promoter methylation. An M value cutoff of 0.358, which empirically maximised the sum of sensitivity and specificity, was then used to discriminate *MGMT* promoter methylation status (**Supplemental Figure 15**).

Non-negative matrix factorisation

Mutations were deconvoluted into separate signatures based on the number of mutations in each of 96 possible trinucleotide contexts. Deconvolution was carried out with a standard NMF method based on Kullback-Leibler divergence using the "NMF" R package[39]. This method is particularly adapted for mutational signature analysis as recent studies demonstrated[40].

A critical parameter in NMF is the estimation of the rank (i.e. the number of expected mutational signatures). To determine this, we performed quality measures on a range of ranks (n=2 to 10) for the 900 CRC exomes in the NHS/HPFS cohorts. This showed a sharp increase in the cophenetic (i.e. the stability of the NMF classes) and dispersion (i.e. the reproducibility of the class assignments) metrics after rank=7. For this rank, we also observed that the residual sum of squares (RSS) reaches a lower plateau (**Supplemental Figure 1**). A similar rank survey on an independent cohort of 540 CRC exomes from the TCGA (**Supplemental Figure 4**) reveals the same dispersion and cophenetic peaks at rank=7 and a lower plateau RSS. For the rest of the analysis, we consequently used rank=7. We confirmed the robustness of these 7 signatures by running NMF with different variant allele frequencies (VAF) cutoffs (**Supplemental Figure 16**). This demonstrates that the signature discovery is not affected by low VAF mutations, which are more likely to represent sequencing artefacts, such as those due to FFPE preservation.

SigProfiler was ran on NHS/ HPFS and TCGA CRC exomes as previously described[3].

Under sampling simulations

To show that the difference in sample size between TCGA (n = 540) and NHS/ HPFS (n = 900) can explain the presence of SBS30 instead of SBS11 in the former cohort, we (i) Randomly sampled 540 patients of the 900 from NHS/ HPFS (ii) Extracted seven signatures from the 540 patients and found their closest fit among SBS1 (aging signature), SBS10a and SBS10b (POLE signatures), SBS15 and SBS26 (dMMR signatures), SBS11 and SBS30.(iii) Repeated steps (i) and (ii) a hundred times.

Crypt mutational signature analysis

Mutational signatures from normal colonic crypts[14] were used in our analysis. These signatures were extracted from WGS data from 571 crypts from 42 individuals from the European Genome-phenome Archive[14]. Deconvolution was performed using a hierarchical Dirichlet process (HDP) which produces similar results as NMF[14]

Analysis of recurrent hotspot mutations

To compute the relative likelihood of mutational processes to target a specific hotspot, we (i) localised the trinucleotide context of the hotspot (ii) extracted the signatures contribution for the specific trinucleotide context and (iii) normalised the contribution of each signature, such that the sum becomes one. Recurrent hotspots were defined as specific point mutations occurring in at least 25 patients.

TCGA germline polymorphisms analysis

TCGA genotyping data (Affymetrix SNP 6.0 array platform) were used to select germline variants from genes in the BER, FA and TLS pathways extracted from the GSEA database[41,42] (https://www.gsea-msigdb.org/gsea/msigdb/). We imputed autosomal variants for TCGA samples using IMPUTE2[43], with haplotypes of 1000 Genomes Phase 3[44] as reference panel. We used the following criteria to select SNPs with the plink software[45]: (i) imputation confidence score, INFO ≥ 0.4, (ii) minor allele frequency (MAF) ≥ 5%, (iii) SNP missing rate < 5% for best-guessed genotypes at

posterior probability ≥ 0.9 and (iv) Hardy–Weinberg Equilibrium P-value > 1 × 10−6. After imputation, 2041 variants were included in our subsequent analysis. We tested for an additive effect (genotype 0,1,2 as a continuous variable) for each SNP and found no association with the alkylating signature (**Supplemental Figure 7 and Supplemental Figure 9**, FDR adjusted p-value (*q-value*) less than 0.1 for all SNPs tested).

Statistical analysis

We used R version 3.6.2 to perform statistical analyses. Significance for two-group comparisons were evaluated by a one-sided Mann–Whitney U test unless otherwise indicated. $P < 0.05$ was considered statistically significant. For the comparisons of the alkylating signature by age in the NHS/HPFS cohorts and TCGA CRC database, the patients' median age (70 and 67 years respectively) was used as the cutoff.

882 patients with available CRC survival data were subsequently used for survival analyses. Univariable and multivariable-adjusted Cox proportional hazards regression analysis were used to calculate hazard ratio (HR) of colorectal cancer-specific survival and overall survival according to ordinal alkylating mutational signature quartiles (Q1-Q4). The multivariable Cox regression model initially included sex (female vs. male), age at diagnosis (<60, 60-64, 65-69, and ≥70 years), year of diagnosis (1995 or before, 1996-2000, 2001-2005, and 2006-2014), family history of colorectal cancer (present vs. absent), current smoking status (never smoking, past smoking, 1-14 pack-years, 15-24 pack-years, ≥25 pack-years), alcohol consumption (women: 0 to <0.15, 0.15 to <2.0, 2.0 to <7.5, and ≥7.5 g/d; men: 0 to <1, 1 to <6, 6 to <15, and ≥15 g/d), tumour location (proximal colon vs. distal colon vs. rectum), CpG island methylator phenotype (high vs. low/negative)[46], *KRAS* mutation (mutant vs. wild-type)[47], *BRAF* mutation (mutant vs. wild-type)[47], tumour differentiation (well to moderate vs. poor), disease stage (I/II vs. III/IV), microsatellite instability status (MSI-high vs. non-MSI-high)[46], and long-interspersed nucleotide element-1 methylation level (continuous)[48]. A backward elimination with a threshold P of 0.05 was used to select variables for the final models. Cases with missing data were assigned to the majority category of a given categorical covariate to limit the degrees of freedom, except for cases with missing LINE-1

methylation, for which we assigned a separate indicator variable. We confirmed that excluding the cases with missing information in any of the covariates did not substantially alter results.

Data availability

Whole-exome sequence data have been deposited in dbGAP (accession number phs000722). WES quality metrics and a subset of clinical annotations are included in this article. Additional clinical and epidemiology data from the NHS I, NHS II and HPFS can be requested through the NHS/HPFS consortia.

Code availability statement

All analysis scripts are available upon request.

References

1. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.

2. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999;401:788–91.

3. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578:94–101.

4. Grolleman JE, de Voer RM, Elsayed FA, Nielsen M, Weren RDA, Palles C, et al. Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype. Cancer Cell. 2019;35:256–66.e5.

5. Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, et al. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019;177:821–36.e16.

6. Pleguezuelos-Manzano C, Puschhof J, Huber AR, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. Nature. 2020;580(7802):269-273.

7. Platz EA, Willett WC, Colditz GA, Rimm EB, Spiegelman D, Giovannucci E. Proportion of colon cancer risk that might be preventable in a cohort of middle-aged US men. Cancer Causes Control. 2000;11:579–88.

8. Bouvard V, Loomis D, Guyton KZ, Grosse Y, Ghissassi FE, Benbrahim-Tallaa L, et al. Carcinogenicity of consumption of red and processed meat. Lancet Oncol. 2015;16:1599–600.

9. Bernstein AM, Song M, Zhang X, Pan A, Wang M, Fuchs CS, et al. Processed and Unprocessed Red Meat and Risk of Colorectal Cancer: Analysis by Tumor Location and Modification by Time. PLoS One. 2015;10:e0135959.

10. Larsson SC, Rafter J, Holmberg L, Bergkvist L, Wolk A. Red meat consumption and risk of cancers of the proximal colon, distal colon and rectum: the Swedish Mammography Cohort. Int J Cancer. 2005;113:829–34.

11. Bastide NM, Pierre FHF, Corpet DE. Heme iron from meat and risk of colorectal cancer: a meta-analysis and a review of the mechanisms involved. Cancer Prev Res . 2011;4:177–84.

12. Liao X, Lochhead P, Nishihara R, Morikawa T, Kuchiba A, Yamauchi M, et al. Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. N Engl J Med. 2012;367:1596–606.

13. Lubbe SJ, Di Bernardo MC, Chandler IP, Houlston RS. Clinical implications of the colorectal cancer risk associated with MUTYH mutation. J Clin Oncol. 2009;27:3975–80.

14. Lee-Six H, Olafsson S, Ellis P, Osborne RJ, Sanders MA, Moore L, et al. The landscape of somatic mutation in normal colorectal epithelial cells. Nature. 2019;574:532–7.

15. Zhang J, Stevens MFG, Bradshaw TD. Temozolomide: mechanisms of action, repair and resistance. Curr Mol Pharmacol. 2012;5:102–14.

16. Fahrer J, Frisch J, Nagel G, Reißig S, Waisman A, Samson LD, et al. Dose–response of alkylation-induced colorectal carcinogenesis in MGMT-proficient and -deficient mice [Internet]. Toxicology Letters. 2013. page S71. Available from: http://dx.doi.org/10.1016/j.toxlet.2013.05.054

17. Povey AC, Badawi AF, Cooper DP, Hall CN, Harrison KL, Jackson PE, et al. DNA alkylation and repair in the large bowel: animal and human studies. J Nutr. 2002;132:3518S – 3521S.

18. Bingham SA, Day NE, Luben R, Ferrari P, Slimani N, Norat T, et al. Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. Lancet. 2003;361:1496–501.

19. Billson HA, Harrison KL, Lees NP, Hall CN, Margison GP, Povey AC. Dietary variables associated with DNA N7-methylguanine levels and O6-alkylguanine DNA-alkyltransferase activity in human colorectal mucosa. Carcinogenesis. 2009;30:615–20.

20. Chao A, Thun MJ, Connell CJ, McCullough ML, Jacobs EJ, Flanders WD, et al. Meat consumption and risk of colorectal cancer. JAMA. 2005;293:172–82.

21. Brink M, Weijenberg MP, de Goeij AFPM, Roemen GMJM, Lentjes MHFM, de Bruïne AP, et al. Meat consumption and K-ras mutations in sporadic colon and rectal cancer in The Netherlands Cohort Study. Br J Cancer. 2005;92:1310–20.

22. Gilsing AMJ, Fransen F, de Kok TM, Goldbohm AR, Schouten LJ, de Bruïne AP, et al. Dietary heme iron and the risk of colorectal cancer with specific mutations in KRAS and APC. Carcinogenesis. 2013;34:2757–66.

23. Song M, Vogelstein B, Giovannucci EL, Willett WC, Tomasetti C. Cancer prevention: Molecular and epidemiologic consensus. Science. 2018;361:1317–8.

24. Nishihara R, Lochhead P, Kuchiba A, Jung S, Yamauchi M, Liao X, et al. Aspirin use and risk of colorectal cancer according to BRAF mutation status. JAMA. 2013;309:2563–71.

25. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. Cell Rep. 2016;17:1206.

26. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol. 1985;122:51–65.

27. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and Validity of an Expanded Self-Administered Semiquantitative Food Frequency Questionnaire among Male Health Professionals [Internet]. American Journal of Epidemiology. 1992. page 1114–26. Available from: http://dx.doi.org/10.1093/oxfordjournals.aje.a116211

28. Ogino S, Kawasaki T, Brahmandam M, Cantor M, Kirkner GJ, Spiegelman D, et al. Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. J Mol Diagn. 2006;8:209–17.

29. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213–9.

30. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, Getz G. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics. 2011;27:2601–2.

31. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. Bioinformatics. Narnia; 2012;28:1811–7.

32. Taylor-Weiner A, Stewart C, Giordano T, Miller M, Rosenberg M, Macbeth A, et al. DeTiN: overcoming tumor-in-normal contamination. Nat Methods. 2018;15:531–4.

33. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell. 2013;152:714–26.

34. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014;505:495–501.

35. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

36. Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, Marlow S, et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. Nat Med. 2014;20:682–8.

37. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 2013;41:e67.

38. Bady P, Sciuscio D, Diserens A-C, Bloch J, van den Bent MJ, Marosi C, et al. MGMT methylation analysis of glioblastoma on the Infinium methylation BeadChip identifies two distinct CpG regions associated with gene silencing and outcome, yielding a prediction model for comparisons across datasets, tumor grades, and CIMP-status [Internet]. Acta Neuropathologica. 2012. page 547–60. Available from: http://dx.doi.org/10.1007/s00401-012-1016-2

39. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010;11:367.

40. Degasperi A, Amarante TD, Czarnecki J, Shooter S, Zou X, Glodzik D, et al. A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies [Internet]. Nature Cancer. 2020. page 249–63. Available from: http://dx.doi.org/10.1038/s43018-020-0027-5

41. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.

42. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. Nature Publishing Group; 2003;34:267–73.

43. van Leeuwen EM, Kanterakis A, Deelen P, Kattenberg MV, Genome of the Netherlands Consortium, Slagboom PE, et al. Population-specific genotype imputations using minimac or IMPUTE2. Nat Protoc. 2015;10:1285–96.

44. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.

45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

46. Ogino S, Kawasaki T, Kirkner GJ, Kraft P, Loda M, Fuchs CS. Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. J Mol Diagn. 2007;9:305–14.

47. Nosho K, Kawasaki T, Ohnishi M, Suemoto Y, Kirkner GJ, Zepf D, et al. PIK3CA mutation in colorectal cancer: relationship with genetic and epigenetic alterations. Neoplasia. 2008;10:534–41.

48. Ogino S, Kawasaki T, Nosho K, Ohnishi M, Suemoto Y, Kirkner GJ, et al. LINE-1 hypomethylation is inversely associated with microsatellite instability and CpG island methylator phenotype in colorectal cancer. Int J Cancer. 2008;122:2767–73.

Acknowledgements

We thank N. Abdennur and S. Abraham for technical feedback, as well as W.L. Chiu, J. Elhai, and L. Fossecave for useful comments.

This work was supported by the U.S. National Institutes of Health (NIH) grants P01 CA87969 to M.J. Stampfer; UM1 CA186107 to M.J. Stampfer; P01 CA55075 to W.C. Willett; UM1 CA167552 to W.C. Willett; U01 CA167552 to W.C. Willett and L.A. Mucci; U01 CA176726 to W.C. Willett, U54 HG003067 to E.S. Lander and S. B. Gabriel; P50 CA127003 to C.S.F.; P30CA016359 to C.S.F.; R01 CA118553 to C.S.F.; R01 CA169141 to C.S.F.; R35 CA197735 to S.O.; R01 CA151993 to S.O.; K07 CA190673 to R.Nishihara.; K07 CA188126 to X.Z.; R21 CA238651 to X.Z.; R03 CA197879 to K.W.; R21 CA222940 to K.W. and M.G.; and R21 CA230873 to K.W. and S.O.; by Cancer Research UK Grand Challenge Award (UK C10674/A27140 to M.G. and S.O.); by Nodal Award (2016-02) from the Dana-Farber Harvard Cancer Center (to S.O.); by the Stand Up to Cancer Colorectal Cancer Dream Team Translational Research Grant (SU2C-AACR-DT22-17 to C.S.F. and M.G.), administered by the American Association for Cancer Research, a scientific partner of SU2C; and by grants from the Project P Fund, The Friends of the Dana-Farber Cancer Institute, Bennett Family Fund, and the Entertainment Industry Foundation through National Colorectal Cancer Research Alliance. Stand Up To Cancer is a division of the Entertainment Industry Foundation. K.H. was supported by fellowship grants from the Uehara Memorial Foundation and the Mitsukoshi Health and Welfare Foundation. X.Z. was supported by American Cancer Society Research Scholar Grant (RSG NEC-130476). X.Z. was supported by the Dana-Farber Harvard Cancer Center (DF/HCC) GI SPORE Developmental Research Project Award (P50CA127003), and DF/HCC Nodal Award (Cancer Center Support Grant, P30CA006516-55), the Karin Grunebaum Cancer Research Foundation, as well as the Zhu Family PEER Award. J.A.M. research is supported by the Douglas Gray Woodruff Chair fund, the Guo Shu Shi Fund, Anonymous Family Fund for Innovations in Colorectal Cancer, Project P fund, and the George Stone Family Foundation. M.G. was supported by a Conquer Cancer Foundation of ASCO Career Development Award. T.U. was supported by a grant from Overseas Research Fellowship (201960541) from Japan

Society for the Promotion of Science. R.Z. was supported by a fellowship grant from Huazhong University of Science and Technology, Wuhan, Hubei, China. We would like to thank the participants and staff of the Nurses' Health Studies and the Health Professionals Follow-up Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

Author contributions

K.W., S.O., and M.G. designed research. C.G., R.Z., K.H., Y.L., L.S., B.R., T.U. and X.Z. performed research, H.L., A.D.C., M.S., E.M.V., J.A.M., J.A.N, E.L.G., C.S.F., K.W., S.O. and M.G. contributed data, reagents and resources. C.G and M.G wrote the manuscript and generated the figures, which all authors reviewed.

Corresponding author

Correspondence to Marios_Giannakis@dfci.harvard.edu

Author Manuscript Published OnlineFirst on June 17, 2021; DOI: 10.1158/2159-8290.CD-20-1656
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

Figure legends

Figure 1: *De novo* signature deconvolution in NHS/HPFS CRCs

(A) Cohort and data overview. NHS: Nurses' Health Studies (NHS I and NHS II). HPFS: Health Professionals Follow-up Study. (B) Quality measures for NMF in NHS/ HPFS, Arrows indicate the estimated rank of mutational signatures. rss: residual sum of squares (C) The consensus seven signatures found by NMF in NHS/HPFS .

Figure 2: Active mutational signatures in colonic cells

Proportion of mutations assigned to *de novo* extracted signatures in CRCs from NHS/ HPFS (A) and TCGA (B), segregated by *MGMT* promoter methylation status, *POLE* exonuclease mutations, microsatellite instability and age at diagnosis. Boxplot outliers not shown. (C) Heatmap of the similarity scores between colorectal tumour (from TCGA and NHS/ HPFS) signatures - clustered on the y axis- and reference COSMIC signatures, clustered on the x axis. COSMIC signatures found in either NHS/ HPFS or TCGA are bolded. The alkylating normal colon (from EGA) signature is also shown. Clustering has been performed according to cosine similarity. EGA: European Genome-phenome Archive.

Figure 3: Epidemiology and distribution of alkylating damage

(A) Proportion of mutations assigned to alkylating damage in NHS/ HPFS, segregated by intake (top decile in grams per day versus the rest) of overall, processed and unprocessed red meat, as well as chicken and fish. (B) Proportion of mutations assigned to alkylating damage in CRC and normal colon, segregated by tumour location. Boxplot outliers not shown.

Figure 4: Carcinogenic potency of alkylating damage

(A) Relative likelihood of mutational processes to target recurrent hotspots in non-hypermutated CRC. As hotspots we considered all point mutations that were present in at least 25 patients with non-hypermutated (non-MSI-high, non-POLE mutated) CRC. Each stacked bar represents the relative likelihood of a given signature to target a given hotspot. (B) Proportion of mutations assigned to alkylating damage in NHS/ HPFS, TCGA CRCs, segregated by *KRAS* G12D/ *KRAS* G13D/ PIK3CA E545K mutation status. Boxplot outliers not shown. (C) Kaplan–Meier plot illustrating colorectal cancer-specific survival of the patients stratified into quartiles of alkylating signature contribution. (D) Forest plot of the association between the colorectal specific survival and quartiles of alkylating signature contribution in univariable and multivariable Cox regression models.

Figure 1

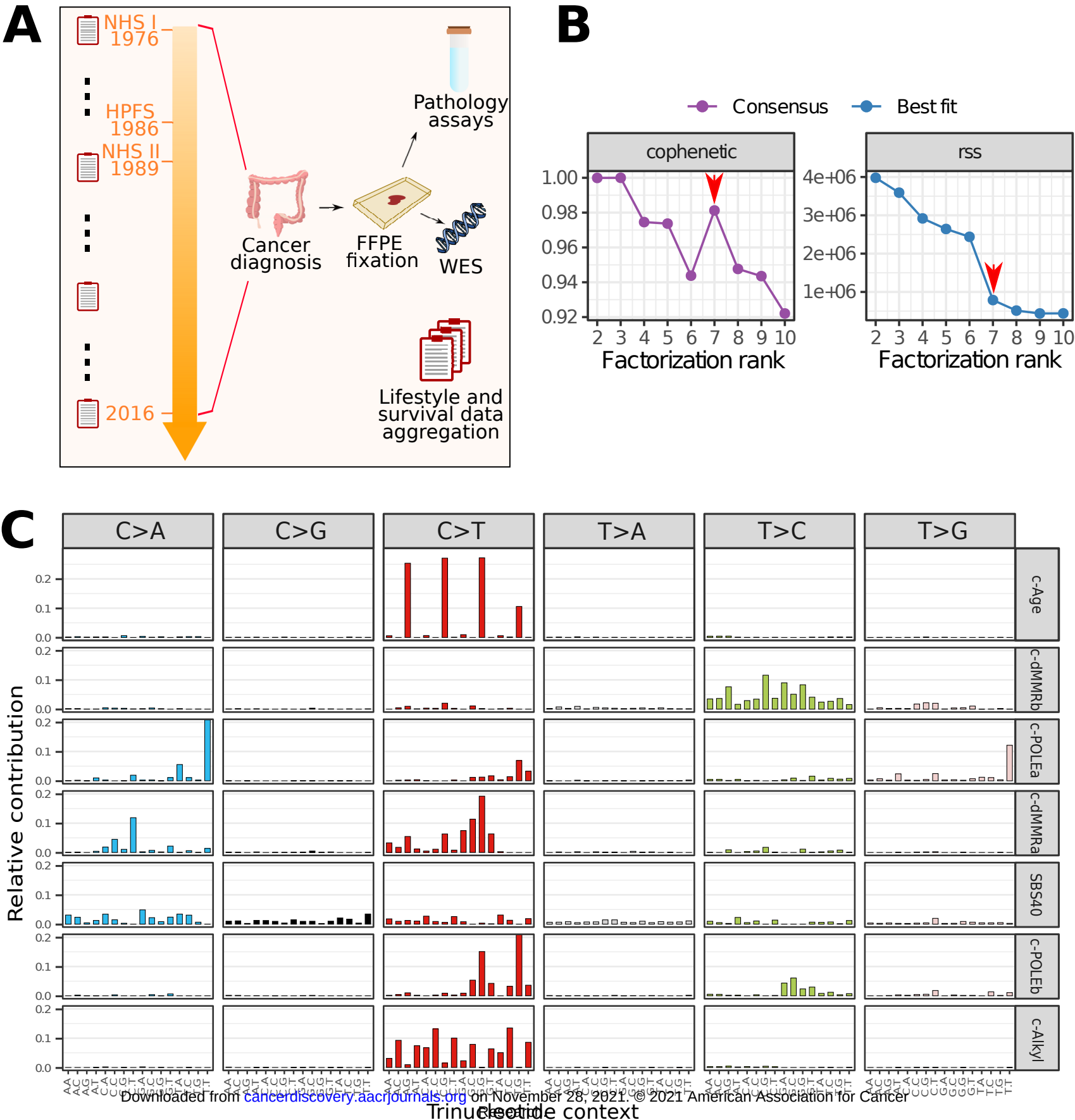


Figure 2

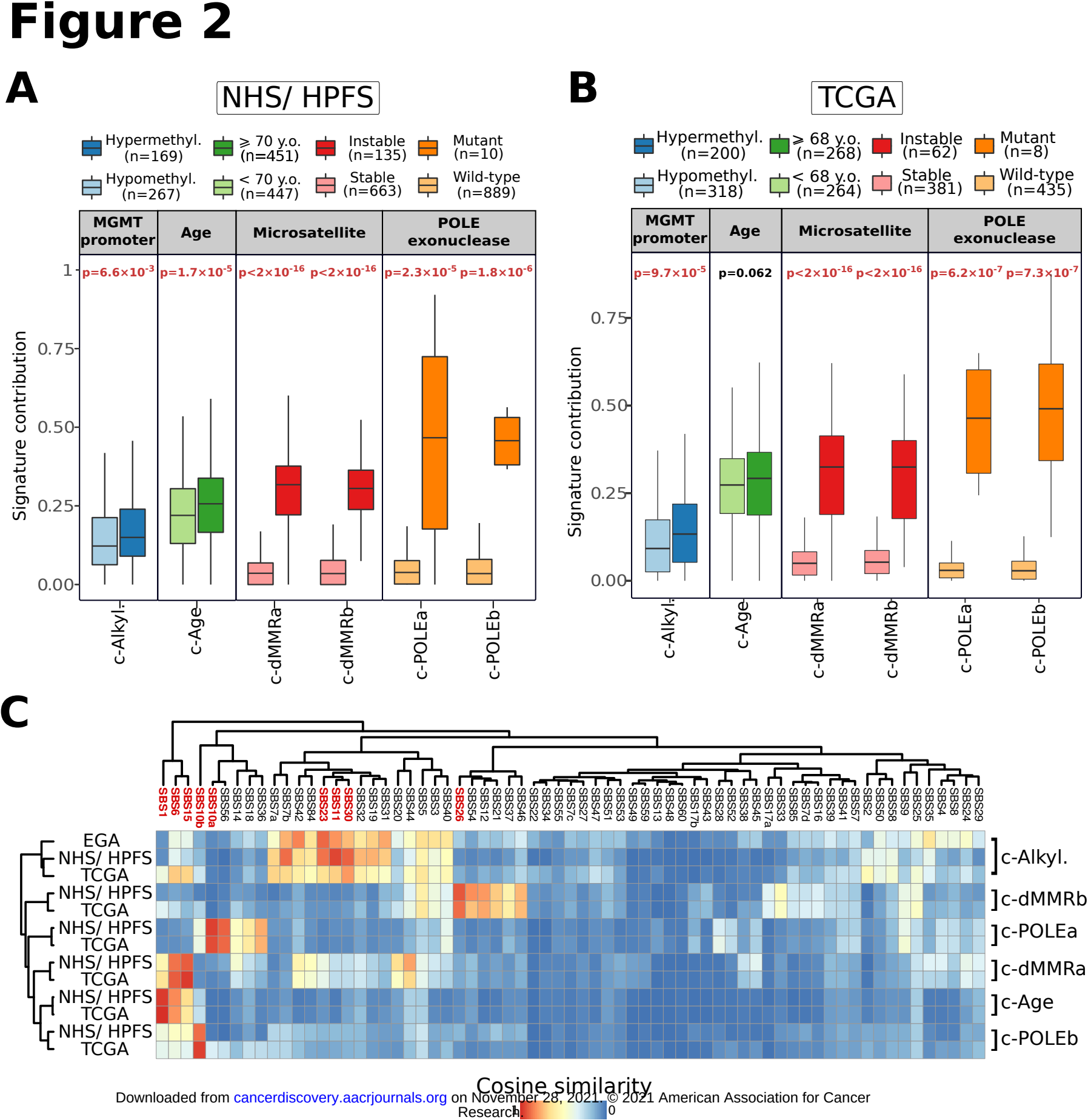


Figure 3

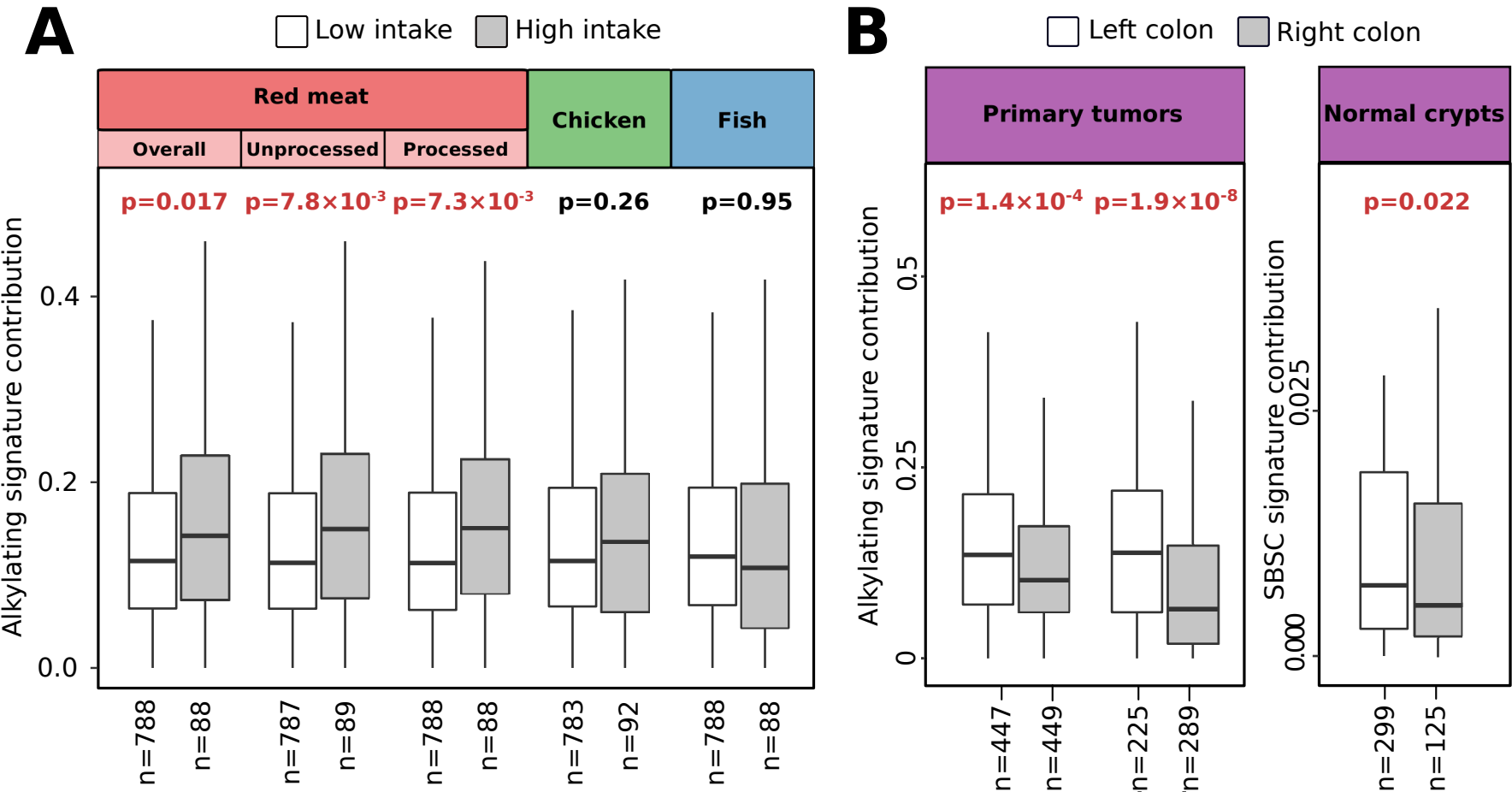
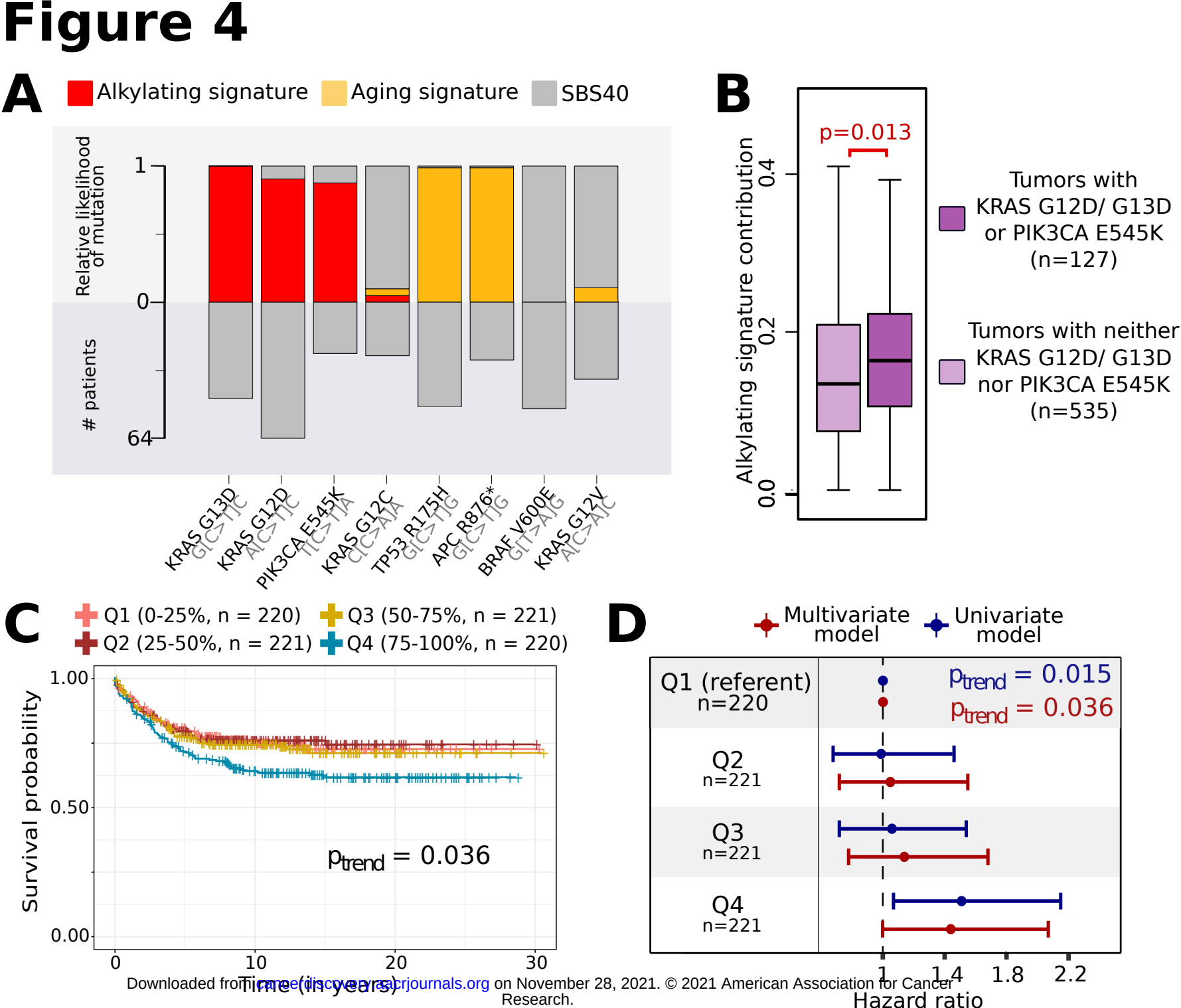


Figure 4



AACR American Association for Cancer Research

CANCER DISCOVERY

Discovery and features of an alkylating signature in colorectal cancer

Carino Gurjao, Rong Zhong, Koichiro Haruki, et al.

Cancer Discov Published OnlineFirst June 17, 2021.