5-Fluorouracil treatment induces characteristic T>G mutations in human cancer

Sharon Christensen[1], Bastiaan Van der Roest[1], Nicolle Besselink[1], Roel Janssen[1], Sander Boymans[1], John Martens[2], Marie-Laure Yaspo[3], Peter Priestley[4], Center for Personalized Cancer Treatment[5], Ewart Kuijk[1], Edwin Cuppen[1,4,*], Arne Van Hoeck[1]

*corresponding author: ecuppen@umcutrecht.nl

[1] Center for Molecular Medicine and Oncode Institute, University Medical Center Utrecht, The Netherlands
[2] Erasmus Medical Center, Rotterdam, The Netherlands
[3] Max Planck Institute for Molecular Genetics, Berlin, Germany
[4] Hartwig Medical Foundation, Amsterdam, The Netherlands
[5] CPCT consortium, Rotterdam, The Netherlands

Abstract:

5-Fluorouracil (5-FU) is a chemotherapeutic drug component that is commonly used for the treatment of solid cancers. It is proposed that 5-FU possesses anticancer properties via the interference with nucleotide synthesis and incorporation into DNA. As both mechanisms may have a mutational impact on both surviving tumor and healthy cells, we treated intestinal organoids with 5-FU followed by whole genome sequencing analysis and uncovered a highly characteristic mutational pattern that is dominated by T>G substitutions in a CTT context. Analysis of tumor whole genome sequencing data confirmed that this signature can also be identified *in vivo* in colorectal and breast cancer patients that have undergone treatment with 5-FU. We also found that more 5-FU mutations are induced in *TP53* null backgrounds which may be of clinical relevance. Taken together, our results demonstrate that 5-FU is mutagenic and may drive tumor evolution and increase the risk of secondary malignancies. Furthermore, the identified signature shows a strong resemblance to COSMIC signature 17, the hallmark signature of treatment-naive esophageal and gastric tumors, which indicates that distinct endogenous and exogenous triggers can converge onto highly similar mutational signatures.

Introduction

The use of 5-Fluorouracil (5-FU) as an anticancer agent became routine practice soon after its primary synthesis in 1957, and remains essential in many chemotherapeutic regimens today (Longley, Harkin, and Johnston 2003). The fluoropyrimidines, especially 5-FU, capecitabine, tegafur and cytarabine, are currently the third most commonly used anticancer drug in the treatment of solid cancers, including colorectal and breast cancers, and over two million patients are estimated to be treated with fluoropyrimidines each year (Ezzeldin and Diasio 2004). Response rates of 5-FU as a single drug are 10-15%, but increase drastically (>50% response) when given in combination therapies with leucovorin together with oxaliplatin or irinotecan (i.e. FOLFOX and FOLFIRI, respectively) (de Gramont et al. 2000; Boige et al. 2010; Cameron, Gabra, and Leonard 1994).

The antifolate property of fluoropyrimidines is thought to be the principal mechanism of action. Fluoropyrimidines are intracellularly converted into the antifolate 5-fluorodeoxyuridine monophosphate (5-FdUMP) that can form a covalent intermediate with the folate-dependent enzyme thymidylate synthase (TYMS) (Sommer and Santi 1974). Consequently, the formation of dTMP from dUMP is inhibited which results in an imbalance of the nucleotide pool that affects DNA synthesis, possibly through incorporation of uracil, and impairs genome replication, with negative consequences for rapidly dividing cells such as cancer cells. Moreover, it has been proposed that 5-fluorodeoxyuridine triphosphate (5-FdUTP) can be directly incorporated into genomic DNA as well (Pettersen et al. 2011; Huehls et al. 2016). Considering these properties, it is conceivable that fluoropyrimidines have mutagenic potential, although the mutational consequences of 5-FU treatments are still poorly understood.

In cancer, systematic analysis of genome-wide mutation catalogs has revealed a number of characteristic mutational patterns or "mutational signatures" (L. B. Alexandrov et al. 2013). Some of these signatures have been linked to perturbed endogenous processes like deficient DNA repair, or exogenous challenges, like exposure to UV-light or mutagenic chemicals. Such information thus provides insight into the mutational processes that have been active during tumorigenesis and which could potentially be used for prevention strategies or personalized treatment strategies. Previously, it has been shown that certain anticancer treatments can be associated with characteristic mutational signatures, such as alkylating agents (L. B. Alexandrov et al. 2013; Phillips 2018), cisplatin (Boot, Huang, et al. 2018; Meier et al. 2014) and ionizing radiation (Behjati et al. 2016; Davidson et al. 2017). Unlike these anticancer treatments, and in spite of its mutagenic potential, 5-FU could thus far not be linked to any mutational signature using these systematic cancer cohort analyses.

Here we assessed the mutational consequences of fluoropyrimidines by exposing organoids of healthy intestinal stem cells to 5-FU followed by genome-wide analysis of single cells. For this, we used a previously described highly sensitive approach based on clonal expansion of individual cells followed by whole genome sequencing for mutational spectrum analysis (Jager et al. 2018; Blokzijl et al. 2016). *In vitro* findings were subsequently validated by exploration of mutational patterns in breast and colorectal cancer patients who have had previous fluoropyrimidine treatments. Our results demonstrate that 5-FU induces both *in vitro* in organoids and *in vivo* in cancer cells a similar mutational pattern that is reminiscent of COSMIC signature 17.

Results

Characterization of 5-FU mutational effect in vitro

We have set up human small intestinal (SI) isogenic organoid cultures which were exposed to 5-FU for 3 days followed by 4 days of recovery (Figure 1A). This treatment procedure was repeated 5 times, which allowed the organoids to survive the exposure conditions and to accumulate a sufficient number of mutations. Then, individual organoid cells from the 5-FU exposed cultures were manually picked, expanded and analyzed by whole genome sequencing (WGS) with a read coverage-depth of ~30X. Somatic mutations were called against the original isogenic organoid line which was also sequenced at ~30X. Lastly, mutations which arose after the single-cell-step were filtered out based on low variant allele frequencies (S Figure 1). A total of 1,324 highly confident induced single base substitutions (SBSs) were identified in the autosomal genome that were accumulated during 5-FU treatment (n = 2 organoid lines). Organoids grown in parallel, but not exposed to 5-FU, served as control (n = 6 organoid lines). Not unexpectedly, untreated control organoids were found to proliferate faster than treated organoids, which makes it impossible to accurately determine the mutation accumulation load per cell division, although qualitative aspects and relative mutation contributions can still be interpreted.

To dissect active mutational processes, we analyzed the 96 mutational spectra of the obtained SBSs with trinucleotide context in more detail. We observed a distinct mutation profile for 5-FU exposed organoids when compared to the background *in vitro* mutation spectrum of untreated control SI organoids (Pearson correlation = 0.26; cosine sim = 0.57) (Figure 1B). The most striking differences are the T>G mutations in a CTT trinucleotide context (further referred as C[T>G]T mutations) and, to a lesser extent, C[T>C]T and G[T>G]T mutations, which together account for more than half of the total mutation profile of 5-FU-treated organoids. This illustrates that 5-FU induces a characteristic mutational pattern *in vitro* that is driven by a mutational process that generates SBSs with a chance of ~35% being a CTT>CGT mutation.

5-FU-induced mutational pattern in human cancer

To assess if the observed 5-FU mutational consequences can also be detected *in vivo* in human cancer samples, we explored cancer whole genome sequencing data from metastatic cancer patients (Hartwig Medical Foundation database) for which treatment data is also available (Priestley et al. 2018). 65% of colorectal (n=352) and 36% of the breast (n=450) cancer patients in this data set underwent 5-FU based treatment (i.e. 5-fluorouracil, fluoropyrimidine, capecitabine or tegafur - further referred to as 5-FU) at any time prior to biopsy and WGS. We performed an unbiased *de novo* mutational signature analysis using non-negative matrix factorization (NMF) (Gaujoux and Seoighe, 2010) on both cohorts with inclusion of the 5-FU exposed organoid data. NMF identified sixteen mutational signatures which showed all high similarity with well-described signatures in human cancer (Figure 2A, S Table 2) (http://cancer.sanger.ac.uk/cosmic/signatures, (L. Alexandrov et al. 2018) (Lee-Six et al. 2018) (Boot, Ng, et al. 2018). Interestingly, a signature that was highly similar to the 5-FU *in vitro* mutation spectrum was found in the set of the *de novo* extracted signatures (Pearson correlation = 0.98; cosine sim = 0.98) (Figure 2A). This signature, further referred as "5-FU signature" (Figure 2B), is predominated by C[T>G]T mutations (36%) which is almost equal to the 5-FU *in vitro* mutation spectrum (35% of C[T>G]T mutations). Ranking by the total mutational load of this 5-FU signature illustrates that patients who display a prominent

contribution of this pattern were treated with 5-FU (S Figure 2). These results indicate that 5-FU has the same mutagenic effect *in vivo* as *in vitro*.

5-FU signature contribution in human cancer

To quantify the mutational contribution of the 5-FU signature we compared 5-FU pretreated and non-5-FU pretreated patients (including a treatment-naive primary colorectal (Schütte et al. 2017) and breast cancer cohort (Nik-Zainal et al. 2016) as additional controls). The relative contribution of the 5-FU signature was calculated and compared for each patient to adjust for differences in tumor mutational burden (TMB - number of SBSs per Mbp) between primary and metastatic cohort (Yates et al. 2017). In line with our previous results, 5-FU pretreated patients showed a significantly higher 5-FU signature contribution compared to 5-FU untreated patients in both the colon and breast cancer cohort (both P<0.05, Wilcoxon rank-sum test) (Figure 2C). No significant differences were found between the 5-FU untreated patients and the treatment-naive cohorts. Examining the absolute mutational contribution for all extracted signatures shows that only the 5-FU signature is increased in contribution illustrating that 5-FU does not have a measurable impact on other signatures (P<0.05, Wilcoxon rank-sum test, S Figure 3). While 5-FU is most commonly used to treat breast and colon cancer patients, it is often also administered to patients with more rare cancer indications including pancreas (n=11), biliary tract (n=6) and head and neck (n=5). In these cancer types, we identified the same 5-FU mutagenic effect as in breast and colon cancer, although not significant due to the low number of patients, which demonstrates that the 5-FU mutational process is tissue independent (S Figure 4).

We observed an extensive variation in the number of 5-FU mutations per 5-FU treated patient ranging from 0 to roughly 15,000 mutations in both colon and breast cancer patients (S Figure 2). This may be explained by variation in pharmacodynamics between patients, differences in the dosing and the duration of 5-FU treatment schedules (Grem 2000), as well as by the evolution dynamics, but potentially also by other characteristics of the tumor. Indeed, analysis of tumor driver and suppressor genes (n=378) uncovered that *TP53* mutated cancers accumulated more 5-FU mutations than *TP53* wild type cancers, both in colon and breast (P<0.05, Wilcoxon rank-sum test, Figure 2D). Also, fluorouracil and capecitabine were both found to be mutagenic in colon cancer, while in breast cancer only capecitabine showed an increased mutagenic effect (S Figure 5), which might reflect differences between both tissues in drug uptake and treatment schemes. Notwithstanding the high variation in 5-FU signature contributions between patients, we observed that colon cancers overall have a higher 5-FU signature contribution than breast cancers, with a median mutation count of 1180 and 139 mutations, respectively.

The underlying clonal architecture of mutational events can be inferred from the variant allele frequency (VAF) and provides more insight into the timing of the activity of specific mutational processes. In comparison to clonal mutations, we found approximately a three-fold increase in the relative mutational contribution of the 5-FU signature for the subclonal mutations (P<0.05, Wilcoxon rank-sum test, S Figure 6). This points out that the 5-FU induced mutagenic activity is more profound in the metastatic colonies and therefore occurred at a later stage in tumor development, which is in line with the time of cancer diagnosis and subsequent 5-FU treatment.

5-FU mutations in paired biopsies

In the studied metastatic cancer patient cohort, 53 patients underwent two or more serial biopsies, which can be used to provide a more direct approach to study the chronological timing of the activity of mutational processes. This group of patients with multiple biopsies consisted of different cancer types of which 8 patients (colorectal cancer (n=4) and breast cancer (n= 4)) that received a systemic 5-FU related treatment after the first biopsy and before one of the following biopsies. For every patient, we determined the mutation profiles of both biopsies and examined the difference in mutation numbers for each of the 96 mutation types, reasoning that 5-FU characteristic mutation types - particularly C[T>G]T mutations - would increase in mutational load. A mixed-effect regression analysis indeed revealed a positive correlation between the normalized absolute count of C[T>G]T mutations from the first biopsy compared to the second biopsy in patients treated with 5-FU (ANOVA linear mixed model; $P<0.05$) (Figure 3). Moreover, iterating this statistical analysis on each of the 96 possible mutation types resulted in significant P-values for all mutation types that are dominating the previously identified 5-FU signature (Figure 3). Of note, no correlations were found between 5-FU characteristic mutation types and any other administered treatment drug (Carboplatin, Cisplatin, Oxaliplatin, Pazopanib, Pembrolizumab and Pemetrexed) demonstrating that the signature is highly specifically induced by 5-FU (S Figure 7).

5-FU signature resembles COSMIC signature 17

We compared the obtained 5-FU signature to the known COSMIC signatures and found a high similarity (Pearson correlation = 0.97; cosine sim = 0.97) with COSMIC signature 17 (Figure 4A), which is predominantly found in treatment-naive esophagus and gastric cancer. Recent work has split COSMIC signature 17 into two constituent signatures (SBS17a, predominantly characterized by T>C mutations and SBS17b, characterized by T>G mutations) (L. Alexandrov et al. 2018), suggesting two distinct mutational processes. However, the here obtained 5-FU *in vitro* mutation spectrum showed both T>C and T>G mutations as in COSMIC signature 17, and thus our findings provide no evidence that COSMIC signature 17 exhibit a pattern of two independent mutational processes.

Next, we investigated whether the 5-FU signature also encompasses more detailed molecular features that are characteristic for COSMIC signature 17. In agreement with signature 17 (Morganella et al. 2016; Secrier et al. 2016), we also found a seven-base mutation context for C[T>G]T mutations in 5-FU pretreated colon and breast cancer patients which is predominated by A/T bases at the -4, -3 and -2 positions from the mutated base position (Figure 4C). Furthermore, COSMIC signature 17 has been shown to display a higher mutation rate on the lagging strand (Letouzé et al. 2017; Tomkova et al. 2018). Consistent with these reports, we observed a strong replication strand bias towards the lagging strand for C[T>G]T mutations types in 5-FU pretreated colon and breast cancer samples (Figure 4B). In addition, we also noted a minor transcriptional strand bias in the colon samples for C[T>G]T mutations (S Figure 8). Given this strong overlap in characteristics between both signatures, we conclude that the identified 5-FU signature is the same as COSMIC signature 17 and does not represent a novel signature.

Impact on tumorigenesis

We observed an average increase (~20%) in the overall TMB for 5-FU treated cancers, at least for the colon cancer patients ($P<0.05$, Wilcoxon rank-sum test) (Figure 2E). However, the 5-

FU contribution on the TMB differs extensively per patient (S Figure 9) where most 5-FU pretreated cancer patients (65% and 85% for colon and breast, respectively) show a limited impact of 5-FU on the TMB (<10%) and only a few patients (6% and 3% for colon and breast, respectively) demonstrate a substantial 5-FU contribution that affect the TMB with at least 30%. To investigate the impact of these 5-FU mutations on tumor evolution and disease progression, we selected all subclonal synonymous and non-synonymous mutations that were most likely induced by 5-FU exposure for each patient (see methods) to quantify oncogenic driver mutations induced by 5-FU (S Figure 10). We observed no increase in the number of validated oncogenic drivers (Tamborero et al. 2018) in the 5-FU pretreated colon (5 driver mutations) and breast (5 driver mutations) cancer patients compared to non 5-FU pretreated colon (2 driver mutations, P=0.56, Fisher exact test) and breast (5 driver mutations, P=0.26, Fisher exact test) cancer patients (S table 2).

In an attempt to characterize genes that may have contributed to 5-FU resistance, we performed a dN/dS analysis in which all single nucleotide mutations and small insertions and deletions (INDELS) were included, but revealed no significantly mutated genes in contrast to resistance to hormonal therapies (e.g. ESR1 for breast and AR for prostate (Zehir et al. 2017; Priestley et al. 2018)) and targeted treatments (e.g. secondary BRAF mutations for melanoma treated with vemurafenib (Poulikakos et al. 2011) and secondary EGFR mutations treated with EGFR inhibitors (Morgillo et al. 2016)).

Next, we investigated loss-of-function (LOF) and gain-of-function (GOF) events of key enzymes of the pyrimidine metabolic pathways. TYMS is considered as the key therapeutic target for 5-FU and overexpression of its gene has been linked to 5-FU resistance in *in vitro* as well as in *in vivo* experiments (Watson et al. 2010; Intuyod et al. 2018). *TYMS* showed no LOF mutations in the breast and colorectal cohort, supporting the findings that *TYMS* is an essential gene (Blomen et al.). On the other hand, GOF events of *TYMS* gene by means of copy number gains were found in 5-FU pretreated colon cancer patients (n=44 out of 231) versus untreated patients (n=8 out of 121) (P<0.05, Fisher exact test) (S Figure 11), although this was not observed for breast cancer patients. This indicates a selective pressure towards increased levels of TYMS activity after 5-FU administration. The copy number level of *TYMS* seems to be inversely correlated with the absolute contribution of 5-FU pattern (S Figure 11), which may suggest that TYMS overexpression can block the 5-FU mutational process by overcoming binding of 5-FdUMP by sheer number of TYMS protein.

It is interesting to note that, as we have shown with the organoid experiments, normal cells also accumulate 5-FU mutations. Consequently, it can be postulated that not only cancer cells, but any other cell in the body exposed to 5-FU may accumulate mutations that lead to the onset of secondary malignancies. To quantify this risk, we modelled the chance of introducing a cancer driver mutation resulting from 5-FU treatment, using the 5-FU specific mutation context and *in vivo* observed average mutation rate (S Figure 12). This model estimates that about 300 oncogenic mutations are introduced *in vivo* in 10^8 colon stem cells per 5-FU treatment, which is 50-fold higher than under normal conditions as a result of *in vivo* mutational processes associated with aging. One full cycle of 5-FU treatment therefore reflects 'normal' mutation accumulation in colon stem cells of about 20 years (Blokzijl et al. 2016). As such, the consequences of 5-FU administration may be limited for patients with age above 60-70 years, but can be significant for cancer patients at a relatively young age (20-30 years old). Furthermore, patients carrying

germline predisposition variants (*e.g. APC* mutation in FAP syndrome resulting in the development of tumors at a relatively young age) are at increased risk for acquiring a second hit and may be a contraindication for 5-FU treatment. We modelled this scenario as well and found a 20-fold increase in risk as compared to non-treated patients, which is equivalent to reducing the average age of onset for tumor development in FAP patients with 10 years.

Discussion

Here, we demonstrate a causal relationship between 5-FU treatment and COSMIC signature 17, characterized by C[T>G]T base substitutions.

This finding differs from a previous study that did not find a measurable mutagenic effect of 5-FU exposure of cultured chicken lymphoblasts (Szikriszt et al. 2016). This discrepancy might be due to differences in experimental conditions (5-FU dosage, mutation detection) or the *in vitro* models used. Indeed, non-human cell lines are known to differ in DNA damage susceptibility (MacRae et al. 2015), e.g. exposing aflatoxin to cell lines, mouse tumors and human tumors results in great diversity in mutation profiles (Huang et al. 2017). Likewise, cisplatin signatures characterized with cell lines of different model organisms (Szikriszt et al. 2016; Meier et al. 2014) do not recapitulate the cisplatin patterns recently found in human cancer (Boot, Huang, et al. 2018; L. Alexandrov et al. 2018).

Since 5-FU is structurally similar to thymidine and uracil nucleotides and has previously been shown to interfere with nucleotide biosynthesis and nucleotide pools (Myers, Young, and Chabner 1975; Peters et al. 1989; Berger and Berger 2006), a mutagenic effect of 5-FU was anticipated. However, the strong resemblance with a previously described signature that was already linked to a different potentially underlying mechanism was surprising. COSMIC signature 17 is the hallmark signature of esophageal and gastric cancers and the presence of gastric refluxate has been suggested to be the responsible mutagen in these cancer types. High COSMIC signature 17 contributions are occasionally found in non-5-FU treated patients diagnosed with other cancer types as well (L. B. Alexandrov et al. 2013; Nik-Zainal et al. 2016). For instance, a comprehensive study dissected the intratumor heterogeneity of three treatment naive colorectal tumors, of which one displayed extensive signature 17 contribution (Roerink et al. 2018). Thus, signature 17 reflects the consequences of a mutational process that can be instigated by multiple triggers including 5-FU exposure.

Recent work has proposed that COSMIC signature 17 reflects the mutagenic consequences of the presence of oxidized dGTP nucleotides in the nucleotide pool (Tomkova et al. 2018). Indeed, a number of studies have reported that the presence of oxidized guanine nucleotides (8-oxo-dGTP) increases the T>G mutation rate (Hidaka et al. 2008; Inoue et al. 1998). Accordingly, inhibition of enzymes responsible for the removal of oxidized nucleotides, such as MTH1, MTH2, and NUDT5, have been shown to promote T>G mutations as well (Suzuki and Kamiya 2017). Also, the flanking sequence context of the dominant mutation type of Signature 17 mirrors the context of the dominant mutation type of Signature 18. This mutational process has been linked to direct oxidation of guanine located inside the DNA (Poetsch, Boulton, and Luscombe 2018; Pilati et al. 2017). It is therefore tempting to speculate that the oxidation of dGTPs in the nucleotide pool underlies Signature 17. As such, the presence of bile refluxate would be a plausible explanation for the elevated levels of 8-oxo-dGTP in esophagus cancer (Dvorak et al. 2007). However, a recent study showed that bile refluxate alone does not generate 8-oxo-dGTPs, but that bile acid also requires an acidic environment to promote the production of 8-oxo-dGTP. This was only found in the epithelial cells of premalignant Barrett's esophageal cells, which gained transporters for bile acids, potentially clarifying why healthy esophageal cells do not show Signature 17 mutations (Yokoyama et al. 2019; Martincorena et al. 2018; Dvorak et al. 2007). Based on this, one could hypothesize that 5-FU exposure induces a similar oxidative stress environment in the cell that generates 8-oxo-dGTP thereby stimulating T>G mutations in a

C[T>G]T context. In line with this, 5-FU treatment is less cytotoxic when combined with antioxidants (Fu et al. 2014) and ROS production is directly correlated with treatment with 5-FU (Focaccetti et al. 2015; Negrei et al. 2016).

An alternative explanation of the underlying mutational process of Signature 17 observed in 5-FU treated patients can be attributed to an imbalance of the nucleotide pool by TYMS inhibition, which is considered to be the major drug target of 5-FU. The 5-FU metabolite 5-FdUMP hampers the synthesis of dTMP which results in a depletion of dTTPs in the nucleotide pool (An et al. 2007; Parker and Stivers 2011) and impaired dTMP biosynthesis results in accelerated rates of genomic deoxyuridine triphosphate (dUTP) incorporation (MacFarlane et al. 2011); (Vértessy and Tóth 2009). Next to dUTPs, also the 5-FU related byproduct 5-FdUTP can be incorporated during replication, which results in the accumulation of U:A and 5-FU:A base pairs (Parker and Stivers 2011). These mutation types largely recapitulate Signature 17 and for this reason nucleotide imbalance by TYMS inhibition is a plausible cause for the here observed 5-FU mutations as well, although the strong similarity with the process active in esophageal cancer is not easily explained. In any case, further experimental follow-up will be required to dissect the underlying molecular mechanisms and to conclude whether one mutational mechanism is responsible for 5-FU specific mutation accumulation or that the 5-FU signature is the result of multiple mutational processes operating simultaneously on the genome (e.g. 8-oxo-dGTP, dUTPs and 5-FdUTPs) that are accompanied by DNA repair mechanisms (e.g. uracil removal by uracil-DNA glycosylase [UDG]). Indeed, recent work revealed that the base excision DNA repair machinery selectively corrects Signature 17 mutations depending on its position around the nucleosome (Pich et al. 2018). The involvement of DNA repair might also explain why tumors deficient in the p53 DNA damage checkpoint regulatory pathway accumulate more 5-FU mutations, as shown here for the first time. Interestingly, breast tumors with high contribution of Signature 17 mutations were recently shown to have poor prognosis (Bertucci et al. 2019).

Nevertheless, we found that the mutation contribution of 5-FU administration does not have a great impact on the total tumor mutational burden and the driver landscape of the cancer in the majority of the patients. However, as the mechanisms driving 5-FU resistance remains largely to be elucidated, it cannot be excluded that induced mutations contribute to this process.

Furthermore, we calculated that young cancer survivors exhibit an increased risk for developing chemotherapy-related second malignancies as 5-FU can accelerate the rate of introducing novel oncogenic mutations in normal cells. Therefore treatment decision makers must be aware of the increased risk factors of 5-FU administration to cancer patients at a relatively young age (Wright et al. 2015; Gladsjo et al. 2009).

Here, we have shown that the administration of fluoropyrimidines activates a mutational process that results in a highly characteristic mutational signature and as such, contributes to the mutational landscape of human (cancer) cells. Moreover, our results indicate that distinct triggers or processes can be at the origin of highly similar mutational signatures. Insights from this study could serve as a basis for future research to elucidate when and how these mutagenic agents converge on similar molecular mechanisms.

Materials and methods

Patient cohort

We selected patients of the CPCT-02 (NCT01855477) and DRUP (NCT02925234) clinical studies, which were approved by the medical ethical committees (METC) of the University Medical Center Utrecht and the Netherlands Cancer Institute, respectively. This national initiative consists of nearly 50 oncology centers from The Netherlands and aims to improve personalized cancer. To this end, Hartwig Medical Foundation sequences and characterizes the genomic landscape for a large number of patients. Furthermore, genomics data is integrated with clinical data which consists of primary tumor type, biopsy location, gender, pretreatment type before biopsy, and treatment type after biopsy. A detailed description of the consortium and the whole patient cohort has been described in detail in (Priestley et al. 2018). For this study, we selected cancers with primary tumor location in the breast, colon and esophagus. Next, we also included all sample IDs, irrespective of the primary tumor location, which underwent at least 2 biopsies. Samples for which pretreatment was not documented (hasSystemicPreTreatment = NA) were excluded from this study.

Organoid culturing

Isogenic healthy human small intestinal organoids were cultured as described previously (Jager et al. 2018). In short, organoids were grown on Complete Human Intestinal Organoid (CHIO) medium, supplemented with 30% Adv+++ (Advanced DMEM F12 [Thermofisher], supplemented with glutamax [1%, Thermofisher], hepes [10 mM, Thermofisher], penicillin/streptomycin [1%, Thermofisher]), in house produced Wnt (50%) (Broutier et al. 2016) and R-spondin (20%) (Broutier et al. 2016), B27 supplement (1x, Thermofisher), nicotinamide (10 mM Sigma), N-acetylcysteine (1.25 mM, Sigma), Primocin (0.1 mg/ml, Invivogen), A83-01 (0.5 μM, Tocris Bioscience), recombinant noggin (0.1 μg/ml, Peprotech), SB202190 (10 μM, Sigma) and hEGF (50 ng/ml, Peprotech). Organoids were embedded in matrigel and medium was refreshed every 2-3 days. A titration series was performed ranging from 0 to 100 uM 5-FU (0, 3.13, 6.25, 12.5, 25, 50, and 100 uM). The selected concentration of 6.25 uM was where roughly 50% of organoids grew out further after the 5 cycles of treatment. The selected concentration (i.e. 6.25 μM) is lower than often used in acute dosing experiments as these conditions were found to killing or senescence of all cells. CHIO medium containing 6.25 uM 5-FU was added to the organoids 5 days post seeding, for a period of 3 days, after which the 5-FU-containing medium was refreshed with 5-FU-free CHIO medium for two consecutive days. The organoids were then left to rest for 2 days. This 7-day treatment cycle was repeated for 5 weeks after which the medium was changed to standard medium again and the organoids were left to rest for an additional day. The organoids were then dissociated into single cells by trypsinization and plated in a limited-dilution series. This was supplemented with CHIO medium containing ROCK inhibitor (10 μM, Abmole) and hES Cell Cloning & Recovery Supplement (1x, Tebu-Bio). Subsequently, individual clonal organoids were manually picked and expanded to gain enough material for WGS. A signed approval was obtained by the medical ethical committee UMC Utrecht (METC UMCU) for using the human small intestinal organoid lines.

DNA isolation and WGS of organoid lines
Organoids were dissociated and DNA was isolated using the QiaSymphony DSP DNA mini kit (Qiagen, cat. No. 937236). Libraries were prepared using the Truseq DNA nano library prep kit (Illumina, cat. No. 20015964). Paired-end sequencing was performed (2 x 150 bp) on the generated libraries with 30x coverage using the Illumina HiSeq Xten at the Hartwig Medical Foundation.

Somatic mutation calling
Somatic mutation data of the CPCT and DRUP project were kindly shared by HMF on September 1, 2018. To exclude differences in accuracy and sensitivity from somatic calling workflows between *in vivo* and *in vitro* data, we pulled the HMF somatic mutation workflow from https://github.com/hartwigmedical/pipeline and installed the pipeline locally using GNU Guix with the recipe from https://github.com/UMCUGenetics/guix-additions. Full pipeline description is explained in (Priestley et al. 2018), and details and settings of all the tools can be found at their Github page. Briefly, sequence reads were mapped against human reference genome GRCh37 using Burrows-Wheeler Alignment (BWA-MEM) v0.7.5a (Li and Durbin 2009). Subsequently, somatic single base substitutions (SBSs) and small insertions and deletions (INDELS) were determined by providing the genotype and tumor (or organoid for in-vitro analysis) sequencing data to Strelka v1.0.14 (Saunders et al. 2012) with adjustments as described elsewhere (Priestley et al. 2018). To obtain high-quality somatic mutations that can be attributed to 5-FU exposure in the organoid lines, we characterized the mutations that have accumulated between the sequential clonal expansion step. As such, we only considered somatic mutations with a variant allele frequency between 0.3 and 0.7 as mutations that fall outside this range were potentially induced *in vitro* after the clonal step.

Mutational signature analysis
De novo mutational signature extraction was performed using the NMF package (v0.21.0) with 100 iterations (Gaujoux and Seoighe 2010). Non-negative matrix factorization (NMF) is an unsupervised approach that decompose high-dimensional datasets in a reduced number of meaningful patterns. For *in vivo* samples, we ran NMF on the colon and breast cancer cohort including the two organoid lines exposed to 5-FU and six organoid lines that were cultured in identical medium for 140-146 days. In order to characterize the optimal number of patterns, we compared the cophenetic correlation coefficient over the range of possible signatures and assigned sixteen *de novo* signatures. This set of *de novo* extracted signatures were compared to the COSMIC cancer mutational signatures (http://cancer.sanger.ac.uk/cosmic/signatures), to the expanded list of mutational signatures (L. Alexandrov et al. 2018), and signatures from other studies (Lee-Six et al. 2018) (Boot, Ng, et al. 2018) using the cosine similarity from the Mutational Patterns R package as a measure of closeness (Blokzijl et al. 2018). We also used Mutational Patterns to determine the absolute contributions of each *de novo* obtained signature for the metastatic and primary cohorts. Briefly, a vector of 96 trinucleotide context counts for each sample was fitted using non-negative least squares regression to a 96 x n (where n is the number of signatures) matrix consisting of the trinucleotide context probabilities for each signature. The relative contribution of each signature was calculated by dividing the absolute counts by the total mutation count (i.e. tumor mutational burden) of the sample.

Paired biopsies

To test whether the number of 5-FU specific mutations was higher in the sample biopsied after 5-FU treatment than in the sample before the treatment, we first determined the 96-mutation count table for each sample. Next, we normalized the absolute mutation count for each set of paired samples per patient using the median ratio algorithm from the Deseq2 package (Love, Huber, and Anders 2014). Subsequently, we performed a linear mixed effect analysis using nlme R package (Ezzet and Pinheiro, n.d.) on each mutation type to assess the relationship between the normalized mutation count for each mutation type and treatment. We entered all the different treatment drugs into the model that were administered to at least 3 patients after biopsy one (5-FU, Carboplatin, Cisplatin, Oxaliplatin, Pazopanib, Pembrolizumab and Pemetrexed), and added random effects to correct for exposure time and dose for each treatment drug as well as the pharmacogenetics on patient level. We repeated this analysis using the relative mutation count of each mutation type.

Ploidy and copy number analysis

We used PURPLE (Priestley et al. 2018) to obtain high quality somatic ploidy and copy number (CN) regions (https://github.com/hartwigmedical/hmftools/tree/master/purity-ploidy-estimator). Briefly, this tool combines B-allele frequency (BAF), read depth and structural variants to estimate the purity and CN profile of a tumor sample.

Clonality

The determination of the clonality of each mutation was adopted from (Priestley et al. 2018). Briefly, the local ploidy level of each variant was calculated by multiplying the tumor adjusted variant allele score, obtained from PURPLE, with the local copy number level. All variants with a score above 1 are considered as clonal. Variants exhibiting a score lower than 1 were searched for a subclonal peak using a kernel density estimation using a kernel bandwidth of 0.05 after plotting the variant ploidy scores of all variants of a sample. All variants present in the peaks below the peak of ploidy = 1 were considered as subclonal mutations. Samples having at least 500 subclonal mutations and show an overall 5-FU signature contribution (at least 5%) were included for the subclonal analysis.

Estimation of tumor mutational burden

The mutation rate per megabase (Mb) of genomic DNA was calculated as the total genome-wide amount of SBSs divided over the total amount of mappable nucleotides (ACTG) in the human reference genome (hg19) FASTA sequence file:

$$TMB = \frac{(SBS_g)}{\left(\frac{2858674662}{10^6}\right)} \tag{1}$$

In this study, we excluded hypermutant samples (>10 mutations/Mbp), as determined by (Campbell et al. 2017), as hypermutant samples have an impact on both absolute and relative mutation contribution analysis.

Detection of significantly mutated genes

Using all SBS and INDEL variants from protein coding genes, we ran dNdScv (Martincorena et al. 2018) to find significantly mutated genes using all SBSs and INDELs variants from protein coding genes. This model can test the normalized ratio of each non-synonymous mutation type individually (missense, nonsense and splicing) over background (synonymous) mutations whilst correcting for sequence composition and mutational signatures. A global q-value ≤ 0.1 was used to identify statistically significant driver genes. A post hoc Fisher's exact test was performed to evaluate whether the number of mutations of individual genes were enriched between two cohorts.

Transcription and replication strand bias

To compare the replication and transcription strand bias between cohorts, we selected samples with a high COSMIC signature 17 contribution (absolute contribution > 2000 mutations and relative contribution > 25% (5-FU pretreated colon n = 41, 5-FU pretreated breast n = 9, not 5-FU pretreated esophagus n = 34). Next, we selected all the point mutations bearing a C[N>N]T context where N can be any nucleotide, reasoning that the majority of the C[T>G]T mutations can be attributed 5-FU exposure in colon and breast cancer and 5-FU independent mutational processes in esophagus cancer. Mutation types other than C[T>G]T can thus be considered as control.

To assess DNA replication strand, we downloaded replication sequencing (Replic-Seq) data from (Tomkova et al. 2018) that characterized the replication timing profiles from (Haradhvala et al. 2016). As in Tomkova et al, we used replication strand information of 1 Mbp regions near the left and right of each origin. Next, we generated a mutation count matrix 12 (6 trinucleotides × 2 strands) for each sample with replication strand information using Mutational Patterns R package (Blokzijl et al. 2018). After counting the number of mutations on each strand per cancer type and mutation type, a Poisson test for strand asymmetry was performed to test for significance. Similarly, a mutation count matrix of 12 was generated containing transcription strand information of all point mutations with a C[N>N]T context that fall within a gene body. The transcribed units of all protein coding genes are based on Ensembl v75 (hg19) including the introns and untranslated regions. After estimating the mutation rate on the transcribed and non-transcribed strands, also a Poisson test for strand asymmetry was performed to test for significance. This package contains also functions to determine the replication timing. In brief, all point mutations were checked whether these were located in an intermediate, early or late replicating region. Enrichment or depletion analysis of point mutations in these genomic regions was performed using genomic distribution functions from Mutational Patterns R package (Blokzijl et al. 2018).

Association of point mutations with mutational patterns

We estimated which mutational process was most likely at the origin of each point mutation as previously done in (Letouzé et al. 2017). In doing so, we considered the mutation category (substitution type and trinucleotide context (TNC)) and the relative contribution of each mutational signature from each tumor sample. The likelihood of a point mutation, with a certain 96 trinucleotide context (TNC), induced by mutation signature X from a sample Y can be expressed as follows:

$$rel\ TNC_{Sample\ y}^{Sig\ x} = \frac{abs\ Sample^{Sig\ x} * rel\ TNC^{Sig\ x}}{\sum_{Sig}\ abs\ TNC_{sample\ y}^{Sig}} \qquad (2)$$

Where $abs\ Sample^{Sig\ x}$ is the absolute mutation contribution of signature X for that sample; $rel\ TNC^{Sig\ x}$ is the mutation type probability for a given TNC of signature X divided by the sum of the mutation type probability for that TNC of all mutation signatures; and $\sum_{Sig}\ abs\ TNC_{sample\ y}^{Sig}$ is the sum of absolute mutation contribution of that TNC for every signature in sample Y. Overall, the sum of $rel\ TNC_{Sample\ y}^{Sig\ x}$ for every signature of one point mutation from one sample is equal to 1. Subsequently, the relative contribution of a mutational signature to all mutations from multiple samples was retrieved as the cumulative $rel\ TNC_{Sample\ y}^{Sig\ x}$ likelihoods of every mutation of the whole cohort. All mutations with a score of higher than 0.5 for a given signature were considered to be originated from that signature and were fed into dNdSCV for selection analysis.

5-FU induced cancer driver mutation risk

We used quantitative *in vivo* data and qualitative mutational characteristics to model the number of oncogenic mutations as a function of the number of cells, in the absence of negative selection. We applied the following formula:

$$M^{\text{active}}(N) = 0.015 \cdot dp \cdot \mu \cdot N \cdot \sum_{\substack{X \in \{C,T\} \\ Y \in \{A,C,G,T\} \\ X \neq Y}} \left(P_{X \to Y} \cdot \frac{n_{X \to Y}}{L} \right)$$

where M^{active} is number of mutations that activate driver genes, *dp* is depletion in coding sequence (CDS), *μ* is the mutation rate, *N* is number of cells, $P_{X>Y}$ is chance on X>Y mutation based on the mutation spectrum, $n_{X>Y}$ is the number of positions where X>Y mutation result in oncogene activation and *L* is the length of CDS.

We used the following parameters: 1.5% of the genome is exon coding; Mutational depletion (likely due to repair) from the coding sequence is 0.3094464 (results obtained from (Blokzijl et al. 2016)); On average 2000 extra mutations with 5-FU signature per year accumulate in tumors due to 5-FU treatment (data based from this study) - 40 mutations accumulate per year in absence of 5-FU (normal in vivo mutation spectrum, 25% ~ signature 1 & 75% signature 5 - results obtained from (Blokzijl et al. 2016)); Colon cancer originates in one of the 10^8 colon stem cells (Frank 2010); Signature 17 mutation chance with inclusion of trinucleotide context (5-FU pretreated) and signature 1 (25%) + signature 5 (75%) for non 5-FU treated model; List of validated oncogenic mutations (exists of roughly 10,000 tumor suppressor and driver variants, obtained from (Tamborero et al. 2018)); Coding sequence length of small intestinal cells: 22563618 bp; The average duration of a 5-FU treatment regime is 24 weeks (12 cycles consisting of 2 weeks).

Comparison with treated naive cancer cohorts

For the primary breast cancer cohort, we used the publicly available somatic mutations from BASIS cohort which were downloaded from the ICGC data portal on August 2, 2017. This cohort consists of 560 primary breast cancers and has previously been characterized in detail (Nik-Zainal

et al. 2016). Somatic mutations of 41 primary colon cancer samples were kindly shared by Max-Planck-Institute with a signed agreement for data and sample transfer. The SBSs were called using Varscan 2.0 and post filtered with a QSS score above 30. Full description of this cohort can be found in (Schütte et al. 2017). Both cohorts comprise of treatment naive cancer patients.

Statistics

Unless otherwise stated, we performed a Wilcoxon rank-sum test to compare continuous variables (for instance the relative or absolute contribution of mutational signatures versus treated and not treated) and a Fisher's exact test was used to evaluate categorical data (treatment versus the occurrence of a certain mutation). All statistical tests were two-sided and considered statistically significant when $P < 0.05$. R version 3.4.4 was used for the statistical analyses.

Code availability

All code and filtered vcf files from 5-FU treated organoid lines are freely available at https://github.com/UMCUGenetics/5FU.

Data availability

WGS data and corresponding clinical data have been obtained from the Hartwig Medical Foundation and provided under data request number DR-047. Both WGS and clinical data is freely available for academic use from the Hartwig Medical Foundation through standardized procedures and request forms can be found at https://www.hartwigmedicalfoundation.nl. The human sequencing data of the 5-FU treated and control organoid lines have been deposited at the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/) under accession numbers EGAS00001003592 and EGAS00001002955, respectively.

Acknowledgements
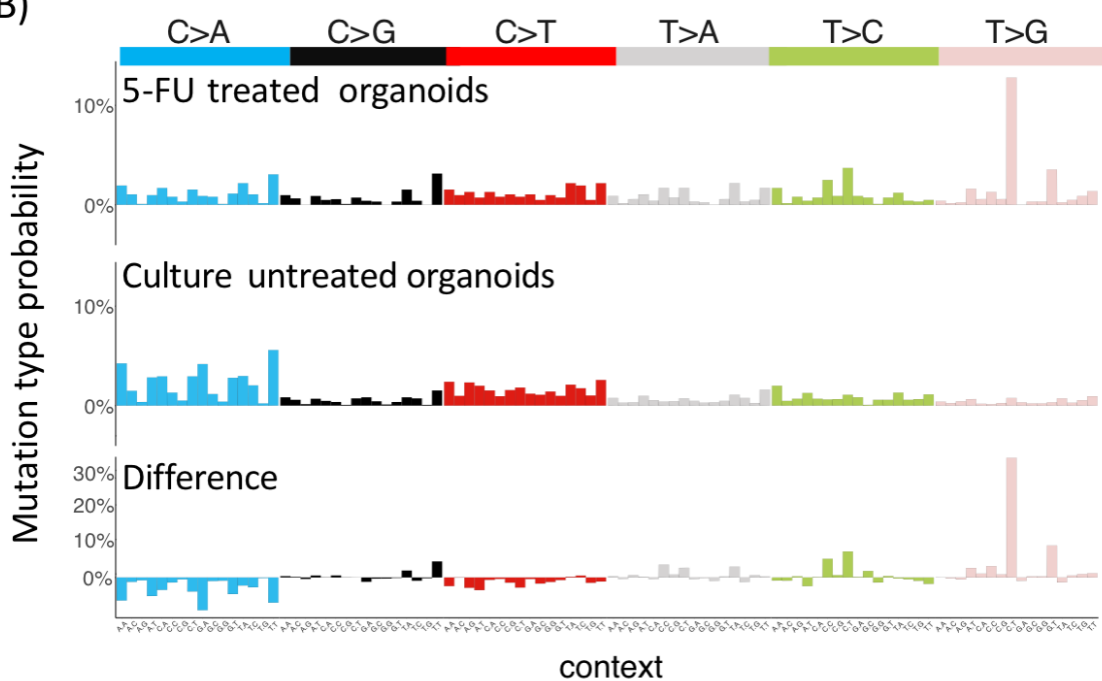
Grants

Figures

(A)
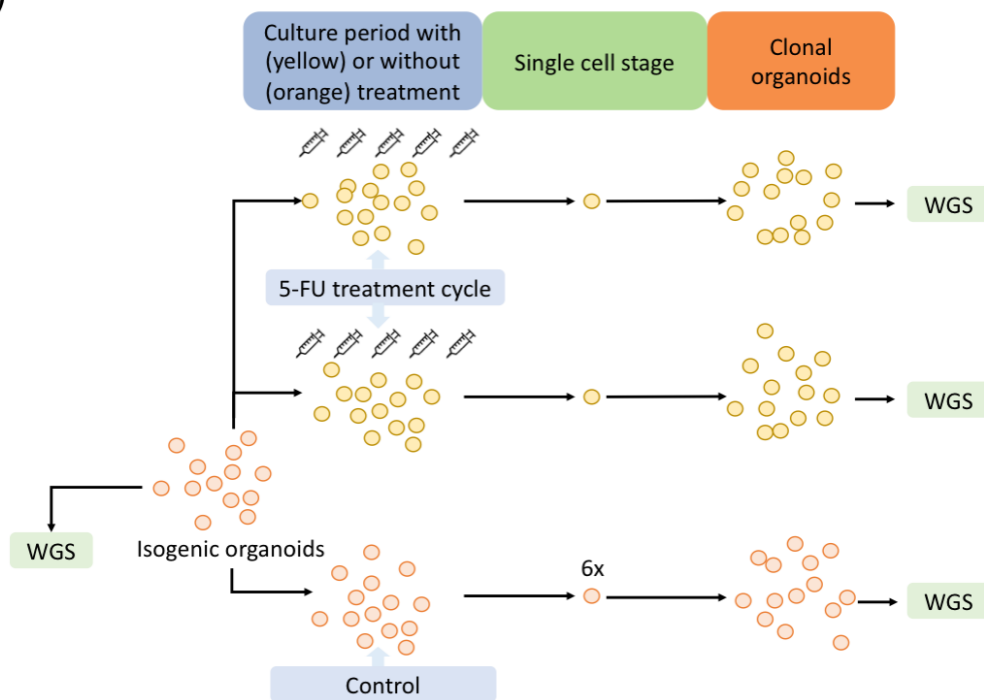
(B)

Fig 1. **5-FU induces context dependent T>G mutations *in vitro*.** (A) Schematic overview of the experimental setup used to determine the 5-FU mutation spectrum in two independent human small intestinal organoid experiments. 6.25 μM 5-FU was added to isogenic organoids for 3 days, followed by a 4 day rest period. This cycle was repeated 5 times. Subsequently, organoids were made single cell and expanded further into clonal organoids to obtain sufficient DNA for WGS. Controls were cultured in 5-FU-free medium. The WGS data of the original isogenic organoid line served as reference sample. (B) The experimentally derived mutation spectra from 5-FU treated organoid lines (upper) and untreated organoid lines (middle). Each spectrum shows the mutation probability of each indicated context-dependent base substitution type. The spectrum below shows the difference between the 5-FU (positive values) and the *in-vitro* (negative values) mutation spectrum.
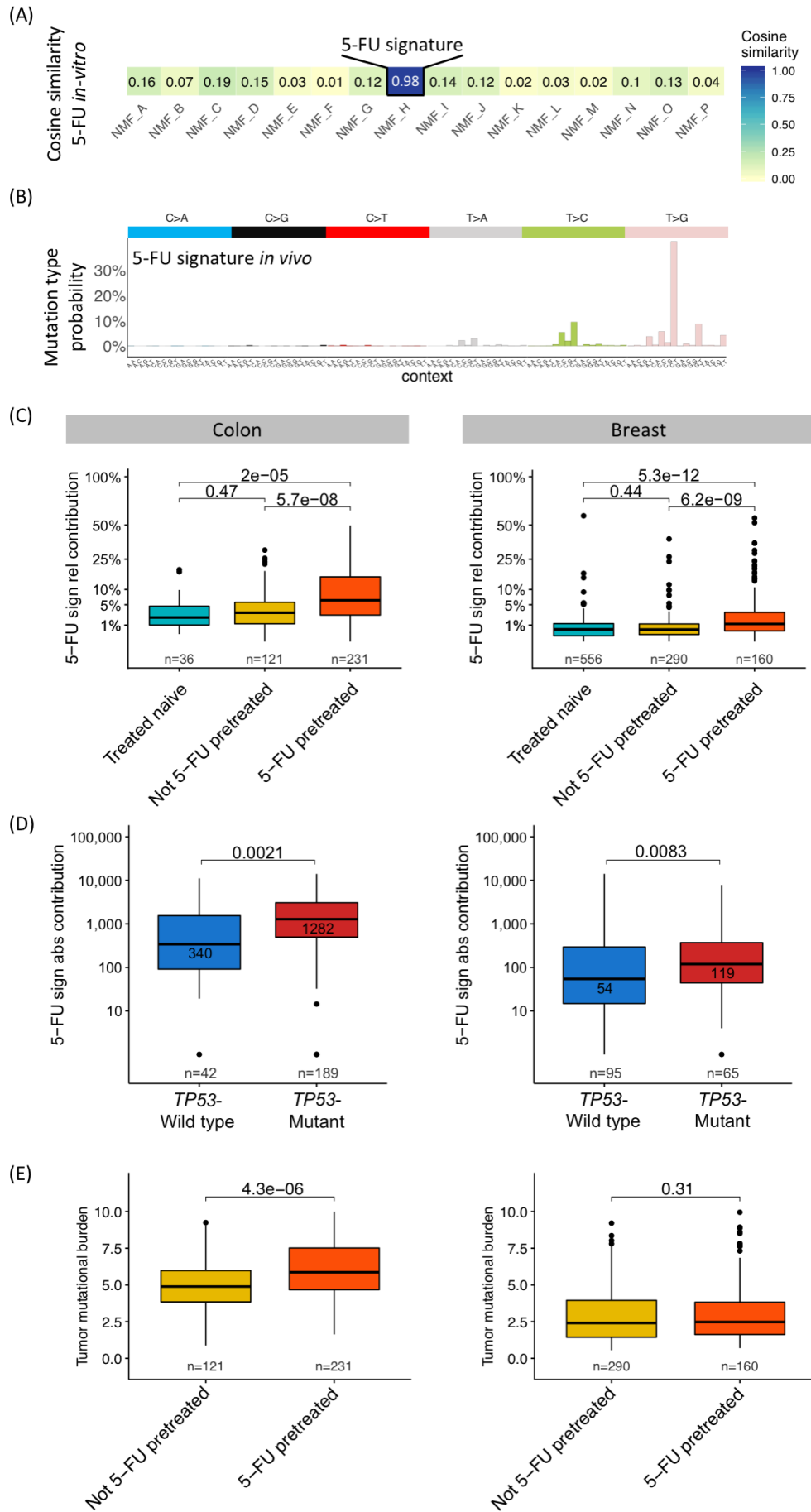
Fig 2. **5-FU mutational pattern and its contribution in human cancer.** (A) Heatmap showing the cosine similarity scores for each *de novo* extracted signature with the *in vitro* experimental obtained 5-FU mutation spectrum. NMF_H resembles the 5-FU experimental mutation spectrum (cos sim = 0.98) and is further assigned as the "5-FU signature" in the main text. (B) 5-FU mutation signature showing the mutation type probability for each context-dependent base substitution type. (C) Box-and whisker plots indicating the relative contribution of the 5-FU signature between 5-FU pretreated and not 5-FU pretreated colon (left) and breast (right) cancer patients with inclusion of the treatment naive cancer cohort. (D) Box-and whisker plots showing the 5-FU mutational load between *TP53*-wild type and *TP53*-mutant cancers in 5-FU pretreated colon (left) and breast (right) patients. (E) Box-and whisker plots showing the tumor mutational burden (number of SBSs per Mbp) between 5-FU pretreated and not 5-FU pretreated cancer patients for the colon (left) and breast (right) cancer patients. For all plots, a Wilcoxon rank-sum test between every cohort was performed and the p-value is illustrated at the top of the plots.
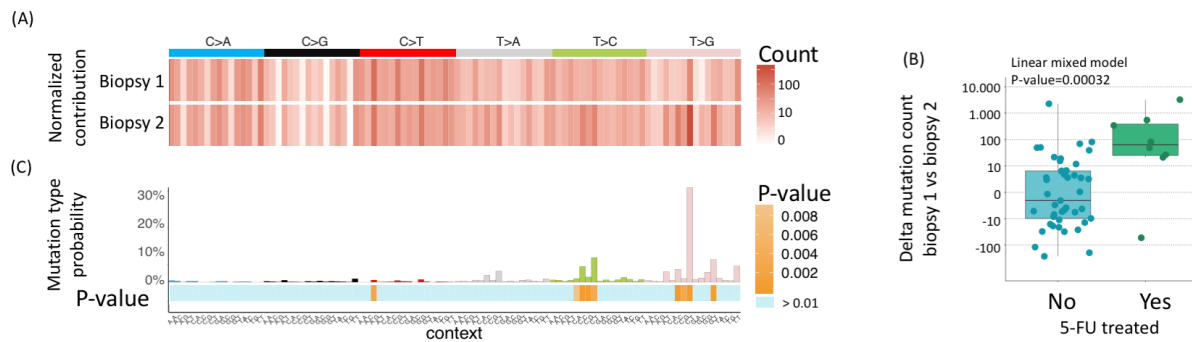
Fig 3. **Mutational enrichment analysis for patients with multiple biopsies in 5-FU treated and 5-FU untreated patients.** (A) Example heat map of one patient showing the normalized mutation count of every mutation type from the first (above) and second (below) biopsy. This normalization step was performed on both samples of each patient. (B) Linear mixed model regression analysis on the normalized mutation counts of one mutation type (here T[T>G]C mutations) between patients that received a 5-FU treatment between the two biopsies and patients not treated with 5-FU between two biopsies (see also S Fig 7). In the model, we controlled for exposure dose and time as well as other therapies that were administered to the patient between the first and second biopsy. P-values were obtained by performing an ANOVA test on the regression model. (C) Bar plot showing the mutation type probability for COSMIC signature 17 with below the obtained p-values from the linear mixed model for every mutation type. Note that most of the mutation types that characterizes COSMIC signature 17 show a significant increase in normalized mutation count for patients treated with 5-FU between both biopsies.
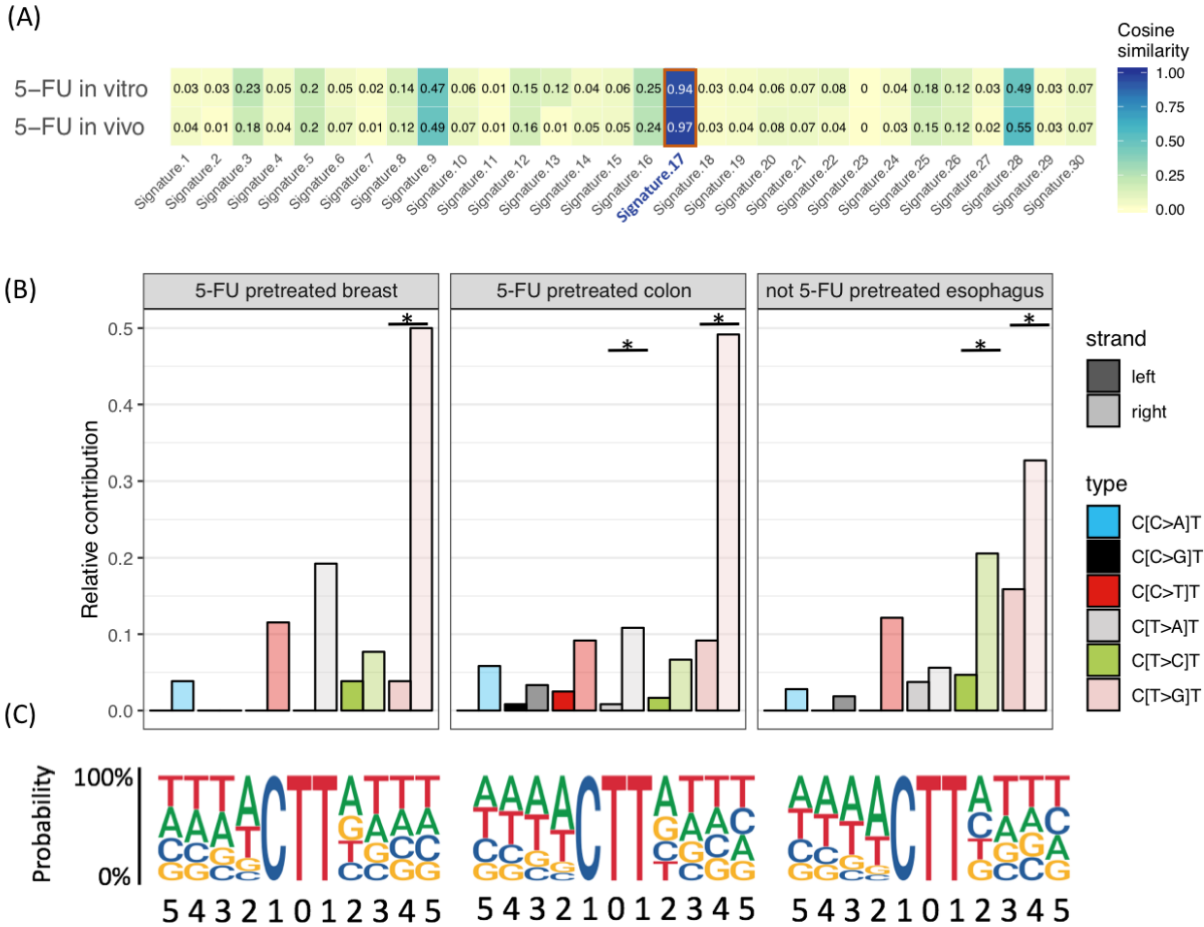
Fig 4. **Comparison between the 5-FU signature and COSMIC Signature 17.** (A) Heatmap showing the cosine similarity scores for the *in-vitro* 5-FU mutation spectrum and the *in-vivo* obtained 5-FU signature with the COSMIC signatures. Both patterns show a strong resemblance with COSMIC Signature 17. (B) Replication strand bias of C[N>N]T mutations in 5-FU pretreated colon and breast samples and not 5-FU pretreated esophagus samples. Relative levels of each base substitution type in the left (leading) and right (lagging) DNA strands are shown for each cohort. Asterisks indicate a significant difference (P < 0.05, two-sided Poisson test). (C) The eleven-base signature context of C[T>G]T mutations are presented as Logo plots. The mutated T is centered in each plot with fixed positions left (5' direction) and right (3' direction) from the mutation position.

Supplementary tables and figures

signature_denovo	COSMIC	Cosine similarity	Alexandrov 2018	Cosine similarity	Boot.et.al	Cosine similarity
NMF_A		NA	SBS41	0,71	Ant	0,9
NMF_B	Signature.6	0,76	SBS44	0,87		
NMF_C	Signature.15	0,81	SBS15	0,71		
NMF_D		NA	SBS34	0,62		
NMF_E	Signature.1	0,97	SBS1	0,97		
NMF_F	Signature.18	0,93	SBS18	0,96		
NMF_G	Signature.3	0,9	SBS37	0,89		
NMF_H	Signature.17	0,97	SBS17b	0,97		
NMF_I	Signature.8	0,78	SBS8	0,81		
NMF_J	Signature.13	1	SBS13	0,95		
NMF_K	Signature.2	1	SBS2	1		
NMF_L	Signature.20	0,77	SBS44	0,87		
NMF_M	Signature.7	0,97	SBS7a	0,98		
NMF_N	Signature.10	0,92	SBS10a	0,78		
NMF_O	Signature.5	0,51	SBS5	0,51		
NMF_P	Signature.29	0,82	SBS29	0,82		

S Table 1. **Comparison of *de novo* signatures with COSMIC signatures.** *De novo* signatures are compared with the highest cosine similarity to the i) COSMIC signatures (first column), ii) signatures as reported in Alexandrov et al., (second column) and iii) signature as reported in Boot et al., (last column).

Cohort	chromosome	position	ref	alt	gene	clonality	trinucl_context	type	canonical	ref_original	alt_original	moa	driver
5-FU pretreated COLON	chr12	25380275	T	G	KRAS	SUBCLONAL	CTT	SNP	missense variant	T	G	Act	Yes
5-FU pretreated COLON	chr3	178936091	C	T	PIK3CA	SUBCLONAL	TCA	SNP	missense variant	G	A	Act	Yes
5-FU pretreated COLON	chr3	178936095	T	G	PIK3CA	SUBCLONAL	CTG	SNP	missense variant	A	C	Act	Yes
5-FU pretreated COLON	chr4	153247366	C	T	FBXW7	SUBCLONAL	TCG	SNP	missense variant	C	T	LoF	Yes
5-FU pretreated COLON	chr5	112173704	C	T	APC	SUBCLONAL	TCG	SNP	stop gained	C	T	LoF	Yes
Not 5-FU pretreated COLON	chr17	7577539	C	T	TP53	SUBCLONAL	CCG	SNP	missense variant	G	A	LoF	Yes
Not 5-FU pretreated COLON	chr4	153249385	C	T	FBXW7	SUBCLONAL	GCG	SNP	missense variant	G	A	LoF	Yes
5-FU pretreated BREAST	chr3	138417865	C	T	PIK3CB	SUBCLONAL	TCA	SNP	missense variant	C	T	Act	Yes
5-FU pretreated BREAST	chr3	178936082	C	T	PIK3CA	SUBCLONAL	TCA	SNP	missense variant	G	A	Act	Yes
5-FU pretreated BREAST	chr3	178952085	T	C	PIK3CA	SUBCLONAL	ATG	SNP	missense variant	A	G	Act	Yes
5-FU pretreated BREAST	chr6	152419920	T	G	ESR1	SUBCLONAL	CTC	SNP	missense variant	T	G	Act	Yes
5-FU pretreated BREAST	chr6	152419926	T	C	ESR1	SUBCLONAL	GTC	SNP	missense variant	A	G	Act	Yes
Not 5-FU pretreated BREAST	chr10	89711893	C	T	PTEN	SUBCLONAL	TCA	SNP	stop gained	C	T	LoF	Yes
Not 5-FU pretreated BREAST	chr3	178916891	C	T	PIK3CA	SUBCLONAL	CCG	SNP	missense variant	G	A	Act	Yes
Not 5-FU pretreated BREAST	chr3	178936091	C	T	PIK3CA	SUBCLONAL	TCA	SNP	missense variant	G	A	Act	Yes
Not 5-FU pretreated BREAST	chr6	152419923	T	G	ESR1	SUBCLONAL	ATA	SNP	missense variant	A	C	Act	Yes
Not 5-FU pretreated BREAST	chr6	152419926	T	C	ESR1	SUBCLONAL	GTC	SNP	missense variant	A	G	Act	Yes

S Table 2. **Subclonal driver mutations caused by the 5-FU mutational process.** Patient cancer type and 5-FU treatment status are indicated in the first column, followed by the chromosomal position (columns 2 and 3), the mutation (with reference in column 4 and mutation in column 5) and cancer driver gene (column 6). The clonality, trinucleotide context and mutation type are summarized in columns 7-14.
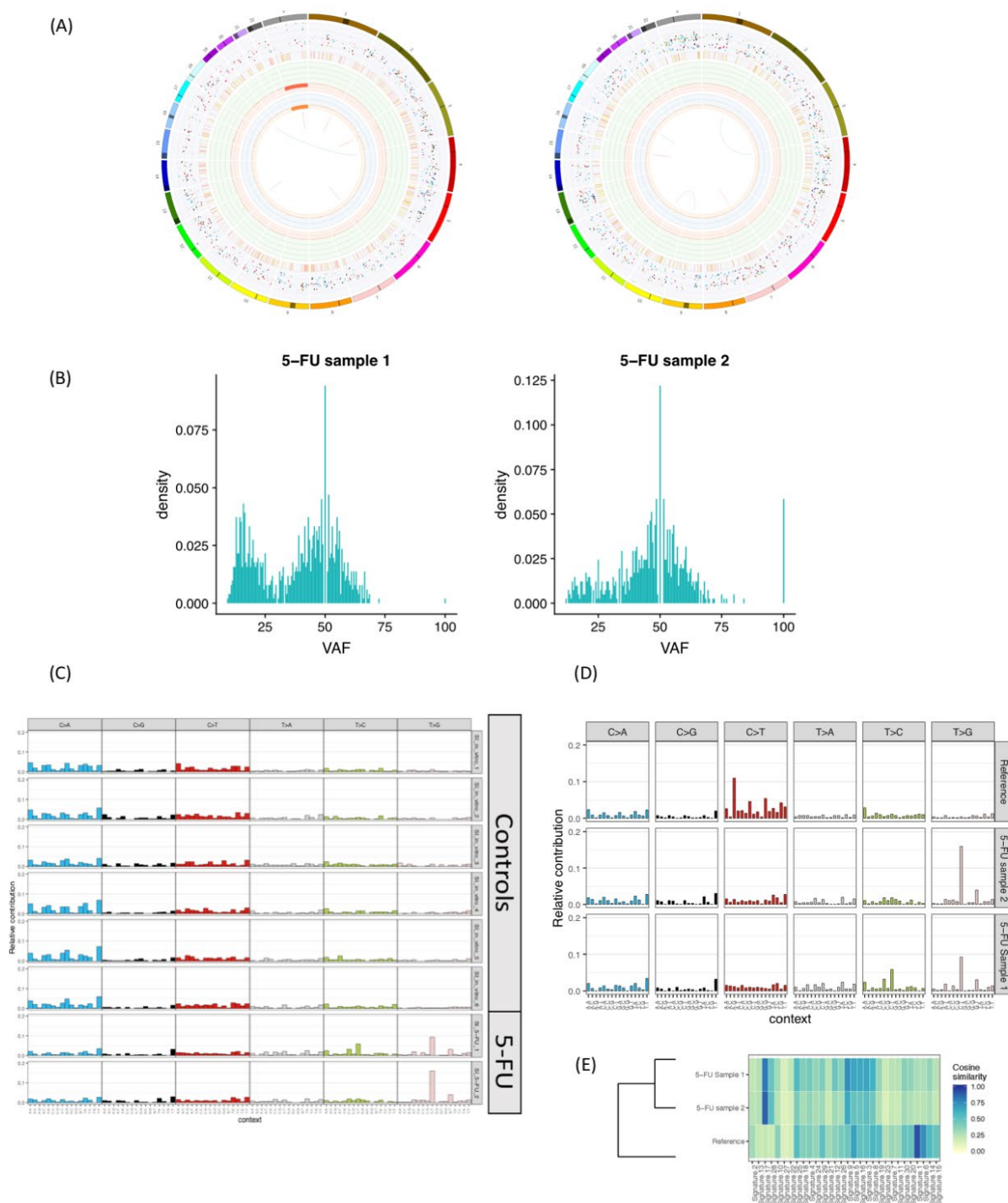
S Fig 1. **Analysis of mutations in 5-FU treated organoids.** (A) Circos plots showing the somatic events of the two organoid lines treated with 5-FU. The outer first circle shows the chromosomes. The second circle shows the somatic variants (incl. exon, intron and intergenic regions). Somatic variants are further divided into an outer ring of SBSs and an inner ring of INDELs. Each dot represents a single somatic variant scaled from to 100% by its allele frequency score. SBSs are colored according to the type of base change (e.g. C>T/G>A in red)

and are in concordance with the coloring used in Alexandrov et al. 2013 Nature paper that describes the use of mutational signatures. INDELs are colored yellow and red for insertions and deletions respectively. The third circle shows all observed tumor purity adjusted copy number changes. Copy number losses are indicated in red, green shows regions of copy number gain. The scale ranges from 0 (complete loss) to 6 (high level gains). If the absolute copy number is > 6 it is shown as 6 with a green dot on the diagram. The fourth circle represents the observed 'minor allele copy numbers' across the chromosome. The range of the chart is from 0 to 3. The expected normal minor allele copy number is 1, and anything below 1 is shown as a loss (orange) and represents a LOH event. Minor allele copy numbers above 1 (blue) indicate amplification events of both A and B alleles at the indicated locations. The innermost circle displays the observed structural variants within or between the chromosomes. Translocations are indicated in blue, deletions in red, insertions in yellow, tandem duplications in green and inversions in black. (B) Histogram showing the SBS allele frequencies of all SBSs. SBSs with a VAF score between 0.3 and 0.7 are considered as clonal which were used for *de novo* mutational pattern characterization. (C) Mutational spectra of all SBSs for each human intestinal organoid line used for 5-FU mutational pattern characterization. The six upper samples are non-exposed control organoid lines, while the two samples below are 5-FU exposed organoids. Different mutation types and the direct sequence context are indicated. (D) Mutational spectra of all SBSs for the isogenic organoid line (using matching blood DNA as reference) and the 5-FU exposed organoid lines. (E) Heat map showing the cosine similarity scores for each indicated sample and all COSMIC signatures.
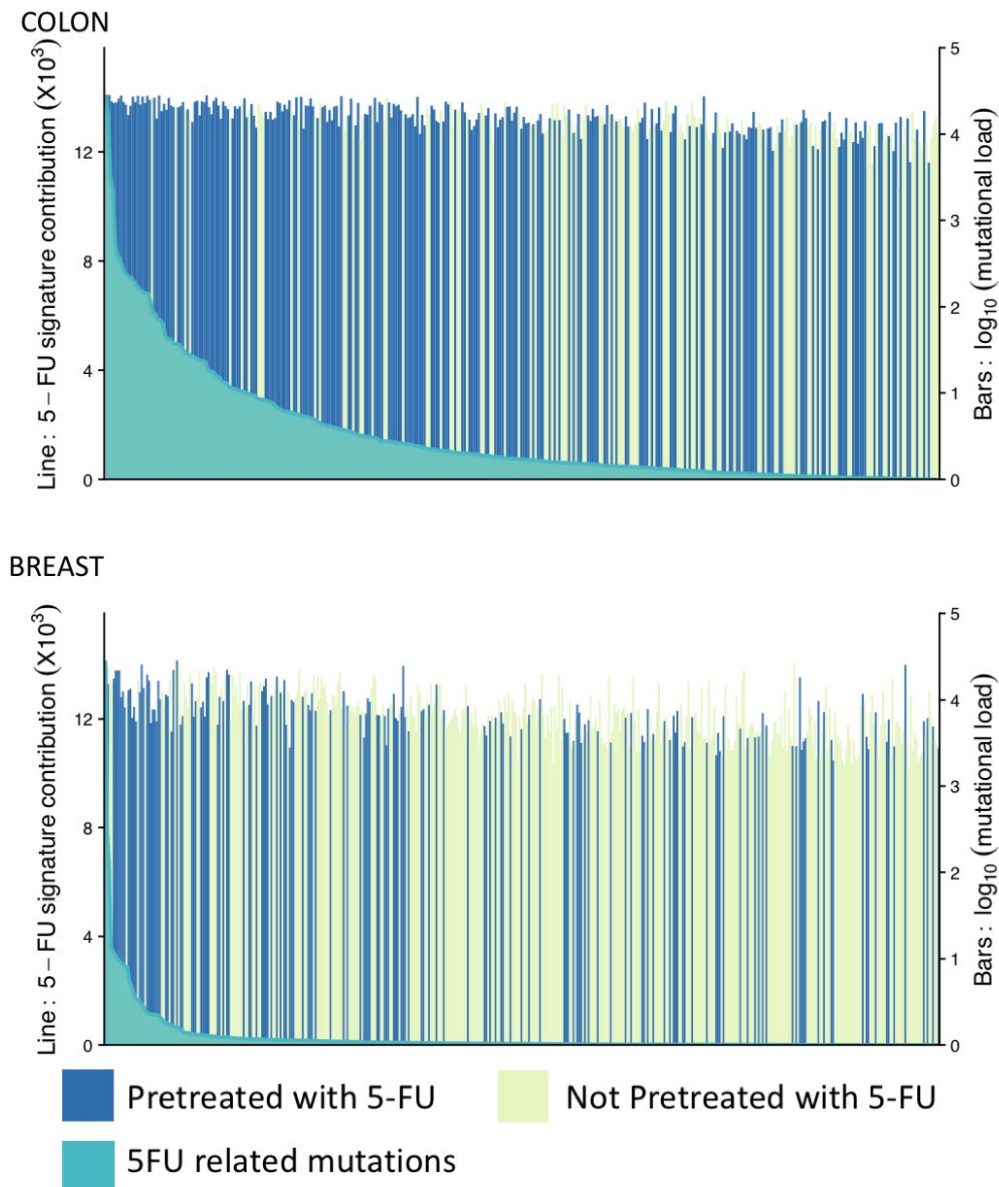
S Fig 2. **5-FU signature contributions in colon and breast cancer cohort.** Mutational distribution plot of each sample for the colon (upper) and breast (lower) cancer cohort. The green bars show the absolute mutation contribution of the 5-FU signature whereas the blue (5-FU pretreated patients) and yellow (not 5-FU pretreated patients) bars illustrate the overall tumor mutational burden for each sample. The samples have been ordered according to their 5-FU signature contribution.
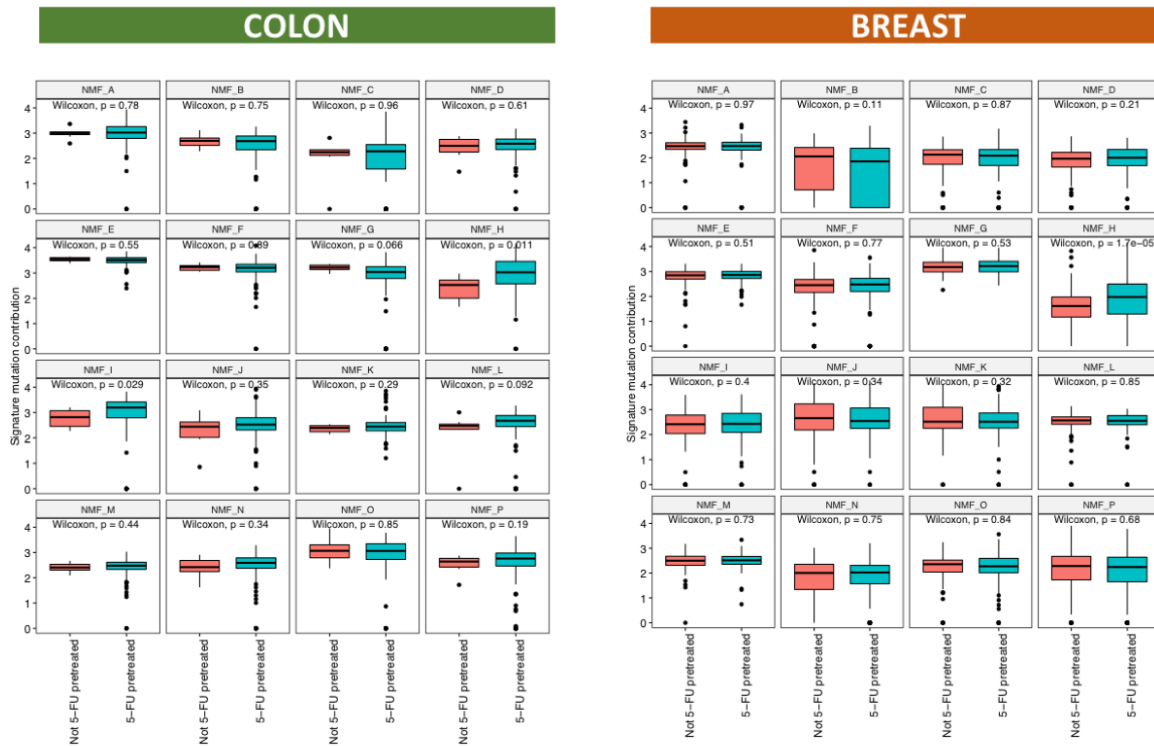
S Fig 3. **Absolute contributions of *de novo* signatures extracted using NMF.** Absolute mutational contribution for each *de novo* signature in the colon (left) and breast (right) cancer cohort. NMF_H, the only signature significantly increased upon 5-FU treatment in both cohorts, resembles the 5-FU *in-vitro* mutation spectrum and is assigned to "5-FU signature" in the main text.
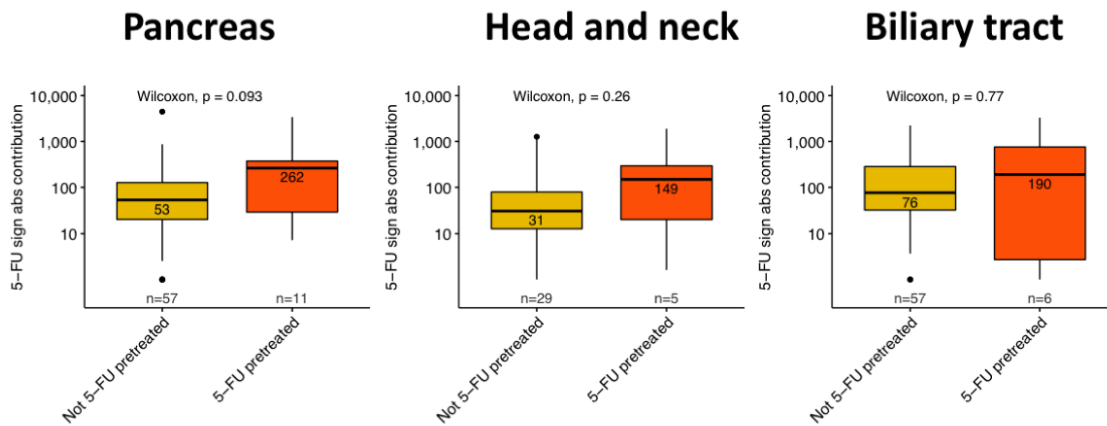
S Fig 4. Comparison of 5-FU and capecitabine treated patients with untreated patients. Box- and whisker plots indicating the absolute contribution for the 5-FU obtained *de novo* pattern between 5-FU pretreated and not 5-FU pretreated cancer patients for different cancer types. A Wilcoxon rank-sum test between every cohort was performed and each p-value is illustrated at the top of the plots.
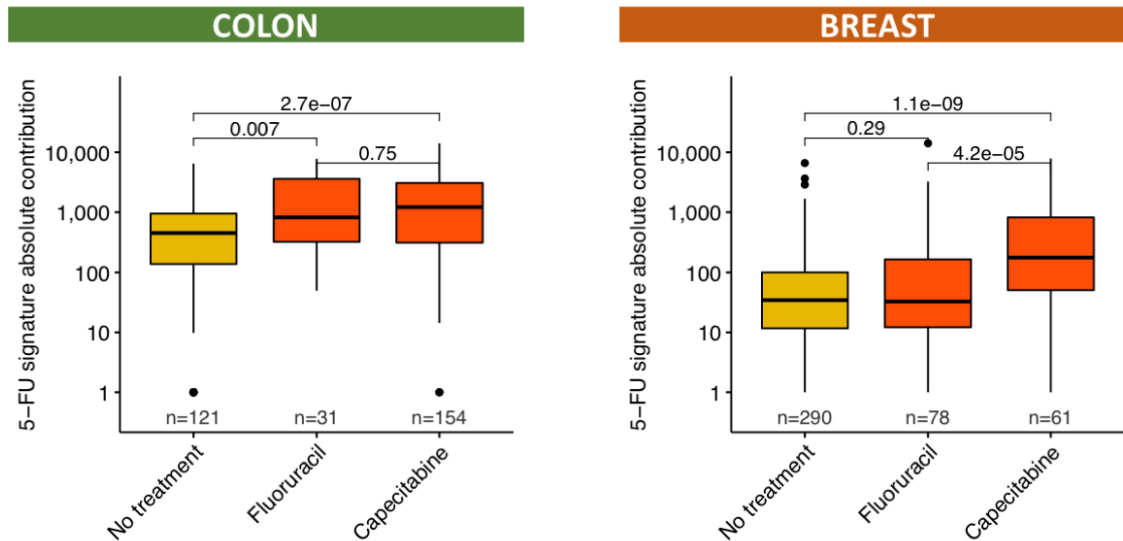
S Fig 5. **Comparison of 5-FU and capecitabine treated patients with untreated patients.**
Box- and whisker plot indicating the absolute contribution for the 5-FU obtained *de novo* pattern between patients pretreated with fluorouracil or capecitabine and not pretreated cancer patients. A Wilcoxon rank-sum test between every cohort was performed and each p-value is illustrated at the top of the plots.
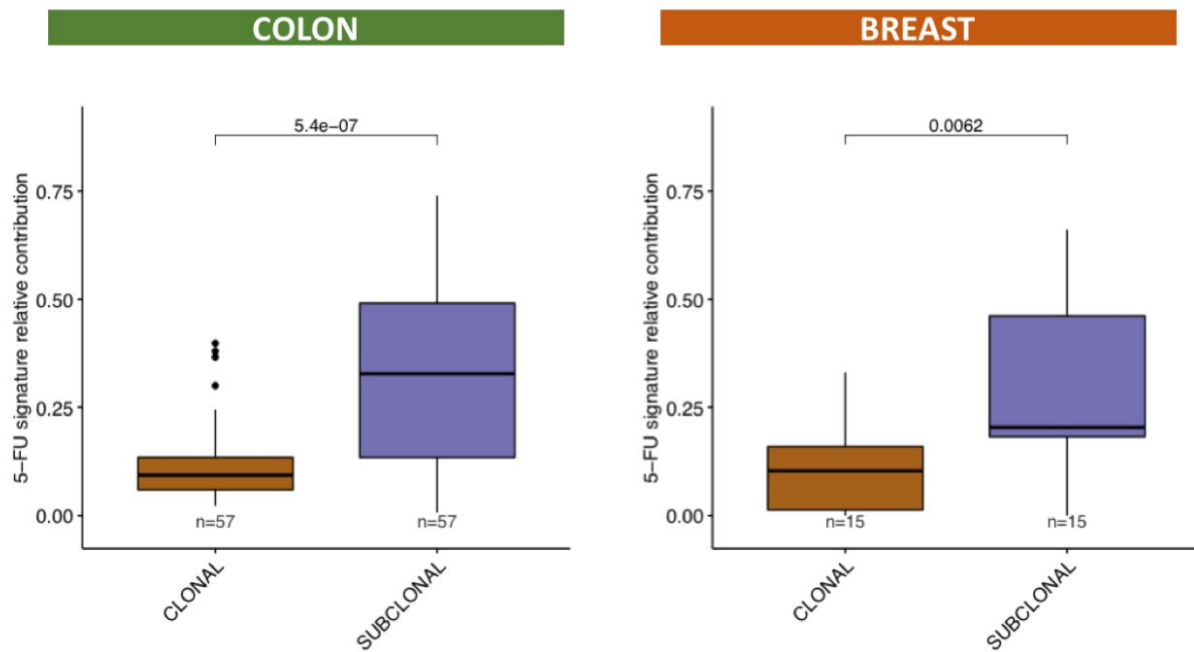
S Fig 6. **Clonal and subclonal analysis for 5-FU mutations.** Box-and whisker plot indicating the relative contribution of 5-FU signature between clonal and subclonal mutations from 5-FU pretreated cancer patients. The left plot is obtained from the colon cancer cohort while the plot illustrated on the right is obtained from breast cancer cohort. A Wilcoxon rank-sum test between every cohort was performed and each p-value is illustrated at the top of the plots.
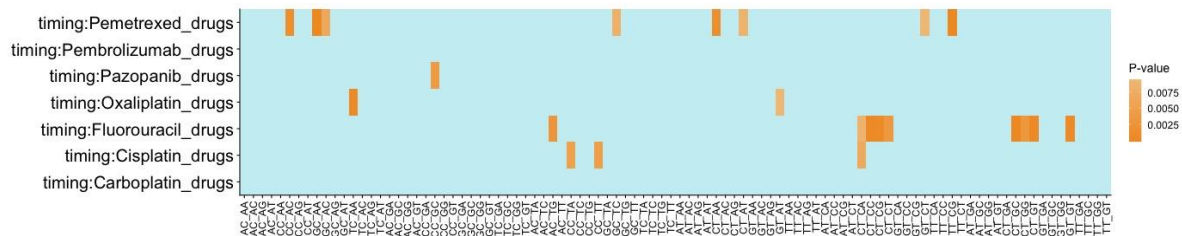
S Fig 7. **Significance of mutations in patients treated with several drugs between biopsies.** (A) Heatmap showing the p-values of the linear mixed model regression analysis on the normalized mutation counts of each mutation type (x-axis) and each treatment type (y-axis) between patients who received 5-FU treatment between the two biopsies and patients not treated with 5-FU between two biopsies. In the model, we controlled for exposure dose and time as well as other therapies that were administered to the patient between the first and second biopsy. P-values were obtained by performing an ANOVA test on the regression model.
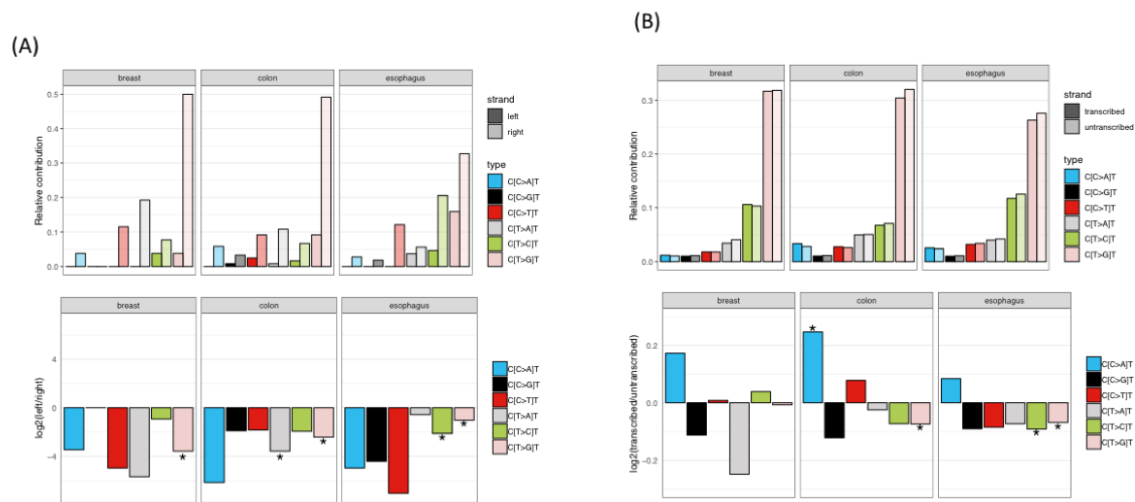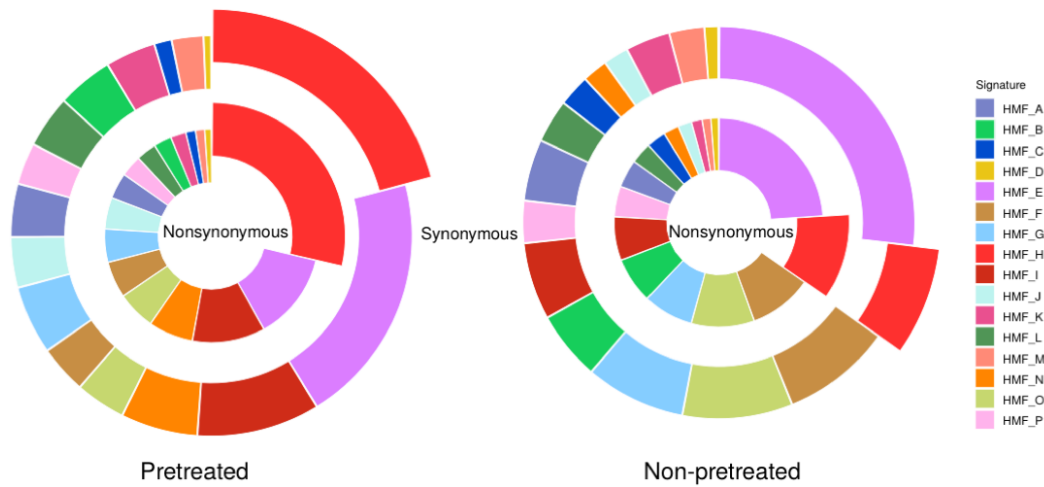
S Fig 8. **Comparison of signature 17-like signatures in 5-FU treated and untreated cancers.** (A) Replication strand bias of C[N>N]T mutations in 5-FU pretreated colon and breast samples and not 5-FU pretreated esophagus samples. Relative levels of each base substitution type in the left (leading) and right (lagging) DNA strands are shown for each cohort. The log2 ratio of the number of SBSs on the left and right strand shows the effect size. Asterisks indicate a significant difference (P < 0.05, two-sided Poisson test). (B) Transcription strand bias of C[N>N]T mutations in 5-FU pretreated colon and breast samples and not 5-FU pretreated esophagus samples. Relative levels of each base substitution type in the transcribed and untranscribed DNA strands are shown for each cohort. The log2 ratio of the number of the number of SBSs on the transcribed and untranscribed strand shows the effect size. Asterisks indicate a significant difference (P < 0.05, two-sided Poisson test).
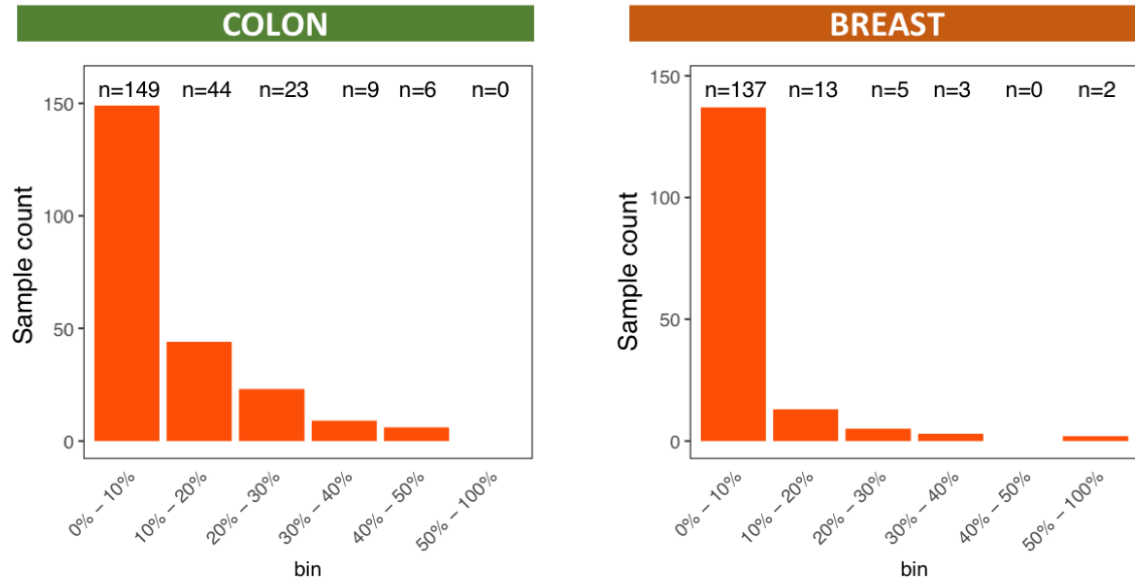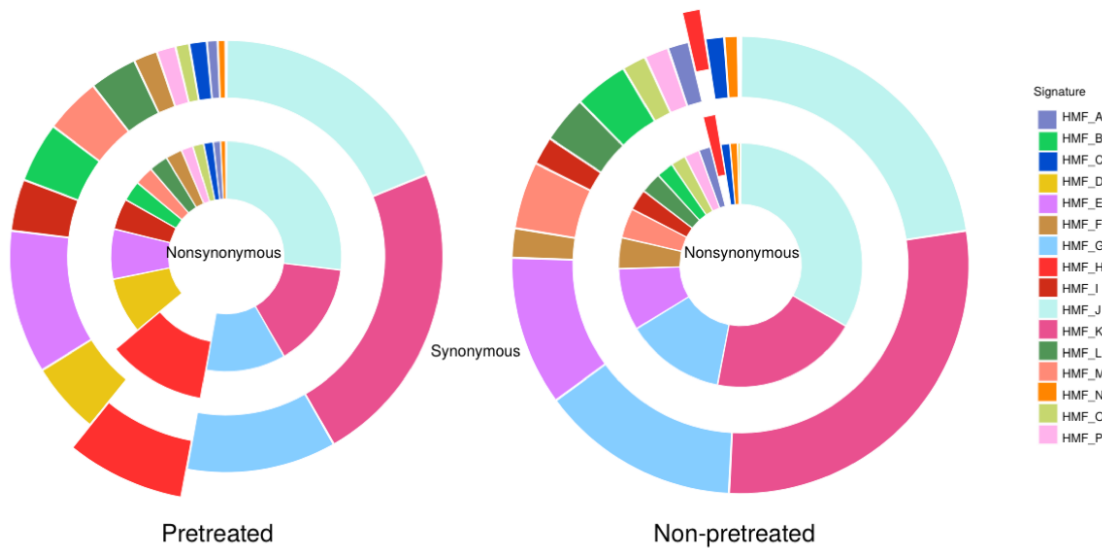
S Fig 9. **Impact of 5-FU treatment on the tumor mutational burden.** Histograms illustrating the relative 5-FU contribution to its respectively tumor mutational burden for all the colon (left) and breast (right) cancer patients.

S Fig 10. **Distribution of the SBSs associated with mutational patterns.** The probability of each SBS being due to each mutational process was estimated. The probabilities over all the SBSs were summed per mutational pattern to obtain the cumulative probabilities across all mutational patterns illustrated in pie charts. NMF_H was identified as 5-FU signature. The selected mutations, which are considered to be induced by 5-FU treatment, are used for dN/dS analysis.
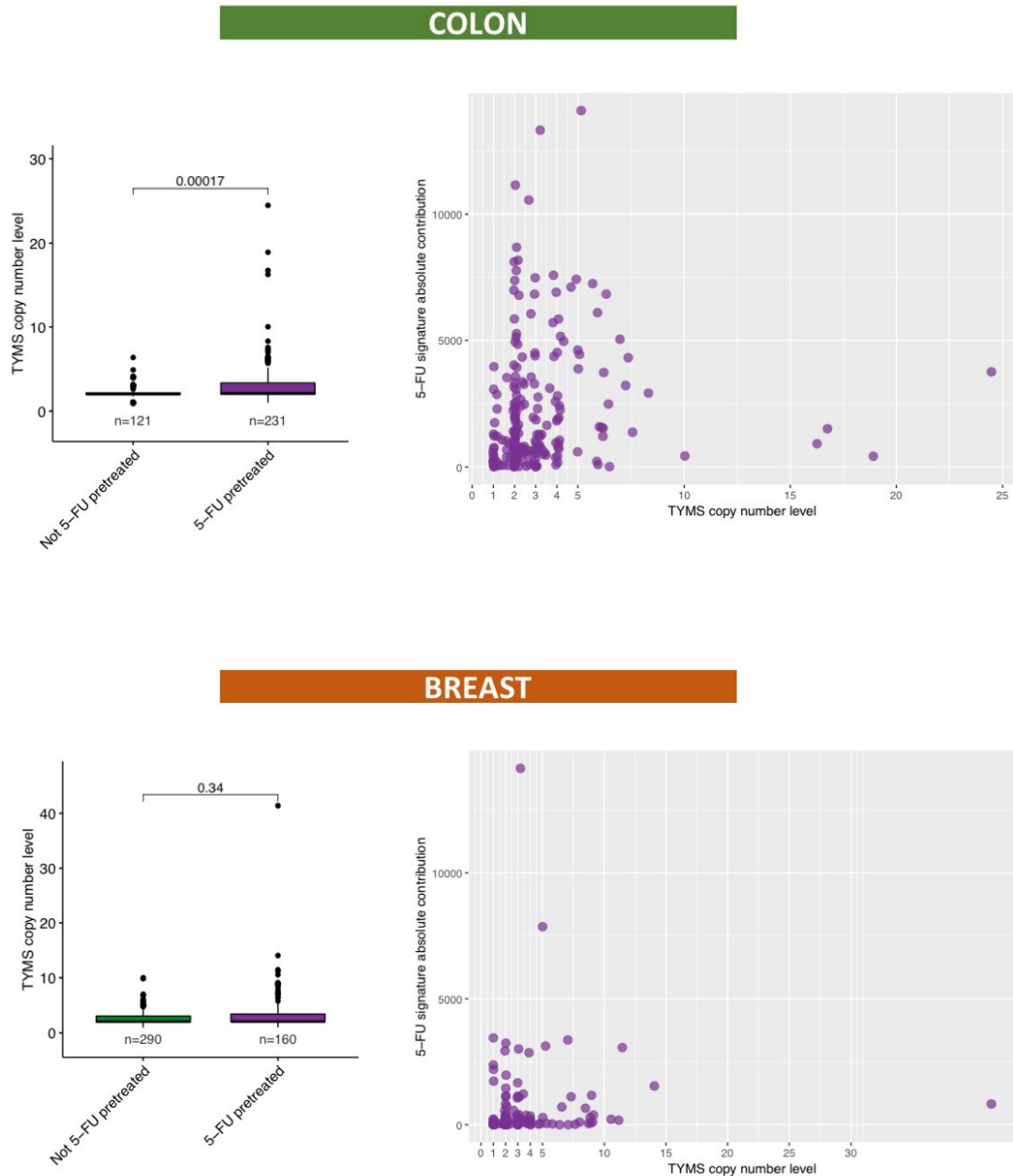
S Fig 11. **Analyses of TYMS copy number related to 5-FU treatment and NMF_H.** (A) Box-and whisker plot indicating the TYMS copy number levels between 5-FU pretreated and not 5-FU pretreated cancer patients. A Wilcoxon rank-sum test between every cohort was performed and each p-value is illustrated at the top of the plots. (B) Scatterplot showing the TYMS copy number levels to the absolute contribution of NMF_5FU.
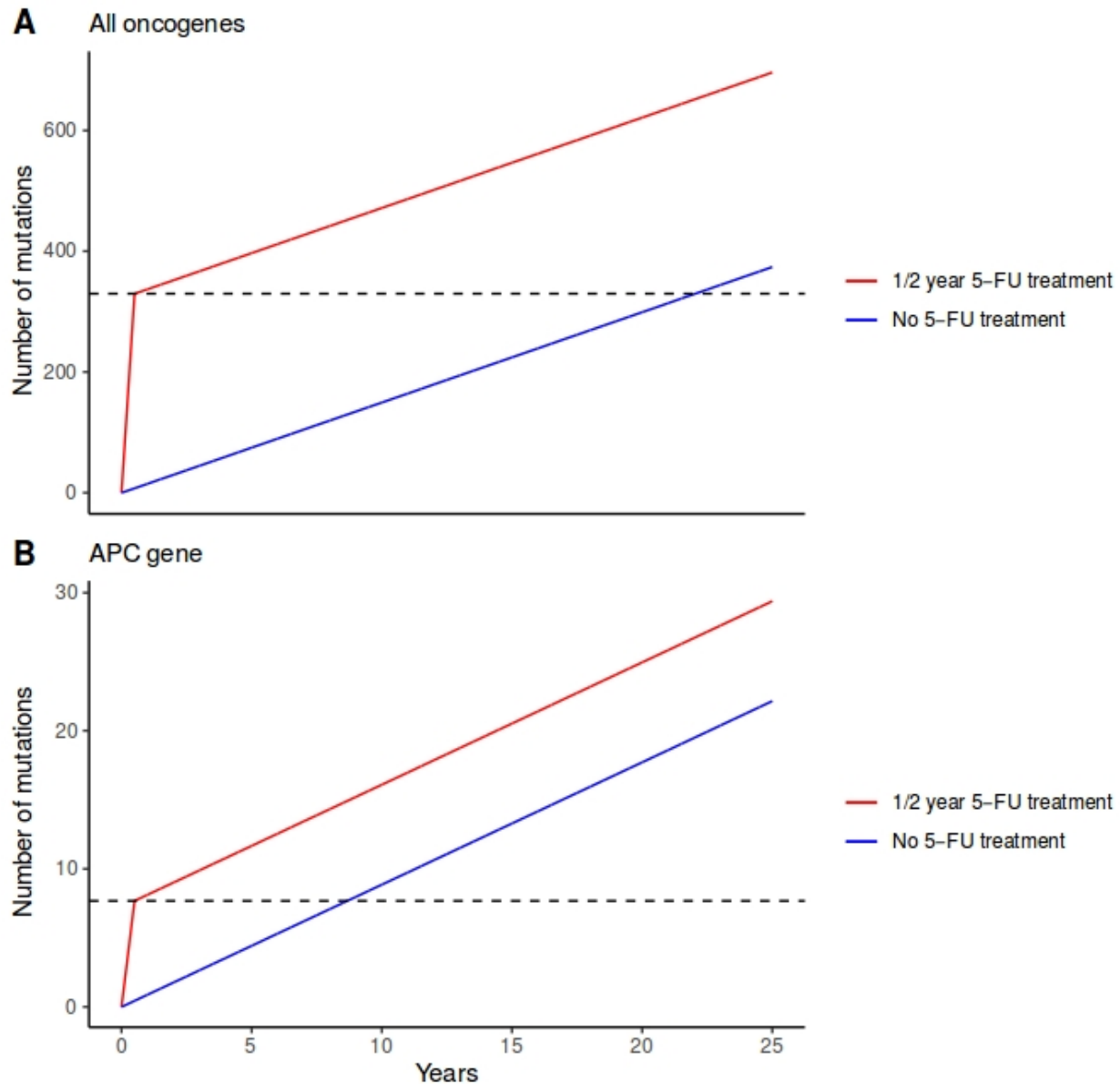
S Fig 12. **Occurrences of oncogenic mutations after half a year of 5-FU treatment.** (A) Oncogenic mutation accumulation in colon stem cells of one full cycle of 5-FU treatment (red, half a year treatment [until dashed line] and subsequent 'normal' oncogenic mutation accumulation) versus 'normal' oncogenic mutation accumulation (blue). Half a year treatment of mutation accumulation is worth 20 years of 'normal' mutation accumulation, represented by the dashed line. (B) Oncogenic mutation accumulation in *APC* of one full cycle of 5-FU treatment (red, half a year treatment [until dashed line] and subsequent 'normal' oncogenic mutation accumulation) reflects 'normal' oncogenic mutation accumulation (blue) in this gene of colon stem cells of about 10 years.

Citations

Alexandrov, Ludmil B., Serena Nik-Zainal, David C. Wedge, Samuel A. J. R. Aparicio, Sam Behjati, Andrew V. Biankin, Graham R. Bignell, et al. 2013. "Signatures of Mutational Processes in Human Cancer." *Nature* 500 (7463): 415–21.

Alexandrov, Ludmil, Jaegil Kim, Nicholas J. Haradhvala, Mi Ni Huang, Alvin W. T. Ng, Arnoud Boot, Kyle R. Covington, et al. 2018. "The Repertoire of Mutational Signatures in Human Cancer." https://doi.org/10.1101/322859.

An, Qian, Peter Robins, Tomas Lindahl, and Deborah E. Barnes. 2007. "5-Fluorouracil Incorporated into DNA Is Excised by the Smug1 DNA Glycosylase to Reduce Drug Cytotoxicity." *Cancer Research* 67 (3): 940–45.

Behjati, Sam, Gunes Gundem, David C. Wedge, Nicola D. Roberts, Patrick S. Tarpey, Susanna L. Cooke, Peter Van Loo, et al. 2016. "Mutational Signatures of Ionizing Radiation in Second Malignancies." *Nature Communications* 7 (September): 12605.

Berger, Franklin G., and Sondra H. Berger. 2006. "Thymidylate Synthase as a Chemotherapeutic Drug Target: Where Are We after Fifty Years?" *Cancer Biology & Therapy* 5 (9): 1238–41.

Bertucci, François, Charlotte K. Y. Ng, Anne Patsouris, Nathalie Droin, Salvatore Piscuoglio, Nadine Carbuccia, Jean Charles Soria, et al. 2019. "Genomic Characterization of Metastatic Breast Cancers." *Nature* 569 (7757): 560–64.

Blokzijl, Francis, Roel Janssen, Ruben van Boxtel, and Edwin Cuppen. 2018. "MutationalPatterns: Comprehensive Genome-Wide Analysis of Mutational Processes." *Genome Medicine* 10 (1): 33.

Blokzijl, Francis, Joep de Ligt, Myrthe Jager, Valentina Sasselli, Sophie Roerink, Nobuo Sasaki, Meritxell Huch, et al. 2016. "Tissue-Specific Mutation Accumulation in Human Adult Stem Cells during Life." *Nature* 538 (7624): 260–64.

Boige, Valérie, Jean Mendiboure, Jean-Pierre Pignon, Marie-Anne Loriot, Marine Castaing, Michel Barrois, David Malka, et al. 2010. "Pharmacogenetic Assessment of Toxicity and Outcome in Patients with Metastatic Colorectal Cancer Treated with LV5FU2, FOLFOX, and FOLFIRI: FFCD 2000-05." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 28 (15): 2556–64.

Boot, Arnoud, Mi Ni Huang, Alvin W. T. Ng, Szu-Chi Ho, Jing Quan Lim, Yoshiiku Kawakami, Kazuaki Chayama, Bin Tean Teh, Hidewaki Nakagawa, and Steven G. Rozen. 2018. "In-Depth Characterization of the Cisplatin Mutational Signature in Human Cell Lines and in Esophageal and Liver Tumors." *Genome Research* 28 (5): 654–65.

Boot, Arnoud, Alvin W. T. Ng, Fui Teen Chong, Daniel S. W. Tan, N. Gopalakrishna Iyer, and Steven G. Rozen. 2018. "Mutational Signature Analysis of Asian OSCCs Reveals Novel Mutational Signature with Exceptional Sequence Context Specificity." https://doi.org/10.1101/368753.

Broutier, Laura, Amanda Andersson-Rolf, Christopher J. Hindley, Sylvia F. Boj, Hans Clevers, Bon-Kyoung Koo, and Meritxell Huch. 2016. "Culture and Establishment of Self-Renewing Human and Mouse Adult Liver and Pancreas 3D Organoids and Their Genetic Manipulation." *Nature Protocols* 11 (9): 1724–43.

Cameron, D. A., H. Gabra, and R. C. Leonard. 1994. "Continuous 5-Fluorouracil in the Treatment of Breast Cancer." *British Journal of Cancer* 70 (1): 120–24.

Campbell, Brittany B., Nicholas Light, David Fabrizio, Matthew Zatzman, Fabio Fuligni, Richard de Borja, Scott Davidson, et al. 2017. "Comprehensive Analysis of Hypermutation in Human Cancer." *Cell* 171 (5): 1042–56.e10.

Davidson, Philip R., Amy L. Sherborne, Barry Taylor, Alice O. Nakamura, and Jean L. Nakamura. 2017. "A Pooled Mutational Analysis Identifies Ionizing Radiation-Associated Mutational Signatures Conserved between Mouse and Human Malignancies." *Scientific*

Reports 7 (1): 7645.

Dvorak, Katerina, Claire M. Payne, Melissa Chavarria, Lois Ramsey, Barbora Dvorakova, Harris Bernstein, Hana Holubec, et al. 2007. "Bile Acids in Combination with Low pH Induce Oxidative Stress and Oxidative DNA Damage: Relevance to the Pathogenesis of Barrett's Oesophagus." *Gut* 56 (6): 763–71.

Ezzeldin, Hany, and Robert Diasio. 2004. "Dihydropyrimidine Dehydrogenase Deficiency, a Pharmacogenetic Syndrome Associated with Potentially Life-Threatening Toxicity Following 5-Fluorouracil Administration." *Clinical Colorectal Cancer* 4 (3): 181–89.

Ezzet, Farkad, and José C. Pinheiro. n.d. "Linear, Generalized Linear, and Nonlinear Mixed Effects Models." In *Pharmacometrics*, 103–35.

Focaccetti, Chiara, Antonino Bruno, Elena Magnani, Desirée Bartolini, Elisa Principi, Katiuscia Dallaglio, Eraldo O. Bucci, et al. 2015. "Effects of 5-Fluorouracil on Morphology, Cell Cycle, Proliferation, Apoptosis, Autophagy and ROS Production in Endothelial Cells and Cardiomyocytes." *PloS One* 10 (2): e0115686.

Frank, Steven A. 2010. "Evolution in Health and Medicine Sackler Colloquium: Somatic Evolutionary Genomics: Mutations during Development Cause Highly Variable Genetic Mosaicism with Risk of Cancer and Neurodegeneration." *Proceedings of the National Academy of Sciences of the United States of America* 107 Suppl 1 (January): 1725–30.

Fu, Y., G. Yang, F. Zhu, C. Peng, W. Li, H. Li, H-G Kim, A. M. Bode, Z. Dong, and Z. Dong. 2014. "Antioxidants Decrease the Apoptotic Effect of 5-Fu in Colon Cancer by Regulating Src-Dependent Caspase-7 Phosphorylation." *Cell Death & Disease* 5 (1): e983–e983.

Gaujoux, Renaud, and Cathal Seoighe. 2010. "A Flexible R Package for Nonnegative Matrix Factorization." *BMC Bioinformatics* 11 (July): 367.

Gladsjo, Julie Akiko, Alessandra B. Alió Sáenz, James Bergman, Greg Kricorian, and Bari B. Cunningham. 2009. "5% 5-Fluorouracil Cream for Treatment of Verruca Vulgaris in Children." *Pediatric Dermatology* 26 (3): 279–85.

Gramont, A. de, A. Figer, M. Seymour, M. Homerin, A. Hmissi, J. Cassidy, C. Boni, et al. 2000. "Leucovorin and Fluorouracil with or without Oxaliplatin as First-Line Treatment in Advanced Colorectal Cancer." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 18 (16): 2938–47.

Grem, J. L. 2000. "5-Fluorouracil: Forty-plus and Still Ticking. A Review of Its Preclinical and Clinical Development." *Investigational New Drugs* 18 (4): 299–313.

Haradhvala, Nicholas J., Paz Polak, Petar Stojanov, Kyle R. Covington, Eve Shinbrot, Julian M. Hess, Esther Rheinbay, et al. 2016. "Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair." *Cell* 164 (3): 538–49.

Hidaka, Katsuhiko, Masami Yamada, Hiroyuki Kamiya, Chikahide Masutani, Hideyoshi Harashima, Fumio Hanaoka, and Takehiko Nohmi. 2008. "Specificity of Mutations Induced by Incorporation of Oxidized dNTPs into DNA by Human DNA Polymerase Eta." *DNA Repair* 7 (3): 497–506.

Huang, Mi Ni, Willie Yu, Wei Wei Teoh, Maude Ardin, Apinya Jusakul, Alvin Wei Tian Ng, Arnoud Boot, et al. 2017. "Genome-Scale Mutational Signatures of Aflatoxin in Cells, Mice, and Human Tumors." *Genome Research* 27 (9): 1475–86.

Huehls, Amelia M., Catherine J. Huntoon, Poorval M. Joshi, Carly A. Baehr, Jill M. Wagner, Xiaoxiao Wang, Marietta Y. Lee, and Larry M. Karnitz. 2016. "Genomically Incorporated 5-Fluorouracil That Escapes UNG-Initiated Base Excision Repair Blocks DNA Replication and Activates Homologous Recombination." *Molecular Pharmacology* 89 (1): 53–62.

Inoue, M., H. Kamiya, K. Fujikawa, Y. Ootsuyama, N. Murata-Kamiya, T. Osaki, K. Yasumoto, and H. Kasai. 1998. "Induction of Chromosomal Gene Mutations in Escherichia Coli by Direct Incorporation of Oxidatively Damaged Nucleotides. New Evaluation Method for Mutagenesis by Damaged DNA Precursors in Vivo." *The Journal of Biological Chemistry* 273 (18): 11069–74.

Intuyod, Kitti, Paula Saavedra-García, Stefania Zona, Chun-Fui Lai, Yannasittha Jiramongkol, Kulthida Vaeteewoottacharn, Chawalit Pairojkul, et al. 2018. "FOXM1 Modulates 5-Fluorouracil Sensitivity in Cholangiocarcinoma through Thymidylate Synthase (TYMS): Implications of FOXM1–TYMS Axis Uncoupling in 5-FU Resistance." *Cell Death & Disease* 9 (12). https://doi.org/10.1038/s41419-018-1235-0.

Jager, Myrthe, Francis Blokzijl, Valentina Sasselli, Sander Boymans, Roel Janssen, Nicolle Besselink, Hans Clevers, Ruben van Boxtel, and Edwin Cuppen. 2018. "Measuring Mutation Accumulation in Single Human Adult Stem Cells by Whole-Genome Sequencing of Organoid Cultures." *Nature Protocols* 13 (1): 59–78.

Lee-Six, Henry, Peter Ellis, Robert J. Osborne, Mathijs A. Sanders, Luiza Moore, Nikitas Georgakopoulos, Franco Torrente, et al. 2018. "The Landscape of Somatic Mutation in Normal Colorectal Epithelial Cells." https://doi.org/10.1101/416800.

Letouzé, Eric, Jayendra Shinde, Victor Renault, Gabrielle Couchy, Jean-Frédéric Blanc, Emmanuel Tubacher, Quentin Bayard, et al. 2017. "Mutational Signatures Reveal the Dynamic Interplay of Risk Factors and Cellular Processes during Liver Tumorigenesis." *Nature Communications* 8 (1): 1315.

Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 25 (14): 1754–60.

Longley, Daniel B., D. Paul Harkin, and Patrick G. Johnston. 2003. "5-Fluorouracil: Mechanisms of Action and Clinical Strategies." *Nature Reviews. Cancer* 3 (5): 330–38.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550.

MacFarlane, Amanda J., Donald D. Anderson, Per Flodby, Cheryll A. Perry, Robert H. Allen, Sally P. Stabler, and Patrick J. Stover. 2011. "Nuclear Localization of de Novo Thymidylate Biosynthesis Pathway Is Required to Prevent Uracil Accumulation in DNA." *The Journal of Biological Chemistry* 286 (51): 44015–22.

MacRae, Sheila L., Matthew Mcknight Croken, R. B. Calder, Alexander Aliper, Brandon Milholland, Ryan R. White, Alexander Zhavoronkov, et al. 2015. "DNA Repair in Species with Extreme Lifespan Differences." *Aging* 7 (12): 1171–84.

Martincorena, Iñigo, Keiran M. Raine, Moritz Gerstung, Kevin J. Dawson, Kerstin Haase, Peter Van Loo, Helen Davies, Michael R. Stratton, and Peter J. Campbell. 2018. "Universal Patterns of Selection in Cancer and Somatic Tissues." *Cell* 173 (7): 1823.

Meier, Bettina, Susanna L. Cooke, Joerg Weiss, Aymeric P. Bailly, Ludmil B. Alexandrov, John Marshall, Keiran Raine, et al. 2014. "C. Elegans Whole-Genome Sequencing Reveals Mutational Signatures Related to Carcinogens and DNA Repair Deficiency." *Genome Research* 24 (10): 1624–36.

Morganella, Sandro, Ludmil B. Alexandrov, Dominik Glodzik, Xueqing Zou, Helen Davies, Johan Staaf, Anieta M. Sieuwerts, et al. 2016. "The Topography of Mutational Processes in Breast Cancer Genomes." *Nature Communications* 7 (May): 11383.

Morgillo, Floriana, Carminia Maria Della Corte, Morena Fasano, and Fortunato Ciardiello. 2016. "Mechanisms of Resistance to EGFR-Targeted Drugs: Lung Cancer." *ESMO Open* 1 (3): e000060.

Myers, C. E., R. C. Young, and B. A. Chabner. 1975. "Biochemical Determinants of 5-Fluorouracil Response in Vivo. The Role of Deoxyuridylate Pool Expansion." *The Journal of Clinical Investigation* 56 (5): 1231–38.

Negrei, Carolina, Ariana Hudita, Octav Ginghina, Bianca Galateanu, Sorina Nicoleta Voicu, Miriana Stan, Marieta Costache, Concettina Fenga, Nikolaos Drakoulis, and Aristidis M. Tsatsakis. 2016. "Colon Cancer Cells Gene Expression Signature As Response to 5-Fluorouracil, Oxaliplatin, and Folinic Acid Treatment." *Frontiers in Pharmacology* 7 (June): 172.

Nik-Zainal, Serena, Helen Davies, Johan Staaf, Manasa Ramakrishna, Dominik Glodzik,

Xueqing Zou, Inigo Martincorena, et al. 2016. "Landscape of Somatic Mutations in 560 Breast Cancer Whole-Genome Sequences." *Nature* 534 (7605): 47–54.

Parker, Jared B., and James T. Stivers. 2011. "Dynamics of Uracil and 5-Fluorouracil in DNA." *Biochemistry* 50 (5): 612–17.

Peters, Godefridus J., Emile J. Laurensse, Cees J. van Groeningen, Sybren Meijer, and Herbert M. Pinedo. 1989. "In Vitro and In Vivo Inhibition of Thymidylate Synthase of Human Colon Cancer by 5-Fluorouracil." In *Purine and Pyrimidine Metabolism in Man VI: Part A: Clinical and Molecular Biology*, edited by Kiyonobu Mikanagi, Kusuki Nishioka, and William N. Kelley, 439–45. Boston, MA: Springer US.

Pettersen, Henrik Sahlin, Torkild Visnes, Cathrine Broberg Vågbø, Eva K. Svaasand, Berit Doseth, Geir Slupphaug, Bodil Kavli, and Hans E. Krokan. 2011. "UNG-Initiated Base Excision Repair Is the Major Repair Route for 5-Fluorouracil in DNA, but 5-Fluorouracil Cytotoxicity Depends Mainly on RNA Incorporation." *Nucleic Acids Research* 39 (19): 8430–44.

Phillips, David H. 2018. "Mutational Spectra and Mutational Signatures: Insights into Cancer Aetiology and Mechanisms of DNA Damage and Repair." *DNA Repair* 71 (November): 6–11.

Pich, Oriol, Ferran Muiños, Radhakrishnan Sabarinathan, Iker Reyes-Salazar, Abel Gonzalez-Perez, and Nuria Lopez-Bigas. 2018. "Somatic and Germline Mutation Periodicity Follow the Orientation of the DNA Minor Groove around Nucleosomes." *Cell*. https://doi.org/10.1016/j.cell.2018.10.004.

Pilati, Camilla, Jayendra Shinde, Ludmil B. Alexandrov, Guillaume Assié, Thierry André, Zofia Hélias-Rodzewicz, Romain Ducoudray, et al. 2017. "Mutational Signature Analysis Identifies MUTYH Deficiency in Colorectal Cancers and Adrenocortical Carcinomas." *The Journal of Pathology* 242 (1): 10–15.

Poetsch, Anna R., Simon J. Boulton, and Nicholas M. Luscombe. 2018. "Genomic Landscape of Oxidative DNA Damage and Repair Reveals Regioselective Protection from Mutagenesis." *Genome Biology* 19 (1): 215.

Poulikakos, Poulikos I., Yogindra Persaud, Manickam Janakiraman, Xiangju Kong, Charles Ng, Gatien Moriceau, Hubing Shi, et al. 2011. "RAF Inhibitor Resistance Is Mediated by Dimerization of Aberrantly Spliced BRAF(V600E)." *Nature* 480 (7377): 387–90.

Priestley, Peter, Jonathan Baber, Martijn Lolkema, Neeltje Steeghs, Ewart de Bruijn, Korneel Duyvesteyn, Susan Haidari, et al. 2018. "Pan-Cancer Whole Genome Analyses of Metastatic Solid Tumors." https://doi.org/10.1101/415133.

Roerink, Sophie F., Nobuo Sasaki, Henry Lee-Six, Matthew D. Young, Ludmil B. Alexandrov, Sam Behjati, Thomas J. Mitchell, et al. 2018. "Intra-Tumour Diversification in Colorectal Cancer at the Single-Cell Level." *Nature* 556 (7702): 457–62.

Saunders, Christopher T., Wendy S. W. Wong, Sajani Swamy, Jennifer Becq, Lisa J. Murray, and R. Keira Cheetham. 2012. "Strelka: Accurate Somatic Small-Variant Calling from Sequenced Tumor-Normal Sample Pairs." *Bioinformatics* 28 (14): 1811–17.

Schütte, Moritz, Thomas Risch, Nilofar Abdavi-Azar, Karsten Boehnke, Dirk Schumacher, Marlen Keil, Reha Yildiriman, et al. 2017. "Molecular Dissection of Colorectal Cancer in Pre-Clinical Models Identifies Biomarkers Predicting Sensitivity to EGFR Inhibitors." *Nature Communications* 8 (February): 14262.

Secrier, Maria, Xiaodun Li, Nadeera de Silva, Matthew D. Eldridge, Gianmarco Contino, Jan Bornschein, Shona MacRae, et al. 2016. "Mutational Signatures in Esophageal Adenocarcinoma Define Etiologically Distinct Subgroups with Therapeutic Relevance." *Nature Genetics* 48 (10): 1131–41.

Sommer, H., and D. V. Santi. 1974. "Purification and Amino Acid Analysis of an Active Site Peptide from Thymidylate Synthetase Containing Covalently Bound 5-Fluoro-2'-Deoxyuridylate and Methylenetetrahydrofolate." *Biochemical and Biophysical Research*

Communications 57 (3): 689–95.

Suzuki, Tetsuya, and Hiroyuki Kamiya. 2017. "Mutations Induced by 8-Hydroxyguanine (8-Oxo-7,8-Dihydroguanine), a Representative Oxidized Base, in Mammalian Cells." *Genes and Environment : The Official Journal of the Japanese Environmental Mutagen Society* 39: 2.

Szikriszt, Bernadett, Ádám Póti, Orsolya Pipek, Marcin Krzystanek, Nnennaya Kanu, János Molnár, Dezső Ribli, et al. 2016. "A Comprehensive Survey of the Mutagenic Impact of Common Cancer Cytotoxics." *Genome Biology* 17 (May): 99.

Tamborero, David, Carlota Rubio-Perez, Jordi Deu-Pons, Michael P. Schroeder, Ana Vivancos, Ana Rovira, Ignasi Tusquets, et al. 2018. "Cancer Genome Interpreter Annotates the Biological and Clinical Relevance of Tumor Alterations." *Genome Medicine* 10 (1): 25.

Tomkova, Marketa, Jakub Tomek, Skirmantas Kriaucionis, and Benjamin Schuster-Böckler. 2018. "Mutational Signature Distribution Varies with DNA Replication Timing and Strand Asymmetry." *Genome Biology* 19 (1): 129.

Vértessy, Béata G., and Judit Tóth. 2009. "Keeping Uracil out of DNA: Physiological Role, Structure and Catalytic Mechanism of dUTPases." *Accounts of Chemical Research* 42 (1): 97–106.

Watson, Roshawn G., Filipe Muhale, Leigh B. Thorne, Jinsheng Yu, Bert H. O'Neil, Janelle M. Hoskins, Michael O. Meyers, et al. 2010. "Amplification of Thymidylate Synthetase in Metastatic Colorectal Cancer Patients Pretreated with 5-Fluorouracil-Based Chemotherapy." *European Journal of Cancer* 46 (18): 3358–64.

Wright, Karen D., Vinay M. Daryani, David C. Turner, Arzu Onar-Thomas, Nidal Boulos, Brent A. Orr, Richard J. Gilbertson, Clinton F. Stewart, and Amar Gajjar. 2015. "Phase I Study of 5-Fluorouracil in Children and Young Adults with Recurrent Ependymoma." *Neuro-Oncology* 17 (12): 1620–27.

Yates, Lucy R., Stian Knappskog, David Wedge, James H. R. Farmery, Santiago Gonzalez, Inigo Martincorena, Ludmil B. Alexandrov, et al. 2017. "Genomic Evolution of Breast Cancer Metastasis and Relapse." *Cancer Cell* 32 (2): 169–84.e7.

Zehir, Ahmet, Ryma Benayed, Ronak H. Shah, Aijazuddin Syed, Sumit Middha, Hyunjae R. Kim, Preethi Srinivasan, et al. 2017. "Erratum: Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients." *Nature Medicine* 23 (8): 1004.