**SUPPLEMENTARY METHODS**

**Whole Exome Sequencing**

Formalin-Fixed Paraffin Embedded (FFPE) tissue from CRCs were macrodissected and DNA extracted using QIAamp DNA FFPE Tissue kit (Qiagen, Hilden, Germany) using standard protocols. Peripheral blood-derived DNAs were extracted using DNeasy blood and tissue kit (Qiagen) and sequenced as germline references. Capture of the whole exome was performed using Agilent Clinical Research Exome V2 (Agilent, Santa Clara, CA) with sequencing performed on an Illumina NovaSeq 6000 (San Diego, CA) comprising 150bp paired-end reads. Mean on-target coverage was 431.8 ± 88.3 (mean ± SD) for FFPE tumour DNA samples and 158.7 ± 89.6 for blood-derived DNA samples.

**Bioinformatics Pipeline and Analysis**

Adapter sequences were trimmed from the raw FASTQ files using trimmomatic 0.38 [1], then aligned to the GRCh37 human reference genome using BWA 0.7.12. Somatic single-nucleotide variants and short insertions and deletions (indels) were called with Strelka 2.9.2 [2] using Illumina’s recommended workflow. Signatures were calculated using the pre-defined set of 67 SBS, 11 doublet, and 17 ID signatures published on the COSMIC website as version 3 [3]. Doublet counts were too low to reliably reconstruct signatures, so were not analysed further in this study (**Supplementary Table 1**). Details of the somatic pipeline and mutational signature calculation are available at <https://github.com/supernifty/somatic_pipeline> and <https://github.com/supernifty/mutational_signature>, respectively.

The impact and suitability of experimental settings was explored by filtering variants based on depth of coverage (DP) in the tumour (calculated by the GATK tool AnnotateVcfWithBamDepth) and the variant allele fraction (VAF). Signatures were calculated for each sample at each filter point. We quantitatively evaluated the ability of mutational signatures to separate classes of samples at different filtering settings, for each relevant signature. We calculated Fisher’s discriminant [4], which measures the ratio of “between-group” and “within-group” variability to find filtering settings that produce tightly clustered groups that are well-separated. We calculated the absolute margin separating the two groups, and the difference in the means of the two groups. P-values were calculated using either a one-sided t-test or, for groups with a single sample (N01), a Grubbs outlier test [5] to evaluate separability.

**P-value calculations**

The method of calculating p-values when determining the discriminatory utility of mutational signatures (**Table 2, Supplementary Table 3**) is dependent on the number of samples in the syndrome of interest – i.e. *MUTYH* biallelic (M01-M10), *NTHL1* biallelic (N01), or Lynch syndrome (L01-L14). If the number of samples was greater than one, a one-sided t-test was used, since we were only interested in signatures showing presence of a signal (higher than average), rather than absence of a signal. When the number of samples was one, a t-test could not be performed. A null hypothesis test measuring the probability of observing a value equal to or higher than the one observed was performed, assuming the comparison group is normally distributed.

When indicating whether a p-value is significant, we used a threshold of 0.05, with Bonferroni correction applied [6]. We performed 756 individual tests for significance; consequently, the adjusted significance threshold is 6.6 × 10-5. **Supplementary Table 3** indicates all signatures measured to have significant unadjusted p-values, with adjusted significant p-values highlighted.

**Separating biallelic samples from monoallelic samples**

**Figure 2** shows signatures found to be significant when separating biallelic *MUTYH* samples and biallelic *NTHL1* samples from all other samples, including their monoallelic counterparts. To justify grouping the monoallelic samples with the non-syndromic samples, p-values were calculated comparing monoallelic *MUTYH* samples to the rest of the non-*MUTYH* samples, which did not find a significant association in any previously reported associated signature (**Supplementary Table 3**). Similarly, comparing *NTHL1* monoallelic samples to the non-*NTHL1* samples did not reveal any significant association in previously reported BER-related signatures. Consequently, when identifying significant signatures and calculating confidence levels, monoallelic samples were included in the non-syndromic group.

**Distribution of samples in a group**

To determine 5th, 50th and 95th percentiles shown on the discrimination graphs (Figures 3c, 3d, 4c, 4d, 6a, 6b, 6c, and 6d), the calculated signatures from each CRC in the group were fitted to a beta distribution. Since signatures are normalized to a percentage, the possible range of values is bounded, the beta distribution is appropriate. When sums of signatures were used, the input to the distribution fitting calculation was normalized to one.

**Determining confidence levels**

Most discrimination graphs comparing groups of samples include 1% and 99% confidence levels, indicating that if a sample with unknown status exhibited this specific TMS percentage, based on our cohort of samples, we would be 1% or 99% confident that it belonged to the group of interest.

An equivalent question is, given a measured signature *x*, how likely is that point to have been generated by each of the two possible source distributions: the distribution of signatures from CRCs *with* the syndrome of interest, or the distribution of signatures from CRCs *without* the syndrome of interest? Here, we label the distribution of values with the syndrome and those without the syndrome *D1* and *D2* respectively. We assume that both *D1* and *D2* follow a beta distribution as discussed previously: *D1 ~ Beta(a1, b12)* and *D2 ~ Beta(a2, b22)*.

Given the two overlapping possible distributions from which the point may have originated, we aim to calculate *P(D=D1|X=x)*. Given that the probability of observing an exact value is zero, we instead calculate the probability of observing *x* within a range, that is *P(X=x±δ|D=D1)*, which can be calculated as the difference in the cumulative distribution function (cdf) at *x+δ* and *x-δ*. i.e. *P(X=x±δ|D=D1)=cdf(x+δ, D1) – cdf(x-δ, D1).*

The application of Bayes’ theorem enables the calculation of *P(X=x±δ|D=D1)* in terms of the cdf of each distribution:

This approach requires prior probabilities to be assigned indicating the likelihood of a sample belonging to each of the two distributions, given no prior information. For this analysis we assume an unbiased prior probability of .

We then calculate the confidence levels of interest (usually 1% or 99%) by considering the probability at values of *x* across each possible signature value. We select the first (lowest) value of *x* that crosses the confidence level of interest. Note that when both SBS and ID signatures are included in a combination, the sum can exceed 100%.

The calculation of a distribution’s cdf requires the distribution to have a non-zero measure of spread, but for groups with a single sample representing the syndrome (*NTHL1*), we cannot calculate a measure of spread. Instead, we assign the same standard deviation as that calculated for the non-syndromic samples and assume that both groups have the same standard deviation.

**Fisher’s Linear Discriminant**

Table 2 includes various measures of separation between groups, including Fisher’s Linear Discriminant (LD), which is based on the distance between the means measured in terms of the average of the standard deviations of the two groups. i.e. given groups A and B with means μA and μB and standard deviations σA and σB, LD is given by:

Since the *NTHL1* cohort consists of a single sample, LD cannot be calculated. Instead, we show the number of standard deviations away the sample of interest is from the rest of the samples in the group [5].

**Selection of the best combination of signatures**

We applied a greedy algorithm based on stepwise forward selection [7] to find the best combination of signatures while reducing the likelihood of overfitting.

Starting with each individual signature, we select the signature with the highest AUC that also has a minimum difference in means between the groups of at least 10%. We then iterate by assessing the AUC when each possible signature is added to the current best combination, provided it improves the AUC by at least 10%, and the difference in means between the groups. If the AUC has reached the maximum possible value (100%), then instead the margin between the groups must improve by at least 10%. These requirements ensure that only signatures with a strong association with the syndrome are selected. Although it is probable that signatures with legitimate but weaker associations are excluded with this method, this stringency reduces the likelihood of spurious signatures arising due to overfitting.

**SUPPLEMENTARY TABLES**

**Supplementary Table 1.** Metadata and clinical observations associated with each CRC.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Gender** | **Age Dx** | **Category** | **Gene** | **Mutation** | **MMR IHC Status** | **Tumour site** | **Histological Type** | **Grade** | **Margin** | **MSISensor** | **TMB** | **SBS Count** | **ID Variants** | **DB Count** | **STAGE\_CLINICALTX** | **STAGE\_MALIG** | **SBS.R1** | **SBS.V1** | **SBS.R2** | **SBS.V2** | **SBS.R3** | **SBS.V3** | **ID.R1** | **ID.V1** | **ID.R2** | **ID.V2** | **ID.R3** | **ID.V3** |
| M01 | Male | 62 | MUTYH biallelic | MUTYH | MUTYH: c.536A>G p.Tyr179Cys; c.536A>G p.Tyr179Cys | Proficient | ascending | Adenocarcinoma | Moderate | Expanding | 0.96 | 14.00 | 942 | 77 | 7 |  | TNM IIImx | SBS36 | 39% | SBS1 | 13% | SBS18 | 12% | ID12 | 20% | ID2 | 15% | ID16 | 13% |
| M02 | Male | 62 | MUTYH biallelic | MUTYH | MUTYH: c.536A>G p.Tyr179Cys; c.536A>G p.Tyr179Cys | Proficient | ascending | Adenocarcinoma | Moderate | Infiltrating | 0.67 | 10.74 | 723 | 57 | 1 |  | TNM Imx | SBS18 | 29% | SBS36 | 28% | SBS45 | 10% | ID12 | 32% | ID2 | 14% | ID1 | 13% |
| M03 | Male | 62 | MUTYH biallelic | MUTYH | MUTYH: c.536A>G p.Tyr179Cys; c.536A>G p.Tyr179Cys | Proficient | caecum | Adenocarcinoma | Moderate | Expanding | 0.91 | 15.41 | 1037 | 89 | 3 |  | TNM Imx | SBS36 | 42% | SBS18 | 10% | SBS1 | 10% | ID2 | 29% | ID12 | 19% | ID16 | 10% |
| M04 | Female | 56 | MUTYH biallelic | MUTYH | MUTYH: c.536A>G p.Tyr179Cys; c.536A>G p.Tyr179Cys | Proficient | ascending | Adenocarcinoma | Moderate | Infiltrating | 1.22 | 14.96 | 1007 | 83 | 3 |  | TNM IIImx | SBS36 | 38% | SBS18 | 11% | SBS52 | 8% | ID12 | 30% | ID2 | 28% | ID1 | 10% |
| M05 | Female | 56 | MUTYH biallelic | MUTYH | MUTYH: c.536A>G p.Tyr179Cys; c.536A>G p.Tyr179Cys | Proficient | ascending | Adenocarcinoma | Moderate | Infiltrating | 2.17 | 12.07 | 812 | 70 | 4 |  | TNM IIImx | SBS36 | 41% | SBS1 | 15% | SBS18 | 11% | ID2 | 35% | ID12 | 18% | ID1 | 13% |
| M06 | Male | 64 | MUTYH biallelic | MUTYH | MUTYH: c.1187G>A p.Gly396Asp; c.1187G>A p.Gly396Asp | Proficient | caecum | Adenocarcinoma | Unknown | Expanding | 0.46 | 6.98 | 470 | 23 | 2 |  | TNM IIImx | SBS18 | 57% | SBS1 | 10% | SBS20 | 7% | ID12 | 39% | ID2 | 28% | ID14 | 9% |
| M07 | Male | 64 | MUTYH biallelic | MUTYH | MUTYH: c.1187G>A p.Gly396Asp; c.1187G>A p.Gly396Asp | Proficient | unknown | Adenocarcinoma | Moderate | Expanding | 0.46 | 6.21 | 418 | 21 | 3 |  | TNM IIImx | SBS18 | 35% | SBS1 | 13% | SBS45 | 11% | ID2 | 25% | ID12 | 25% | ID11 | 19% |
| M08 | Male | 59 | MUTYH biallelic | MUTYH | MUTYH: c.1187G>A p.Gly396Asp; c.1187G>A p.Gly396Asp | Proficient | rectal | Adenocarcinoma | Moderate | Expanding | 0.41 | 5.72 | 385 | 37 | 0 | TNM III | TNM IIImx | SBS18 | 21% | SBS1 | 16% | SBS36 | 8% | ID12 | 34% | ID2 | 21% | ID7 | 14% |
| M09 | Male | 54 | MUTYH biallelic | MUTYH | MUTYH: c.536A>G p.Tyr179Cys; c.545G>A p.Arg182His | Proficient | sigmoid | Adenocarcinoma | Moderate | Infiltrating | 3.30 | 16.41 | 1104 | 45 | 5 |  | Missing N | SBS32 | 19% | SBS1 | 19% | SBS36 | 15% | ID1 | 18% | ID2 | 17% | ID12 | 14% |
| M10 | Male | 39 | MUTYH biallelic | MUTYH | MUTYH: c.536A>G p.Tyr179Cys; c.734G>A p.Arg245His | Proficient | ascending | Adenocarcinoma | Moderate | Expanding | 1.88 | 22.97 | 1546 | 119 | 4 | TNM III | TNM IIImx | SBS36 | 39% | SBS18 | 17% | SBS45 | 11% | ID2 | 34% | ID1 | 23% | ID12 | 19% |
| M11 | Female | 64 | MUTYH monoallelic | MUTYH | MUTYH: c.1187G>A p.Gly396Asp | Proficient | sigmoid | Mucinous | Moderate | Expanding | 1.36 | 7.80 | 525 | 113 | 1 |  | TNM IImx | SBS1 | 22% | SBS6 | 16% | SBS39 | 11% | ID2 | 30% | ID12 | 25% | ID1 | 13% |
| M12 | Male | 45 | MUTYH VUS | MUTYH | MUTYH: c.1187G>A p.Gly396Asp; c.912C>G p.Ser304Arg | Proficient | sigmoid | Adenocarcinoma | Moderate | Expanding | 0.99 | 4.65 | 313 | 47 | 3 |  | TNM IIImx | SBS5 | 27% | SBS1 | 21% | SBS39 | 8% | ID2 | 42% | ID12 | 22% | ID1 | 14% |
| M13 | Female | 35 | MUTYH monoallelic | MUTYH | MUTYH: c.1187G>A p.Gly396Asp | Proficient | sigmoid | Adenocarcinoma | Moderate | Infiltrating | 1.30 | 8.19 | 551 | 76 | 5 | TNM III | TNM IIImx | SBS1 | 17% | SBS15 | 8% | SBS39 | 7% | ID1 | 29% | ID2 | 20% | ID4 | 9% |
| M14 | Male | 49 | MUTYH monoallelic | MUTYH | MUTYH: c.1187G>A p.Gly396Asp | Proficient | sigmoid | Adenocarcinoma | Moderate | Infiltrating | 4.58 | 12.44 | 837 | 268 | 5 |  | TNM IIImx | SBS1 | 21% | SBS5 | 19% | SBS16 | 8% | ID1 | 21% | ID12 | 20% | ID2 | 18% |
| N01 | Female | 61 | NTHL1 biallelic | NTHL1 | NTHL1: c.268C>T p.Gln90Ter; c.235\_236insG, p.Ala79fs | Proficient | transverse colon | Adenocarcinoma | Moderate | Expanding | 0.86 | 8.84 | 595 | 31 | 2 |  |  | SBS30 | 47% | SBS6 | 15% | SBS42 | 8% | ID12 | 38% | ID2 | 27% | ID1 | 22% |
| N02 | Female | 61 | NTHL1 monoallelic | NTHL1 | NTHL1: c.859C>T p.Gln287Ter | Proficient | rectal | Mucinous | Poor | Infiltrating | 0.94 | 1.93 | 130 | 35 | 0 |  | TNM IImx | SBS1 | 15% | SBS39 | 13% | SBS10b | 8% | ID12 | 39% | ID2 | 32% | ID1 | 10% |
| N03 | Male | 73 | NTHL1 monoallelic | NTHL1 | NTHL1: c.859C>T p.Gln287Ter | Proficient | caecum/sigmoid | Adenocarcinoma | N/A | N/A | 2.37 | 14.19 | 955 | 66 | 5 | TNM II | TNM II | SBS6 | 21% | SBS1 | 15% | SBS39 | 13% | ID2 | 30% | ID1 | 12% | ID17 | 12% |
| N04 | Male | 53 | NTHL1 monoallelic | NTHL1 | NTHL1: c.268C>T p.Gln90Ter | Proficient | rectal | Mucinous | N/A | N/A | 1.65 | 6.87 | 462 | 38 | 4 |  |  | SBS1 | 24% | SBS6 | 19% | SBS31 | 7% | ID2 | 37% | ID14 | 12% | ID10 | 10% |
| N05 | Male | 43 | NTHL1 monoallelic | NTHL1 | NTHL1: c.268C>T p.Gln90Ter | Proficient | rectal | Adenocarcinoma | Moderate | Infiltrating | 0.83 | 4.90 | 330 | 36 | 2 | TNM III | TNM IIImx | SBS39 | 17% | SBS24 | 17% | SBS1 | 15% | ID2 | 42% | ID12 | 23% | ID3 | 16% |
| K01 | Female | 72 | MMRd control | MLH1 | MLH1 methylation | Deficient | N/A | N/A | N/A | N/A | 26.89 | 28.63 | 1927 | 2727 | 25 |  |  | SBS1 | 29% | SBS6 | 18% | SBS15 | 16% | ID2 | 77% | ID7 | 9% | ID12 | 8% |
| K02 | Female | 47 | MMRd control | MLH1 | MLH1 methylation | Deficient | transverse colon | Adenocarcinoma | N/A | Infiltrating | 35.28 | 42.36 | 2851 | 2797 | 21 | TNM III | TNM IIImx | SBS6 | 21% | SBS15 | 16% | SBS1 | 14% | ID2 | 69% | ID12 | 13% | ID7 | 11% |
| K03 | Female | 62 | MMRd control | MLH1 | MLH1 methylation | Deficient | N/A | N/A | N/A | N/A | 5.41 | 7.04 | 474 | 532 | 4 |  |  | SBS6 | 31% | SBS54 | 13% | SBS39 | 9% | ID2 | 70% | ID7 | 15% | ID12 | 11% |
| K04 | Male | 59 | MMRd control | MLH1 | MLH1 methylation | Deficient | caecum | Adenocarcinoma | Moderate | Expanding | 22.04 | 176.73 | 11893 | 2999 | 240 | TNM I | TNM Imx | SBS26 | 18% | SBS46 | 17% | SBS37 | 15% | ID2 | 52% | ID16 | 14% | ID12 | 11% |
| K05 | Female | 59 | MMRd control | MLH1 | MLH1 methylation | Deficient | transverse colon | Mucinous | Moderate | Expanding | 26.84 | 32.99 | 2220 | 3043 | 20 | TNM II | TNM IImx | SBS15 | 34% | SBS1 | 26% | SBS6 | 11% | ID2 | 66% | ID7 | 16% | ID12 | 14% |
| K06 | Female | 57 | MMRd control | MLH1 | MLH1 methylation | Deficient | caecum | Adenocarcinoma | Moderate | Expanding | 25.04 | 31.15 | 2096 | 2476 | 22 | TNM II | TNM IImx | SBS6 | 24% | SBS15 | 19% | SBS1 | 18% | ID2 | 61% | ID7 | 22% | ID12 | 12% |
| K07 | Female | 47 | MMRd control | MLH1 | MLH1 methylation | Deficient | transverse colon | Mucinous | Poor | Expanding | 33.31 | 74.54 | 5016 | 1994 | 44 |  | TNM IImx | SBS15 | 71% | SBS54 | 6% | SBS1 | 5% | ID2 | 72% | ID7 | 14% | ID12 | 10% |
| K08 | Female | 53 | MMRd control | MLH1 | MLH1 methylation | Deficient | caecum | Mucinous | Poor | Expanding | 46.73 | 66.01 | 4442 | 4129 | 28 | TNM II | TNM IImx | SBS6 | 21% | SBS21 | 14% | SBS15 | 13% | ID2 | 59% | ID7 | 21% | ID12 | 15% |
| K09 | Female | 55 | MMRd control | MLH1 | MLH1 methylation | Deficient | caecum | Adenocarcinoma | Moderate | Unknown | 43.35 | 40.00 | 2692 | 2150 | 26 | TNM I | TNM Imx | SBS1 | 30% | SBS15 | 30% | SBS54 | 9% | ID2 | 56% | ID7 | 33% | ID12 | 7% |
| L01 | Male | 32 | Lynch | MSH2 | MSH2: del x8 | Deficient | transverse colon | Mucinous | Poor | Expanding | 14.30 | 9.67 | 651 | 714 | 5 |  | TNM IIImx | SBS1 | 29% | SBS6 | 23% | SBS54 | 8% | ID2 | 67% | ID7 | 20% | ID12 | 10% |
| L02 | Male | 52 | Lynch | MSH2 | MSH2: c.2074\_2081delGGGTGTTT p.Gly692CysfsX4 | Deficient | rectal | Mucinous | Moderate | Expanding | 23.61 | 97.20 | 6541 | 2502 | 46 |  | TNM IImx | SBS6 | 33% | SBS1 | 22% | SBS54 | 11% | ID2 | 46% | ID7 | 46% | ID1 | 4% |
| L03 | Male | 31 | Lynch | MSH6 | MSH6: c.1422\_1423delGCinsAT p.Gln475Ter | Deficient | caecum | Adenocarcinoma | Moderate | Expanding | 6.54 | 45.16 | 3039 | 655 | 9 | TNM II | TNM IImx | SBS1 | 34% | SBS6 | 20% | SBS15 | 16% | ID2 | 71% | ID7 | 19% | ID12 | 6% |
| L04 | Male | 22 | Lynch | MSH6 | MSH6: c.3311\_3312delTT p.Phe1104TrpfsX3 | Deficient | caecum | Adenocarcinoma | Poor | Expanding | 16.28 | 52.32 | 3521 | 1852 | 22 |  | TNM IIImx | SBS1 | 29% | SBS6 | 24% | SBS15 | 14% | ID2 | 77% | ID12 | 13% | ID1 | 7% |
| L05 | Male | 35 | Lynch | MSH6 | MSH6: c.1957delG p.Val653Ter | Deficient | sigmoid | Mucinous | Moderate | Expanding | 21.68 | 79.16 | 5327 | 2699 | 19 |  | TNM IImx | SBS1 | 29% | SBS6 | 16% | SBS15 | 16% | ID2 | 77% | ID12 | 13% | ID1 | 5% |
| L06 | Male | 41 | Lynch | MSH6 | MSH6: c.3958\_3961delGCAA p.Ala1320GlufsX6 | Deficient | ascending | Mucinous | Moderate | Expanding | 1.73 | 18.38 | 1237 | 379 | 4 |  | TNM IImx | SBS1 | 40% | SBS6 | 16% | SBS15 | 14% | ID2 | 77% | ID7 | 9% | ID12 | 8% |
| L07 | Male | 32 | Lynch | MLH1 | MLH1: c.672delT p.Ser225ValfsX4 | Deficient | caecum | Mucinous | Moderate | Infiltrating | 34.88 | 40.76 | 2743 | 4729 | 29 |  | TNM IV | SBS1 | 25% | SBS6 | 18% | SBS15 | 17% | ID2 | 66% | ID7 | 15% | ID12 | 14% |
| L08 | Male | 40 | Lynch | MLH1 | MLH1: c.76C>T p.Gln26Ter | Deficient | transverse colon | Adenocarcinoma | Moderate | Infiltrating | 38.84 | 36.18 | 2435 | 4038 | 32 |  | TNM IImx | SBS15 | 31% | SBS1 | 23% | SBS54 | 10% | ID2 | 73% | ID12 | 15% | ID7 | 7% |
| L09 | Male | 27 | Lynch | MLH1 | MLH1: c.1852\_1854delAAG p.Lys618del | Deficient | sigmoid | Adenocarcinoma | Moderate | Expanding | 31.42 | 64.13 | 4316 | 1400 | 29 | TNM III | TNM IIImx | SBS15 | 76% | SBS1 | 8% | SBS54 | 7% | ID2 | 49% | ID7 | 40% | ID12 | 6% |
| L10 | Male | 29 | Lynch | MLH1 | MLH1: c.672del p.(Ser225ValfsTer4) | Deficient | caecum | Adenocarcinoma | Poor | Expanding | 20.47 | 13.33 | 897 | 1887 | 4 | TNM I | TNM Imx | SBS1 | 26% | SBS6 | 18% | SBS15 | 16% | ID2 | 76% | ID12 | 12% | ID1 | 6% |
| L11 | Male | 43 | Lynch | MLH1 | MLH1: c.1459C>T p.Arg487Ter | Deficient | transverse colon | Mucinous | Moderate | Expanding | 35.03 | 39.27 | 2643 | 3224 | 40 |  | TNM IImx | SBS1 | 30% | SBS15 | 27% | SBS6 | 15% | ID2 | 69% | ID7 | 15% | ID12 | 11% |
| L12 | Female | 45 | Lynch | MLH1 | MLH1: c.1644C>A p.Tyr548Ter | Deficient | caecum | Adenocarcinoma | Poor | Expanding | 43.23 | 106.20 | 7147 | 6609 | 69 |  | TNM IIImx | SBS6 | 28% | SBS15 | 18% | SBS1 | 14% | ID2 | 61% | ID7 | 23% | ID12 | 9% |
| L13 | Female | 44 | Lynch | MLH1 | MLH1: c.350C>T p.Thr117Met | Deficient | ascending | Mucinous | Moderate | Expanding | 30.87 | 45.22 | 3043 | 5177 | 14 |  | TNM IImx | SBS15 | 38% | SBS1 | 19% | SBS6 | 10% | ID2 | 66% | ID12 | 14% | ID7 | 14% |
| L14 | Female | 51 | Lynch | MSH2 | MSH2: c.2005+3\_2005+14del12 | Deficient | ascending | Adenocarcinoma | Moderate | Expanding | 40.05 | 58.81 | 3958 | 3372 | 38 | TNM II | TNM IImx | SBS1 | 40% | SBS6 | 22% | SBS15 | 15% | ID2 | 60% | ID7 | 18% | ID12 | 14% |
| C01 | Male | 57 | MMRp control | CON | MMR-prof | Proficient | rectal | Adenocarcinoma | Poor | Expanding | 6.27 | 32.45 | 2184 | 77 | 44 |  | TNM Imx | SBS46 | 12% | SBS6 | 11% | SBS54 | 11% | ID16 | 29% | ID13 | 15% | ID15 | 10% |
| C02 | Male | 69 | MMRp control | CON | MMR-prof | Proficient | ascending | Adenocarcinoma | Moderate | Infiltrating | 1.00 | 6.15 | 414 | 43 | 1 |  | TNM Imx | SBS1 | 28% | SBS6 | 20% | SBS39 | 13% | ID12 | 32% | ID2 | 19% | ID10 | 17% |
| C03 | Male | 44 | MMRp control | CON | MMR-prof | Proficient | rectal | Adenocarcinoma | Well | Expanding | 1.16 | 3.77 | 254 | 53 | 2 | TNM I | TNM Imx | SBS1 | 27% | SBS6 | 18% | SBS39 | 10% | ID12 | 52% | ID2 | 27% | ID10 | 7% |
| C04 | Male | 41 | MMRp control | CON | MMR-prof | Proficient | sigmoid | Signet ring | Poor | Infiltrating | 1.68 | 3.67 | 247 | 46 | 4 | TNM III | TNM IIImx | SBS1 | 22% | SBS46 | 10% | SBS5 | 10% | ID2 | 42% | ID12 | 22% | ID17 | 14% |
| C05 | Male | 37 | MMRp control | CON | MMR-prof | Proficient | ascending | Adenocarcinoma | Moderate | Infiltrating | 2.64 | 7.71 | 519 | 90 | 1 | TNM III | TNM IIImx | SBS6 | 24% | SBS1 | 11% | SBS39 | 8% | ID2 | 29% | ID12 | 17% | ID14 | 15% |
| C06 | Female | 51 | MMRp control | CON | MMR-prof | Proficient | rectal | Adenocarcinoma | Moderate | Expanding | 0.68 | 4.09 | 275 | 67 | 2 | TNM II | TNM IImx | SBS1 | 24% | SBS6 | 16% | SBS39 | 11% | ID12 | 30% | ID2 | 21% | ID1 | 14% |
| C07 | Male | 21 | MMRp control | CON | MMR-prof | Proficient | descending | Adenocarcinoma | Moderate | Expanding | 2.13 | 9.54 | 642 | 81 | 17 |  | TNM IImx | SBS39 | 21% | SBS4 | 12% | SBS1 | 11% | ID2 | 30% | ID12 | 27% | ID5 | 12% |
| C08 | Female | 21 | MMRp control | CON | MMR-prof | Proficient | sigmoid | Adenocarcinoma | Poor | Infiltrating | 0.80 | 2.27 | 153 | 42 | 0 |  | TNM IIImx | SBS6 | 14% | SBS15 | 11% | SBS57 | 9% | ID2 | 45% | ID6 | 15% | ID12 | 10% |
| C09 | Male | 27 | MMRp control | CON | MMR-prof | Proficient | transverse colon | Adenocarcinoma | Poor | Unknown | 0.67 | 4.58 | 308 | 26 | 6 |  | TNM IIImx | SBS39 | 23% | SBS60 | 9% | SBS6 | 9% | ID12 | 45% | ID2 | 15% | ID9 | 11% |
| C10 | Male | 46 | MMRp control | CON | MMR-prof | Proficient | ascending | Adenocarcinoma | Moderate | Expanding | 0.17 | 6.24 | 420 | 81 | 0 |  | TNM IIImx | SBS39 | 22% | SBS15 | 11% | SBS1 | 10% | ID2 | 23% | ID3 | 22% | ID12 | 18% |
| C11 | Female | 47 | MMRp control | CON | MMR-prof | Proficient | caecum | Adenocarcinoma | Moderate | Unknown | 0.92 | 5.42 | 365 | 68 | 4 |  | TNM Imx | SBS1 | 23% | SBS6 | 19% | SBS39 | 11% | ID2 | 53% | ID12 | 24% | ID1 | 11% |
| C12 | Female | 46 | MMRp control | CON | MMR-prof | Proficient | transverse colon | Adenocarcinoma | Moderate | Infiltrating | 1.20 | 5.47 | 368 | 80 | 1 |  | TNM IImx | SBS1 | 23% | SBS39 | 15% | SBS6 | 14% | ID2 | 33% | ID1 | 19% | ID12 | 16% |
| C13 | Male | 55 | MMRp control | CON | MMR-prof | Proficient | transverse colon | Adenocarcinoma | Moderate | Expanding | 0.65 | 10.19 | 686 | 85 | 9 |  | TNM IIImx | SBS37 | 15% | SBS39 | 14% | SBS1 | 11% | ID12 | 26% | ID4 | 14% | ID16 | 12% |
| C14 | Male | 59 | MMRp control | CON | MMR-prof | Proficient | transverse colon | Adenocarcinoma | Moderate | Infiltrating | 0.70 | 9.38 | 631 | 140 | 7 | TNM III | TNM IIImx | SBS1 | 23% | SBS39 | 16% | SBS54 | 10% | ID12 | 21% | ID4 | 21% | ID10 | 13% |
| C15 | Male | 48 | MMRp control | CON | MMR-prof | Proficient | sigmoid | Adenocarcinoma | Moderate | Expanding | 2.02 | 4.70 | 316 | 51 | 4 |  | TNM IIImx | SBS6 | 24% | SBS1 | 17% | SBS39 | 13% | ID12 | 33% | ID2 | 20% | ID1 | 13% |
| C16 | Female | 48 | MMRp control | CON | MMR-prof | Proficient | rectal | Undifferentiated | Poor | Expanding | 0.56 | 9.82 | 661 | 27 | 5 |  | TNM Imx | SBS2 | 24% | SBS13 | 23% | SBS10b | 13% | ID12 | 38% | ID17 | 14% | ID9 | 13% |
| C17 | Female | 33 | MMRp control | CON | MMR-prof | Proficient | transverse colon | Adenocarcinoma | Poor | Infiltrating | 3.42 | 9.50 | 639 | 65 | 8 |  | TNM IIImx | SBS1 | 19% | SBS6 | 11% | SBS39 | 10% | ID2 | 34% | ID12 | 26% | ID14 | 14% |
| C18 | Female | 44 | MMRp control | CON | MMR-prof | Proficient | transverse colon | Adenocarcinoma | Moderate | Expanding | 1.54 | 14.61 | 983 | 113 | 7 |  | TNM IV | SBS39 | 25% | SBS5 | 17% | SBS8 | 8% | ID6 | 22% | ID12 | 20% | ID2 | 12% |

**Supplementary Table 2.** Analysis outputs associated with each CRC. The top three SBS and IDS TMS are indicated for each CRC, as well as overall mutation counts in each of the substitution, indel and doublet categories. MSISensor [8] indicates MMRd status; CRCs with values greater than 3.5 are considered likely to be MSI-H [8].

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **MSISensor** | **TMB** | **SBS Count** | **ID Variants** | **DB Count** | **SBS.R1** | **SBS.V1** | **SBS.R2** | **SBS.V2** | **SBS.R3** | **SBS.V3** | **ID.R1** | **ID.V1** | **ID.R2** | **ID.V2** | **ID.R3** | **ID.V3** |
| M01 | 0.96 | 14.00 | 942 | 77 | 7 | SBS36 | 39% | SBS1 | 13% | SBS18 | 12% | ID12 | 20% | ID2 | 15% | ID16 | 13% |
| M02 | 0.67 | 10.74 | 723 | 57 | 1 | SBS18 | 29% | SBS36 | 28% | SBS45 | 10% | ID12 | 32% | ID2 | 14% | ID1 | 13% |
| M03 | 0.91 | 15.41 | 1037 | 89 | 3 | SBS36 | 42% | SBS18 | 10% | SBS1 | 10% | ID2 | 29% | ID12 | 19% | ID16 | 10% |
| M04 | 1.22 | 14.96 | 1007 | 83 | 3 | SBS36 | 38% | SBS18 | 11% | SBS52 | 8% | ID12 | 30% | ID2 | 28% | ID1 | 10% |
| M05 | 2.17 | 12.07 | 812 | 70 | 4 | SBS36 | 41% | SBS1 | 15% | SBS18 | 11% | ID2 | 35% | ID12 | 18% | ID1 | 13% |
| M06 | 0.46 | 6.98 | 470 | 23 | 2 | SBS18 | 57% | SBS1 | 10% | SBS20 | 7% | ID12 | 39% | ID2 | 28% | ID14 | 9% |
| M07 | 0.46 | 6.21 | 418 | 21 | 3 | SBS18 | 35% | SBS1 | 13% | SBS45 | 11% | ID2 | 25% | ID12 | 25% | ID11 | 19% |
| M08 | 0.41 | 5.72 | 385 | 37 | 0 | SBS18 | 21% | SBS1 | 16% | SBS36 | 8% | ID12 | 34% | ID2 | 21% | ID7 | 14% |
| M09 | 3.30 | 16.41 | 1104 | 45 | 5 | SBS32 | 19% | SBS1 | 19% | SBS36 | 15% | ID1 | 18% | ID2 | 17% | ID12 | 14% |
| M10 | 1.88 | 22.97 | 1546 | 119 | 4 | SBS36 | 39% | SBS18 | 17% | SBS45 | 11% | ID2 | 34% | ID1 | 23% | ID12 | 19% |
| M11 | 1.36 | 7.80 | 525 | 113 | 1 | SBS1 | 22% | SBS6 | 16% | SBS39 | 11% | ID2 | 30% | ID12 | 25% | ID1 | 13% |
| M12 | 0.99 | 4.65 | 313 | 47 | 3 | SBS5 | 27% | SBS1 | 21% | SBS39 | 8% | ID2 | 42% | ID12 | 22% | ID1 | 14% |
| M13 | 1.30 | 8.19 | 551 | 76 | 5 | SBS1 | 17% | SBS15 | 8% | SBS39 | 7% | ID1 | 29% | ID2 | 20% | ID4 | 9% |
| M14 | 4.58 | 12.44 | 837 | 268 | 5 | SBS1 | 21% | SBS5 | 19% | SBS16 | 8% | ID1 | 21% | ID12 | 20% | ID2 | 18% |
| N01 | 0.86 | 8.84 | 595 | 31 | 2 | SBS30 | 47% | SBS6 | 15% | SBS42 | 8% | ID12 | 38% | ID2 | 27% | ID1 | 22% |
| N02 | 0.94 | 1.93 | 130 | 35 | 0 | SBS1 | 15% | SBS39 | 13% | SBS10b | 8% | ID12 | 39% | ID2 | 32% | ID1 | 10% |
| N03 | 2.37 | 14.19 | 955 | 66 | 5 | SBS6 | 21% | SBS1 | 15% | SBS39 | 13% | ID2 | 30% | ID1 | 12% | ID17 | 12% |
| N04 | 1.65 | 6.87 | 462 | 38 | 4 | SBS1 | 24% | SBS6 | 19% | SBS31 | 7% | ID2 | 37% | ID14 | 12% | ID10 | 10% |
| N05 | 0.83 | 4.90 | 330 | 36 | 2 | SBS39 | 17% | SBS24 | 17% | SBS1 | 15% | ID2 | 42% | ID12 | 23% | ID3 | 16% |
| K01 | 26.89 | 28.63 | 1927 | 2727 | 25 | SBS1 | 29% | SBS6 | 18% | SBS15 | 16% | ID2 | 77% | ID7 | 9% | ID12 | 8% |
| K02 | 35.28 | 42.36 | 2851 | 2797 | 21 | SBS6 | 21% | SBS15 | 16% | SBS1 | 14% | ID2 | 69% | ID12 | 13% | ID7 | 11% |
| K03 | 5.41 | 7.04 | 474 | 532 | 4 | SBS6 | 31% | SBS54 | 13% | SBS39 | 9% | ID2 | 70% | ID7 | 15% | ID12 | 11% |
| K04 | 22.04 | 176.73 | 11893 | 2999 | 240 | SBS26 | 18% | SBS46 | 17% | SBS37 | 15% | ID2 | 52% | ID16 | 14% | ID12 | 11% |
| K05 | 26.84 | 32.99 | 2220 | 3043 | 20 | SBS15 | 34% | SBS1 | 26% | SBS6 | 11% | ID2 | 66% | ID7 | 16% | ID12 | 14% |
| K06 | 25.04 | 31.15 | 2096 | 2476 | 22 | SBS6 | 24% | SBS15 | 19% | SBS1 | 18% | ID2 | 61% | ID7 | 22% | ID12 | 12% |
| K07 | 33.31 | 74.54 | 5016 | 1994 | 44 | SBS15 | 71% | SBS54 | 6% | SBS1 | 5% | ID2 | 72% | ID7 | 14% | ID12 | 10% |
| K08 | 46.73 | 66.01 | 4442 | 4129 | 28 | SBS6 | 21% | SBS21 | 14% | SBS15 | 13% | ID2 | 59% | ID7 | 21% | ID12 | 15% |
| K09 | 43.35 | 40.00 | 2692 | 2150 | 26 | SBS1 | 30% | SBS15 | 30% | SBS54 | 9% | ID2 | 56% | ID7 | 33% | ID12 | 7% |
| L01 | 14.30 | 9.67 | 651 | 714 | 5 | SBS1 | 29% | SBS6 | 23% | SBS54 | 8% | ID2 | 67% | ID7 | 20% | ID12 | 10% |
| L02 | 23.61 | 97.20 | 6541 | 2502 | 46 | SBS6 | 33% | SBS1 | 22% | SBS54 | 11% | ID2 | 46% | ID7 | 46% | ID1 | 4% |
| L03 | 6.54 | 45.16 | 3039 | 655 | 9 | SBS1 | 34% | SBS6 | 20% | SBS15 | 16% | ID2 | 71% | ID7 | 19% | ID12 | 6% |
| L04 | 16.28 | 52.32 | 3521 | 1852 | 22 | SBS1 | 29% | SBS6 | 24% | SBS15 | 14% | ID2 | 77% | ID12 | 13% | ID1 | 7% |
| L05 | 21.68 | 79.16 | 5327 | 2699 | 19 | SBS1 | 29% | SBS6 | 16% | SBS15 | 16% | ID2 | 77% | ID12 | 13% | ID1 | 5% |
| L06 | 1.73 | 18.38 | 1237 | 379 | 4 | SBS1 | 40% | SBS6 | 16% | SBS15 | 14% | ID2 | 77% | ID7 | 9% | ID12 | 8% |
| L07 | 34.88 | 40.76 | 2743 | 4729 | 29 | SBS1 | 25% | SBS6 | 18% | SBS15 | 17% | ID2 | 66% | ID7 | 15% | ID12 | 14% |
| L08 | 38.84 | 36.18 | 2435 | 4038 | 32 | SBS15 | 31% | SBS1 | 23% | SBS54 | 10% | ID2 | 73% | ID12 | 15% | ID7 | 7% |
| L09 | 31.42 | 64.13 | 4316 | 1400 | 29 | SBS15 | 76% | SBS1 | 8% | SBS54 | 7% | ID2 | 49% | ID7 | 40% | ID12 | 6% |
| L10 | 20.47 | 13.33 | 897 | 1887 | 4 | SBS1 | 26% | SBS6 | 18% | SBS15 | 16% | ID2 | 76% | ID12 | 12% | ID1 | 6% |
| L11 | 35.03 | 39.27 | 2643 | 3224 | 40 | SBS1 | 30% | SBS15 | 27% | SBS6 | 15% | ID2 | 69% | ID7 | 15% | ID12 | 11% |
| L12 | 43.23 | 106.20 | 7147 | 6609 | 69 | SBS6 | 28% | SBS15 | 18% | SBS1 | 14% | ID2 | 61% | ID7 | 23% | ID12 | 9% |
| L13 | 30.87 | 45.22 | 3043 | 5177 | 14 | SBS15 | 38% | SBS1 | 19% | SBS6 | 10% | ID2 | 66% | ID12 | 14% | ID7 | 14% |
| L14 | 40.05 | 58.81 | 3958 | 3372 | 38 | SBS1 | 40% | SBS6 | 22% | SBS15 | 15% | ID2 | 60% | ID7 | 18% | ID12 | 14% |
| C01 | 6.27 | 32.45 | 2184 | 77 | 44 | SBS46 | 12% | SBS6 | 11% | SBS54 | 11% | ID16 | 29% | ID13 | 15% | ID15 | 10% |
| C02 | 1.00 | 6.15 | 414 | 43 | 1 | SBS1 | 28% | SBS6 | 20% | SBS39 | 13% | ID12 | 32% | ID2 | 19% | ID10 | 17% |
| C03 | 1.16 | 3.77 | 254 | 53 | 2 | SBS1 | 27% | SBS6 | 18% | SBS39 | 10% | ID12 | 52% | ID2 | 27% | ID10 | 7% |
| C04 | 1.68 | 3.67 | 247 | 46 | 4 | SBS1 | 22% | SBS46 | 10% | SBS5 | 10% | ID2 | 42% | ID12 | 22% | ID17 | 14% |
| C05 | 2.64 | 7.71 | 519 | 90 | 1 | SBS6 | 24% | SBS1 | 11% | SBS39 | 8% | ID2 | 29% | ID12 | 17% | ID14 | 15% |
| C06 | 0.68 | 4.09 | 275 | 67 | 2 | SBS1 | 24% | SBS6 | 16% | SBS39 | 11% | ID12 | 30% | ID2 | 21% | ID1 | 14% |
| C07 | 2.13 | 9.54 | 642 | 81 | 17 | SBS39 | 21% | SBS4 | 12% | SBS1 | 11% | ID2 | 30% | ID12 | 27% | ID5 | 12% |
| C08 | 0.80 | 2.27 | 153 | 42 | 0 | SBS6 | 14% | SBS15 | 11% | SBS57 | 9% | ID2 | 45% | ID6 | 15% | ID12 | 10% |
| C09 | 0.67 | 4.58 | 308 | 26 | 6 | SBS39 | 23% | SBS60 | 9% | SBS6 | 9% | ID12 | 45% | ID2 | 15% | ID9 | 11% |
| C10 | 0.17 | 6.24 | 420 | 81 | 0 | SBS39 | 22% | SBS15 | 11% | SBS1 | 10% | ID2 | 23% | ID3 | 22% | ID12 | 18% |
| C11 | 0.92 | 5.42 | 365 | 68 | 4 | SBS1 | 23% | SBS6 | 19% | SBS39 | 11% | ID2 | 53% | ID12 | 24% | ID1 | 11% |
| C12 | 1.20 | 5.47 | 368 | 80 | 1 | SBS1 | 23% | SBS39 | 15% | SBS6 | 14% | ID2 | 33% | ID1 | 19% | ID12 | 16% |
| C13 | 0.65 | 10.19 | 686 | 85 | 9 | SBS37 | 15% | SBS39 | 14% | SBS1 | 11% | ID12 | 26% | ID4 | 14% | ID16 | 12% |
| C14 | 0.70 | 9.38 | 631 | 140 | 7 | SBS1 | 23% | SBS39 | 16% | SBS54 | 10% | ID12 | 21% | ID4 | 21% | ID10 | 13% |
| C15 | 2.02 | 4.70 | 316 | 51 | 4 | SBS6 | 24% | SBS1 | 17% | SBS39 | 13% | ID12 | 33% | ID2 | 20% | ID1 | 13% |
| C16 | 0.56 | 9.82 | 661 | 27 | 5 | SBS2 | 24% | SBS13 | 23% | SBS10b | 13% | ID12 | 38% | ID17 | 14% | ID9 | 13% |
| C17 | 3.42 | 9.50 | 639 | 65 | 8 | SBS1 | 19% | SBS6 | 11% | SBS39 | 10% | ID2 | 34% | ID12 | 26% | ID14 | 14% |
| C18 | 1.54 | 14.61 | 983 | 113 | 7 | SBS39 | 25% | SBS5 | 17% | SBS8 | 8% | ID6 | 22% | ID12 | 20% | ID2 | 12% |

**Supplementary Table 3**. Each individual mutational signature was tested for its ability to identify inherited CRC syndromes. CRCs included in each test are shown in the “positive” and “negative” columns. Each TMS found to have an unadjusted significant p-value is included, with adjusted significant TMS (p-value < 6.6 x 10-5) highlighted (with \*). Where LDA is not shown it could not be calculated due to all values from one group exhibiting the same signature value.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Positive** | **Negative** | **Classifier** | **AUC** | **LDA** | **Margin** | **MeanDiff** | **PValue** | **Sig** |
| M01-M10 | N01-N05,K01-K09,L01-L14,C01-C18 | SBS18 | 0.9913 | 1.724 | -0.04 | 0.1998 | 6.11E-12 | \* |
|  |  | SBS24 | 0.6902 | -1 | -0.167 | 0.0207 | 0.0192069 |  |
|  |  | SBS36 | 0.9478 | 2.6478 | -0.019 | 0.2542 | 1.88663E-15 | \* |
|  |  | SBS38 | 0.6326 | -1 | -0.04 | 0.0126 | 0.0021529 |  |
|  |  | SBS42 | 0.6185 | -1 | -0.083 | 0.0112 | 2.18E-02 |  |
|  |  | SBS45 | 0.8424 | 0.7891 | -0.045 | 0.0405 | 5.71368E-07 | \* |
|  |  | SBS48 | 0.6054 | -1 | -0.01 | 0.0017 | 0.0215872 |  |
|  |  | SBS52 | 0.712 | -1 | -0.023 | 0.0151 | 4.96E-04 |  |
|  |  | SBS53 | 0.7826 | -1 | -0.045 | 0.0163 | 4.28788E-05 | \* |
|  |  | SBS59 | 0.5815 | -1 | -0.024 | 0.011 | 0.00113945 |  |
|  |  | ID1 | 0.8217 | 0.6826 | -0.14 | 0.0519 | 6.95E-04 |  |
|  |  | ID10 | 0.7467 | -1 | -0.17 | 0.0287 | 0.0278568 |  |
|  |  | ID12 | 0.7391 | 0.2553 | -0.387 | 0.0707 | 0.0369198 |  |
|  |  | ID8 | 0.6663 | -1 | -0.101 | 0.0237 | 0.0194396 |  |
| M11,M13,M14 | N01-N05,K01-K09,L01-L14,C01-C18 | SBS12 | 0.8071 | -1 | -0.045 | 0.0097 | 0.0188222 |  |
|  |  | SBS16 | 0.6033 | -1 | -0.075 | 0.0151 | 0.0461037 |  |
|  |  | SBS17b | 0.8614 | -1 | -0.083 | 0.0177 | 0.0218226 |  |
|  |  | SBS40 | 0.625 | -1 | 0 | 0.0032 | 0.000181955 |  |
|  |  | SBS5 | 0.8451 | 1.267 | -0.167 | 0.1217 | 4.53577E-07 | \* |
|  |  | SBS50 | 0.9402 | -1 | -0.035 | 0.0181 | 0.000112049 |  |
|  |  | SBS56 | 0.8614 | -1 | -0.013 | 0.0118 | 8.26579E-09 | \* |
|  |  | SBS58 | 0.8098 | -1 | -0.034 | 0.015 | 1.29E-03 |  |
|  |  | SBS7a | 0.7935 | -1 | -0.064 | 0.01 | 0.0473572 |  |
|  |  | SBS8 | 0.8125 | -1 | -0.077 | 0.0222 | 0.0032041 |  |
|  |  | SBS9 | 0.7092 | -1 | -0.059 | 0.0117 | 1.99E-02 |  |
|  |  | ID1 | 0.9647 | 2.5586 | -0.083 | 0.1214 | 2.77593E-06 | \* |
|  |  | ID4 | 0.8533 | 0.7035 | -0.188 | 0.0473 | 2.15E-02 |  |
| M01-M10 | M11,M13,M14,N01-N05,K01-K09,L01-L14,C01-C18 | SBS18 | 0.9918 | 1.7317 | -0.04 | 0.2001 | 1.29309E-12 | \* |
|  |  | SBS24 | 0.6806 | -1 | -0.167 | 0.0202 | 0.0190494 |  |
|  |  | SBS36 | 0.948 | 2.6487 | -0.019 | 0.2542 | 2.60E-16 | \* |
|  |  | SBS38 | 0.6337 | -1 | -0.04 | 0.0127 | 0.00153043 |  |
|  |  | SBS42 | 0.6204 | -1 | -0.083 | 0.0113 | 0.0173037 |  |
|  |  | SBS45 | 0.8378 | 0.7756 | -0.045 | 0.0402 | 3.73E-07 | \* |
|  |  | SBS48 | 0.6041 | -1 | -0.01 | 0.0018 | 0.0178493 |  |
|  |  | SBS52 | 0.7041 | -1 | -0.023 | 0.0149 | 0.000379759 |  |
|  |  | SBS53 | 0.7765 | -1 | -0.045 | 0.0161 | 3.26E-05 | \* |
|  |  | SBS59 | 0.5765 | -1 | -0.024 | 0.0107 | 0.00129947 |  |
|  |  | ID1 | 0.7776 | 0.3623 | -0.211 | 0.0434 | 0.0132651 |  |
|  |  | ID10 | 0.7378 | -1 | -0.17 | 0.0276 | 0.0303096 |  |
|  |  | ID12 | 0.7327 | 0.2607 | -0.387 | 0.0705 | 0.0341691 |  |
|  |  | ID8 | 0.6622 | -1 | -0.101 | 0.024 | 1.57E-02 |  |
| N01 | M01-M14,K01-K09,L01-L14,C01-C18 | SBS23 | 1 | 432.6818 | 0.061 | 0.0785 | 1.6615E-190 | \* |
|  |  | SBS30 | 1 | 524.148 | 0.379 | 0.4611 | 2.86E-230 | \* |
|  |  | SBS35 | 0.9818 | 5.7103 | -0.002 | 0.0373 | 0.000363188 |  |
|  |  | SBS42 | 1 | 22.4655 | 0.016 | 0.0797 | 1.02E-11 | \* |
|  |  | SBS7b | 0.9818 | 4.3483 | -0.006 | 0.047 | 0.00159397 |  |
|  |  | ID1 | 0.9636 | 2.8017 | -0.071 | 0.1308 | 0.00896248 |  |
|  |  | ID12 | 0.9455 | 1.6927 | -0.141 | 0.1944 | 0.0328861 |  |
| N01 | N02-N05,M01-M14,K01-K09,L01-L14,C01-C18 | SBS23 | 1 | 463.48 | 0.061 | 0.0785 | 6.7624E-204 | \* |
|  |  | SBS30 | 1 | 560.0535 | 0.379 | 0.4614 | 7.04E-246 | \* |
|  |  | SBS35 | 0.9831 | 6.1323 | -0.002 | 0.0377 | 0.000230807 |  |
|  |  | SBS42 | 1 | 24.1093 | 0.016 | 0.0799 | 1.91E-12 | \* |
|  |  | SBS7b | 0.9831 | 3.5811 | -0.006 | 0.0457 | 0.00372279 |  |
|  |  | ID1 | 0.9661 | 2.915 | -0.071 | 0.1309 | 0.00787722 |  |
|  |  | ID12 | 0.9322 | 1.6202 | -0.141 | 0.1943 | 0.035923 |  |
| N02-N05 | M01-M14,K01-K09,L01-L14,C01-C18 | SBS10b | 0.75 | 0.6906 | -0.129 | 0.0326 | 0.00452259 |  |
|  |  | SBS12 | 0.5568 | -1 | -0.035 | 0.0085 | 0.0399199 |  |
|  |  | SBS16 | 0.7591 | -1 | -0.081 | 0.0203 | 0.00660571 |  |
|  |  | SBS17a | 0.7659 | -1 | -0.042 | 0.0097 | 0.0341816 |  |
|  |  | SBS24 | 0.4955 | 0.1889 | -0.094 | 0.0326 | 0.0122078 |  |
|  |  | SBS25 | 0.6045 | -1 | -0.017 | 0.0044 | 1.05E-02 |  |
|  |  | SBS27 | 0.6114 | -1 | -0.002 | 0.0024 | 0.000095365 |  |
|  |  | SBS31 | 0.7932 | -1 | -0.072 | 0.0283 | 0.036112 |  |
|  |  | SBS34 | 0.6727 | -1 | -0.008 | 0.004 | 0.000735335 |  |
|  |  | SBS39 | 0.7955 | 0.539 | -0.211 | 0.0607 | 0.0431214 |  |
|  |  | SBS48 | 0.7227 | -1 | -0.009 | 0.0028 | 0.0156059 |  |
|  |  | SBS7b | 0.725 | -1 | -0.068 | 0.0198 | 0.0125894 |  |
|  |  | ID8 | 0.6182 | -1 | -0.107 | 0.0285 | 4.46E-02 |  |
| L01-L14,K01-K09 | M01-M14,N01-N05,C01-C18 | SBS1 | 0.7027 | 0.3374 | -0.231 | 0.0713 | 1.06E-03 |  |
|  |  | SBS14 | 0.7004 | -1 | -0.032 | 0.0044 | 4.49E-03 |  |
|  |  | SBS15 | 0.9066 | 1.0698 | -0.107 | 0.1956 | 4.01709E-08 | \* |
|  |  | SBS20 | 0.869 | 1.1993 | -0.065 | 0.0519 | 1.35E-08 | \* |
|  |  | SBS21 | 0.9054 | -1 | -0.027 | 0.0264 | 6.31697E-07 | \* |
|  |  | SBS35 | 0.7744 | -1 | -0.046 | 0.0099 | 0.000566542 |  |
|  |  | SBS43 | 0.7156 | -1 | -0.04 | 0.0042 | 0.0491042 |  |
|  |  | SBS54 | 0.8649 | 1.0452 | -0.099 | 0.035 | 9.03E-07 | \* |
|  |  | SBS57 | 0.7515 | -1 | -0.087 | 0.0278 | 2.23E-04 |  |
|  |  | SBS6 | 0.7538 | 0.513 | -0.24 | 0.0842 | 1.65E-04 |  |
|  |  | SBS7c | 0.728 | -1 | -0.05 | 0.0049 | 0.0478156 |  |
|  |  | ID2 | 0.9965 | 7.5541 | -0.074 | 0.4055 | 1.68E-20 | \* |
|  |  | ID7 | 0.9759 | 1.9931 | -0.114 | 0.155 | 9.71E-12 | \* |
| L01-L14 | M01-M14,N01-N05,C01-C18 | SBS1 | 0.8456 | 1.0758 | -0.2 | 0.1122 | 4.96E-06 | \* |
|  |  | SBS14 | 0.7471 | -1 | -0.032 | 0.0062 | 0.00129999 |  |
|  |  | SBS15 | 0.888 | 1.0619 | -0.107 | 0.189 | 1.11E-07 | \* |
|  |  | SBS20 | 0.888 | 1.3589 | -0.065 | 0.0551 | 1.35525E-08 | \* |
|  |  | SBS21 | 0.9054 | -1 | -0.027 | 0.0223 | 8.44363E-10 | \* |
|  |  | SBS35 | 0.7481 | -1 | -0.046 | 0.011 | 0.00149904 |  |
|  |  | SBS54 | 0.918 | 1.6539 | -0.049 | 0.0372 | 2.40E-06 | \* |
|  |  | SBS57 | 0.7181 | -1 | -0.087 | 0.0255 | 2.63E-03 |  |
|  |  | SBS6 | 0.7838 | 0.6868 | -0.24 | 0.0941 | 0.000243831 |  |
|  |  | SBS7c | 0.7452 | -1 | -0.05 | 0.0068 | 0.0306141 |  |
|  |  | ID2 | 0.9961 | 7.4562 | -0.074 | 0.4138 | 6.90164E-16 | \* |
|  |  | ID7 | 0.9749 | 1.6442 | -0.114 | 0.158 | 8.82128E-10 | \* |
| L01-L14 | K01-K09 | SBS1 | 0.7619 | 0.6345 | -0.228 | 0.1045 | 0.00928907 |  |
|  |  | ID1 | 0.7183 | -1 | -0.035 | 0.0095 | 0.0431202 |  |

**Figures**

**Supplementary Figure 1.** Assessment of SBS and ID mutational signatures while varying the DP threshold for samples L01 (a, b), L04 (c, d), and L10 (e, f), representing tumours with germline PVs in *MSH2*, *MSH6*, and *MLH1* respectively. SBS signatures previously associated with MMRd are prevalent across most thresholds. ID signatures ID2 (green) and ID7 (purple) dominate all three examples at most DP settings.

**Supplementary Figure 2.** Assessment of somatic loss of heterozygosity (LOH) across genes of interest *MUTYH* and *NTHL1* in control samples C19 (a) and C17 (b), as well as monoallelic *MUTYH* samples M11-M14 (c, d, e, f), and monoallelic *NTHL1* samples N02-N05 (g, h, i, j). Each dot indicates a variant seen at a given allele frequency in the tumour. LOH is unlikely in regions containing heterozygous variants (red), while somatic homozygous variants seen as heterozygous in the germline sequence are indicative of LOH (green). Samples M11-M14 did not exhibit LOH, nor any somatic pathogenic variant in *MUTYH*. Similarly, N02-N05 did not exhibit LOH, or any somatic pathogenic variant across *NTHL1*.

**Bibliography**

1 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20. doi:10.1093/bioinformatics/btu170

2 Saunders CT, Wong WSW, Swamy S, *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 2012;**28**:1811–7. doi:10.1093/bioinformatics/bts271

3 Wellcome Sanger Institute. COSMIC Signatures of Mutational Processes in Human Cancer. Signatures of Mutational Processes in Human Cancer. 2019.https://cancer.sanger.ac.uk/cosmic/signatures (accessed 31 May2019).

4 Webb AR. *Statistical Pattern Recognition*. 2nd ed. West Sussex, England: : Wiley 2002.

5 Grubbs FE. Sample Criteria for Testing Outlying Observations. *Ann Math Statist* 1950;**21**:27–58. doi:10.1214/aoms/1177729885

6 Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995;**310**:170.

7 de Carvalho JF, Draper NR, Smith H. Applied Regression Analysis (2nd ed). *J Am Stat Assoc* 1981;**76**:1012. doi:10.2307/2287608

8 Niu B, Ye K, Zhang Q, *et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* 2014;**30**:1015–6. doi:10.1093/bioinformatics/btt755