nature communications

Article

https://doi.org/10.1038/s41467-025-59775-2

# Unraveling mutagenic processes influencing the tumor mutational patterns of individuals with constitutional mismatch repair deficiency

Received: 22 November 2023

Accepted: 5 May 2025

Published online: 14 May 2025

Check for updates

Dilys D. Weijers[1], Snežana Hinić[2,10], Emma Kroeze[1,10], Mark AJ Gorris[3,4], Gerty Schreibelt[3], Sjors Middelkamp[1,5], Arjen R. Mensenkamp[2], Reno Bladergroen[1], Kiek Verrijp[3,4,6], Nicoline Hoogerbrugge[2], Pieter Wesseling[1,7], Rachel S. van der Post[6], Jan LC Loeffen[1], Corrie EM Gidding[1], Mariëtte CA van Kouwen[8], I. Jolanda M. de Vries[3], Ruben van Boxtel[1,5], Richarda M. de Voer[2], Marjolijn CJ Jongmans[1,9] & Roland P. Kuiper[1,9]

Constitutional mismatch repair deficiency (CMMRD), caused by bi-allelic germline variants in mismatch repair (MMR) genes, is associated with high cancer incidence early in life. A better understanding of mutational processes driving sequential CMMRD tumors can advance optimal treatment. Here, we describe a genomic characterization on a representative collection of CMMRD-associated tumors consisting of 41 tumors from 17 individuals. Mutational patterns in these tumors appear to be influenced by multiple factors, including the affected MMR gene and tumor type. Somatic polymerase proofreading mutations, commonly present in brain tumors, are also found in a T-cell lymphoblastic lymphoma displaying associated mutational patterns. We show prominent mutational patterns in two second primary hematological malignancies after temozolomide treatment. Furthermore, an indel signature, characterized by one-base pair cytosine insertions in cytosine homopolymers, is found in 54% of tumors. In conclusion, analysis of sequential CMMRD tumors reveals diverse mutational patterns influenced by the affected MMR gene, tumor type and treatment history.

DNA mismatch repair (MMR) protects the genomic integrity of eukaryotic cells by repairing mismatched bases and small insertions and deletions (indels), primarily during replication[1]. Monoallelic germline pathogenic variants (gPVs) in one of the MMR genes *MLH1*, *MSH2*, *MSH6* or *PMS2* cause Lynch syndrome, which predisposes primarily to adult-onset cancers in the gastrointestinal- (GI) and genitourinary tract[2]. When both alleles of an MMR gene are affected by a gPV, this results in constitutional mismatch repair deficiency (CMMRD). CMMRD is a rare cancer predisposition syndrome (CPS), associated with a high incidence of cancer early in life. Individuals with CMMRD predominantly develop tumors in the brain, hematological system and GI tract[3,4]. These tumors likely develop as a consequence of the increased mutation rate due to MMR deficiency in all cells[5]. Overall survival among individuals with CMMRD is poor, but intensive surveillance protocols geared towards early, asymptomatic detection of tumors can improve outcome considerably, at least for brain tumors and non-hematological solid tumors. The early detection of tumors opens up possibilities for full resection and less aggressive

A full list of affiliations appears at the end of the paper. e-mail: r.kuiper@prinsesmaximacentrum.nl

treatment[6,7]. The latter is important, because tumors of individuals with CMMRD regularly accumulate high numbers of additional somatic mutations due to their resistance to a subset of widely used mutagenic agents[8]. There is also still a need for specialized treatments for individuals with CMMRD. To advance the optimal treatment for individuals with CMMRD, it is important to understand the mechanisms driving tumor development.

The mechanisms driving the development of MMR-deficient tumors in general, and CMMRD tumors in particular, have been studied using mutational patterns[5,9–11]. MMR-deficient tumors present with a high tumor mutational load (TML) when compared to sporadic tumors of the same tissue type, and most of them are hypermutated (>10 mutations/Mb)[12]. MMR deficiency-related mutations are typical with respect to mutation type and sequence contexts, and present as recognizable mutational signatures[13]. In the Catalogue of Somatic Mutations in Cancer (COSMIC) database (v3.2), five single base substitution (SBS) signatures (SBS6, SBS15, SBS21, SBS26 and SBS44) and three indel signatures (ID1, ID2 and ID7) have thus far been associated with MMR deficiency[14]. Together, these signatures can adequately explain a substantial proportion of the mutational patterns in MMR-deficient tumors[10,15]. However, additional internal and external factors can also influence the mutagenesis in MMR-deficient tumors. For example, somatic mutations in polymerase genes *POLE* and *POLD1* often result in ultrahypermutation (>100 mutations/Mb) and are regularly encountered in MMR-deficient tumors and, more specifically, malignant brain tumors of individuals with CMMRD[12,16]. SBS signatures SBS14 and SBS20 are associated with these combined MMR and polymerase defects[14]. Furthermore, cancer therapy itself can influence the mutational patterns in MMR-deficient tumors[17–19]. A comprehensive overview of mutational patterns in tumors of individuals with CMMRD can help to understand the impact of prior therapy on cancer risk and guide treatment decisions.

Here, we present a cohort of 17 individuals with CMMRD and study mutational patterns in 41 tumors diagnosed in these individuals. The tumors are a representative reflection of CMMRD-associated tumor types and include primary and second (primary) tumors, with multiple sequential tumors in the same individual. We show the presence of specific somatic mutational processes related to tumor type, affected MMR gene and exposure to prior cancer treatment in tumors of individuals with CMMRD.

## Results

### Cohort description

We collected clinicopathological information for 17 individuals with CMMRD from 14 families diagnosed in the Netherlands. These 17 individuals presented with bi-allelic gPVs in *PMS2* ($n = 10$), *MSH6* ($n = 5$) and *MSH2* ($n = 2$) (Fig. 1 and Table S1). Collectively, these individuals developed 52 malignancies (median 3, range 1-7) and during surveillance, liver and GI adenomas were detected in 12 individuals on 39 occasions (Fig. 1 and Table S2). Most adenomas developed in the colorectum. The malignancies developed in the brain ($n = 18$), GI-tract ($n = 18$) and hematological system ($n = 15$), and one individual (C16)

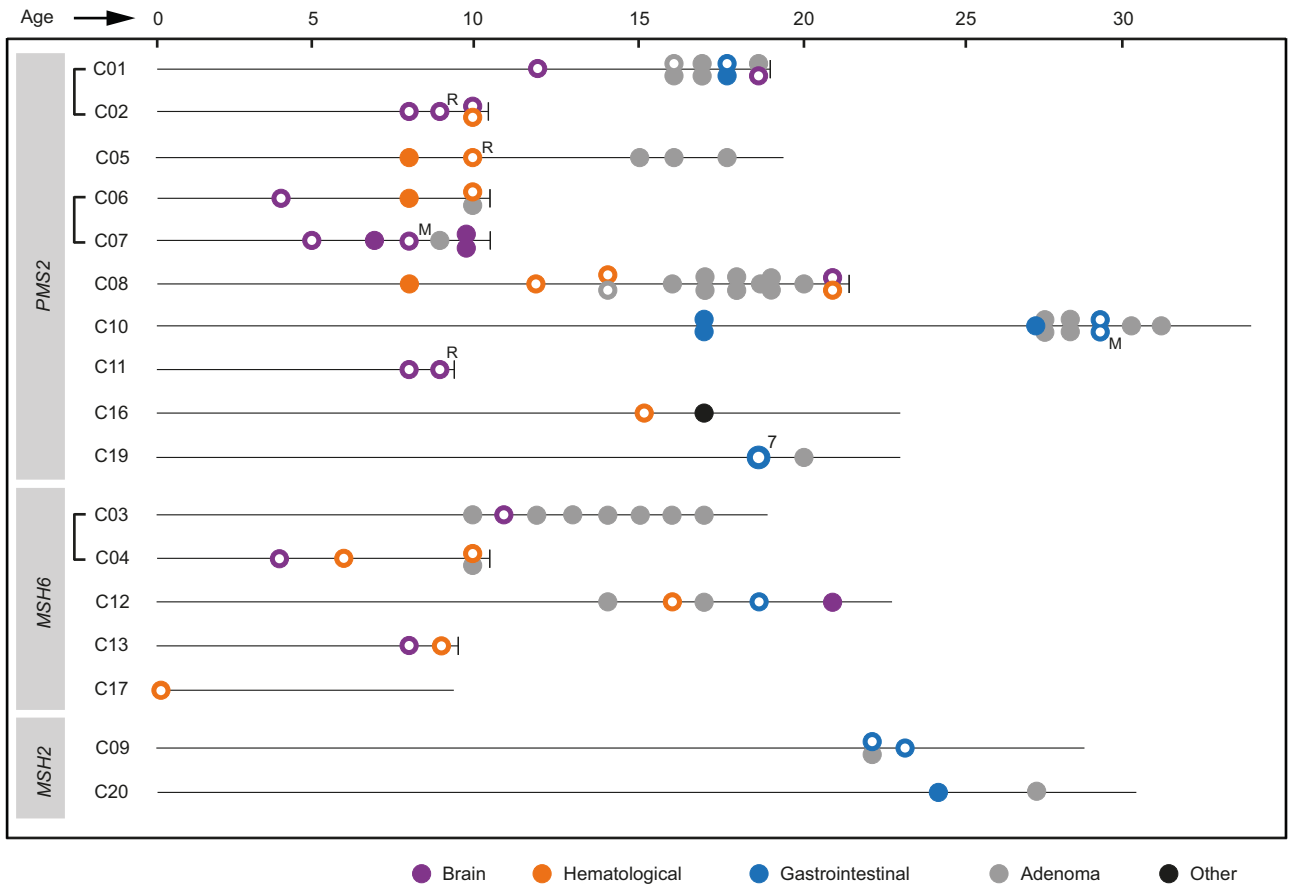


**Fig. 1 | Schematic overview of tumors developed by individuals in the CMMRD cohort.** Siblings are connected by a bracket. For each individual, malignancies, including relapses (R) and metastases (M), and the occasions at which liver or GI adenomas were detected during surveillance are indicated on a timeline according to age at diagnosis. The death of an individual is indicated by a vertical dash at the end of the timeline. For tumors marked with a white dot, material was collected and sequenced. Individual C19 was diagnosed with seven simultaneous colorectal adenocarcinomas, all of which were sequenced. Previous studies have described individuals C03 and C04[62], C01 and C06[63], C09 and C20[64], C02, C04, C05, C06, C08, C12, C16 and C17[65], C11[66,67] and C10[7]. Source data are provided as a Source Data file.

developed a squamous cell carcinoma of the skin. Most malignant brain tumors were glioblastomas or high-grade gliomas in the cerebral hemisphere ($n = 10$) and most GI cancers were colorectal adenocarcinomas (CRC) ($n = 10$), while most hematological malignancies were T-cell lymphoblastic lymphomas (T-LBL) ($n = 10$) (Table S2). We observed age differences between the types of malignancies, with malignant brain tumors (median 9, range 4–21 years) and hematological malignancies (median 10, range 0-21 years) developing at an earlier age than GI cancers (median 19, range 17–29 years) ($p < 0.0001$ and $p < 0.001$ respectively, Kruskal-Wallis test followed by a Dunn's test with Bonferroni adjustment).

### Hypomorphic gPVs in *MSH2* in two individuals with later-onset cancer

Individuals with CMMRD due to biallelic gPVs in *MSH2* or *MLH1* generally develop cancer at a younger age compared to individuals with gPVs in *MSH6* or *PMS2*, and often decease before reaching adulthood[4]. However, individuals C09 and C20, who both have bi-allelic gPVs in *MSH2*, developed GI tumors only at an adult age, suggesting that these variants might be hypomorphic and result in residual MSH2 activity. Variant databases and prior literature gave no or variable interpretations of pathogenicity for two of the gPVs in *MSH2* (Fig. S1a). The homozygous gPV c.2727dup (p.Gln910Thrfs*3) in C09 is located at the far end of *MSH2* and, following the ACMG guidelines, it is classified as variant of uncertain significance. Functional testing of gPV c.2727dup was performed through re-expression of $MSH2^{c.2727dup}$ in an MSH2KO cell line. This revealed that $MSH2^{c.2727dup}$ could not rescue the increased mutation accumulation in this cell line (Fig. S1b, c), thus confirming that the variant is deleterious. We nevertheless reasoned that (part of) this protein could have retained some functionality, for example through alternative splicing. We therefore performed RNA sequencing of a non-malignant sample of individual C09 and detected an alternatively spliced transcript containing a downstream alternative last exon (Fig. S1d). This transcript has previously been reported in a study by Thompson et al under the name 16A[20]. Presence of this alternative transcript may partially rescue MSH2 function in this individual. Individual C20 is compound heterozygous for a large pathogenic deletion of the 5′ end of *MSH2*, including exon 1, and an exon 11 splice region variant (c.1759+3 A > G). The latter has conflicting interpretations of pathogenicity in ClinVar but was recently found to introduce a frameshift due to exon 11 skipping (Fig. S1a). RNA sequencing of a non-malignant sample of individual C20 confirmed the exon 11 skipping, but also revealed the presence of correctly spliced transcripts (Fig. S1d). With the other allele not being expressed, this observation illustrates that the splice site variant does not completely alter *MSH2* splicing, and wildtype MSH2 is still expressed. Overall, our findings in transcriptome data are in line with the attenuated CMMRD phenotype of individuals C09 and C20, a phenomenon that was recently described in other CMMRD individuals as well[21].

### Tumor mutational load and mutational signatures in CMMRD tumors

For 39 malignancies in the CMMRD cohort (75%), including 14 malignant brain tumors, 13 GI cancers and 12 hematological malignancies, we collected and sequenced genomic DNA. In addition, we sequenced one high-grade and one low-grade GI adenoma (Fig. 1, Table S3). Of these 41 tumors, 13 were whole genome sequenced and 28 were whole exome sequenced and, to enable comparison, downstream analysis was performed on the overlapping regions between exome captures (Methods). A Fisher-Freeman-Halton asymptotic test for associations between MMR gene, tumor type and polymerase status of the tumor showed that only tumor type and polymerase status of the tumor were significantly associated ($p = 0.0002$, Table S4). Two-sided pairwise Fisher's Exact tests show that the association is only significant for brain tumors (Table S4), which has been described elaborately in prior studies as well. Eighteen tumors were found to be ultrahypermutated (median 234.9 mut/Mb, range 122.8-624.5) and 17 were hypermutated (median 22.8 mut/Mb, range 12–89.7) (Fig. 2 and Table S5)[12]. We performed a de novo signature extraction with all 41 sequenced CMMRD tumors (Fig. S2, Fig. S3 and Supplementary note 1). Seven COSMIC (v3.2) SBS signatures[14] contributed to the mutational profile of the CMMRD tumors, each accounting for >15% of the mutations in these tumors (Fig. 2). The MMR deficiency-associated signatures SBS15 and SBS26 were most abundantly present in the CMMRD cohort, contributing to 34 and 14 tumors, respectively. We detected SBS14, associated with combined MMR and POLE proofreading deficiency, in six tumors and SBS20, associated with combined MMR and POLD1 proofreading deficiency, in three tumors. Another three tumors had contributions of SBS11, which is associated with temozolomide treatment in an MMR-deficient context[18]. SBS1, known as a clock-like signature, but also regularly seen in MMR-deficient tumors[5,10,22], was detected in 12 tumors. MMR-associated signature SBS44 was not detected, which may be due to its similarity to other MMR-associated signatures, causing it to be redundant in our model. Of the six de novo extracted indel signatures, three were highly similar to COSMIC (v3.2) signatures[14] ID1, ID2, and ID12, which contributed to >15% of the mutations in 28, 24 and 5 tumors in our cohort, respectively. ID1, characterized by one-bp thymine insertions, and ID2, characterized by one-bp thymine deletions, are both associated with MMR deficiency. As a validation for mutational signature assignment in the CMMRD tumors, we repeated the de novo extraction with SigProfiler, which confirmed the contributions and distributions of the seven SBS signatures and indel signatures ID1 and ID2 to the mutational profile in CMMRD tumors (Fig. S4).

### Indel signature IDA

In addition to the detection of known signatures, we extracted indel signature IDA ($IDA^{SP}$ in the SigProfiler extraction), which is not included in COSMIC v3.2 (Fig. S5a, b). Signature IDA accounted for >15% of mutations in 22 tumors, both treatment-naïve and post-treatment (Fig. 2). The most prominent feature of signature IDA is a one-bp cytosine insertion in homopolymers of five, six or seven cytosines (InsC5 + ; Fig. S5c). Signature IDA was not significantly associated with a specific MMR gene or tumor type, although the contributions appeared to be higher in PMS2-deficient tumors and lower in malignant brain tumors (Fig. S5d–g). Absolute contributions of IDA in malignant brain tumors were lower compared to GI tumors, with a $p$-value just below the threshold ($p = 0.0492$). We found a significantly higher relative (mean 32.6 vs 12.0, $p = 0.006$, two-sided Mann-Whitney U test) and absolute contribution (mean 182.0 vs 161.7, $p = 0.02$, two-sided Mann-Whitney U test) of signature IDA in tumors that did not have a (likely) pathogenic somatic mutation in *POLD1* or *POLE*, compared to those that did (Fig. S5h, i). The previously reported signature MS-Sig4, which was found to be associated with polymerase proofreading-deficient tumors, is almost exclusively hallmarked by InsC5+ [23], the prominent feature of IDA and $IDA^{SP}$. We therefore analyzed the distribution of Ins5C+ in the cohort of CMMRD tumors separately. Again, we observed higher relative and absolute contributions in PMS2-deficient tumors ($n = 30$) compared to MSH2-deficient ($n = 2$) (mean 0.12 vs 0.02 and 120.0 vs 7.5, $p = 0.02$ and ns) and MSH6-deficient tumors ($n = 9$) (mean 0.12 vs 0.05 and 120.0 vs 9.3, $p < 0.01$ and $p < 0.0001$, Kruskal-Wallis test followed by a Dunn's test with Bonferroni adjustment) and no difference in relative and absolute contributions between tumor types (Fig. S5j–m). In polymerase proofreading-mutated tumors, absolute contributions of InsC5+ were significantly higher than in tumors without mutations in the polymerase proofreading domain of *POLD1* or *POLE* (mean 151.7 vs 42.1, $p < 0.001$, two-sided Mann-Whitney U test), while relative contributions of InsC5+ did not differ (Fig. S5n, o).

https://doi.org/10.1038/s41467-025-59775-2

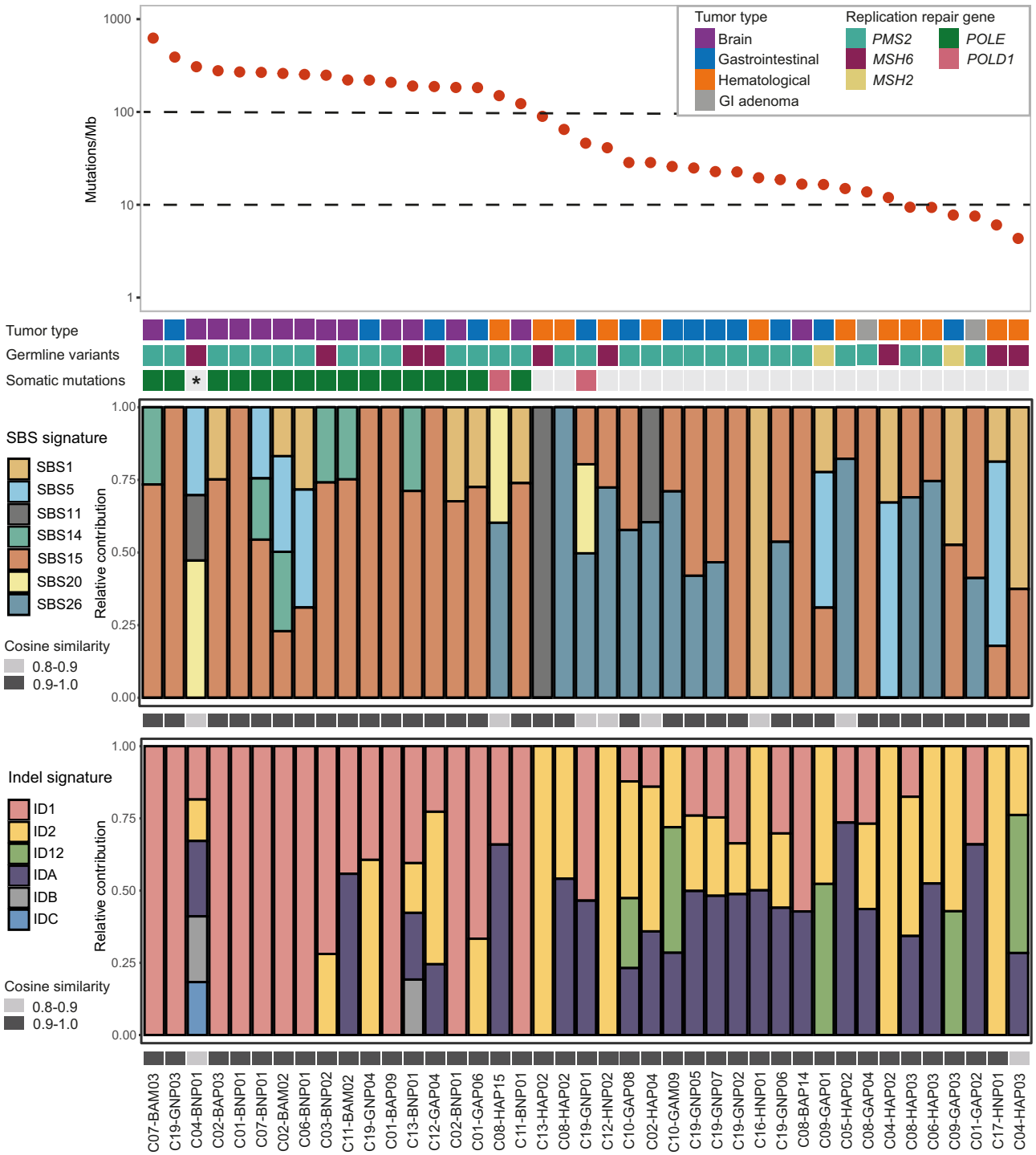


**Fig. 2 | Overview of mutational patterns in all 41 CMMRD tumors sequenced.** Total TML in each tumor is reported, the ultrahypermutation ( >100mutations/Megabase(Mb)) and hypermutation ( >10mutations/Megabase(Mb)) thresholds are indicated by dashed lines. Tumor type, affected MMR gene and presence of somatic driver mutations in POLE or POLD1 are reported below. The contributions of COSMIC and de novo single base substitution (SBS) and indel mutational signatures indicated result from the signature extraction with MutationalPatterns. Mutational signatures are associated with various processes, SBS1 and SBS5: clock-like signatures, related to aging, SBS11: prior temozolomide treatment, SBS14: combined MMR and POLE deficiency, SBS15: MMR deficiency, SBS20: combined MMR and POLD1 deficiency, SBS26: MMR deficiency, ID1: aging and MMR deficiency, ID2: aging and MMR deficiency, ID12: unknown aetiology[14], IDA-IDC: de novo extracted signatures. The cosine similarity between the original and reconstructed mutational profile is specified below the signature contributions. Tumors are identified by the individual ID, followed by tumor type (B=brain, G = gastrointestinal (GI), H=hematological), treatment-status (N= treatment-naïve, A= after treatment), tumor status (P=primary tumor, M=metastasis or relapse) and the order in which they occurred (Fig. 1 and Table S2). *In C04-BNP01, a somatic POLD1 mutation was validated in transcriptomic data only. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-025-59775-2

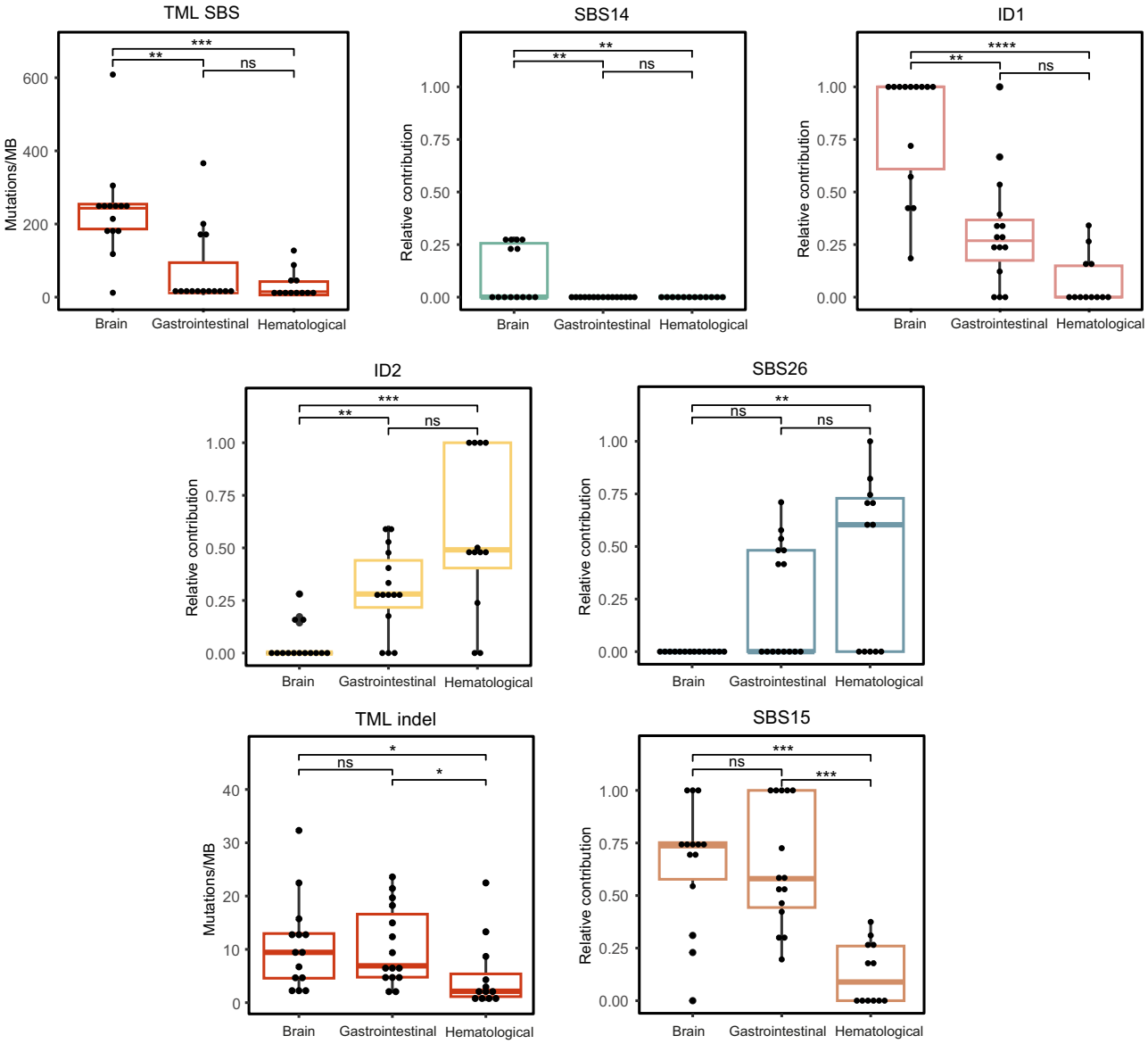


**Fig. 3 | Tumor mutational load (TML) and mutational signatures in CMMRD tumors differ depending on the tumor type.** The single base substitution (SBS) TML and relative contribution of SBS14 and ID1 are significantly higher in malignant brain tumors ($n = 14$) than in gastrointestinal tumors ($n = 15$) and hematological malignancies ($n = 12$). Conversely, compared to malignant brain tumors, the relative contribution of ID2 is significantly higher in gastrointestinal tumors and hematological malignancies, and the relative contribution of SBS26 is significantly higher in hematological malignancies. The indel TML and relative contribution of SBS15 is significantly higher in malignant brain tumors and gastrointestinal tumors than in hematological malignancies. The center line of the box displays the median, the box limits represent the first and third quartiles and the whiskers indicate the minimum and maximum, or, in case of outliers, 1.5 times the interquartile range. *P*-values were calculated with a Kruskal-Wallis test, followed by a Dunn's test with Bonferroni adjustment. ns = not significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$. Means and *p*-values are reported in Table S6. Source data are provided as a Source Data file.

## Tumor type-associated mutational profiles in CMMRD tumors

TML and the relative contribution of known mutational signatures differed significantly between tumor types in the CMMRD cohort (Fig. 3 and Table S6). Malignant brain tumors presented with a higher SBS TML and a higher relative contribution of signatures SBS14 and ID1 than GI tumors and hematological malignancies. Conversely, compared to malignant brain tumors, we found a higher contribution of ID2 in GI tumors and hematological malignancies and a higher contribution of signature SBS26 in hematological malignancies (Fig. 3). These differences could largely be explained by somatic *POLE* mutations that, in our cohort, occurred more often in malignant brain tumors (12 out of 14) than in GI tumors (four out of 15) and were absent in hematological malignancies ($n = 12$) (Fig. 2, Table S7). For 14 of the *POLE*-mutated tumors, the *POLE* mutations affected the exonuclease domain. Two tumors had a p.Glu978Gly mutation which affects the polymerase domain, but nevertheless has been shown to drive ultrahypermutation[12,24]. For 15 of these 16 *POLE* mutations, we found evidence for pathogenicity in prior studies (Table S7). POLE proofreading deficiency is known to increase SBS TML and lead to higher contributions of signatures SBS14 and ID1[12,23,25]. As a consequence, the relative contribution of other signatures, such as SBS26 and ID2, was found to be decreased in POLE proofreading-deficient tumors. Finally, we detected a higher indel TML and relative contribution of SBS15 in malignant brain tumors and GI tumors when compared to hematological malignancies (Fig. 3). This appears to be related to polymerase proofreading deficiency as well, as both the indel TML and relative contribution of SBS15 are significantly higher in polymerase proofreading-deficient tumors compared to polymerase

https://doi.org/10.1038/s41467-025-59775-2

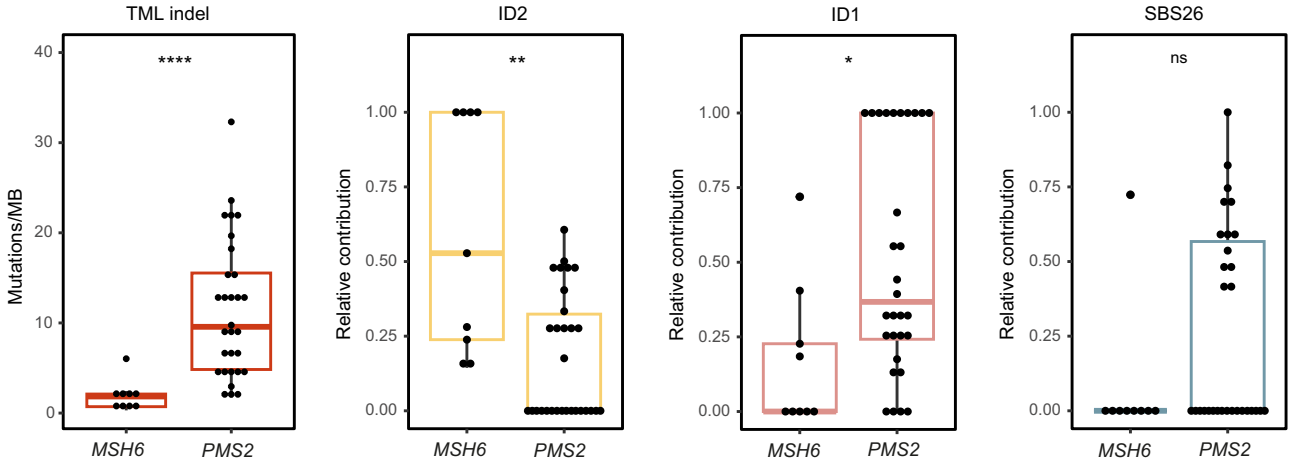


**Fig. 4 | Tumor mutational load (TML) and mutational signature contributions differ between CMMRD tumors with different affected MMR genes.** The indel TML is significantly lower and the relative contribution of signature ID2 is significantly higher in MSH6-deficient tumors ($n = 9$). PMS2-deficient tumors (n = 30) have significantly higher contributions of ID1 and, although not significant, higher contributions of SBS26. The center line of the box displays the median, the box limits represent the first and third quartiles and the whiskers indicate the minimum and maximum, or, in case of outliers, 1.5 times the interquartile range. *P*-values were calculated with a two-sided Mann-Whitney U test. ns = not significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{****}p < 0.0001$. Means and *p*-values are reported in Table S8. Source data are provided as a Source Data file.

proofreading-proficient tumors (mean 14.17 vs 4.91 and 0.67 vs 0.36, $p < 0.001$ and $p < 0.01$, two-sided Mann-Whitney U test). Other known signatures did not have notably different relative contributions between tumor types (Fig. S6 and Table S6).

### POLD1 proofreading deficiency

While POLE proofreading deficiency occurred in 14/41 CMMRD tumors in our cohort, predominantly brain tumors, POLD1 proofreading deficiency seemed to be less frequent. The POLD1 proofreading deficiency-associated signature SBS20 was detected in three tumors (Fig. 2). In the GI tumor C19-GNP01, the presence of SBS20 could be explained by a confirmed pathogenic mutation, leading to p.Asp402Asn in the POLD1 exonuclease domain (Table S7). MSH6-deficient high-grade glioma C04-BNP01 presented with high SBS TML (305 mut/Mb) and 47% of the mutations were attributed to SBS20, strongly suggesting a POLD1 exonuclease defect. Notably, WGS of the tumor did not reveal a POLD1 exonuclease domain mutation, but in RNA sequencing data, a c.1504 G > T variant leading to p.Asp502Tyr in the exonuclease domain was detected (Fig. S7). Although this variant was not previously reported as pathogenic and it could not be validated in tumor DNA, mutation of the orthologous position in yeast (Asp506) has been shown to increase mutagenicity in an MSH6-deficient context[26], and this variant may therefore explain the POLD1 proofreading deficiency-associated characteristics in this tumor.

The third tumor with contributions of SBS20 was a T-LBL from a 21-year-old individual (C08-HAP15). As shown in our cohort and in previous studies[12,16], somatic polymerase proofreading deficiency is common in malignant brain tumors of individuals with CMMRD, but for CMMRD-associated hematological malignancies, cases with polymerase proofreading deficiency have not been described so far. We screened for somatic mutations in *POLD1* in this tumor and detected a clonal missense mutation leading to p.Asp515Gly in the exonuclease domain of POLD1, which was absent in the normal samples and three other malignancies of this individual. The consensus classification in ClinVar[27] is variant of uncertain significance. However, Asp515 is one of four highly conserved, negatively charged amino acids that attract magnesium ions to catalyze exonuclease activity[28]. Alteration of the orthologous amino acid in yeast (Asp520) was found to lead to a significant decrease in exonuclease activity and increased mutation rates[26,29,30]. Additionally, an increased mutation rate and contributions of SBS20 have been shown in tumors where the other three highly conserved amino acids were affected[25,28,31]. It is therefore likely that the missense variant leading to p.Asp515Gly is in fact pathogenic and caused the ultrahypermutation phenotype and contributions of SBS20 in tumor C08-HAP15. Ultrahypermutated tumors are associated with a favorable response to PD1-inhibitors[32]. Although additional studies in hematological malignancies are required, this may open up new treatment possibilities for this tumor type in individuals with CMMRD.

### MMR gene-associated mutational profiles

In addition to tumor type-associated mutational profiles, TML and the relative contribution of mutational signatures also differed between CMMRD tumors with different affected MMR genes (Fig. 4 and Table S8). We compared PMS2- and MSH6-deficient tumors because these groups had sufficient sample sizes ($n = 30$ and 9, respectively). In the CMMRD cohort, MSH6-deficient tumors had a significantly lower indel TML and higher relative contribution of signature ID2 than PMS2-deficient tumors. The PMS2-deficient tumors, on the other hand, had significantly higher relative contributions of signature ID1 (Fig. 4). These findings were consistent with previous in vitro studies of MSH6- and PMS2-deficient cell lines[11]. In addition, there was a trend towards higher relative SBS26 contribution in PMS2-deficient tumors, in line with previous findings in PMS2-deficient tumors and cell lines[11,18]. SBS TML and relative contributions of other identified signatures were not notably different between tumors with different affected MMR genes (Fig. S6 and Table S8).

### Discrepant mutational profiles in two T-LBLs

In T-LBLs C12-HNP02 and C16-HNP01, the mutational profiles were different from what was expected based on the underlying gPVs. Both malignancies were primary and not treated with neo-adjuvant therapy, excluding the possibility of prior therapy influences. MSH6-deficient C12-HNP02 presented with high contributions of PMS2 deficiency-associated SBS26, while PMS2-deficient C16-HNP01 lacked contributions of SBS26 (Fig. 2). For C12-HNP02, further dissection of the mutational profile showed a smaller contribution of SBS15 as well (Fig. 2). The VAFs of the SBS26-like T > C mutations and SBS15-like C > T mutations in this tumor were similar and indicated that most mutations were present in the majority of tumor cells (Fig. S8). The SBS TML was higher than in other CMMRD T-LBLs in our cohort, (39.6 mut/Mb vs median 12.1 mut/Mb; $n = 8$) and was largely determined by T > C mutations (23.3 Mut/Mb). The low indel TML and predominance of

https://doi.org/10.1038/s41467-025-59775-2

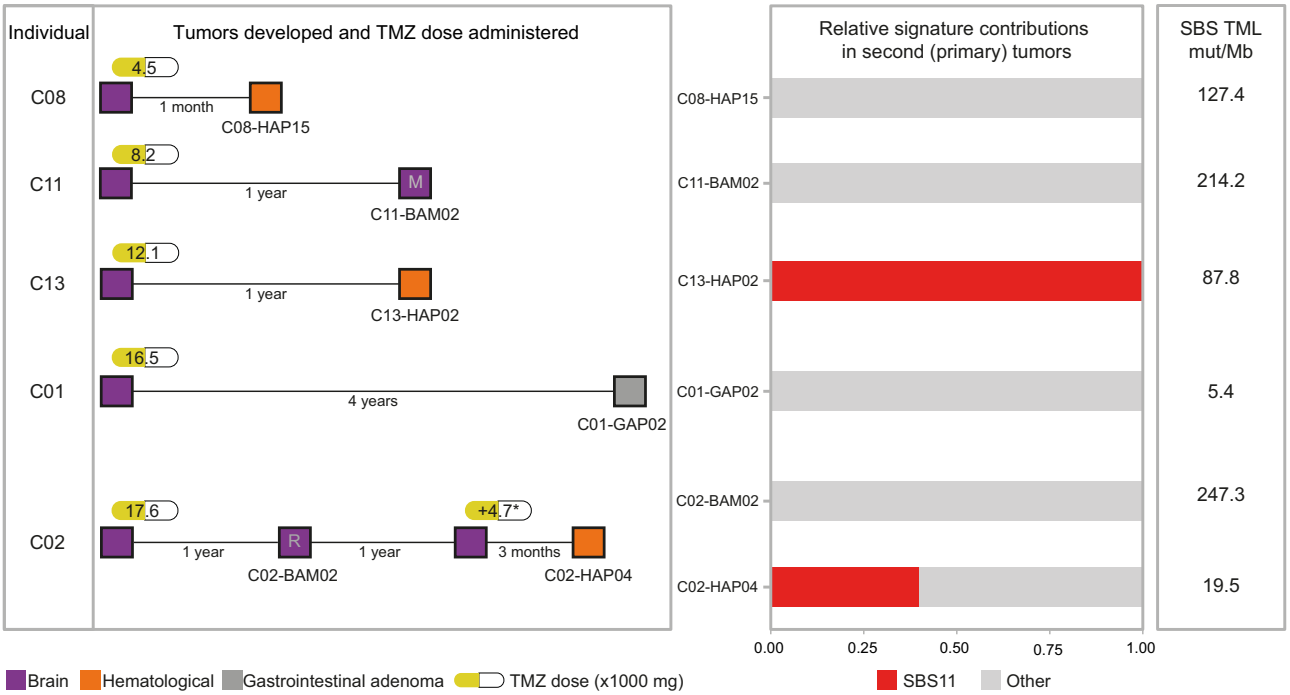


**Fig. 5 | Contributions of signature SBS11 in second (primary) tumors of individuals with CMMRD who were treated with temozolomide (TMZ).** Five individuals in the CMMRD cohort were treated with temozolomide for a total of six primary malignant brain tumors. The cumulative temozolomide dose is indicated for each primary malignant brain tumor. Individual C02 was treated with temozolomide for two primary brain tumors, further increasing the cumulative dose (milligrams (mg)). After variable timeframes following the start of temozolomide treatment, these individuals developed a second primary tumor, or a metastasis (M) or relapse (R) of the primary malignant brain tumor. In these subsequent tumors, the single base substitution (SBS) tumor mutational load (TML) ranges between 5.4 and 214.2 mutations/megabase (mut/Mb) and in two hematological malignancies, SBS11 contributed >35% of mutations to the mutational profile. *Dose administered during a second treatment protocol with temozolomide. Source data are provided as a Source Data file.

ID2-like thymine deletions were in line with the underlying MSH6 deficiency (Fig. S8). Using targeted single molecule molecular inversion probe (smMIP)-based next generation sequencing and multiplex ligation-dependent probe amplification (MLPA), the tumor was screened for (likely) pathogenic mutations or deletions in *PMS2*, but none were detected. Although we cannot exclude that rare intronic pathogenic mutations or hypermethylation of the *PMS2* locus may have been missed, a similar SBS26-like pattern was seen previously in an MSH6-deficient germinal center sample[5], suggesting that T > C mutational patterns in MSH6-deficient tumors may be a tissue-specific consequence of MSH6 deficiency. For C16-HNP01, we found that the mutational profile differed between clonal and subclonal mutations (Fig. S9), with the former consisting of SBS1-like C > T mutations and the latter consisting of C > T mutations in a broad mutational context. The indel mutational profile showed a predominance of ID2-like thymine deletions, where ID1-like thymine insertions were expected in this PMS2-deficient tumor. The SBS and indel TML were not markedly increased compared to other CMMRD T-LBLs in our cohort (16.9 and 2.4 mut/Mb vs median 12.1 and 2.3 mut/Mb, $n = 8$). No (likely) pathogenic alterations were detected in other MMR genes. These results show that tumor C16-HNP01 displayed mutational processes that were different from the typical PMS2 deficiency-associated mutagenesis.

### Effects of temozolomide treatment in second tumors

The inclusion of 22 tumors that developed after prior cancer treatment or that were resected after neo-adjuvant treatment allowed us to investigate whether treatment-associated mutagenesis influences the mutational patterns in second (primary) CMMRD tumors. The only treatment-associated mutational signature found to contribute to tumors of the CMMRD cohort was SBS11, related to temozolomide treatment in an MMR-deficient context[18]. We identified contributions of SBS11 in three tumors, two of which in individuals who received temozolomide during treatment of an earlier brain tumor (Fig. 2 and Fig. S4). The third case was a first primary tumor (C04-BNP01), which developed in an individual that had never been exposed to cancer treatment. These mutations could therefore not have been caused by temozolomide, and further analyses indicated that the attribution of SBS11 contribution to this sample was less reliable (Fig. S10). To investigate under which conditions CMMRD tumors accumulate temozolomide-induced mutations, we identified all individuals in our cohort that were treated with this agent and analyzed the mutational processes in the tumors they developed afterwards. We identified five individuals who were treated with temozolomide for a total of six primary malignant brain tumors; individual C02 received temozolomide for two primary brain tumors. The cumulative dose of temozolomide differed between individuals, depending on the protocol, age, and treatment interruptions due to side-effects or new tumor development. Within one month to four years after the start of temozolomide treatment, each of these individuals was diagnosed with a metastasis, relapse, or new primary tumor with TMLs ranging between 5.4 and 247.3 mut/Mb. These included two malignant brain tumors, one high-grade GI adenoma and three hematological malignancies. There does not appear to be an association between contributions of SBS11 and the cumulative temozolomide dose, the time between treatment initiation and new tumor development, or the TML. Furthermore, based on the mutational context of driver gene mutations, evidence for a direct causal link between temozolomide treatment and subsequent tumor development could not be found. However, we did detect contributions of SBS11 only in the second primary hematological malignancies C02-HAP04 and C13-HAP02 (Fig. 5). Our findings show that in individuals with CMMRD, temozolomide can influence the mutational profile of unrelated second primary malignancies and suggest that the hematological system could be more sensitive to this type of mutagenesis.

## Discussion

Individuals with CMMRD lack DNA mismatch repair in all cells of their body, resulting in rapid mutation accumulation and the development of multiple tumors early in life. In this study we have characterized the mutational patterns in subsequent tumors from individuals with CMMRD to investigate the impact of the affected MMR gene, tumor type, and prior treatment. The 41 selected tumors were a good representation of tumor types that arise in individuals with CMMRD, and included 12 hematological malignancies, which have not yet been extensively studied in the context of CMMRD. Signatures associated with MMR deficiency were detected in all tumors and in all tumor types, supporting the assumption that MMR deficiency drives tumorigenesis in individuals with CMMRD. Furthermore, we found that mutational patterns in CMMRD tumors are associated with tumor type, underlying MMR deficiencies, and somatic polymerase proofreading deficiency, which is in line with other studies[5,12,18,23,25]. Hematological CMMRD malignancies showed a lower indel TML and lower relative contributions of SBS15-associated mutations compared to the other tumor types, which is likely to be an indirect effect of the prevalent polymerase proofreading mutations and subsequent ultra-hypermutation in GI tumors and, particularly, brain tumors. Apart from these general associations, exceptional patterns were also observed in individual tumors, including a unique case of an ultra-hypermutated T-LBL with contributions of a POLD1 proofreading deficiency-associated mutational signature. This wide variety in contributions of MMR- and polymerase proofreading-associated signatures between individual tumors revealed a complex interplay of processes contributing to the mutation accumulation in CMMRD tumors.

In 22 of the 41 CMMRD tumors in the CMMRD cohort, we detected signature IDA, characterized by one-bp cytosine insertions in five-to-seven-cytosine homopolymers (InsC5+). One-bp indels are characteristic for MMR deficiency; they accumulate after polymerase slippage during replication and are preserved in the absence of functional MMR[14]. Insertions and deletions of thymine residues in T homopolymers, ID1 and ID2, are indels for which an association with MMR deficiency is well-known, but this is not the case for cytosine homopolymer alterations. Whereas the indel signature IDA is not yet reported in COSMIC, the predominance of InsC5+ resembles signature MS-sig4, which was previously reported by Chung et al., and found to be associated with combined MMR- and polymerase proofreading deficiency. Based on their findings, these authors proposed a model where microsatellite indel loops on the nascent strand are repaired by DNA polymerase proofreading, with polymerase proofreading deficiency resulting in the accumulation of insertions[23]. We observed a reversed correlation for IDA, with higher contributions in tumors without proofreading deficiency. This could, at least in part, be due to contamination of IDA with other signatures, which could not be individually extracted with the number of tumors included in our cohort. When focusing only on InsC5+, the key feature of IDA and MS-Sig4, our data show that absolute contributions of InsC5+ in polymerase proofreading-deficient tumors are higher than in polymerase proofreading-proficient tumors. Since relative contributions were not different, the ultra-high TML in polymerase proofreading-deficient tumors must underly this enrichment. In addition, we found that relative contributions of InsC5+ were significantly higher in PMS2-deficient tumors compared to the other MMR-deficient tumors, similar to what was observed for ID1. Together, these findings suggests that the pattern of one-bp insertions in homopolymers of >5 bp in polymerase proofreading-deficient and PMS2-deficient tumors is not limited to thymines, but also includes cytosines, and that the InsC5+ pattern should be considered as an MMR-associated signature.

We found large contributions of temozolomide-associated signature SBS11 in second primary tumors from individuals with CMMRD in our cohort. Temozolomide induces a specific type of DNA damage that starts a futile cycle of MMR repair, ending in apoptosis[17]. In MMR-deficient tumor cells, the specific pattern of DNA damage remains. Contributions of signature SBS11 have mostly been reported in relapses or metastases of glioblastomas and melanomas that acquired MMR deficiency while treated with temozolomide[13,18,33]. We did not detect contributions of SBS11 in the relapsed or metastasized brain tumors of individuals treated with temozolomide in the CMMRD cohort. Instead, contributions of SBS11 were detected in two second primary hematological malignancies, which may suggest that temozolomide predominantly affects the hematological system. The finding of SBS11 in second primary tumors raises the question whether temozolomide induced new tumor development, or contributed to the genomic profile of a tumor that was already developing. The analysis of the mutational context of driver gene mutations in the two tumors with contributions of SBS11 in this study was inconclusive and should be repeated on a larger number of retrospective cases to determine whether there is a causal relationship between temozolomide treatment and new tumor development. Because of the presumed risk of resistance in MMR-deficient high-grade gliomas, the use of temozolomide is no longer recommended in individuals with CMMRD[34–37]. Our findings suggest that it is important to exclude CMMRD when treatment with this agent is considered.

Although mutational signatures elucidate ongoing processes in CMMRD tumors, not all associations are completely understood, and the mutational signature contributions identified in some tumors in our cohort could not be fully explained. One of the signatures involved is SBS26, which, based on this study as well as previous work, is strongly associated with PMS2 deficiency[11,18], particularly in the absence of other mutational processes induced by, for example, polymerase proofreading deficiency. In our cohort, we observed contributions of SBS26 in nearly all PMS2-deficient hematological malignancies. However, we also encountered a PMS2-deficient T-LBL lacking SBS26-associated mutations as well as an MSH6-deficient T-LBL with SBS26 contribution, suggesting that other factors can interfere with the etiology of T > C mutations. Of note, the occurrence of SBS26-associated T > C mutations has also been observed in a germinal center sample of an individual with MSH6-deficient CMMRD[5], suggesting that MSH6 deficiency can indeed induce T > C mutations under specific circumstances. These observations are illustrative of the heterogeneity in mutational patterns of CMMRD tumors, which may be influenced by different factors, including tissue specificity. Another example relates to tumor C04-BNP01, in which ultrahypermutation combined with a strong contribution of signature SBS20 led us to suspect a POLD1 proofreading deficiency. However, thorough inspections of whole genome sequencing data did not reveal any mutations that were likely to cause such a deficiency. In RNA sequencing data, we detected transcripts with a c.1504 G > T variant, leading to p.Asp502Tyr in the exonuclease domain, which may have caused polymerase proofreading deficiency. The lack of unambiguous confirmation in genomic DNA may suggest an allele-specific mapping artefact caused by a flanking structural variant, but such a variant was not found. Together, these findings illustrate the importance of mutational signature analysis in the detection of replication repair deficiency in tumors of individuals with CMMRD, while also reflecting gaps in our understanding of how to interpret mismatch repair deficiency signatures.

In conclusion, this study of mutational patterns in sequential tumors in individuals with CMMRD revealed a complex interplay of mutational processes associated with the affected MMR gene, tumor type and prior treatment. We encountered a rare case of a CMMRD T-LBL with somatic polymerase proofreading deficiency, identified indel signature IDA, which contributed to large numbers of cytosine insertions in 54% of tumors in our cohort, and we found temozolomide-induced footprints in two second primary hematological malignancies. With our study, we show that studying a diverse CMMRD cohort, reflecting CMMRD-associated tumor types and

including primary and (second) primary tumors, with multiple sequential tumors in the same individual, can provide insights in the diverse range of mutational processes associated with mismatch repair deficiency. In future studies using large tumor type-specific cohorts, it would therefore be informative to combine a driver mutation analysis with the detection of mutational processes. Such studies may lead to a better understanding of the impact of current therapies and the discovery of new treatment options, ultimately improving quality of life and overall survival of individuals with CMMRD.

## Methods

### Ethics statement

This study was conducted in accordance with Good Clinical Practice guidelines and the Declaration of Helsinki. Patient data was de-identified. For all data originating from the Princess Máxima Center, informed consent has been obtained for all individuals involved in this study through the Máxima biobank [https://research.prinsesmaximacentrum.nl/en/core-facilities/scientific-committee] informed consent procedure and corresponding protocol. The Máxima biobank protocol has been approved by the Medical Ethics Committee of the Erasmus Medical Center in Rotterdam, The Netherlands, under reference number MEC-2016-739. The use of the individual's data within the context of this study has been approved by the Máxima biobank and data access committee, biobank request nr. PMCLAB2018-022 and PMCLAB2022-343. For all data originating from the Radboud university medical center in Nijmegen (Radboudumc), permission was granted by the Central Committee Involving Human Subjects (CCMO-2020-6330). Individuals were not compensated for their participation in the study. We do not report on sex or gender, because we specifically study somatic (tumor) mutations and mutational signatures and prior studies have not shown that sex (or gender) is relevant for these analyses.

### Collection of CMMRD samples and controls

We included 17 individuals with CMMRD diagnosed in the Netherlands between 2010 and 2020 for whom tumor and/or normal samples were available in our cohort. Thirty-one tumor samples and 11 normal samples were collected from the Radboudumc. Material and data of an additional six tumors and three normal samples were collected through the Princess Máxima Center biobanking initiative. For four tumors, material was collected from the Dutch Nationwide Pathology Databank (PALGA)[38] and two additional normal samples were obtained through the Dutch Childhood Oncology Group (DCOG) biobank (Table S3). A control dataset containing whole genome sequencing (WGS) data of 1006 non-CMMRD pediatric tumors from the Princess Máxima Center biobank was used to facilitate mutational signature extraction. After filtering, 658 and 440 tumors remained for SBS and indel mutational signature extraction, respectively. This dataset, control cohort 1, is a representative reflection of the different tumor types that children develop (Table S9). A second control dataset, control cohort 2, containing sequencing data of 1474 adult cancers and 80 pediatric cancers was obtained from The Cancer Genome Atlas (TCGA) PanCancer project [www.cancer.gov/tcga], a study by St Jude Children's Research Hospital and a study by the Hospital for Sick Kids and downloaded from cBioPortal[39–42]. Germline MMR status was investigated with restricted access through request number 106777-1. Replication repair-deficient tumors and replication repair-proficient tumors with a tumor type fitting the CMMRD spectrum were included (Table S10).

### DNA sequencing and mutation calling

For 41 tumors and 16 normal samples from individuals with CMMRD, DNA was isolated and whole exome sequencing was performed on formalin-fixed, paraffin-embedded material, while WGS was performed on fresh or fresh frozen material. Details on library preparation, enrichment and alignment are provided in Table S3. Preprocessing of all sequencing data was performed using the GATK4 preprocessing for variant discovery workflow, including mapping of raw reads to human reference genome hg38 with the Burrows-Wheeler aligner (v0.7.13)[43]. Somatic mutation calling was performed with Mutect2 (GATK v4.2.0.0)[43], using a bed file with overlapping regions of exome captures KAPA MedExome, KAPA HyperExome (Roche, Basel, Switzerland), Agilent SureSelect version 5 and 7 (Agilent Technologies, Santa Clara, California, United States), extended by 200 bp flanking each region, as described previously[22]. For the samples captured with Twist v1 (Twist Bioscience HQ, South San Francisco), an adapted bed file was used for somatic calling, containing the overlapping regions between the four other captures and Twist v1, extended by 200 bp flanking each region. For individual C04, no normal sample was available. Therefore, one of the unrelated tumor samples from this individual was used as normal in the tumor-normal pairs (Table S3).

### Data filtering

Data filtering strategies were optimized for different datasets used and, therefore, slightly different. Filtering was performed using R (v4.2.1) and packages VariantAnnotation (v1.50.0) and GenomeInfoDb (1.40.1). The sequencing data of the 41 CMMRD tumors were derived from different sources and had different qualities. To reduce noise, but also retrieve as many reliable mutations as possible to enable comparison of mutational signatures in these samples, somatic mutations were filtered for a variant allele frequency (VAF) of ≥0.1, ≥5 alternative reads, a VAF of 0 in the normal sample and a population frequency of ≤0.01 in both the GnomAD and GoNL database[44,45].

The mutations in the non-CMMRD pediatric tumors in control cohort 1 were of high-quality and therefore, strict filter settings were applied. Somatic mutations were filtered to have a VAF of ≥0.15, ≥5 alternative reads, a VAF of 0 in the normal sample and a population frequency of ≤0.01 in both the GnomAD and GoNL database[44,45]. Additionally, all included mutations had ≥20 total reads and were located outside centromeric regions. Tumors with <200 SBS mutations were excluded from mutational signature extraction, as this is the minimum number required for a representative mutational profile[46]. For indels, a minimum of 150 mutations was used as a threshold. The somatic mutations in the adult and pediatric tumors of control cohort 2 were downloaded from cBioPortal and not filtered.

### Mutation load and mutational signature analysis

TML was calculated for SBSs, multinucleotide substitutions (MNS) and indels, as well as for the total number of mutations (Table S5), by dividing the number of somatic mutations left after filtering by the size of the bed file that was used for somatic calling. The bed files had a size of 97.2 Mb for tumors captured with Twist v1 and 101.6 Mb for all other tumors (Table S3). To determine the mutational patterns and signature contributions in tumors of the CMMRD cohort, R (v.4.2.1 and v4.1.2) and the R packages MutationalPatterns (version 3.10.0) and BSgenome.Hsapiens.UCSC.hg38 (v1.4.5) were used[47]. Control cohort 1 was used to power the signature extraction. First, an SBS and an indel mutational matrix was constructed containing the mutation type and sequencing context of all mutations in the tumors of the CMMRD- and control cohorts. When both the primary tumor and the corresponding metastasis were in the dataset, only newly acquired mutations in the metastasis were included. For the de novo extraction[48], the function extract_signatures was used, with options rank=12 (SBS) or rank=6 (indels) and nrun=200. To reduce noise and overfitting, a strict refit approach was performed between the CMMRD tumor mutations and the extracted signatures using the function fit_res_strict with max delta=0.033. For the refit, the de novo extracted signatures were replaced by the original COSMIC signatures (v3.2)[14] if the cosine similarity between the two was >0.85 (Fig. S2a, b and Fig. S3a, b). A de novo extracted signature could also be replaced by two or three COSMIC signatures if these could reconstruct the de novo extracted signature with a cosine similarity of >0.8 (Fig. S2c, d and S3c, d). The

extracted signatures that could not be replaced by COSMIC signatures were used as de novo signatures. The refit was performed with a total of 14 signatures for SBS and six signatures for indels. The cosine similarity between the mutational profile and the reconstruction was >0.8 for all CMMRD tumors.

The de novo signature extraction was validated using python3 and SigProfiler in combination with control cohort 2. SigProfilerMatrixGenerator (v1.1.28) was used to generate the mutational matrices for all tumors in the CMMRD and control cohort, and SigProfilerExtractor (v1.1.0) was used with default settings for de novo mutational signature analysis[49,50]. The contribution of each de novo signature and, when applicable, each COSMIC signature contributing to the de novo signatures was determined for all tumors of the CMMRD cohort and control cohort 2. This analysis was performed separately for SBS and indels.

Because a treatment signature, SBS11, was detected in a treatment-naïve tumor, an additional refit with bootstrapping was performed using R (v.4.2.1) and the R package MutationalPatterns (version 3.10.0)[47]. For this, the function fit_to_signatures_bootstrapped with n_boots=200, max_delta=0.033 and method = "strict" was used. We compared the SBS profile of all tumors with contributions of SBS11 by performing a bootstrapped refit with all original COSMIC signatures (v3.2)[14], all extracted signatures (Fig. S2) and all extracted signatures supplemented with MMR deficiency-associated signatures that did not result from the extraction (SBS6, SBS21, SBS44). The results were plotted using the function plot_bootstrapped_contribution, with mode = "relative" and plot_type = "jitter".

Finally, a strict refit using the function fit_res_strict, was performed as described above, with all tumors with contributions of SBS11 and the set of extracted SBS mutational signatures (Fig. S2) including and excluding SBS11. Cosine similarities were calculated to compare between the original and refitted mutational profile using the function plot_original_vs_reconstructed.

### RNA sequencing

For sample C04-BNP01, RNA was isolated using the AllPrep DNA/RNA/miRNA universal kit (Cat. No. 80224, Qiagen, Hilden, Germany) according to the manufacturer's protocol. A total of 160 ng of RNA was used for sequencing[51], for which libraries were prepared using KAPA RNA HyperPrep Kit with RiboErase (HMR) (Cat. No. 08098140702, Roche, Basel, Switzerland) according to manufacturer's protocol. Fragmentation and priming were performed at 94 degrees for 6.5 min and for the adapter ligation, a 7 µM stock of NextFlex DNA barcodes was used (Bioo Scientific, Austin, United States). Six cycli were used for library amplification and purified using a using a 0.8x bead-based cleanup. The High Sensitivity DNA bioanalyzer (Agilent Technologies, Santa Clara, California, United States) was used to determine the library size and the DeNovix dsDNA High Sensitivity Assay (Cat. No. K-30100, DeNovix, Wilmington, United States) was used to measure the library concentration. Libraries were sequenced using an Illumina NextSeq 500 (Illumina, San Diego, United States), resulting in 42-bp paired-end reads. For non-malignant samples of individuals C20 and C09, RNA was isolated using the Qiagen RNeasy mini kit (Cat. No. 74104, Qiagen, Hilden, Germany), according to the manufacturer's protocol. The KAPA RNA HyperPrep Kit with RiboErase (Cat. No. 08098140702, Roche, Basel, Switzerland) was used to prepare the libraries according to the manufacturer's protocol, with 300 ng of RNA as input. Libraries were sequenced using the NovaSeq6000 according to manufacturer's protocol (Illumina, San Diego, United States), resulting in paired-end 2 × 150 bp reads. Preprocessing of all sequencing data was performed using the GATK RNA-seq short variant discovery pipeline, which makes use of the Genome Analysis Toolkit (v4.1.7.0)[43], reads were aligned to a STAR (v.2.7.2 d) indexed genome generated from the human reference genome hg38[52]. Visualizations and sashimi plots of the target regions were produced using the Integrative Genomics Viewer (v2.8.2)[53,54].

### POLD1 PCR validation

To validate a *POLD1* mutation discovered in RNA sequencing data of tumor C04-BNP01, RNA was isolated from new tissue sections using the All Prep DNA/RNA/miRNA universal kit (Cat. No. 80224, Qiagen, Hilden, Germany) according to the manufacturer's protocol. RNA of C04-BNP01 was first converted to cDNA using the High-Capacity cDNA Reverse Transcription Kit (Cat. No. 4368814, ThermoFisher Scientific, Waltham, United States) according to the manufacturer's protocol. A PCR was performed on the cDNA using the PrimeSTAR GXL kit (Cat. No. R050A, Takara, Kusatsu, Japan) according to manufacturer's protocol, using forward primer CAACCCTACCTCCATCCCCA and reverse primer AGGTAGGCATCCTTCAGGCA with 10 s denaturation at 98 degrees, 40 cycles of 10 s denaturation at 98 degrees, 15 s primer annealing at 60 degrees and 30 s annealing at 68 degrees, and 2 min of final extension at 68 degrees. The PCR product was analyzed on a 2% agarose gel at 130 V (7 V/cm) for 75 min. The PCR products were purified using the ExoSAP-IT (Cat. No. 78200.200.UL, ThermoFisher Scientific, Waltham, United States) according to manufacturer's protocol and sent for Sanger sequencing (Macrogen, Amsterdam, the Netherlands). Results were analysed using Geneious Prime (Geneious Solutions, Auckland, New Zealand).

### Targeted copy number analysis and sequencing of *PMS2*

*PMS2* was screened for alterations in tumor C12-HNP02 using targeted smMIP-based next generation sequencing. Probes were designed to sequence both *PMS2* and *PMS2CL*. Library preparation was performed as previously described[55]. Briefly, 100 ng of DNA was used in a reaction including the phosphorylation smMIP pool, 1 unit of Ampligase DNA ligase (Cat. No. A0110K, EpiBio, Madison, United States) with Ampligase Buffer (Cat. No. A1905B, EpiBio, Madison, United States), 3.2 units of Hemo Klentaq (Cat. No. M0332, New England Biolabs, Ipswich, United States), and 8 µmol of dNTPs (Cat. No. 28-4065-20/−12/−22/−32, GE Healthcare, Little Chalfont, UK). This reaction was denatured for 10 min at 95 degrees, incubated at 60 degrees for 18 h for probe hybridization, extension and ligation, and subsequently cooled. Exonuclease I (Cat. No. M0293, New England Biolabs, Ipswich, United States) and III (Cat. No. M0206, New England Biolabs, Ipswich, United States) and Ampligase Buffer were added, and the reaction was incubated at 37 degrees for 45 min and inactivated at 95 degrees for 2 min. 10 ul of the capture was used for PCR, using 25 nmol regular forward primers and barcoded reverse primers and iProof high-fidelity master mix (Cat. No. 1725310, Bio-Rad, Hercules, United States). PCR products were pooled and purified using 0.8× volume of Agencourt Ampure XP Beads (Cat. No. A63881, Beckman Coulter, Brea, United States). After denaturation, the purified libraries were diluted to a concentration of 1.2 pmol/L. A 300 cycles Mid Output sequencing kit, v2 (Illumina, San Diego, United States) was used to sequence the samples on a NextSeq500 (Illumina, San Diego, United States) according to the manufacturer's protocol, resulting in 2 × 150 bp paired-end reads. Sequencing reads were aligned to reference genome hg19. Variants were called and sequencing results were analyzed using Sequence Pilot software (JSI Medical Systems, Ettenheim, Germany). MLPA was performed to detect any structural alterations, using SALSA® MLPA® Probemix P008 PMS2 (Cat. No. P008, MRC Holland, Amsterdam, the Netherlands) according to the manufacturer's protocol.

### MSH2 mutant cell-line

AHH-1 MSH2$^{-/-}$ (MSH2KO) cell lines were generated using CRISPR/Cas9 gene editing[56]. Human AHH-1 cell lines (CRL-8146, ATCC, Manassas, United States) were cultured in RPMI 1640 GlutaMAX medium (Cat. No. 61870036, Thermo Fisher Scientific, Waltham, United States) supplemented with 1% Penicillin-Streptomycin (Cat. No. 15140122, Thermo Fisher Scientific, Waltham, United States) and 10% horse serum (Thermo Fisher Scientific, Waltham, United States). The *MSH2* guide RNA, 5′-GTGCCTTTCAACAACCGGTTG-3′, was cloned into a

pSpCas9(BB)−2A-GFP (PX458) vector (plasmid #48138, Addgene, Watertown, United States). AHH-1 cells were transfected using Lipofectamine 2000 (Cat. No. 11668027, Thermo Fisher Scientific, Waltham, United States). GFP-positive transfected cells were single-cell sorted for clonal expansion on a SH800S Cell Sorter (Sony, Tokyo, Japan) one or two days after transfection. MSH2 inactivation was confirmed using Sanger sequencing, WGS and Western blot. For the latter, the following antibodies were used: rabbit anti-MSH2 (Cat. No. D24B5, 1:2000, Cell Signaling Technology, Danvers, United States) and mouse anti-α-Tubulin (Cat. No. T5168, 1:5000, Sigma-Aldrich, St Louis, United States). As secondary antibodies, anti-rabbit IgG IRDye 800CW (Cat. No. 926-32211, 1:10000, Li-Cor, Lincoln, United States) and anti-mouse IgG IRDye 680RD (Cat. No. 926-68070, 1:10000, Li-Cor, Lincoln, United States) were used. Western blots were imaged on an Odyssey DLx imaging system (Li-Cor, Lincoln, United States).

To create the mutant cell-line, NEBuilder® HiFi DNA Assembly (New England Biolabs, Ipswich, United States) was used to insert a 1 bp duplication in *MSH2* cDNA at position 2727 and to make hPGK-3xFLAG-MSH2c.2727dup-UbC-mNeonGreen and hPGK-MSH2-Clover constructs, which were inserted in a pLL3.7 m lentiviral vector. Lentivirus was produced by transfecting HEK293T cells with the lentiviral transfer plasmids, pMD2.G envelope plasmid and pCMV-dR8.91 packaging plasmid using PEI (Polysciences, Warrington, United States). To re-express $MSH2^{WT}$ or $MSH2^{c.2727dup}$, AHH1-MSH2-/- were transduced with pLL3.7m-hPGK-MSH2-Clover or pLL3.7m-hPGK-3xFLAG-MSH2c.2727 dup-UbC-mNeonGreen lentivirus, respectively. Medium was refreshed one day after transfection and lentivirus-containing medium was harvested 4 days post-transfection. For each lentivirus, two million AHH1-MSH2-/- cells were spin transduced (1000 g at 32 °C for 2 h) with 500 µl 0.45 µm filtered lentivirus supernatant. Single clover- or mNeonGreen-positive cells were flow sorted for clonal expansion five days after transduction on a SH800S Cell Sorter (Sony, Tokyo, Japan). Protein expression of $MSH2^{WT}$ and $MSH2^{c.2727dup}$ after transduction was confirmed using western blot as described above (Fig. S11).

After validating MSH2 knockout or re-expression in the clonal cell lines, a second single cell clonal expansion step was performed after a variable number of days to enable calculations of the mutation rates (Table S11). Cells were harvested for DNA extraction after several days, when the (sub)clonal lines were sufficiently expanded. DNA was isolated using the DNeasy Blood & Tissue Kit (Cat. No. 69504, Qiagen, Hilden, Germany). WGS libraries were generated using the same workflow as for WGS of CMMRD tumors and normal samples (Table S3). WGS libraries were sequenced to ~15x coverage (2 x 150 bp) on an Illumina NovaSeq 6000 at the Hartwig Medical Foundation (Amsterdam, the Netherlands). The WGS data for the AHH-1 cell lines was processed using the Nextflow Illumina Analysis Pipeline [https://github.com/ToolsVanBox/NF-IAP]. The inhouse developed Somatic Mutations Rechecker and Filtering (SMuRF) tool (v3.0.0) [https://github.com/ToolsVanBox/SMuRF] was used to filter somatic variants. For the wildtype, $MSH2^{-/-}$ (MSH2KO), MSH2KO + $MSH2^{WT}$ and MSH2KO + $MSH2^{c.2727dup}$ cell lines, only clonal variants were selected to avoid inclusion of mutations that occurred during clonal expansion[56,57]. Therefore, all somatic variants that had a VAF of less than 0.25, a base coverage of less than 5 reads, a mapping quality of less than 55, a GATK phred-scaled quality score <100 and/or presence in a panel of unmatched normal human genomes were excluded.

### Statistical tests

Pairwise differences were assessed using a two-sided Mann-Whitney U test. Comparisons between three groups were made by first performing a Kruskal-Wallis test, followed by a Dunn's test with multiple comparisons adjustment according to the Bonferroni method. Fisher-Freeman-Halton asymptotic tests and two-sided Fisher's exact tests were used to analyze contingency tables. Statistical analyses were performed using GraphPad Prism (v9.5.0, San Diego, United States) or R (v4.2.1) with packages stats (v4.2.1), FSA (v0.9.5) and contingencytables (v3.0.1)[58–60]. Values of $P < 0.05$ were considered to be significantly different.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The raw sequencing data generated in this study have been deposited in the EGA database under accession codes EGAS00001007660 (containing whole exome, whole genome, and transcriptome sequencing data generated in the Princess Máxima Center), EGAS50000000081 (Whole exome sequencing data from Radboud UMC) and EGAC00001003259 (control cohort 1). These datasets are available under restricted access. Control cohort 2 contains data from TCGA (Study accession number phs000178; https://portal.gdc.cancer.gov), St Jude Children's Research Hospital https://www.cbioportal.org/study/summary?id=all_stjude_2015, and the Hospital for Sick Kids https://www.cbioportal.org/study/summary?id=mbl_sickkids_2016. Details on sequencing methods are available in Table S3 and in the Methods section. A conversion key for the sample IDs, including sample origins and sample types, is available through GitHub [https://github.com/kuiper-lab/Tumor-mutational-patterns-in-CMMRD][61]. The sequencing data are available under restricted access, conditions for access are specified for each dataset in EGA. Access can be obtained by sending an application to the relevant Data Access Committee, contact details are available through EGA. The processed somatic mutation data are available as VCF files through GitHub [https://github.com/kuiper-lab/Tumor-mutational-patterns-in-CMMRD][61]. The mutational matrices of the CMMRD cohort, control cohort 1 and control cohort 2 are available through GitHub [https://github.com/kuiper-lab/Tumor-mutational-patterns-in-CMMRD][61]. Source data are provided with this paper.

## Code availability

Only publicly available tools were used for the analysis in this study. The relevant tools and parameters have been described in the Methods section, and the code can be accessed through GitHub [https://github.com/kuiper-lab/Tumor-mutational-patterns-in-CMMRD][61] under General Public License (GPL). The code of the Nextflow Illumina Analysis Pipeline [https://github.com/ToolsVanBox/NF-IAP] and the in house developed Somatic Mutations Rechecker and Filtering (SMuRF) tool (v3.0.0) [https://github.com/ToolsVanBox/SMuRF] are available under MIT license under their respective repositories.

## References

1. Kunkel, T. A. & Erie, D. A. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev. Genet*. **49**, 291–313 (2015).
2. Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D. & Hitchins, M. P. Milestones of Lynch syndrome: 1895-2015. *Nat. Rev. Cancer* **15**, 181–194 (2015).
3. Wimmer, K. et al. Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'Care for CMMRD' (C4CMMRD). *J. Med. Genet*. **51**, 355 (2014).
4. Ercan, A. B. et al. Clinical and biological landscape of constitutional mismatch-repair deficiency syndrome: an International Replication Repair Deficiency Consortium cohort study. *Lancet Oncol.* **25**, 668–682 (2024).
5. Sanders, M. A., et al. Life without mismatch repair. *bioRxiv*, 2021.2004.2014.437578, https://doi.org/10.1101/2021.04.14.437578 (2021).
6. Durno, C. et al. Survival Benefit for Individuals With Constitutional Mismatch Repair Deficiency Undergoing Surveillance. *J. Clin. Oncol*. **39**, 2779–2790 (2021).

7. Ghorbanoghli, Z. et al. High yield of surveillance in patients diagnosed with constitutional mismatch repair deficiency. *J. Med. Genet.* **60**, 679–684 (2023).
8. Fedier, A. & Fink, D. Mutations in DNA mismatch repair genes: implications for DNA damage signaling and drug sensitivity (review). *Int J. Oncol.* **24**, 1039–1047 (2004).
9. Meier, B. et al. Mutational signatures of DNA mismatch repair deficiency in C. elegans and human cancers. *Genome Res.* **28**, 666–675 (2018).
10. Fang, H. et al. Deficiency of replication-independent DNA mismatch repair drives a 5-methylcytosine deamination mutational signature in cancer. *Sci. Adv.* **7**, eabg4398 (2021).
11. Zou, X. et al. A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. Cancer* **2**, 643–657 (2021).
12. Campbell, B. B. et al. Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* **171**, 1042–1056.e1010 (2017).
13. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
14. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
15. Georgeson, P. et al. Evaluating the utility of tumour mutational signatures for identifying hereditary colorectal cancer and polyposis syndrome carriers. *Gut* **70**, 2138–2149 (2021).
16. Shlien, A. et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat. Genet.* **47**, 257–262 (2015).
17. Daniel, P. et al. Temozolomide Induced Hypermutation in Glioma: Evolutionary Mechanisms and Therapeutic Opportunities. *Front Oncol.* **9**, 41 (2019).
18. Degasperi, A., et al. Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science* **376**, https://doi.org/10.1126/science.abl9283 (2022).
19. Yang, F. et al. Chemotherapy and mismatch repair deficiency cooperate to fuel TP53 mutagenesis and ALL relapse. *Nat. Cancer* **2**, 819–834 (2021).
20. Thompson, B. A., Martins, A. & Spurdle, A. B. A review of mismatch repair gene transcripts: issues for interpretation of mRNA splicing assays. *Clin. Genet.* **87**, 100–108 (2015).
21. Gallon, R. et al. Constitutional mismatch repair deficiency mimicking Lynch syndrome is associated with hypomorphic mismatch repair gene variants. *NPJ Precis Oncol.* **8**, 119 (2024).
22. Weijers, D. D. et al. Molecular analysis of cancer genomes in children with Lynch syndrome: Exploring causal associations. *Int J. Cancer* **154**, 1455–1463 (2024).
23. Chung, J. et al. DNA Polymerase and Mismatch Repair Exert Distinct Microsatellite Instability Signatures in Normal and Malignant Human Cells. *Cancer Discov.* **11**, 1176–1191 (2021).
24. Ostroverkhova, D. et al. DNA polymerase epsilon and delta variants drive mutagenesis in polypurine tracts in human tumors. *Cell Rep.* **43**, 113655 (2024).
25. Haradhvala, N. J. et al. Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. *Nat. Commun.* **9**, 1746 (2018).
26. Herr, A. J. et al. Mutator suppression and escape from replication error-induced extinction in yeast. *PLoS Genet.* **7**, e1002282 (2011).
27. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
28. Wei, C. H., et al POLD1 DEDD Motif Mutation Confers Hypermutation in Endometrial Cancer and Durable Response to Pembrolizumab. *Cancers (Basel)* **15**, https://doi.org/10.3390/cancers15235674 (2023).
29. Jin, Y. H. et al. The 3′->5′ exonuclease of DNA polymerase delta can substitute for the 5′ flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability. *Proc. Natl Acad. Sci. USA* **98**, 5122–5127 (2001).
30. Jin, Y. H. et al. The multiple biological roles of the 3′->5′ exonuclease of Saccharomyces cerevisiae DNA polymerase delta requir switching between the polymerase and exonuclease domains. *Mol. Cell Biol.* **25**, 461–471 (2005).
31. Shinbrot, E. et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res.* **24**, 1740–1750 (2014).
32. Das, A. et al. Genomic predictors of response to PD-1 inhibition in children with germline DNA replication repair deficiency. *Nat. Med.* **28**, 125–135 (2022).
33. Mas-Ponte, D., McCullough, M. & Supek, F. Spectrum of DNA mismatch repair failures viewed through the lens of cancer genomics and implications for therapy. *Clin. Sci. (Lond.)* **136**, 383–404 (2022).
34. Pollack, I. F. et al. Mismatch repair deficiency is an uncommon mechanism of alkylator resistance in pediatric malignant gliomas: a report from the Children's Oncology Group. *Pediatr. Blood Cancer* **55**, 1066–1071 (2010).
35. Westdorp, H. et al. Immunotherapy holds the key to cancer treatment and prevention in constitutional mismatch repair deficiency (CMMRD) syndrome. *Cancer Lett.* **403**, 159–164 (2017).
36. Touat, M. et al. Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* **580**, 517–523 (2020).
37. Abidi, A., et al. Challenges of Neoantigen Targeting in Lynch Syndrome and Constitutional Mismatch Repair Deficiency Syndrome. *Cancers* **13**, https://doi.org/10.3390/cancers13102345 (2021).
38. Casparie, M. et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol.* **29**, 19–24 (2007).
39. Andersson, A. K. et al. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat. Genet.* **47**, 330–337 (2015).
40. Morrissy, A. S. et al. Divergent clonal selection dominates medulloblastoma at recurrence. *Nature* **529**, 351–357 (2016).
41. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
42. Gao, J., et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* **6**, pl1, https://doi.org/10.1126/scisignal.2004088 (2013).
43. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. 1st edn, (O'Reilly, 2020).
44. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
45. Boomsma, D. I. et al. The Genome of the Netherlands: design, and project goals. *Eur. J. Hum. Genet.: EJHG* **22**, 221–227 (2014).
46. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).
47. Manders, F. et al. MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genom.* **23**, 134 (2022).
48. Bertrums, E. J. M. et al. Elevated Mutational Age in Blood of Children Treated for Cancer Contributes to Therapy-Related Myeloid Neoplasms. *Cancer Discov.* **12**, 1860–1872 (2022).
49. Bergstrom, E. N. et al. SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC Genom.* **20**, 685 (2019).
50. Islam, S. M. A. et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genom.* **2**, None, https://doi.org/10.1016/j.xgen.2022.100179 (2022).
51. Wester, R. A. et al. Retinoic acid signaling drives differentiation toward the absorptive lineage in colorectal cancer. *iScience* **24**, 103444 (2021).

52. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
53. Robinson, J. T., Thorvaldsdottir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer Res.* **77**, e31–e34 (2017).
54. Katz, Y. et al. Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics* **31**, 2400–2402 (2015).
55. Eijkelenboom, A. et al. Reliable Next-Generation Sequencing of Formalin-Fixed, Paraffin-Embedded Tissue Using Single Molecule Tags. *J. Mol. Diagn.* **18**, 851–863 (2016).
56. Middelkamp, S. et al. Comprehensive single-cell genome analysis at nucleotide resolution using the PTA Analysis Toolbox. *Cell Genom.* **3**, 100389 (2023).
57. Blokzijl, F. et al. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260–264 (2016).
58. R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2022).
59. FSA: Simple Fisheries Stock Assessment Methods (2023).
60. Fagerland, M. W. *contingencytables: Statistical Analysis of Contingency Tables*, https://CRAN.R-project.org/package=contingencytables (2024).
61. Weijers, D. D. et al. Kuiper RP Unraveling mutagenic processes influencing the tumor mutational patterns of individuals with Constitutional Mismatch Repair Deficiency. *GitHub*, https://doi.org/10.5281/zenodo.14938856 (2025).
62. Jongmans, M. C. et al. Constitutional mismatch repair deficiency syndrome. *Ned. Tijdschr. Geneeskd.* **159**, A8602 (2015).
63. van der Klift, H. M. et al. Comprehensive Mutation Analysis of PMS2 in a Large Cohort of Probands Suspected of Lynch Syndrome or Constitutional Mismatch Repair Deficiency Syndrome. *Hum. Mutat.* **37**, 1162–1179 (2016).
64. Vos, J. R. et al. Evaluation of yield and experiences of age-related molecular investigation for heritable and nonheritable causes of mismatch repair deficient colorectal cancer to identify Lynch syndrome. *Int J. Cancer* **147**, 2150–2158 (2020).
65. Kroeze, E. et al. High Prevalence of Constitutional Mismatch Repair Deficiency in a Pediatric T-cell Lymphoblastic Lymphoma Cohort. *Hemasphere* **6**, e668 (2022).
66. Suerink, M. et al. Constitutional mismatch repair deficiency in a healthy child: On the spot diagnosis? *Clin. Genet.* **93**, 134–137 (2018).
67. Langenberg, K. P. S. et al. Implementation of paediatric precision oncology into clinical practice: The Individualized Therapies for Children with cancer program 'iTHER. *Eur. J. Cancer (Oxf., Engl.: 1990)* **175**, 311–325 (2022).


## Acknowledgements

This work was supported by the Dutch Cancer Society (Grant KWF-12090, received by R.P.K. and R.vB. and KWF-12174, received by R.M.V.) and Stichting Kinderen Kankervrij (Foundation KiKa, project number 359, received by M.C.J.J., G.S., N.H. and I.J.M.V). We thank Dr. Sabine Plaschaert for her help in collecting treatment information, prof. dr. Ronald de Krijger for his assistance in obtaining additional tumor samples and Freerk van Dijk for performing additional bioinformatic analyses on our data.

## Author contributions

Conception and design of the work: D.D.W., S.H., R.M.V., M.C.J.J., R.P.K. Acquisition of the data: D.D.W., S.H., E.K., M.A.J.G., R.B., K.V., G.S., N.H., S.M., I.J.M.V., R.vB., R.M.V., M.C.J.J., R.P.K. Acquisition of clinical information: P.W., R.S.P., J.L.C.L., C.E.M.G., M.C.A.K., M.C.J.J. Analysis and interpretation of the data: D.D.W., S.H., S.M., A.R.M. R.vB., R.M.V., M.C.J.J., R.P.K. Writing the original draft: D.D.W., S.H., M.C.J.J., R.P.K. Revision of the original draft: D.D.W., S.H., R.M.V., M.C.J.J., R.P.K. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-59775-2.

**Correspondence** and requests for materials should be addressed to Roland P. Kuiper.

**Peer review information** *Nature Communications* thanks Katharina Wimmer who co-reviewed with Esther SchamschulaRichard Gallon and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



[1]Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands. [2]Department of Human Genetics, Radboud university medical center, Nijmegen, The Netherlands. [3]Department of Medical BioSciences, Radboud university medical center, Nijmegen, The Netherlands. [4]Division of Immunotherapy, Oncode Institute, Radboud university medical center, Nijmegen, The Netherlands. [5]Oncode Institute, Utrecht, The Netherlands. [6]Department of Pathology, Radboud university medical center, Nijmegen, The Netherlands. [7]Department of Pathology, Amsterdam University Medical Centers/VUmc, Amsterdam, The Netherlands. [8]Department of Gastroenterology and Hepatology, Radboud university medical center, Nijmegen, The Netherlands. [9]Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. [10]These authors contributed equally: Snežana Hinić, Emma Kroeze. ✉e-mail: r.kuiper@prinsesmaximacentrum.nl