nature genetics

Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers

Adam Shlien[1–3], Brittany B Campbell[1,4,5,31], Richard de Borja[1,31], Ludmil B Alexandrov[6], Daniele Merico[1,7], David Wedge[6], Peter Van Loo[6,8], Patrick S Tarpey[6], Paul Coupland[9], Sam Behjati[6], Aaron Pollett[10], Tatiana Lipman[1,4], Abolfazl Heidari[1,4], Shriya Deshmukh[1,4], Na'ama Avitzur[1,4], Bettina Meier[11], Moritz Gerstung[6], Ye Hong[11], Diana M Merino[1], Manasa Ramakrishna[6], Marc Remke[4], Roland Arnold[1], Gagan B Panigrahi[1], Neha P Thakkar[1,12], Karl P Hodel[13], Erin E Henninger[13], A Yasemin Göksenin[13], Doua Bakry[14,15], George S Charames[3,10], Harriet Druker[12,14], Jordan Lerner-Ellis[3,10,16], Matthew Mistry[1,4,5], Rina Dvir[17], Ronald Grant[14,15], Ronit Elhasid[17], Roula Farah[18], Glenn P Taylor[19], Paul C Nathan[14,15], Sarah Alexander[14,15], Shay Ben-Shachar[20], Simon C Ling[15,21], Steven Gallinger[22,23], Shlomi Constantini[24], Peter Dirks[4,25], Annie Huang[4,14,15], Stephen W Scherer[1,7,12,26], Richard G Grundy[27], Carol Durno[21,22], Melyssa Aronson[22], Anton Gartner[11], M Stephen Meyn[1,12,15,28], Michael D Taylor[4,25], Zachary F Pursell[13], Christopher E Pearson[1,12], David Malkin[1,14,15], P Andrew Futreal[6], Michael R Stratton[6], Eric Bouffet[4,14,15], Cynthia Hawkins[3,4,19], Peter J Campbell[6,29] & Uri Tabori[1,4,14,15] for the Biallelic Mismatch Repair Deficiency Consortium[30]

DNA replication–associated mutations are repaired by two components: polymerase proofreading and mismatch repair. The mutation consequences of disruption to both repair components in humans are not well studied. We sequenced cancer genomes from children with inherited biallelic mismatch repair deficiency (bMMRD). High-grade bMMRD brain tumors exhibited massive numbers of substitution mutations (>250/Mb), which was greater than all childhood and most cancers (>7,000 analyzed). All ultra-hypermutated bMMRD cancers acquired early somatic driver mutations in DNA polymerase ε or δ. The ensuing mutation signatures and numbers are unique and diagnostic of childhood germ-line bMMRD ($P < 10^{-13}$). Sequential tumor biopsy analysis revealed that bMMRD/polymerase-mutant cancers rapidly amass an excess of simultaneous mutations (~600 mutations/cell division), reaching but not exceeding ~20,000 exonic mutations in <6 months. This implies a threshold compatible with cancer-cell survival. We suggest a new mechanism of cancer progression in which mutations develop in a rapid burst after ablation of replication repair.

Genetic changes underlie the development of neoplasia and can take many forms, including point mutations, copy number alterations and rearrangements. Irrespective of their type, somatic changes are caused, or allowed to persist, because of deficiencies in DNA repair. However, our understanding of the relationship between specific DNA repair defects and the resultant mutation type is limited. This is primarily because sporadic cancers are heterogeneous and involve dysfunction in multiple DNA-repair defects and types of mutation that accumulate over many years. In contrast, early-onset cancers from patients with inherited DNA-repair deficiency can offer an unobstructed view of the mutation types and secondary pathways that drive carcinogenesis. bMMRD is a childhood cancer syndrome characterized by early-onset cancers in various organs caused by biallelic mutations in the mismatch repair pathway[1]. This is one of two components that prevent point mutations during replication. The second safeguard resides within the intrinsic proofreading ability of the DNA polymerases (ε and δ). Although correction of replication errors has been studied in model systems, the consequences of its complete absence have not been investigated in humans.

To study the secondary alterations and mutation types that lead to bMMRD cancer, we analyzed genomes of 17 inherited cancers (from 12 patients), using genome and exome sequencing and microarrays (**Supplementary Table 1a**). Additionally, we sequenced non-neoplastic tissues from patients for which matched tumor was not available (total of 16 exomes and 1 genome from 18 patients; **Supplementary Table 1b**). We compared the mutational landscape of bMMRD tumors to a reference data set of >7,000 cancers[2].

A full list of author affiliations appears at the end of the paper.

Received 30 October 2014; accepted 5 January 2015; published online 2 February 2015; doi:10.1038/ng.3202

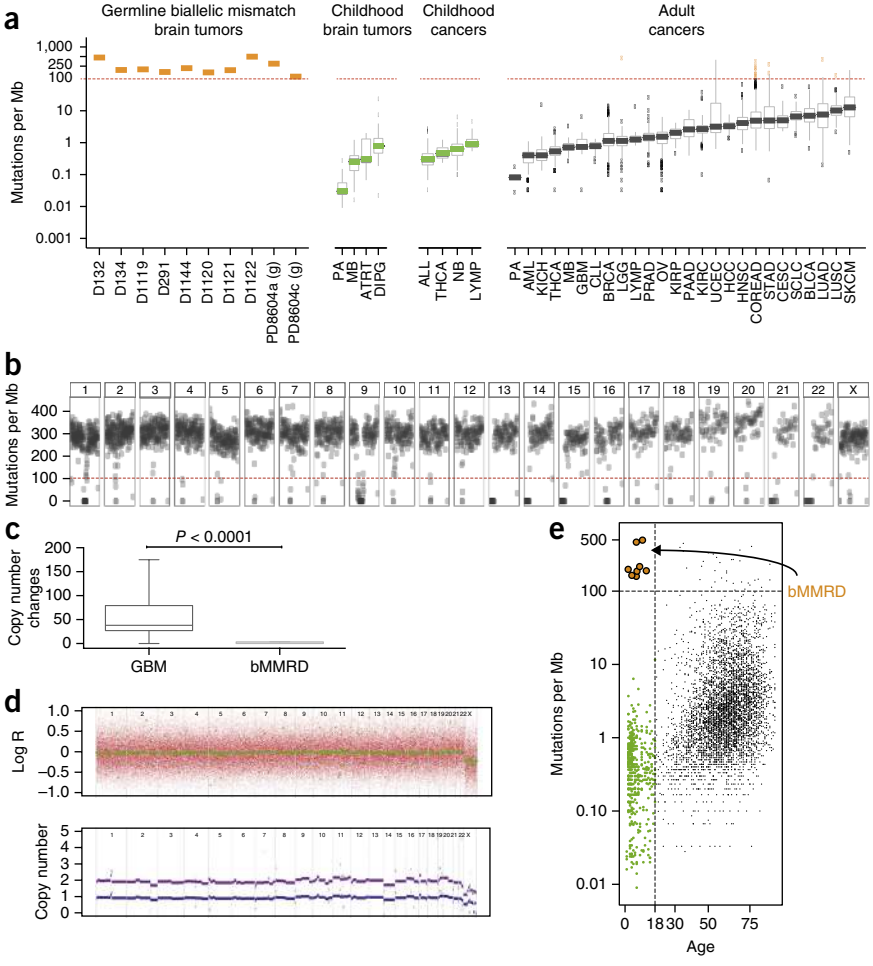


Figure 1 Somatic mutation frequency in bMMRD ultra-hypermutated cancers. (**a**) Mutation frequencies in bMMRD ultra-hypermutated malignant brain tumors (mean = 249 mutations/Mb) compared to a diverse cohort of other childhood brain cancers (<1 mutation/Mb), childhood cancers (<1 mutation/Mb) and adult cancers (<10 mutations/Mb). Data on the *y* axis are log-transformed. bMMRD from exome sequencing unless denoted with "(g)". Cancers with >100 mutations/Mb are highlighted in orange. Cancer-type abbreviations and number of samples per group (representative cancers from ref. 2) are indicated in the **Supplementary Note**. For box plots, the thick horizontal line (green or black) indicate median, and upper and lower hinges correspond to the 25th and 75th percentiles. (**b**) Mutation frequencies, as calculated in 1-Mb bins, are plotted for each chromosome and reveal no evidence of localized hypermutation (kataegis[22]). The red dashed line indicates 100 mutations/Mb. (**c**) Total copy number changes in sporadic glioblastomas ($n = 578$, average = 55.48 changes/sample) and bMMRD glioblastomas ($n = 4$, average = 1.5 changes/sample). The Mann-Whitney non-parametric test was used to calculate *P* values. (**d**) Copy number profile of two bMMRD brain tumors. Chromosomal log R ratios and copy number plots are shown; in each plot, purple indicates total copy number and blue indicates copy number of the minor allele. (**e**) Tumor mutation frequency (log scale) as a function of age. bMMRD cancers are marked in orange. All other pediatric cancers are in green. The probability of observing ultra-hypermutation in a child with sporadic non-bMMRD was $<10^{-13}$.

Of the 17 bMMRD cancers, all 10 malignant brain tumors exhibited an extremely large number of point mutations (average, 7,911 coding mutations; 249 mutations/Mb). This mutation frequency is in stark contrast to that in other pediatric cancers (0.61 mutations/Mb) and in all other sequenced cancers, irrespective of age of onset (**Fig. 1a**), and we therefore refer to these cancers as 'ultra-hypermutated cancers'.

Ultra-hypermutated bMMRD cancers contained an even distribution of mutations throughout the genome (**Fig. 1b**) and displayed other features distinct from other sequenced tumors: they were almost completely devoid of the copy number alterations typically observed in childhood brain cancers (**Fig. 1c**,**d** and **Supplementary Fig. 1**) and were microsatellite stable (unlike mismatch repair (MMR)-mutated sporadic cancers[3]).

These were the only cases of ultra-hypermutation in our analysis of childhood cancers; the probability of observing this staggering number of mutations in a child with sporadic non-bMMRD disease was $<10^{-13}$ (**Fig. 1e**). Indeed, in a previous large profiling study of pediatric high-grade gliomas, three ultra-hypermutated tumors also had been found to harbor germline biallelic mismatch repair mutations[4]. To our knowledge, ours is the first report of a tumor genome profile that can be used to infer germline mutational status.

DNA for non-neoplastic samples from patients with bMMRD (lymphocytes, $n = 16$) and controls had similar numbers of variants (**Supplementary Fig. 2**). This contrasts with the high mutation load observed in non-neoplastic tissues of MMR-deficient mice[5].

To test whether this absence of excessive mutation was a result of residual mismatch repair activity, we evaluated MMR activity in non-neoplastic cells derived from patients with bMMRD using the G•T mismatch assay[6,7]. All cells lacked protein expression of the corresponding mutant MMR gene and were completely deficient in G•T mismatch repair (**Supplementary Figs. 3** and **4**). Therefore, it appears that secondary mutations are required to cause the ultra-hypermutation seen in bMMRD tumors.

We examined each cancer for somatic mutations in the replication repair machinery. All ultra-hypermutated cancers harbored mutations in polymerase ε (Pol ε, *POLE*, 7/10 tumors) or polymerase δ (Pol δ, *POLD1*, 3/10 tumors). Nonmalignant tissue and non–ultra-hypermutated bMMRD cancers lacked mutations in polymerase genes ($n = 17$ and 7 tumors, respectively; **Fig. 2a**). These proofreading polymerases work cooperatively with MMR proteins.

POLE was the most frequently mutated DNA repair gene in bMMRD (**Supplementary Fig. 5**). Nonetheless, with a somatic mutation every ~5 kb, a large proportion of protein-coding genes would be expected to carry mutations and the presence of a high number of polymerase mutations could in theory be due to chance. We therefore undertook several analyses to address the potential role that polymerase mutations might assume in bMMRD cancers.

POLE mutations affected critical amino acid residues. Each bMMRD cancer with a mutation in *POLE* (bMMRD/*POLE* cancer; 7/7 tumors) harbored a mutation affecting the exonuclease domain or domains important to the intrinsic proofreading activity of Pol ε (**Supplementary Fig. 6a** and **Supplementary Table 2**). Residues S459 and S461 (substituted in one tumor and three tumors, respectively) are in the ExoIII exonuclease motif, adjacent to one of the exonuclease

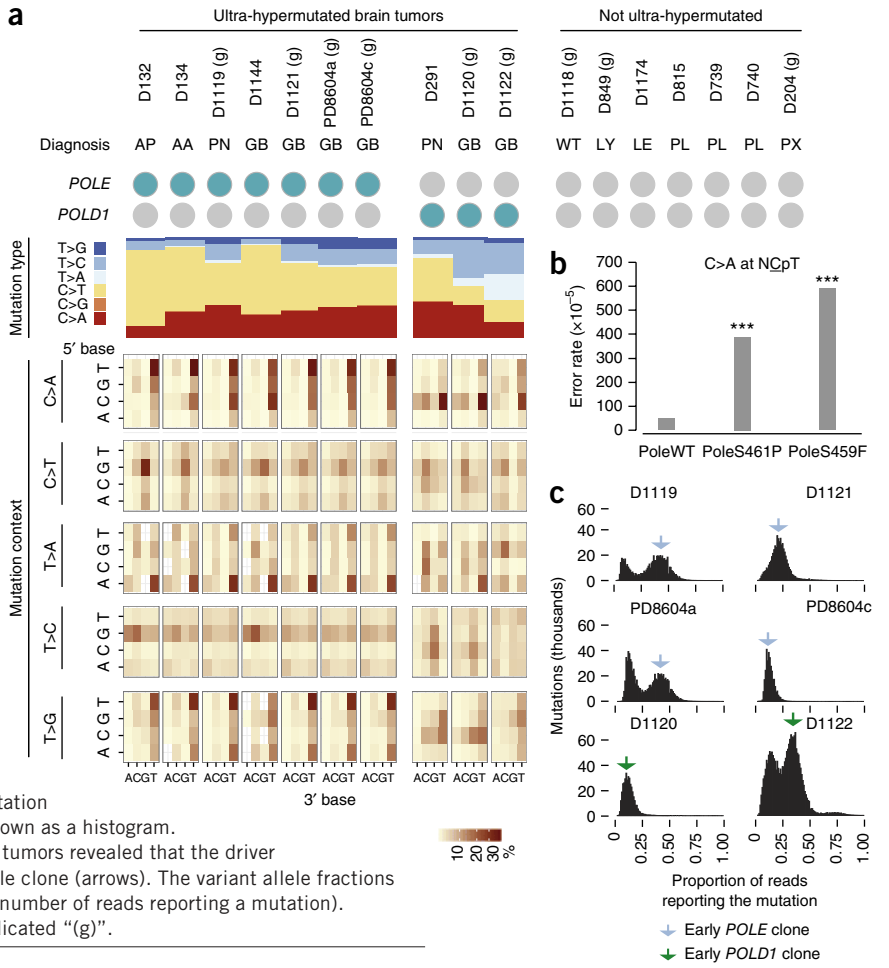


Figure 2 Consequences of polymerase mutations in bMMRD cancers. (**a**) *POLE* and *POLD1* driver mutations (blue circles) found in ultra-hypermutated malignant brain tumors ($n = 10$) but not in low-grade tumors, other cancer types or benign polyps from patients with bMMRD ($n = 7$). Mutation type indicates the simple mutation spectrum of ultra-hypermutated cancers. Mutation context shows base substitution mutation spectra for *POLE* and *POLD1* cancers. Each of the 96 mutated trinucleotides are represented in a heatmap. The base located 5′ to each mutated base is shown on the vertical axis, and the 3′ base is on the horizontal axis. C>G mutations are not included in this plot as there were too few of them. Cancer type abbreviations are indicated in the **Supplementary Note**. (**b**) Pol ε *in vitro* error rates for tumor mutation hotspots. A reversion substrate similar to the CT→AT transversion error hotspot seen in human tumors was generated. This substrate only scores CT→AT transversions. Mutant frequencies were calculated for wild type (3 mutants out of 9,927 plaques scored), S461P and S459F along with error rates for each (***$P \leq 0.0001$). *P* values were calculated using chi-square tests. Error rates are the averages of two experiments, each conducted with independent DNA and enzyme preparations for each construct tested. (**c**) Timing of *POLE* and *POLD1* mutation with respect to all other mutations in the genome, shown as a histogram. Clonality analysis of the ultra-hypermutated bMMRD tumors revealed that the driver polymerase mutations occurred in the earliest possible clone (arrows). The variant allele fractions of somatic mutations per tumor are plotted (i.e., the number of reads reporting a mutation). Samples with whole-genome sequencing data are indicated "(g)".

catalytic residues conserved in all polymerases (D462)[8]. F104 is in an F/YxPYFY motif conserved in both human Pol ε and δ (ref. 9). S297 and P436 closely flank the ExoI and ExoII motifs and are absolutely conserved in all *POLE* orthologs[8].

To assess how the proofreading capability of Pol ε was affected by these *POLE* mutations, we introduced mutations conferring the most frequent substitutions (S459F and S461P) into a construct encoding the Pol ε catalytic subunit[10] and performed an *in vitro* assay measuring mutation accumulation. These mutations resulted in the loss of replication fidelity and a high mutation rate[11] (**Fig. 2b**).

POLD1 mutations also affected conserved domains (**Supplementary Fig. 7**). C319 and L606 were mutated in one and two tumors, respectively. C319 is immediately adjacent to one of the exonuclease catalytic sites in Pol δ (E318) within the ExoI motif. The recurrent L606 substitution (L606M) is in motif A of the polymerase domain[12–14]; the identical substitution in yeast Pol δ (L612M) has been shown to dramatically reduce replication fidelity[15].

POLE and *POLD1* mutations are likely to have occurred as early events in each cancer's life history (**Fig. 2c** and **Supplementary Fig. 8**). All samples harbored vast numbers of genomic mutations at a low allelic fraction (<20%; subclonal variants), indicating a recent and explosive accumulation of mutations after *POLE* or *POLD1* mutation. These data, coupled with the presence of mutator polymerases and high mutation loads, suggest that mutant Pol ε and Pol δ are drivers in bMMRD.

To understand the extent to which polymerase defects affect the overall bMMRD genome, we explored their mutational profiles in greater depth. bMMRD/*POLE* cancers exhibited a mutational 'signature' that is specific to cancers with a mutant gene encoding DNA polymerase ε (ref. 2; as confirmed above, **Fig. 2b**). Of the six classes of base substitution, the mutational landscape of these cancers was characterized by C>T and C>A changes (**Fig. 2a**, mutation type). They also contained very few C>G mutations. We then analyzed the sequence context of each substitution based on the flanking 3′ and 5′ bases (**Fig. 2a**, mutation context). All bMMRD/*POLE* cancers had a common signature with highly distinctive features: most C>A and T>G transversions were followed by a 3′ thymine (>85% of C>A and >70% of T>G substitutions). Notably, these were frequently preceded by a thymine, that is, C>A at TCT and T>G at TTT (>30% of C>A and T>G substitutions). Thus, the genome in tumors with mutant *POLE* incurred a signature mutation spectrum.

bMMRD/*POLD1* cancers displayed their own idiosyncratic mutational pattern, which differed markedly from that of the bMMRD/*POLE* tumors (**Supplementary Fig. 9**). These cancers exhibited many C>A and C>T mutations, as well as an excess of T>A and T>C mutations, especially as compared to the bMMRD/*POLE* cancers (**Fig. 2a**, mutation context). Although C>A changes feature prominently in both *POLE* and *POLD1* cancers, these occur in a completely different sequence context: bMMRD/*POLD1* cancers are characterized by C>A mutations at CCN, with a particular enrichment for C>A at CCT. To our knowledge, this is the first report of somatic *POLD1* driver mutations in ultra-hypermutated cancers. This mutation spectrum was also recently found in engineered yeast with the same mutated residue (pol 3 L612M)[16].

This signature occurred early and matches that in previously described sporadic *POLE*-related cancers[2]. Next, we looked for the

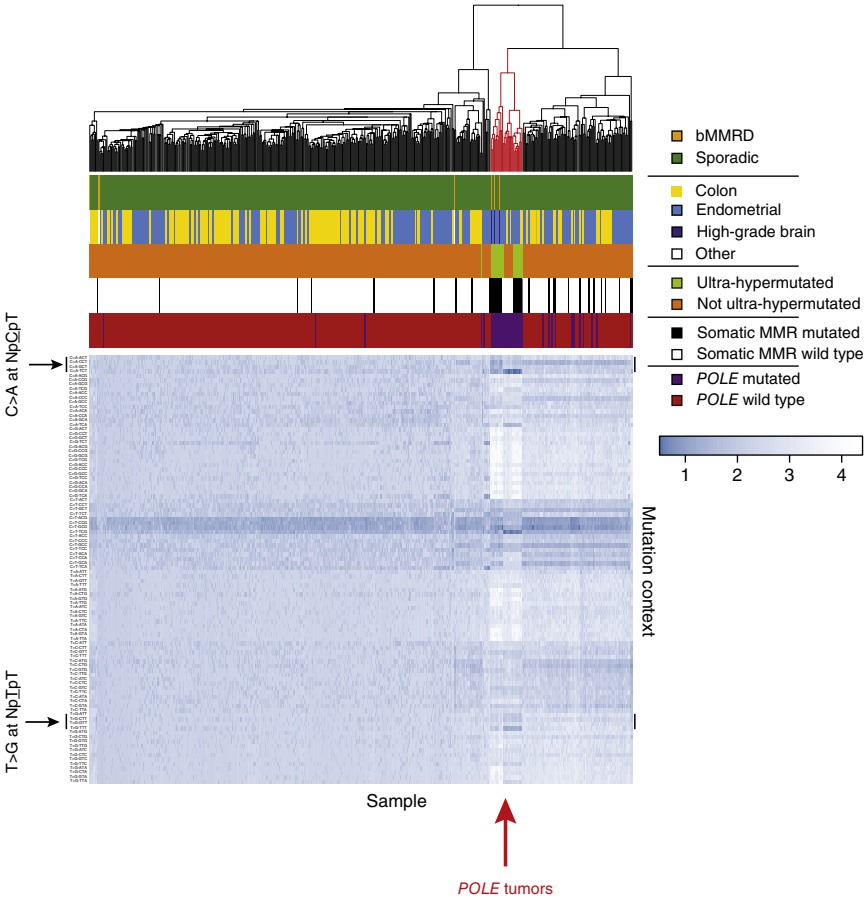
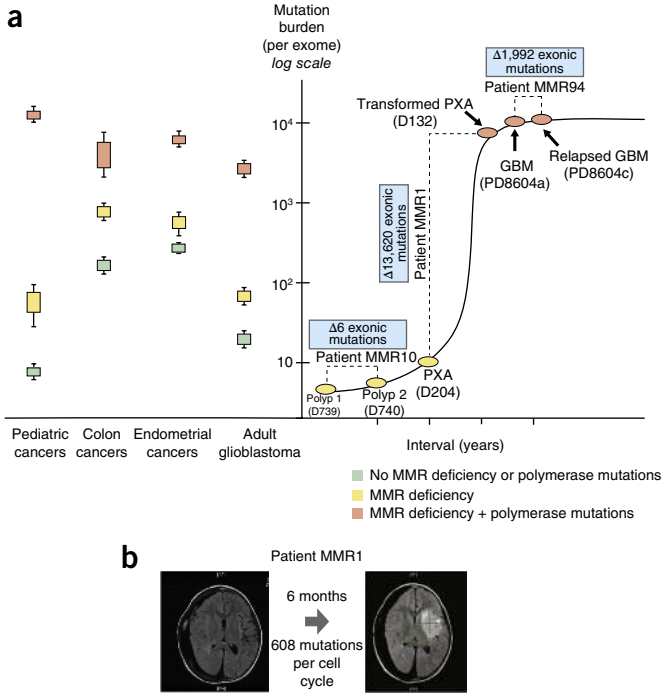


Figure 3 Mutation spectrum of inherited and sporadic cancers. Shown is a cluster analysis based on mutation context of bMMRD cancers and sporadic colorectal and endometrial tumors. The 96 possible trinucleotides of all substitutions are on the *y* axis, and individual samples are on the *x* axis. All bMMRD/*POLE* cancers clustered together with ultra-hypermutated *POLE* endometrial and colorectal cancers of adulthood. Arrows indicate the mutation contexts enriched in *POLE* cancers. In the heatmap, colors represent the proportion of each trinucleotide ($-\log_{10}$ transformed) in that sample, such that the most common mutation types are in dark blue and the least common mutation types are in white.

same signature in other common cancers. Substitutions in *POLE* had a similar effect on bMMRD cancers as tumors with known somatic MMR and *POLE* mutations (colorectal and endometrial tumors[17,18]). Similar to data for our cohort, MMR/*POLE* cancers were hypermutated, contained few copy number changes and were microsatellite stable[17] (**Supplementary Fig. 10**). Lastly, unbiased hierarchical clustering of trinucleotide sequence revealed that all ultra-hypermutated bMMRD/*POLE* tumors grouped into a single cluster with sporadic MMR/*POLE* endometrial and colorectal cancers (**Fig. 3**).

Our data suggest that bMMRD cancers bear the imprint of polymerase defects, in the form of a massive number of highly specific substitutions acquired in a short time. We wondered whether we could use these unique features of bMMRD/polymerase cancers to study the accumulation, frequency and threshold (upper limit) of mutation accrual in cancer.



We compared the mutation load of bMMRD and bMMRD/polymerase cancers with that found in other human cancers (**Fig. 4a**). bMMRD tumors lacking a *POLE* mutation had an approximately 5- to 10-fold increase in mutation load relative to pediatric cancers from the same tissue type with intact MMR, whereas bMMRD/polymerase tumors displayed a 230-fold increase in exonic mutations relative to bMMRD alone. This mutation prevalence is similar to what has been previously reported in model organisms with engineered deficiencies in each pathway[9]. The genomes of inherited and sporadic MMR/polymerase cancers reached the same mutation level and did not exceed it ($1–2 \times 10^4$) despite decades of difference in ages of onset (**Fig. 4**).

Finally, to study the rate of mutation accumulation and to establish the time required to develop bMMRD/polymerase cancer, we used specimens collected as part of our clinical surveillance protocol[19]. Sequential magnetic resonance imaging (MRI) and endoscopies enabled determination of tumor appearance and the collection of multiple specimens from carriers, which we used to measure the accumulation of somatic mutations over time. bMMRD gastrointestinal polyps

Figure 4 Mutation threshold and rate in cancers with mismatch and polymerase mutations. (**a**) Mutation burden in pediatric and adult cancers with and without mutations in MMR genes and/or polymerase defects (left). Box plots indicate the actual number of exonic mutations in human cancers, as determined by exome or genome sequencing. Box plots indicate 25th and 75th percentiles, and whiskers denote the upper and lower number of mutations per exome. Rate of mutation accumulation of serially collected bMMRD tumors (right). For three patients, the mutation frequency of tumor pairs was contrasted with the number of new exonic mutations shown (box) and the tumor type indicated (below the ovals). PXA, pleomorphic xanthoastrocytoma; GBM, glioblastoma multiforme. (**b**) Surveillance MRI scans.

(MMR10; **Fig. 4a**) did not contain a higher mutational load than adult polyps[20]. Normal gastrointestinal mucosa and blood-derived DNA collected at different times contained few new mutations. It is reasonable to suggest that in the absence of secondary polymerase variants cancers mutate steadily, requiring many years to develop sufficient drivers.

In contrast, serial analysis of recurrent brain cancers revealed rapid accumulation of mutations over a very short time. A bMMRD/*POLE* mutant glioblastoma that transformed from a low-grade glioma with wild-type *POLE* (MMR1) had similar mutations (including *TP53* mutation: p53 substitution R273C) but exhibited 13,620 new exonic substitutions. Moreover, although we observed 72,354 new substitutions (by whole-genome sequencing) between a primary and relapsed bMMRD/*POLE* mutant glioblastoma (MMR94), the total amount of mutations remained the same within the threshold (**Fig. 4a**).

To quantify mutation accumulation over time, we used repeated MRI data of four glioblastomas that developed from nonvisible masses (over 4–6 months). For example, a tumor of 381 mm^3 corresponds to ~35 cell doublings and a mutation load of 21,284 mutations; it would therefore have 608 mutations per cell division (**Fig. 4b**). We appreciate that these are conservative estimates because they do not take into account the last few cycles of tumor growth (in which mutations would be below the detection of high-throughput sequencing) and potential loss owing to death of hypermutant cell clones. However, they are consistent between patients and similar to the numbers postulated from our *in vitro* polymerase assay. bMMRD/polymerase-mutant cancer that divides every 5–6 d will accumulate a staggering 250–600 mutations per cell cycle, thereby enabling bMMRD/polymerase cancers to acquire sufficient driver mutations in less than 6 months (**Supplementary Table 3**).

Our data directly reveal the consequences of complete ablation of replication error repair in human cancer. Once the proofreading ability of the DNA polymerases are lost in a mismatch repair–deficient cell, there is no defense against a rapid and catastrophic accumulation of point mutations (**Supplementary Fig. 11**). Despite the extreme consequences of absent DNA replication repair, the resulting signatures are similar and consistent: mutations arise throughout the genome in a specific spectrum in the background of a near-diploid genome, and accumulate to a threshold without surpassing it. Ultra-hypermutated cells mutate continuously, potentially generating multiple independent subclones (**Fig. 2c**), until confronting a threshold. The high mutation load and threshold may be this cancer's Achilles' heel, exploitable for therapeutic intervention.

This is to our knowledge the first description of a massive simultaneous accumulation of point mutations associated with extremely rapid tumor initiation. The ultra-hypermutated phenotype occurs rapidly and is limited to substitutions, making it distinct from other tumors which carry a variety of mutation types that typically accumulate in a slow and stepwise manner to provide sufficient clonal advantage[21]. bMMRD/polymerase-mutant cancers therefore suggest a new and unique mechanism for cancer initiation.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. European Genome-phenome Archive (EGA): EGAD00001000369 and EGAS00001001112.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

U.T. received funding from BRAINchild Canada and the Canadian Institute of Health Research (operating grant MOP123268). C.E.P. received funding from the Canadian Institute of Health Research (operating grant FRN131596). P.J.C. and A.G. are personally funded through Wellcome Trust Senior Clinical and Basic Research Fellowships and are members of the Wellcome-funded COMSIG consortium. S.B. is funded through a Wellcome Trust Research Training Fellowship for Clinicians. B.B.C., M.M. and R.A. are supported by a SickKids Restracomp award. We acknowledge J. Costello for his contribution to the manuscript.

AUTHOR CONTRIBUTIONS

A.S., E.B., C.H., P.J.C. and U.T. designed the study. B.B.C., A.P., T.L., A.H., S.D., N.A., B.M., M.G., Y.H., D.M.M., M.R., Ma.R., G.B.P., N.P.T., K.P.H., E.E.H., A.Y.G., D.B., G.S.C., H.D., J.L.E. and M.M. performed experiments. A.S., B.B.C., R.B., L.B.A., Da.M., D.W., P.V.L., P.S.T., P.C., S.B., R.A., C.D., M.A. and U.T collected and analyzed data. R.G., R.D., Ro.G., R.E., R.F., G.P.T., P.C.N., S.A., S.B.-S., S.C.L., S.C., P.D., A.H. and U.T. provided reagents, tissue and clinical data. A.S., M.S.M., M.D.T., Z.F.P., C.E.P., D.M., P.J.C. and U.T. wrote the manuscript. S.G., S.W.S., C.D., M.A., A.G., M.S.M., M.D.T., Z.F.P., C.E.P., D.M., P.A.F., M.R.S., E.B., C.H. and P.J.C. provided technical support and conceptual advice. All authors approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. 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**, 355–365 (2014).
2. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
3. Liu, B. *et al.* Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat. Genet.* **9**, 48–55 (1995).
4. Wu, G. *et al.* The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat. Genet.* **46**, 444–450 (2014).
5. Narayanan, L., Fritzell, J.A., Baker, S.M., Liskay, R.M. & Glazer, P.M. Elevated levels of mutation in multiple tissues of mice deficient in the DNA mismatch repair gene Pms2. *Proc. Natl. Acad. Sci. USA* **94**, 3122–3127 (1997).
6. Thomas, D.C., Roberts, J.D. & Kunkel, T.A. Heteroduplex repair in extracts of human HeLa cells. *J. Biol. Chem.* **266**, 3744–3751 (1991).
7. Panigrahi, G.B., Slean, M.M., Simard, J.P., Gileadi, O. & Pearson, C.E. Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. *Proc. Natl. Acad. Sci. USA* **107**, 12593–12598 (2010).
8. Henninger, E.E. & Pursell, Z.F. DNA polymerase epsilon and its roles in genome stability. *IUBMB Life* **66**, 339–351 (2014).
9. Preston, B.D., Albertson, T.M. & Herr, A.J. DNA replication fidelity and cancer. *Semin. Cancer Biol.* **20**, 281–293 (2010).
10. Korona, D.A., Lecompte, K.G. & Pursell, Z.F. The high fidelity and unique error signature of human DNA polymerase epsilon. *Nucleic Acids Res.* **39**, 1763–1773 (2011).
11. Bebenek, K. & Kunkel, T.A. Analyzing fidelity of DNA polymerases. *Methods Enzymol.* **262**, 217–232 (1995).
12. Ghodgaonkar, M.M. *et al.* Phenotypic characterization of missense polymerase-delta mutations using an inducible protein-replacement system. *Nat. Commun.* **5**, 4990 (2014).
13. Venkatesan, R.N. *et al.* Mutation at the polymerase active site of mouse DNA polymerase delta increases genomic instability and accelerates tumorigenesis. *Mol. Cell. Biol.* **27**, 7669–7682 (2007).
14. Schmitt, M.W. *et al.* Active site mutations in mammalian DNA polymerase delta alter accuracy and replication fork progression. *J. Biol. Chem.* **285**, 32264–32272 (2010).
15. Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M. & Kunkel, T.A. Division of labor at the eukaryotic replication fork. *Mol. Cell* **30**, 137–144 (2008).
16. Lujan, S.A. *et al.* Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res.* **24**, 1751–1764 (2014).
17. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
18. Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
19. Durno, C.A. *et al.* Oncologic surveillance for subjects with biallelic mismatch repair gene mutations: 10 year follow-up of a kindred. *Pediatr. Blood Cancer* **59**, 652–656 (2012).
20. Nikolaev, S.I. *et al.* A single-nucleotide substitution mutator phenotype revealed by exome sequencing of human colon adenomas. *Cancer Res.* **72**, 6279–6289 (2012).
21. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. USA* **105**, 4283–4288 (2008).
22. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).

[1]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada. [2]Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada. [3]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. [4]The Arthur and Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada. [5]Institute of Medical Science, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. [6]Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK. [7]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada. [8]Department of Human Genetics, University of Leuven, Leuven, Belgium. [9]Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire, UK. [10]Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada. [11]Centre for Gene Regulation and Expression, University of Dundee, Dundee, UK. [12]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [13]Department of Biochemistry & Molecular Biology, Tulane Cancer Center, Tulane University, School of Medicine, New Orleans, Louisiana, USA. [14]Division of Hematology/Oncology, The Hospital for Sick Children, Toronto, Ontario, Canada. [15]Department of Pediatrics, University of Toronto, Ontario, Canada. [16]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [17]Department of Pediatric Hemato-Oncology, Tel Aviv Medical Center, Tel-Aviv, Israel. [18]Saint George Hospital University Medical Center, Beirut, Lebanon. [19]Division of Pathology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada. [20]The Gilbert Israeli Neurofibromatosis Center, Tel Aviv Medical Center, Tel Aviv, Israel. [21]Division of Gastroenterology, Hepatology, and Nutrition, Department of Paediatrics, University of Toronto, The Hospital for Sick Children, Toronto, Ontario, Canada. [22]The Familial Gastrointestinal Cancer Registry at the Zane Cohen Centre for Digestive Disease, Mount Sinai Hospital, Toronto, Ontario, Canada. [23]Department of Surgery, Mount Sinai Hospital, Toronto, Ontario, Canada. [24]Department of Pediatric Neurosurgery, Dana Children's Hospital, Tel Aviv Medical Center, Tel Aviv, Israel. [25]Division of Neurosurgery, The Hospital for Sick Children, Toronto, Ontario, Canada. [26]The McLaughlin Centre, University of Toronto, Toronto, Canada. [27]Children's Brain Tumour Research Centre, University of Nottingham, Nottingham, UK. [28]Division of Clinical and Metabolic Genetics, Department of Paediatrics, The Hospital for Sick Children, Toronto, Ontario, Canada. [29]Department of Haematology, University of Cambridge, Cambridge, UK. [30]A list of contributing members appears in the **Supplementary Note**. [31]These authors contributed equally to this work. Correspondence should be addressed to A.S. (adam.shlien@sickkids.ca), P.J.C. (pc8@sanger.ac.uk) or U.T. (uri.tabori@sickkids.ca).

ONLINE METHODS

Patient and sample collection. Patients were registered as a part of the International Biallelic Mismatch Repair Consortium, which includes multiple centers worldwide. Detailed information on each family and all patients can be found in our previous study[23]. Following Institutional Research Ethics Board approval, all data were centralized in the Division of Haematology/Oncology at The Hospital for Sick Children (SickKids) and the Familial Gastrointestinal Cancer Registry (FGICR) at the Zane Cohen Centre for Digestive Diseases at Mount Sinai Hospital, in Toronto, Canada. Consent forms were obtained from the parents or guardians, or from the patients, where applicable. Family history, demographic and clinical data were obtained from the responsible physician and/or genetic counselor at the corresponding centers. Further information can be found in **Supplementary Table 1**.

Tumor and blood samples were collected from the Sickkids tumor bank. The diagnosis of bMMRD was made when a germ-line biallelic mutation in any of the four MMR genes (*MLH1*, *MSH2*, *MSH6* and *PMS2*) was confirmed by sequencing in a clinically approved laboratory.

The surveillance protocol developed by our group[23] was used to gather clinical information, such as time to tumor development, and tumor samples from biopsies that were used for sequencing (**Fig. 4** and **Supplementary Table 1**).

Microsatellite instability testing. Microsatellite instability testing was performed in a clinically approved laboratory, as described previously[23].

High-throughput sequencing, read mapping and identification of mutations. Tumors were sequenced using Agilent's exome enrichment kit (Sure Select V4; with >50% of baits above 25× coverage) or by whole genome sequencing to a depth ~40× (**Supplementary Fig. 12**). In all cases but one, the matched blood-derived DNA was also sequenced. Base calls and intensities from the Illumina HiSeq 2500 were processed into FASTQ files using CASAVA. The paired-end FASTQ files were aligned to the genome (to UCSC's hg19 GRCh37) with BWA[24] (v0.5.9). Duplicate paired-end sequences were removed using Picard MarkDuplicates (v1.35) to reduce potential PCR bias. Aligned reads were realigned for known insertion/deletion events using SRMA[25] (v0.1.155). Base quality scores were recalibrated using the Genome Analysis Toolkit[26] (v1.1-28). Somatic substitutions were identified using MuTect[27] (v1.1.4) or CaVEMaN[22]. Mutations were then filtered against common single-nucleotide polymorphisms (SNPs) found in dbSNP (v132), the 1000 Genomes Project (Feb 2012), a 69-sample Complete Genomics data set, and the Exome Sequencing Project (v6500). Mutation signature profiles were extracted using the single base substitution and the corresponding tri-nucleotide sequence context (i.e., reference base at mutation position and its 5′ and 3′ neighbors).

Comparison of bMMRD mutation frequency to sporadic cancers. Mutation frequencies (substitutions per Mb) for bMMRD tumors were calculated from genome or exome data as per previous publications[2] and data on sporadic cancer, including age of onset, were obtained from ref. 2. Data shown in **Figure 1** are from ref. 2 and from brain tumors sequenced at SickKids.

Copy number analysis. DNA from bMMRD tumors was hybridized to Affymetrix SNP 6.0 arrays ($n = 4$ tumors). Copy number segmentation was performed using the Single Nucleotide Polymorphism-Fast Adaptive States Segmentation Technique (Biodiscovery Nexus Copy Number 7.5). This hidden Markov model–based approach was used with a significance threshold for segmentation set at 5.0×10^{-7} also requiring a minimum of three probes per segment and a maximum probe spacing of 1,000 kbp between adjacent probes before breaking a segment. The log ratio thresholds for copy gain and copy loss were set at 0.1 and −0.15, respectively. We compared bMMRD tumor copy number profiles to that of 578 glioblastoma samples previously hybridized to the same array platform. To account for possible differences in segmentation algorithms in the two data sets, copy number segments (either gains or losses) smaller than 5 Mb were excluded. The frequency of segments was compared using a Mann-Whitney nonparametric test.

Validation of substitution mutations. Putative driver mutations in *POLE* and *POLD1* were validated by Sanger sequencing (**Supplementary Fig. 13**).

Western blotting for MMR protein expression in non-neoplastic biallelic MMR mutant cells. Cell extracts were prepared as described[7] and 40 µg of HeLa, wild-type lymphoblast, LoVo, MMR8 lymphoblast and MMR10 lymphoblast cell extracts were loaded in each well. Simultaneous western blotting for human MSH2, MSH3, MSH6 and actin was carried out as described[7,28]. Another membrane was simultaneously probed for human PMS2 with 1/100 dilution of anti-PMS2 (BD Pharmingen 556415), human MLH1 with 1/500 dilution of anti-MLH1 (BD Pharmingen 554073) and actin. Both immunoblots were incubated in HRP-conjugated sheep anti-mouse secondary antibody, and chemiluminescence signals were generated using Biorad Clarity Western ECL substrate. Images were captured on VWR CA11006-128 films with multiple exposures.

G•T mismatch repair reactions and repair efficiencies. Repair reactions were carried out as described previously[6,7]. Briefly 20 fmol of circular substrate carrying a G•T mismatch and a nick 5′ to the mismatch was incubated with whole cell extracts (2–4 mg/ml of proteins), NTPs, dNTPs, creatine kinase and creatine phosphate for 1 h at 37 °C. Reactions were stopped in 2 mg/ml proteinase K, 2% SDS, 50 mM EDTA, pH 8.0, for 1 h followed by phenol-chloroform extraction. Mixtures were subjected to enzymatic purification kit of Qiagen and mini elute column. Products were eluted with 15 µl of elution buffer and digested with XmnI to linearize the substrate, and HindIII to assess whether correct repair had occurred. Products were resolved on 1% agarose gels, and probed by Southern blotting for quantitative analysis. Membranes were probed with radioactive probe and quantification was performed with a Typhoon FLA 9500 phosphorimager. Repair efficiency is the proportion of radiointensity of the repair products relative to all fragments.

Purification of polymerase ε. An expression vector encoding residues 1–1,189 of the catalytic subunit of human Pol ε was used in site-directed mutagenesis reactions to change Ser461 to proline. Human Pol ε was prepared as described[10]. Briefly, the human Pol ε was coexpressed in autoinduction medium with pRK603, which allows coexpression of TEV protease, at 25 °C until the culture was saturated. Peak fractions from the HisTrap column were pooled, dialyzed into 50 mM HEPES, pH 7.5, 1 mM DTT, 5% glycerol and bound to SP sepharose. Bound protein was eluted with a 0–1 M with NaCl gradient. Peak fractions were pooled, dialyzed into 50 mM Tris, pH 7.5, 1 mM DTT, 5% glycerol, 100 mM NaCl and bound to Q Sepharose. Bound protein was eluted with a 100 mM–M M NaCl gradient. Peak fractions were pooled, concentrated and passed through a pre-equilibrated Superdex200 size exclusion column. Fractions containing the purified 140 kDa protein were pooled, dialyzed into 50 mM Tris, pH 8.0, 1 mM DTT, 5% glycerol and aliquots were frozen and stored at −80 °C.

The data for S459F have been described previously[29].

***LacZ in vitro* mutant frequency and error-rate calculations.** The *lacZ in vitro* forward mutation assay was performed essentially as described previously[11]. Briefly, double-stranded M13mp2 DNA containing a 407-nt ssDNA gap was used as a substrate in reactions containing 0.15 nM DNA, 50 mM Tris-Cl, pH 7.4, 8 mM $MgCl_2$, 2 mM DTT, 100 µg/ml BSA, 10% glycerol, 250 µM dNTPs and 1.5 nM Pol ε-M630G at 37 °C. Completely filled product was transfected into *Escherichia coli* cells, which were used to determine the frequency of light blue and colorless plaques that occurred as a result of mutations arising during DNA synthesis. In this assay, accurate DNA synthesis yields dark blue plaques.

One of the limitations of the forward assay is that sequence context-specific errors can be underestimated if that context is not well represented. To overcome this limitation, we generated a *lacZ* reversion substrate that only reports on C>A transversions in the CT>AT context. To generate the reversion substrate, we used site-directed mutagenesis to change A_{-11} to C_{-11} and prepared gapped substrate. The C_{-11}–containing substrate gives rise to light blue plaques. $C_{-11} \rightarrow A_{-11}$ transversion mutations are scored as dark blue plaques.

A pilot sequencing study indicated that 100% of these revertant plaques contained the $C_{-11} \rightarrow A_{-11}$ transversion mutation. *LacZ* mutant frequencies were calculated from combining at least two independent experiments. DNA from mutant plaques was purified and the *lacZ* gene was sequenced. Error rates were calculated according to the following equation: error rate (per nucleotide synthesized) = ((number of mutants of a particular class) × (mutant frequency)) / ((number of mutations sequenced) × (0.6) × (number of detectable sites)).

The data for S459F have been described previously[29].

Clustering of cancers by mutation spectra. Data from bMMRD were combined with somatic substitutions from sporadic endometrial cancers (n = 248) and colon cancers (n = 215), obtained from the TCGA (specifically, the Uterine Corpus Endometrial Carcinoma (UCEC) and the Colon Adenocarcinoma (COAD) studies). Only data sequenced on the Illumina platform were included. Only somatic substitutions were included. That is, insertions and deletions were discarded as were point mutations found in the 1000 Genomes Project (Feb 2012), a 69-sample Complete Genomics data set, and the Exome Sequencing Project (v6500). Data was then reannotated with ANNOVAR[2] to remain consistent with the annotations used on the bMMRD samples.

Substitutions were grouped on the basis of their 3′ and 5′ bases into 96 possible trinucleotide categories that were used for mutation spectrum analysis (**Fig. 2**) and clustering (**Fig. 3**). In the clustering analysis, the color of each grid represents the proportion of that trinucleotide in the sample ($-\log_{10}$ transformed). Pairwise comparisons were performed between samples, the Euclidean distance of the trinucleotide proportions was determined, and clustering was performed using the Divisive Analysis (Diana) clustering algorithm. Mutation frequencies were calculated (using 30 Mb as capture size). Samples with greater than 100 mutations/Mb were designated as ultra-hypermutated.

Calculation of mutation rate from repeated brain MRI scans of bMMRD patients. Calculation of mutation per cell cycle were performed based on established formulas. An example for one sample is provide (D132; **Supplementary Table 3**):

Diameter of tumor from MRI = 45 mm
Radius of tumor = 22.5 mm
Estimated tumor volume = $4/3\ \pi\ (0.0075)^3 = 1.8 \times 10^{-6}$ mm^3
Diameter of average animal cell = 15 μm
Radius of average animal cell = 0.0075 mm
Volume of average animal cell = $4/3\ \pi\ (0.0075)^3 = 1.8 \times 10^{-6}$ mm^3
Estimated number of cells in tumor = $44{,}712/1.8 \times 10^{-6} = 2.65 \times 10^{10}$ cells
Estimated number of cell divisions tumor has undergone = $2^x = 2.65 \times 10^{10}$
Where $x = \dfrac{\ln(2.65 \times 10^{10})}{\ln(2)}$ and $x \approx 35$ cell divisions.
Total mutations = 21,284
Number of cell cycles = 35
Mutations per cell cycle = 21,284/35 = 608 mutations per cell cycle

23. Bakry, D. *et al.* Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency consortium. *Eur. J. Cancer* **50**, 987–996 (2014).
24. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
25. Homer, N. & Nelson, S.F. Improved variant discovery through local re-alignment of short-read next-generation sequencing data using SRMA. *Genome Biol.* **11**, R99 (2010).
26. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
27. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
28. Tomé, S. *et al.* Tissue-specific mismatch repair protein expression: MSH3 is higher than MSH6 in multiple mouse tissues. *DNA Repair (Amst.)* **12**, 46–52 (2013).
29. 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).

doi:10.1038/ng.3202