

Published in final edited form as:

*Science*. 2017 October 13; 358(6360): 234–238. doi:10.1126/science.aao3130.

## Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer

Jarno Drost<sup>#1,2,†</sup>, Ruben van Boxtel<sup>#2,3,†</sup>, Francis Blokzijl<sup>2,3</sup>, Tomohiro Mizutani<sup>1,2</sup>, Nobuo Sasaki<sup>1,2</sup>, Valentina Sasselli<sup>1,2</sup>, Joep de Lig<sup>2,3</sup>, Sam Behjati<sup>4,5</sup>, Judith E. Grolleman<sup>6</sup>, Tom van Wezel<sup>7</sup>, Serena Nik-Zainal<sup>4,8</sup>, Roland P. Kuiper<sup>6,9</sup>, Edwin Cuppen<sup>2,3,‡</sup>, and Hans Clevers<sup>1,2,9,‡</sup>

<sup>1</sup>Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center (UMC) Utrecht, 3584CT Utrecht, Netherlands <sup>2</sup>Cancer Genomics Netherlands, UMC Utrecht, 3584CX Utrecht, Netherlands <sup>3</sup>Center for Molecular Medicine, Department of Genetics, UMC Utrecht, 3584CX Utrecht, Netherlands <sup>4</sup>Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK <sup>5</sup>Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK <sup>6</sup>Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, Netherlands <sup>7</sup>Departments of Pathology, Leiden University Medical Center, Leiden, Netherlands <sup>8</sup>East Anglian Medical Genetics Service, Cambridge University Hospitals National Health Service Foundation Trust, Cambridge, UK <sup>9</sup>Princess Máxima Center for Pediatric Oncology, 3584CT Utrecht, Netherlands

# These authors contributed equally to this work.

### Abstract

Mutational processes underlie cancer initiation and progression. Signatures of these processes in cancer genomes may explain cancer etiology and could hold diagnostic and prognostic value. We developed a strategy that can be used to explore the origin of cancer-associated mutational signatures. We used CRISPR-Cas9 technology to delete key DNA repair genes in human colon organoids, followed by delayed subcloning and whole-genome sequencing. We found that mutation accumulation in organoids deficient in the mismatch repair gene *MLH1* is driven by replication errors and accurately models the mutation profiles observed in mismatch repair-deficient colorectal cancers. Application of this strategy to the cancer predisposition gene *NTHL1*, which encodes a base excision repair protein, revealed a mutational footprint (signature 30) previously observed in a breast cancer cohort. We show that signature 30 can arise from germline *NTHL1* mutations.

---

Cancer arises through the sequential accumulation of mutations in somatic cells. The overall mutational burden of somatic cells is determined by a balance between DNA damage and repair activity. Mutational processes leave specific signatures, as defined by systematic

---

exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works.

‡Corresponding author. ecuppen@umcutrecht.nl (E.C.); h.clevers@hubrecht.eu (H.C.).

†Present address: Princess Máxima Center for Pediatric Oncology, 3584CT Utrecht, Netherlands.

analysis of mutation characteristics across many independent cancer sequencing data sets (1). A recent study showed that a set of mutational signatures in genome-wide mutation collections of breast cancers predicts deficiency of *BRCA1* and *BRCA2*, as well as sensitivity to PARP [poly(adenosine diphosphate–ribose) polymerase] inhibition (2). To date, the main approach to link specific mutational signatures to underlying molecular mechanisms has involved associating cancer mutations with defined exposures to carcinogens, such as tobacco smoke (3) and ionizing radiation (4), or with the absence of specific DNA repair proteins, such as components of the DNA mismatch repair (MMR) (1) and nucleotide excision repair (NER) pathways (5). However, multiple processes are simultaneously active in tumors with highly unstable genomes, which makes it difficult to causally link the presence of specific mutational signatures to DNA repair deficiency.

Organoid technology allows for the long-term in vitro expansion of epithelial tissues, starting from a single adult stem cell. Such organoids remain genetically stable over long periods of time (6). We have previously shown that clonal organoid cultures can be used for in-depth pattern analysis of mutations that accumulate throughout life in tissue-specific adult stem cells (7, 8). Moreover, organoids can be readily modified using CRISPR-Cas9 genome editing (9). Mammalian species differ greatly in their DNA repair capacity and will thus respond differently to mutagenic stress (10, 11), and mutation types and numbers will vary over a lifetime (7). We used CRISPR-Cas9 genome editing in human intestinal organoid cultures to systematically decipher the mutational consequences of DNA repair deficiency.

To establish a tool for dissecting mutational signatures in human organoid cultures, we introduced loss-of-function mutations in the MMR gene *MutL homolog 1 (MLH1)*. Inactivating mutations in MMR genes, including *MLH1*, predispose people to colorectal cancer (12, 13). These tumors are characterized by an immense mutational load, such as base substitutions and small insertions and deletions (INDELS) at short repeat sequences in the genome, referred to as micro-satellite instability (MSI) (13). We used organoids derived from human normal colon epithelium, because this allowed us to study the effect of disruption of single DNA repair genes in an otherwise normal genetic background. Moreover, human colonic organoids are as close as possible to the cell-of-origin of colorectal cancer (14). To inactivate *MLH1* in normal human colonic organoids, we used CRISPR-Cas9 technology to insert a puromycin-resistance cassette into the second exon of the *MLH1* gene (Fig. 1A and fig. S1A). After puromycin selection, we clonally expanded single organoids and subsequently genotyped these to confirm correct biallelic targeting (figs. S1, A and B, and S2A). Gene inactivation was verified by quantitative reverse transcription polymerase chain reaction (qRT-PCR), revealing a substantial reduction in *MLH1* mRNA expression in *MLH1* knockout (*MLH1*<sup>KO</sup>) organoids, most likely due to the degradation (through nonsense-mediated decay) of nonsense mRNAs transcribed from the mutant alleles (Fig. 1C). Western blot analysis confirmed the loss of MLH1 protein expression in the *MLH1*<sup>KO</sup> organoids (Fig. 1E).

Next, we passaged the *MLH1*<sup>KO</sup> and parental normal human colon organoids for 2 months to allow cells to accumulate sufficient mutations required for downstream analyses (fig. S1C). We then used flow cytometry to establish subclonal cultures of single cells (fig. S1C) and expanded these until sufficient DNA could be obtained (7). Both clonal and subclonal

cultures were subjected to whole-genome sequencing (WGS) analyses to identify the mutations that accumulated between the two clonal expansion steps and to correlate their appearance with time (Fig. 2A). As expected, *MLH1*<sup>KO</sup> organoids showed an increased base substitution load ( $27.7 \pm 4.9$  mutations per genome per day) compared with normal organoids ( $3.8 \pm 1.2$  mutations per genome per day), as well as a change in type of base substitutions (increase in C>T transitions). Similarly, the *MLH1*<sup>KO</sup> organoids displayed an increased number of INDELs (Fig. 2A), which were predominantly single-base-pair deletions (Fig. 2B) at mononucleotide repeats (Fig. 2C). This confirmed that deletion of *MLH1* is sufficient to generate the mutator phenotype observed in MMR-deficient tumors. As expected, we did not observe an increase in structural variations.

Somatic mutations are unevenly distributed throughout the genomes of most cancers (15, 16) and normal adult stem cells (7). In general, base substitutions are more frequent in heterochromatic and late-replicating regions of the genome. It has been reported that specific DNA repair activity, including MMR (17) and NER (18), underlies this variation in regional mutation rates, as tumors lacking essential components of either of these DNA repair pathways show a less biased genomic distribution. To directly test this, we performed a genome-wide analysis of replication timing (Repli-seq) (19) on human intestinal organoids. We defined early, intermediate, and late replicating genomic regions and determined the relationship between somatic mutational load and replication timing. As observed in most cancers (16), in vitro-accumulated mutations in normal organoids were more frequent in late-replicating DNA (Fig. 3A). This bias was no longer present in *MLH1*<sup>KO</sup> cells, in line with the notion that MMR might be more active in euchromatic early-replicating regions of the genome, thereby creating variation in regional mutation rates (17). However, when we compared the absolute mutational loads in the genomic regions with different replication timing, we found that all regions (not only euchromatic early-replicating regions) showed increased mutation numbers (Fig. 3A). Thus, additional mutagenic processes might be active in the *MLH1*<sup>KO</sup> cells that increase mutational load throughout the genome, thereby removing the bias in genomic distribution.

By removing part of the newly synthesized strand, MMR is able to repair errors—such as base-base mismatches and insertion and/or deletion loops at simple repeats—that are introduced by DNA replication polymerases (20). *MLH1*<sup>KO</sup> cells are characterized by increased levels of both base substitutions and INDELs at simple repeats (Fig. 2). An imbalance in mutation incorporation can arise between the leading and lagging strands during DNA replication, because different DNA polymerases are used with distinct fidelities and proofreading capacities. Additionally, the lagging strand is exposed longer as single-stranded DNA, which is chemically less stable than double-stranded DNA and may thus be more vulnerable to accumulate damage (21). To define leading and lagging strands to test for replication strand asymmetry, we determined the location of replication origins using the Repli-seq data. We observed a significant replication strand asymmetry in *MLH1*<sup>KO</sup> cells for C>A, C>G, C>T, and T>C substitutions ( $P < 0.05$ , two-sided Poisson test) that was not observed in normal cells (Fig. 3B). MSI colorectal tumors, typically resulting from damage to the MMR system, show similar replicative asymmetry (21). Without functional MMR, the DNA polymerase errors cannot be resolved. Therefore, the additional mutagenic process that is active throughout the entire genome of *MLH1*<sup>KO</sup> cells may represent stochastic mistakes

by DNA replication polymerases, which were previously suggested to be important in mutation accumulation in human cancers (22).

We next extracted mutational signatures (fig. S3) and compared them to those described in the Catalogue of Somatic Mutations in Cancer (COSMIC) database, using cosine similarity as a measure of closeness (1, 23). As we previously reported (7), normal organoids typically show a contribution of a mutational signature that resembles COSMIC signature 18 (cosine similarity = 0.870), which is characterized by C>A transversions. Recently, a similar signature was described to be associated with 8-oxoguanine mutagenesis (24). In contrast, *MLH1*<sup>KO</sup> organoids were characterized by the predominant occurrence of a signature that resembles COSMIC signature 20 (cosine similarity = 0.792), suggesting that this signature reflects mistakes made by polymerases during normal DNA replication. As human intestinal stem cells divide approximately once every day (fig. S4), we estimate that ~25 replication errors occur per cell division, which are largely resolved by MMR in normal cells (Fig. 4A). Next, we determined the cosine similarity between the mutational profile of each sample and each COSMIC signature, which reflects how well the mutational profile of a sample can be explained by each signature individually. We included genome-wide mutation data obtained by analysis of individual cancer cells (expanded as clonal organoids before sequencing) from colon cancers derived from three different patients (materials and methods). One of these cancers is driven by *MLH1* promoter hypermethylation and therefore lacks MMR activity, whereas the other two cancers are MMR proficient. Mutation accumulation in our *MLH1*<sup>KO</sup> organoids closely resembled that of MMR-deficient cancer cells and was markedly different from MMR-proficient cancer cells (Fig. 4B).

As our organoid-targeting strategy allowed us to define a very clean mutational signature for MMR deficiency, we next set out to determine signatures caused by deficiency in base excision repair (BER), the role of which in generating cancer-causing mutations is less well established than that of MMR deficiency. To this end, we inactivated the BER gene *NTHL1*, which encodes a DNA glycosylase that is involved in the removal of oxidized pyrimidines through initiation of the BER pathway (25, 26) (Fig. 1B and figs. S1A and S2B). Germline homozygous mutations in *NTHL1* were recently shown to cause adenomatous polyposis and colorectal cancer (27). Upon gene targeting, we confirmed the loss of *NTHL1* expression by qRT-PCR analysis (Fig. 1D) and Western blotting (Fig. 1F). Because we expected a lower mutation frequency with BER deficiency as compared with MMR deficiency (27), we cultured the *NTHL1*<sup>KO</sup> clones for 3 months and subsequently generated subclonal cultures and used WGS to analyze the mutations that specifically accumulated between the two clonal expansion steps (fig. S1C). As expected, the base substitution accumulation rate in *NTHL1*<sup>KO</sup> cells was approximately one-fourth of that in *MLH1*<sup>KO</sup> cells but approximately two times higher than in normal cells (Fig. 2A). Loss of *NTHL1* did not result in increased numbers of INDELS (Fig. 2A) or structural variation.

In contrast to *MLH1*<sup>KO</sup> cells, *NTHL1*<sup>KO</sup> cells retained a nonrandom distribution of mutations throughout the genome, as observed for normal cells (Fig. 3A). Signature 30, characterized by C>T transitions, was the main contributor to the mutation spectrum observed in *NTHL1*<sup>KO</sup> cells (Fig. 4A and fig. S3C). Of note, somatic mutation analysis in the *NTHL1*<sup>KO</sup> clones did not reveal any nonsynonymous or stop-gain mutations in other

DNA repair genes (table S1), supporting the notion that the observed change in mutation accumulation can be solely attributed to *NTHL1* deficiency. In agreement with the observations on the *NTHL1*<sup>KO</sup> clones (Fig. 2A), it was reported previously that colorectal cancers from individuals with biallelic *NTHL1* germline mutations predominantly show C>T transitions in their exomes (27, 28). Signature 30 has previously been identified in one patient with breast cancer (PD13297a) analyzed by WGS (29). We examined tumor and germline sequences of this patient and identified a germline nonsense mutation in *NTHL1* (Fig. 4C), compounded by loss of heterozygosity in the tumor. This observation further corroborates the link between *NTHL1* deficiency and signature 30. Mutation accumulation in our *NTHL1*<sup>KO</sup> organoids closely resembled that in PD13297a (Fig. 4B).

We have shown that mutational signatures can be dissected by characterizing the genomic landscapes of genetically modified human organoid subclones. Engineered *MLH1*<sup>KO</sup> organoids validated our approach, as they accurately model the predominant mutation profiles observed in MMR-deficient colorectal cancers. We subsequently used our approach to demonstrate that a high contribution of signature 30 mutations within a tumor can be indicative of cancer-predisposing germline mutations in the base excision repair gene *NTHL1*. A similar strategy could be exploited to investigate the consequences of other BER gene deficiencies. Although signature 30 has been previously identified in one patient of a large breast cancer cohort, it may be indicative of predisposition to a broader range of cancer types. *NTHL1* germline mutations appear to predispose people to multiple cancer types, including colorectal and breast cancer (27, 28). The strategy we have described can be used to study the mutational consequences of DNA repair knockouts or mutagen exposure, to systematically dissect mutational signatures and potentially unveil their molecular origins.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank M. R. Stratton for scientific discussion and for reading the manuscript. We also thank the members of the contributing labs for helpful suggestions and discussions. This work was supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)–ZonMw Veni grant to J.D. (91614138), a grant from Worldwide Cancer Research to R.v.B. (16-0193), funding from the Wellcome Trust and St. Baldrick's Foundation to S.B., “Sta op tegen Kanker” International Translational Cancer Research grant to T.M. and H.C., and funding from the CancerGenomics.nl (NWO Gravitation) program to E.C. and H.C. S.N.-Z. has a paid consulting relationship with Artios Pharma. S.N.-Z. is an inventor on patent applications (1607629.1, 1607628.3, 1607630.9, and 1607635.8) submitted by the Wellcome Trust Sanger Institute that cover the applications of mutational signatures as a biomarker in clinical contexts. The sequencing data have been deposited to the European Genome-Phenome Archive ([www.ebi.ac.uk/ega/](http://www.ebi.ac.uk/ega/)) under accession numbers EGAS00001001969, EGAS00001001682, and EGAS00001000881.

## References and Notes

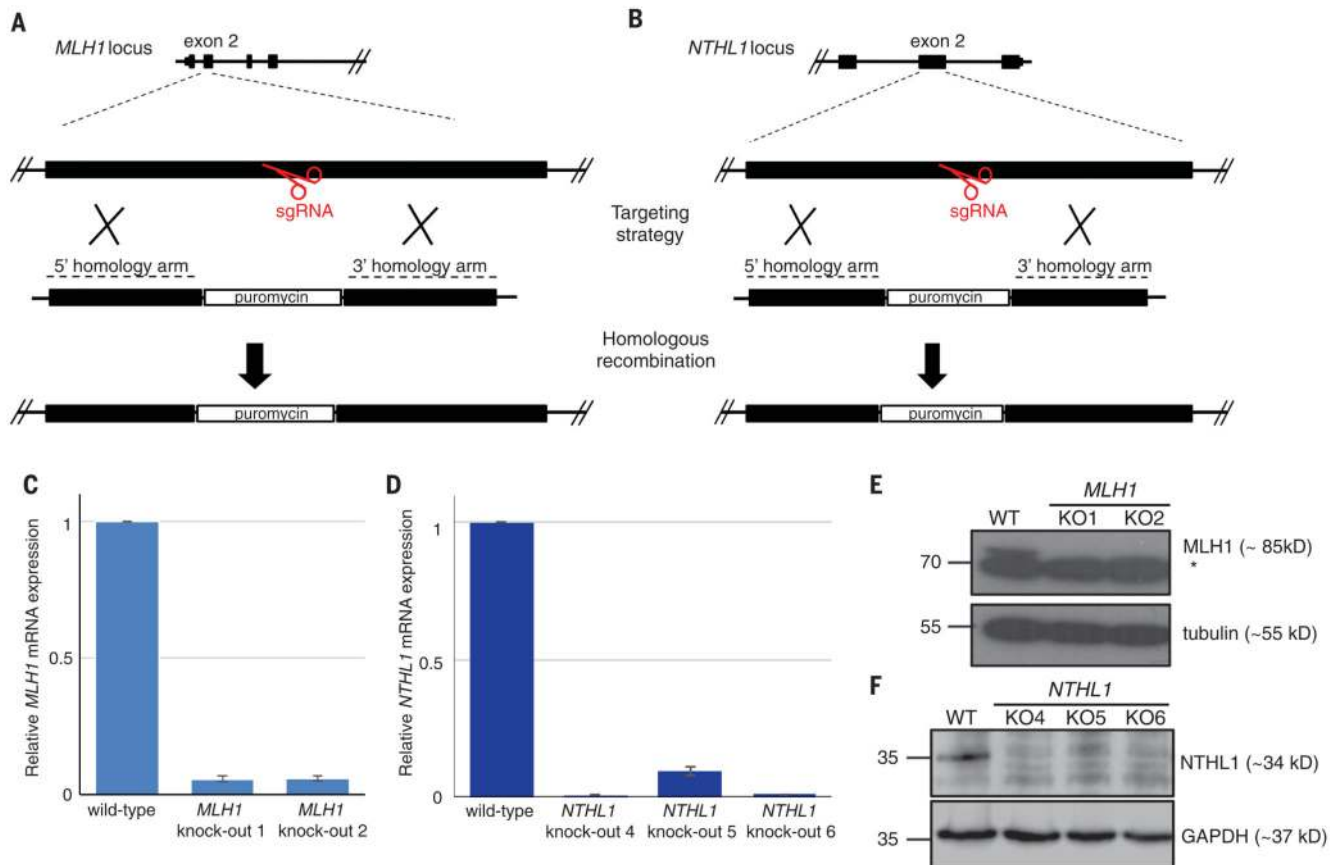
1. Alexandrov LB, et al. *Nature*. 2013; 500:415–421. [PubMed: 23945592]
2. Davies H, et al. *Nat Med*. 2017; 23:517–525. [PubMed: 28288110]
3. Alexandrov LB, et al. *Science*. 2016; 354:618–622. [PubMed: 27811275]
4. Behjati S, et al. *Nat Commun*. 2016; 7:12605. [PubMed: 27615322]
5. Kim J, et al. *Nat Genet*. 2016; 48:600–606. [PubMed: 27111033]

6. Huch M, et al. *Cell*. 2015; 160:299–312. [PubMed: 25533785]
7. Blokzijl F, et al. *Nature*. 2016; 538:260–264. [PubMed: 27698416]
8. Behjati S, et al. *Nature*. 2014; 513:422–425. [PubMed: 25043003]
9. Drost J, et al. *Nature*. 2015; 521:43–47. [PubMed: 25924068]
10. MacRae SL, et al. *Aging*. 2015; 7:1171–1182. [PubMed: 26729707]
11. Parrinello S, et al. *Nat Cell Biol*. 2003; 5:741–747. [PubMed: 12855956]
12. Lynch HT, de la Chapelle A. *N Engl J Med*. 2003; 348:919–932. [PubMed: 12621137]
13. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. *Nat Rev Cancer*. 2015; 15:181–194. [PubMed: 25673086]
14. Barker N, et al. *Nature*. 2009; 457:608–611. [PubMed: 19092804]
15. Peltomäki P. *Hum Mol Genet*. 2001; 10:735–740. [PubMed: 11257106]
16. Schuster-Böckler B, Lehner B. *Nature*. 2012; 488:504–507. [PubMed: 22820252]
17. Supek F, Lehner B. *Nature*. 2015; 521:81–84. [PubMed: 25707793]
18. Zheng CL, et al. *Cell Rep*. 2014; 9:1228–1234. [PubMed: 25456125]
19. Hansen RS, et al. *Proc Natl Acad Sci USA*. 2010; 107:139–144. [PubMed: 19966280]
20. Jiricny J. *Nat Rev Mol Cell Biol*. 2006; 7:335–346. [PubMed: 16612326]
21. Haradhvala NJ, et al. *Cell*. 2016; 164:538–549. [PubMed: 26806129]
22. Tomasetti C, Li L, Vogelstein B. *Science*. 2017; 355:1330–1334. [PubMed: 28336671]
23. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. *Cell Rep*. 2013; 3:246–259. [PubMed: 23318258]
24. Viel A, et al. *EBioMedicine*. 2017; 20:39–49. [PubMed: 28551381]
25. Prasad A, Wallace SS, Pederson DS. *Mol Cell Biol*. 2007; 27:8442–8453. [PubMed: 17923696]
26. Li J, Braganza A, Sobol RW. *Antioxid Redox Signal*. 2013; 18:2429–2443. [PubMed: 23311711]
27. Weren RD, et al. *Nat Genet*. 2015; 47:668–671. [PubMed: 25938944]
28. Rivera B, Castellsagué E, Bah I, van Kempen LC, Foulkes WD. *N Engl J Med*. 2015; 373:1985–1986. [PubMed: 26559593]
29. Nik-Zainal S, et al. *Nature*. 2016; 534:47–54. [PubMed: 27135926]

### A signature event for organoids

Human cancer genomes harbor cryptic mutational signatures that represent the cumulative effects of DNA damage and defects in DNA repair processes. Knowledge of how specific signatures originate could have a major impact on cancer diagnosis and prevention. One approach to address this question is to reproduce the signatures in experimental systems by genetic engineering and then match the signatures to those found in naturally occurring cancers. Drost *et al.* used CRISPR-Cas9 to delete certain DNA repair enzymes from human colon organoids. In a proof-of-concept study, they show that deficiency in base excision repair is responsible for a mutational signature previously identified in cancer genome sequencing projects.

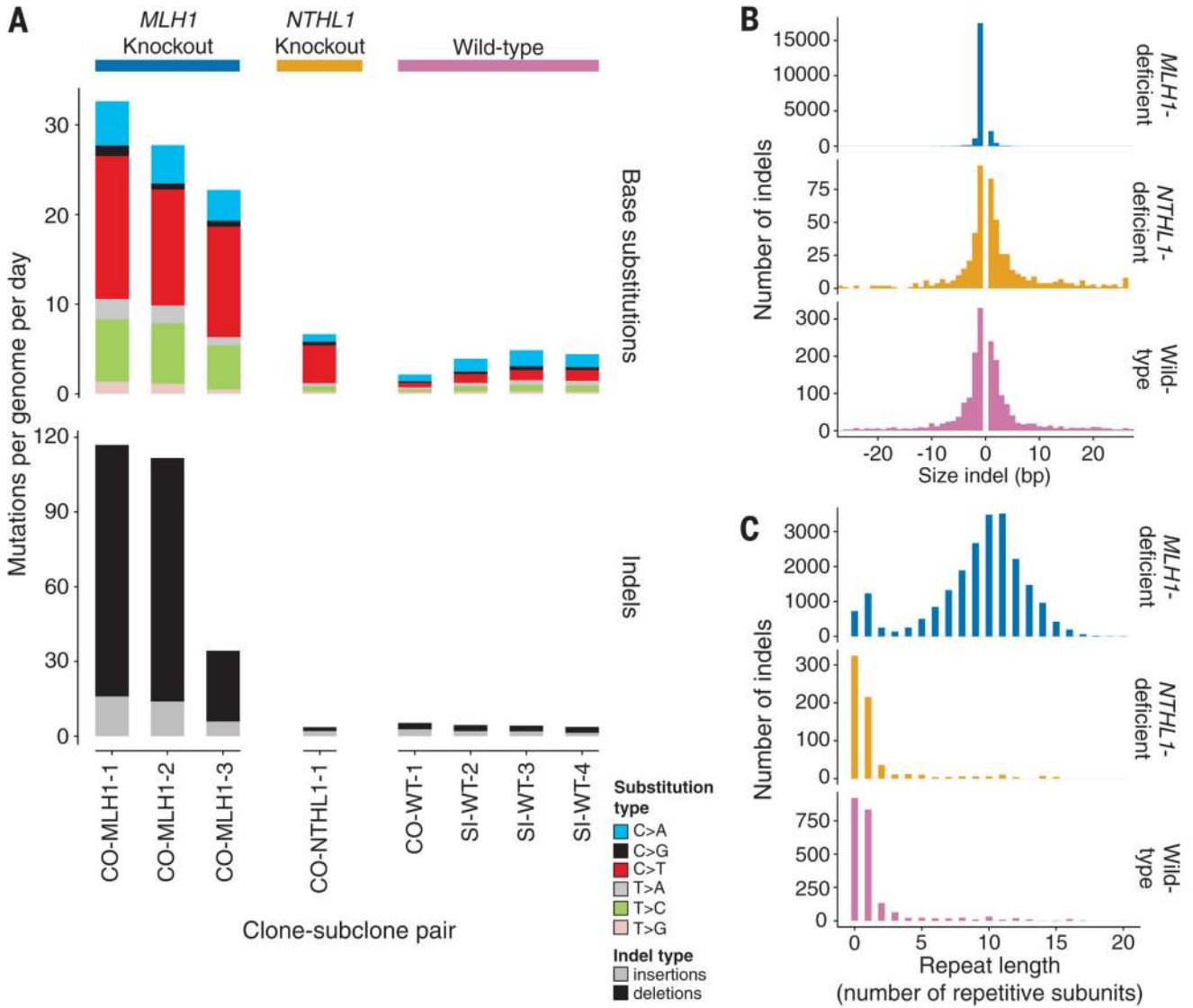
*Science*, this issue p. 234



**Fig. 1. Generation of DNA repair gene knock-outs in human intestinal stem cell cultures.**

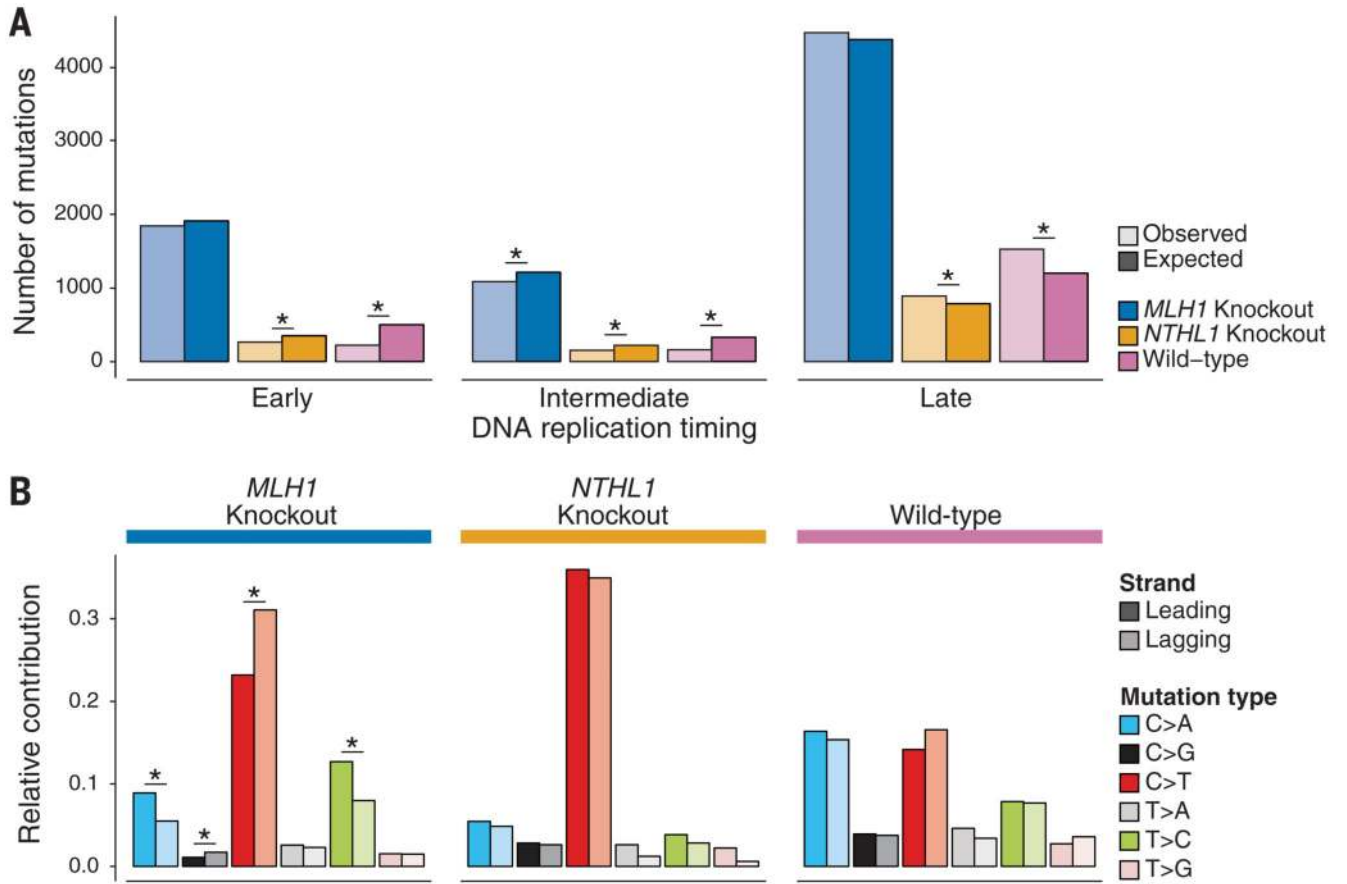
Targeting strategy for the generation of *MLH1* (A) and *NTHL1* (B) knockout organoids using CRISPR-Cas9 genome editing. sgRNA, single guide RNA. (C) qRT-PCR for *MLH1* in normal and *MLH1*<sup>KO</sup> organoids. Expression was normalized to *GAPDH*. Mean and SD (error bars) of  $n = 3$  independent experiments are indicated. (D) Same as in (C), but for *NTHL1*. (E) Western blot analysis of *MLH1* expression in normal and *MLH1*<sup>KO</sup> organoids (representative from  $n = 3$ ). Tubulin was used as a loading control. The asterisk indicates a background band. (F) Same as in (E), but for *NTHL1*. Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) was used as a loading control.



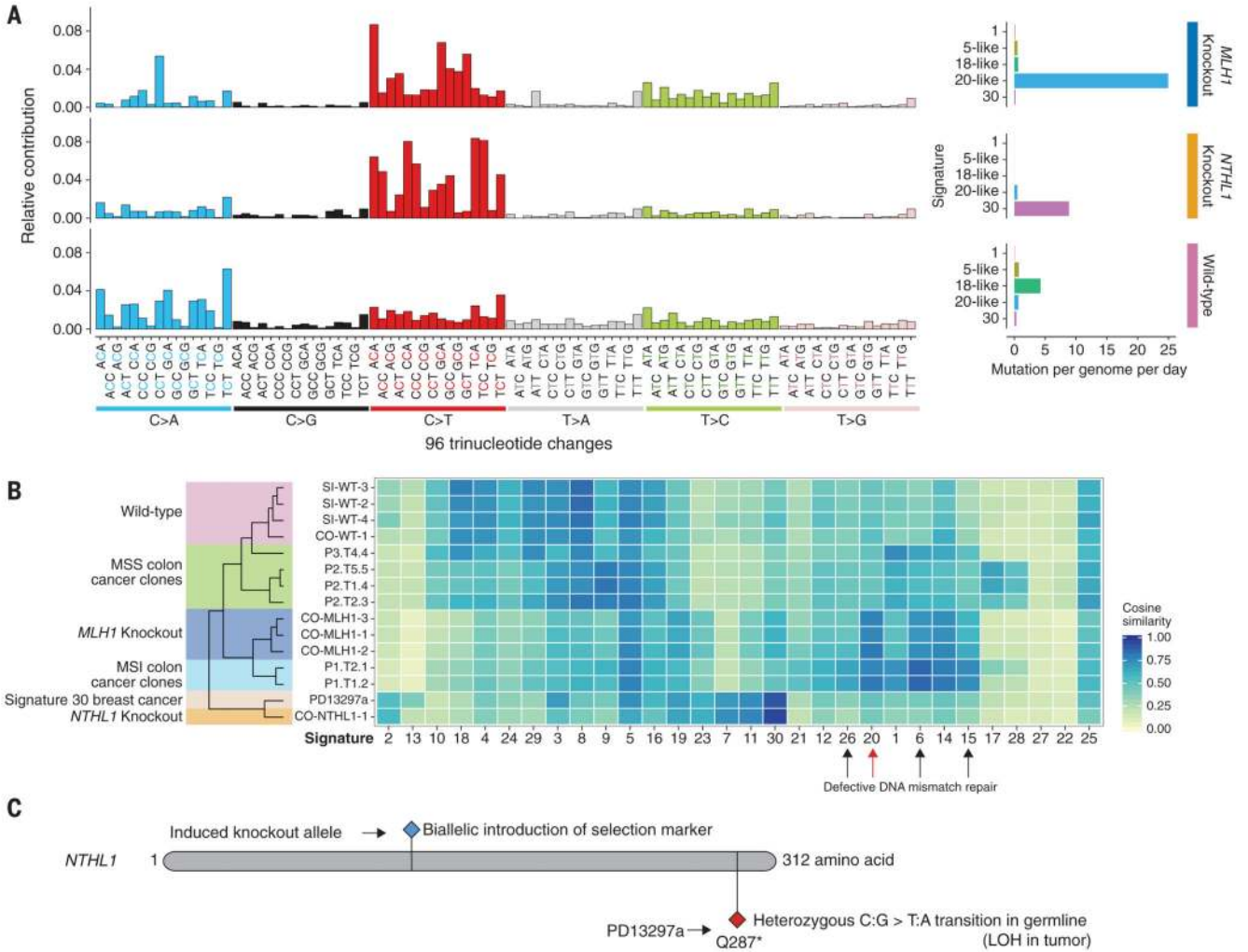


**Fig. 2. Mutational burden in DNA repair-deficient human colonic stem cells.**

(A) Number of mutations accumulated in the absence of the indicated DNA repair proteins per day. Base substitutions subdivided by mutation type and INDELS are shown. (B) Size distribution of the observed INDELS per genotype. A negative value indicates deletions and a positive value indicates insertions. (C) Number of INDELS located in simple repeats per genotype. Indicated are the number of repetitive subunits surrounding an inserted or deleted subunit. A value of 0 indicates that the INDEL is not located within a simple repeat.



**Fig. 3. Nonrandom genomic distribution of base substitutions in DNA repair-deficient organoids.** (A) Shown for each genotype are enrichment and depletion of base substitutions in the genomic regions that are replicated at the indicated stages during S phase of the cell cycle. Asterisks indicate a significant enrichment or depletion ( $P < 0.05$ , one-sided binomial test). (B) Relative levels of each base substitution type in the leading and lagging DNA strands are shown for each genotype. Asterisks indicate a significant difference ( $P < 0.05$ , two-sided Poisson test).



**Fig. 4. Signatures of mutational processes in DNA repair-deficient organoids.**

(A) (Left) Mutational spectra of all base substitutions observed for each genotype. Different mutation types and the direct sequence context are indicated. (Right) Number of mutations per genome per day that can be explained by the indicated mutational signatures for each genotype. (B) Heat map showing the cosine similarity scores for each indicated sample and COSMIC signature. The samples have been clustered according to the similarity score with each signature. The signatures have been ordered according to their similarity, such that very similar signatures cluster together. Arrows indicate signatures that have been associated with deficiency in DNA MMR in pan-cancer analyses (1). MSI, microsatellite instability. (C) Mutations that have been introduced or identified in *NTHL1*. The blue diamond indicates the site where the selection marker was introduced by gene targeting in the organoids. The red diamond denotes the nonsense germline mutation that was identified in a patient with breast cancer (PD13297a). LOH, loss of heterozygosity.