Supplementary file

Mutation rates in normal cells. To investigate whether the expected increase of mutations with age continued throughout life, even in very old individuals, we evaluated the literature on this topic (1-19). The most extensive studies of the relationship between aging and mutations in normal cells have been performed in those of the hematopoietic system (1-5, 7-10, 14, 15, 17, 18). These studies generally concluded that the number of mutations linearly increase with age, with estimates for mutation rates ranging from 0.04 to 0.26 per 10^6 cells per year (see Cole *et al.* (9) for a comprehensive review). Of note, many of these studies *imposed* a linear relationship between age and mutation number *a priori* when analyzing the data. We reanalyzed the original data, when available, without imposing any such linear relationship. When analyzed by local regression, the relationship between the number of mutations and age appeared to be non-linear, with the rate of increase decreasing at older ages (Fig. S1).

Demonstrating statistical significance of a change in slope of curves like those shown in Fig. S1 requires a large number of data points. For this reason it was only possible to perform a test with reasonable power in the dataset provided in Branda *et al.* (7) (Fig. S1B and S3D), where the difference of the slope from a constant (linearity) was not statistically significant ($P<0.41$). Individuals in the 38 to 60 year age group had a significant median increase of $2.6 \cdot 10^{-6}$ *hprt* mutations per cell ($0.6 \cdot 10^{-6} – 4.8 \cdot 10^{-6}$; 95% CI) with respect to individuals in the zero to 37 year age group ($P=0.013$). Normal T-cells of 61 to 65 year-old individuals had a further significant increase of $5.5 \cdot 10^{-6}$ in mutation frequency ($2.8 \cdot 10^{-6} – 8.2 \cdot 10^{-6}$; 95% CI) with respect to individuals in the 38 to 60 year

age group (P=0.00016). The mutation frequency then appeared to stabilize. There was a non-significant (P=0.072) decrease of $2.8 \cdot 10^{-6}$ ($-5.6 \cdot 10^{-6}$ – $0.4 \cdot 10^{-6}$; 95% CI) in the oldest subgroup (66 and older) compared to the 61 to 65 year-old subgroup.

Other studies not depicted in Fig. S1 provided further support for the deceleration; see Fig. 3 of Cole *et al.* (8) and Fig. 1 of Trainor *et al.* (17), where the non-linearity is apparent but not commented upon. Additionally, a recent study on the accumulation of mutations in the exome of normal hematopoietic stem cells of seven healthy human donors is consistent with the proposed deceleration (19). Though the slowing down was not commented upon and the results not statistically significant, the average number of somatic mutations in the normal stem cells of individuals in their 30's, 40's, 50's and 70's was 2, 5.33, 8.33, and 10, respectively (19).

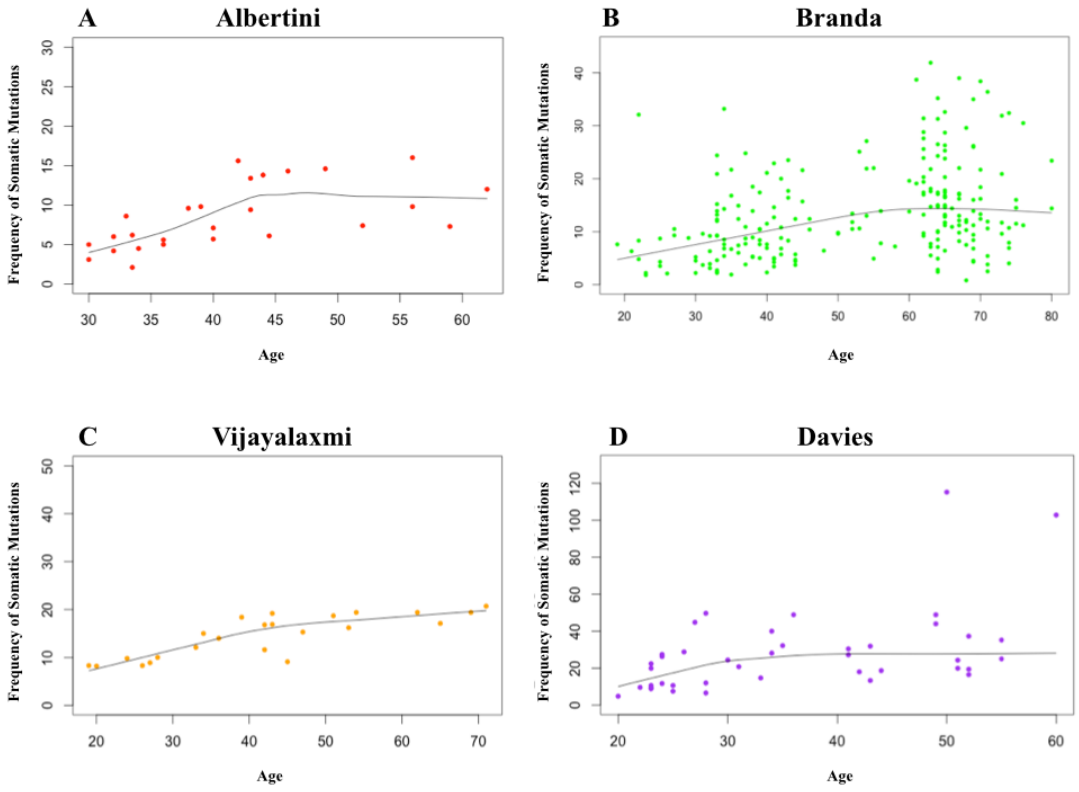

Fig. S1. The nonlinear relationship between the frequency ($x10^{-6}$) of somatic mutations and the age of a person in normal blood cells (A) Albertini *et al.* (2). (B) Branda *et al.* (7). (C) Vijayalaxmi *et al.* (18). (D) Davies *et al.* (10). The black lines represent the average frequency at a given age, as found by local regression.

Mutation rates from cancers in self-renewing tissues. In self-renewing tissues, sequence data derived from *tumor* DNA can be used to assess the accumulation of somatic mutations in *normal* stem cells (16). The reason for this is that cancers are initiated by the first driver gene mutation that occurs in the normal stem cell precursor of the cancer. The cell of origin of cancer may or may not be a stem cell (20-22), but even if not, the cancer precursor cell will be only a few divisions away from its stem cell mother, and thus share the stem cell's mutational load. The older the patient, the more passenger mutations the stem cell precursor would have accumulated prior to that first driver gene mutation. This reasoning is consistent with the fact that the vast majority of somatic mutations in common cancers are passengers rather than drivers (23-26).

We analyzed three whole-exome sequencing datasets publicly available on The Cancer Genome Atlas (TCGA) website: head and neck squamous cell carcinoma (HNSC, 306 patients), kidney clear cell renal carcinoma (KIRC, 184 patients), and colorectal cancer (CRC, 224 patients)(27). Data on chronic lymphocytic leukemia (CLL – 109 patients) are available on the International Cancer Genome Consortium (ICGC) website. These tumor types were chosen because they originate from self-renewing tissues, exome sequencing information was available from a large number of cases of each type, and each cancer type occurred in patients of many different ages.

We had previously analyzed two of these tumor types (CRC and CLL), assuming that mutations increased over time in a linear fashion (16). We reanalyzed by local regression the CRC and CLL datasets, and found that the rate decelerated at older age (Fig. S2A and D). The same trend was found for the two other tumor types (Fig. S2C and D).

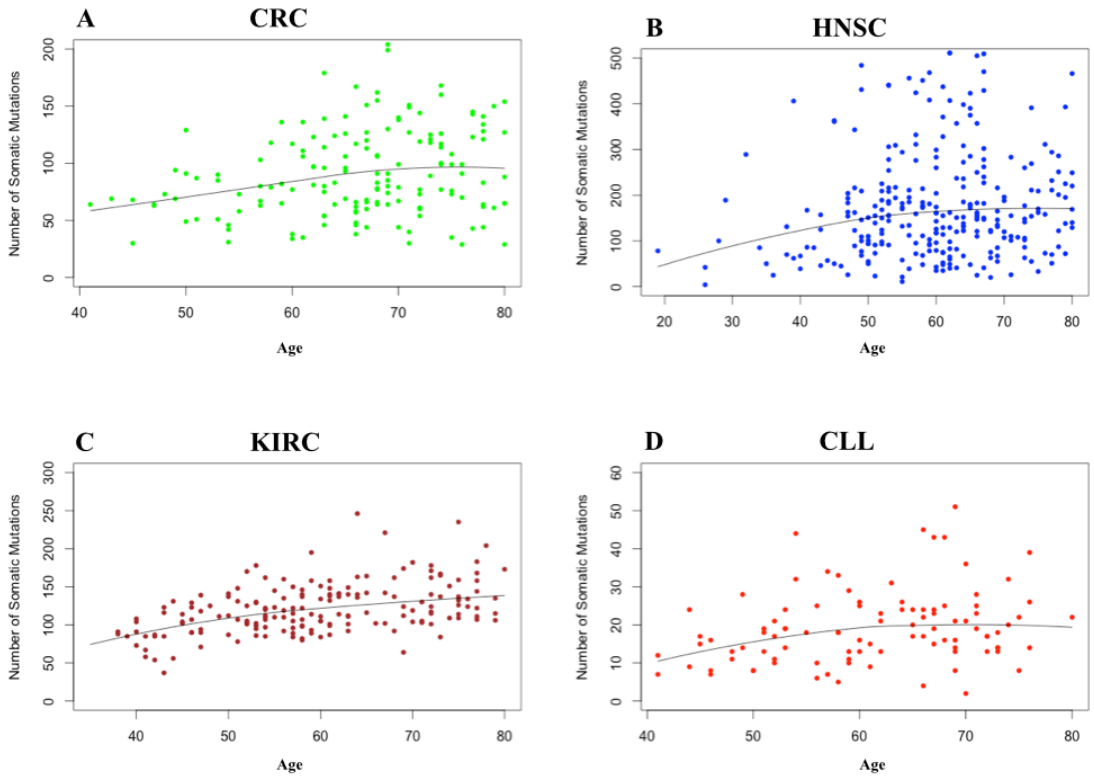

Fig. S2. The nonlinear relationship between the number of somatic mutations and the age at cancer diagnosis in four different cancer types. (A) Colorectal cancers (CRC). (B) Head and neck squamous cell carcinomas (HNSC). (C) Kidney clear cell renal carcinomas (KIRC). (D) Chronic lymphocytic leukemias (CLL). The black lines represent the average number of mutations at a given age, as found by local regression.

Interestingly, the magnitude of this aging effect and its timing varied with the specific tissue type. For example, the decrease in slope occurred at a later age in CRCs than in the other three tumors. In CRCs, there was a rather constant slope until about 70 years of age

when the slope flattened, with an estimated decrease of about 42% thereafter (Fig. S2A and see analysis below). In the other three tumor types, the deceleration appeared to begin between 50 and 60 years of age (Fig. S2B to D). There was no correlation between the stage of cancer and its age of onset in any tumor type. The linear increase found in (16) for uterine corpus endometrioid carcinoma (UCEC), did not appear to decelerate with age.

Given the smaller number of points in the CLL dataset, it was only possible to perform a test with reasonable power in the CRC, HNSC, and KIRC datasets. The difference of the slope from a constant (linearity) was not statistically significant: CRC ($P<0.90$), KIRC ($P<0.13$), and HNSC ($P<0.14$).

To estimate the median increase in mutations with respect to age, patients were divided into age subgroups corresponding to the 0.25, 0.5, 0.75 quartiles of the age distribution in HNSC, KIRC, and CLL (Fig. S3). In CRC, due to the smaller effect size, the dataset was divided into only three age groups, according to the 0.33, and 0.67 quantiles.

In the HNSC dataset (Fig. S3A), tumors from patients in the 53 to 60 year age group had 54 more mutations (20 – 88; 95% CI) than tumors from patients in the zero to 52 year age group ($P=0.001$). The number of mutations then appeared to stabilize, with a non-significant ($P=0.77$) increase of only 6 mutations (-33 – 54; 95% CI) in the subsequent 61 to 68 year age subgroup, and a non-significant decrease of 16 mutations in the oldest age subgroup (-57 to 23; 95% CI). These results are readily apparent in Fig. S2, where the trend towards deceleration can be observed at a rather young age in HNSC patients.

KIRCs from patients in the 52 to 58 year age group (Fig. S3B) had a median of 14 more mutations (3 – 25; 95% CI) than KIRCs from patients in the zero to 51 year age group

(P=0.02). Tumors from patients in the 59 to 69 year age group had a median of 13 more mutations (1 – 25; 95% CI) than tumors from patients in the 52 to 58 year age group (P=0.038). However, the increase then slowed down, with tumors in patients from the oldest age group (70 and older) harboring a non-significant (P=0.18) median increase of 10 mutations (-4 – 23; 95% CI) compared to tumors in the 59 to 69 year age group.

In the CRC dataset (Fig. S3C), tumors from patients in the 65 to 74 year age group had a median of 14 more mutations (1 – 28; 95% CI) than tumors from patients in the zero to 64 year age group (P=0.0372). CRC tumors from patients in the oldest age group (75 and older) harbored a median of only one additional mutation (non-significant, -12 – 16; 95% CI; P=0.8) than found in patients in the 65 to 74 year age group.

In CRC, the deceleration in the accumulation of mutations (additional mutations per year of age) appeared to occur around age 70 (Fig. S2A). To estimate the change in the rate of accumulation, we divided the dataset in the following five age subgroups: 40-49, 50-59, 60-69, 70-79, 80-90, and considered the median 10-year increases observed in each subgroup with respect to the previous one. We obtained an estimated average increase of 0.95 mutations per year in the 40 to 69 year age group and 0.55 mutations per year in the 70 to 90 year age groups, indicating a 42% deceleration.

The analysis in the CLL dataset also indicated a trend toward a decrease in the oldest age group; CLL cells from patients in the 52 to 60 year age group had a median of 3 mutations more than patients in the zero to 51 year age group, and patients in the 61 to 68 year age group had 4 mutations more than patients in the 52 to 60 year age group. CLL cells from patients in the oldest age group (69 years or older) had no more mutations than patients in the 61 to 68 year age group.

Statistical analysis of mutations in the datasets analyzed. Local regression was used for Fig. S1 and S2 via the R function *scatter.smooth*. The regressions for Figure S1 were estimated using all data points within the indicated age intervals. Given the very small number of points in young individuals, and the known high incidence of mismatch repair deficiency in CRC patients younger than 50 years (28), CRC patient outliers were not considered (outliers were defined as patients in their 40's with >175 mutations or any patient with >300 mutations) in Fig. S2. The other regressions of Figure S2 were estimated using all data points within the indicated age intervals.

To determine the statistical significance of a change in slope with age (i.e. a difference of the slope from a constant), each dataset was divided into two age subgroups, based on the median age among patients in the dataset, and a dummy variable introduced in the regression model to allow for a second slope in the older age interval. The test was performed via the likelihood ratio test (with the R function lrtest) on the two regression models (one vs. two slopes).

To compare the median increase in the number of somatic mutations according to age, patients were divided into age subgroups corresponding to the 0.25, 0.5, 0.75 quartiles of the age distribution in HNSC, KIRC, and CLL. In CRC, the smaller effect size required division of the dataset into only three age groups, according to the 0.33, and 0.67 quantiles. To test the null hypothesis that no difference existed in the total number of

somatic mutations among patients of two neighboring age subgroups, a two-sided Wilcoxon-Mann-Whitney U test was used, incorporating its estimated 95% confidence interval for the shift in location. Boxplots were obtained via the *boxplot* R function.

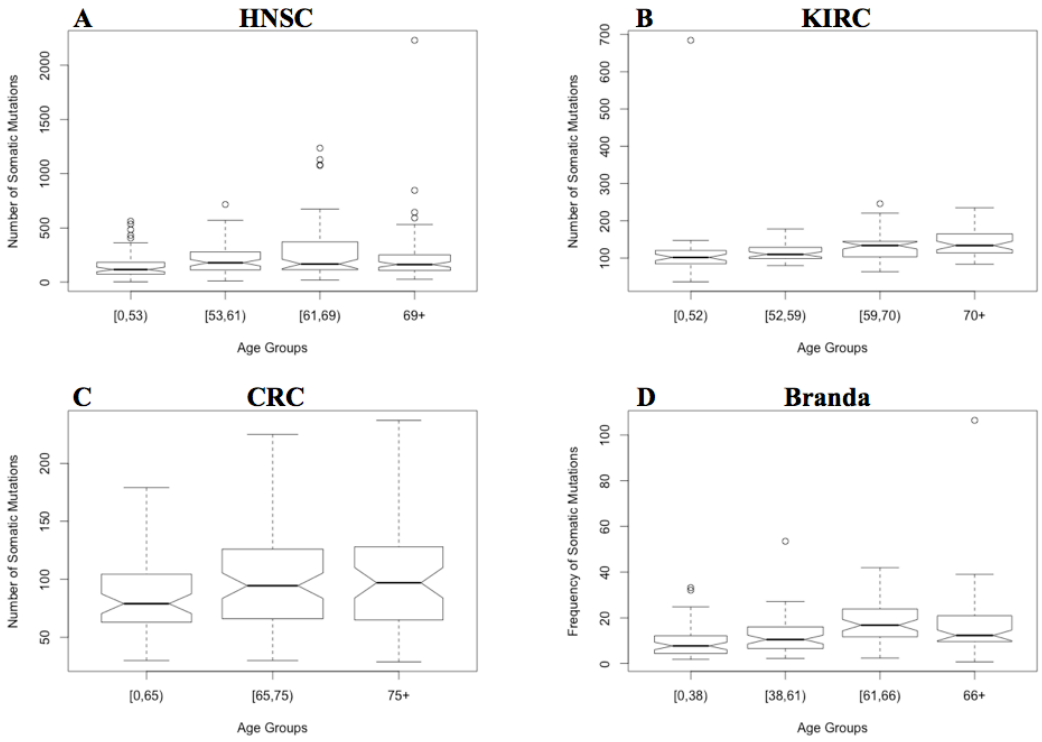

Figure S3. Boxplots for the number of somatic mutations in different age groups in four different datasets: (A) head and neck squamous cell carcinoma (HNSC), (B) kidney clear cell renal carcinoma (KIRC), (C) colorectal cancer (CRC), and (D) Branda et al. (7).

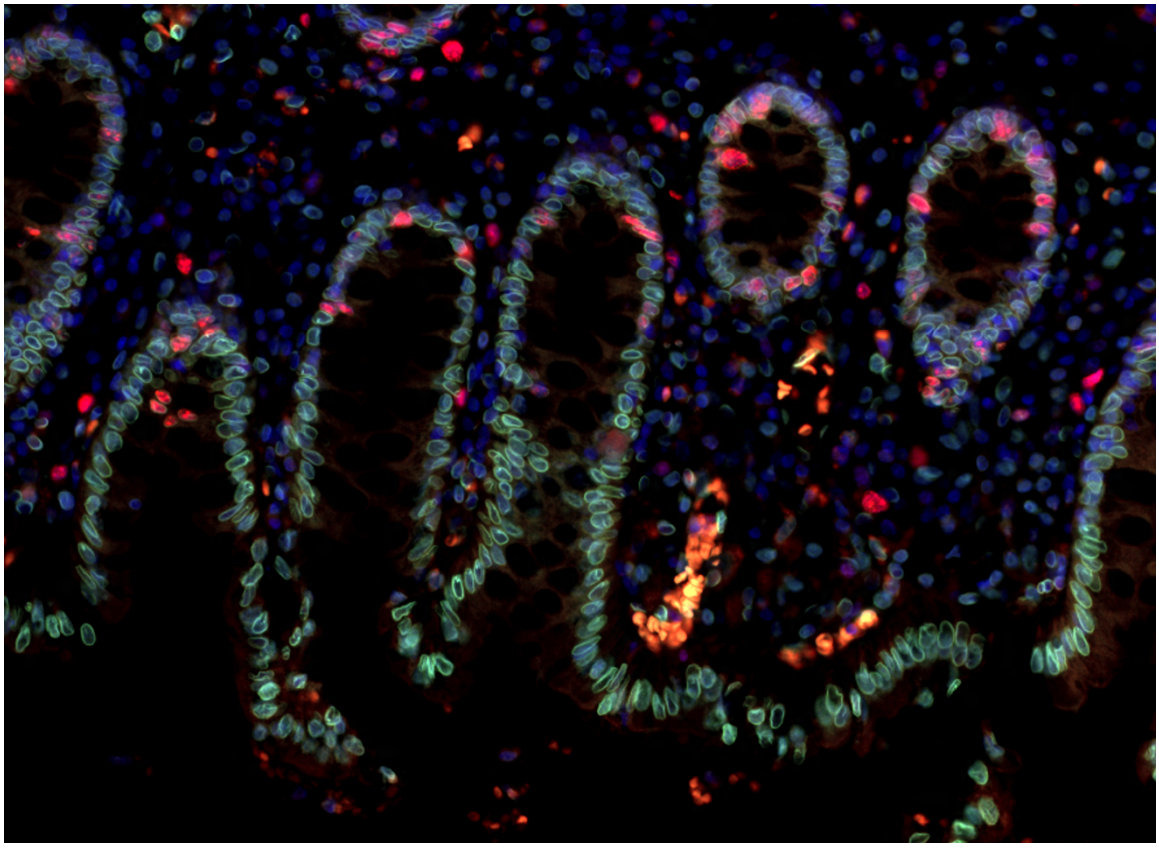

Figure S4. Ki67 plus Lamin A/C labeling of normal colon resections. Shown is an example of the combined Ki67 and Lamin A/C labeling in a colon sample. Color code: green = Lamin A/C, red = Ki67, and blue = nuclei (DAPI) (see Materials and Methods).

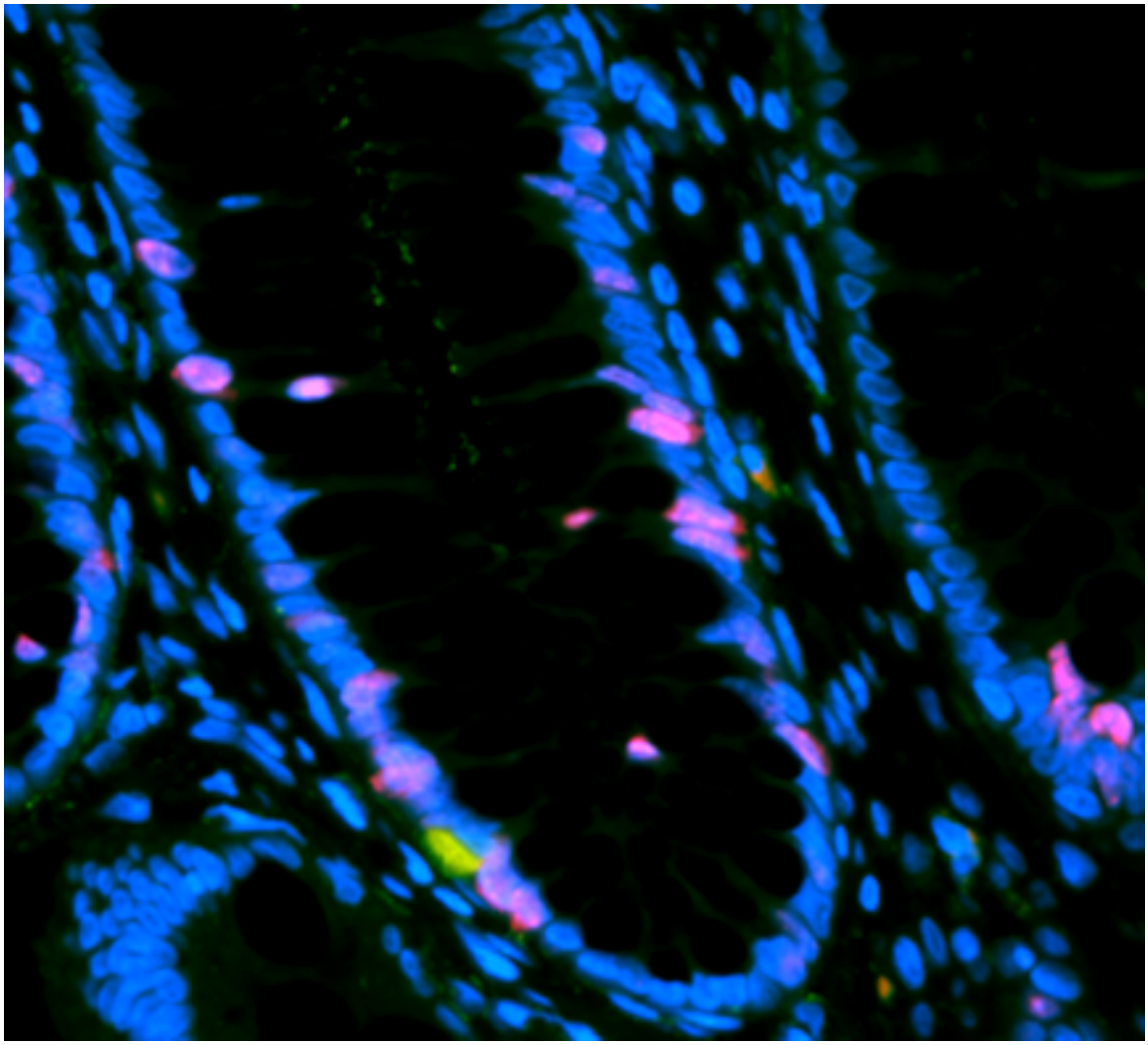

Figure S5. Ki67 plus Lgr5 labeling of normal colon resections. Shown is an example of the combined Ki67 and Lgr5 labeling in a colon sample. Color code: green = Lgr5, red = Ki67, and blue = nuclei (DAPI) (see Materials and Methods).

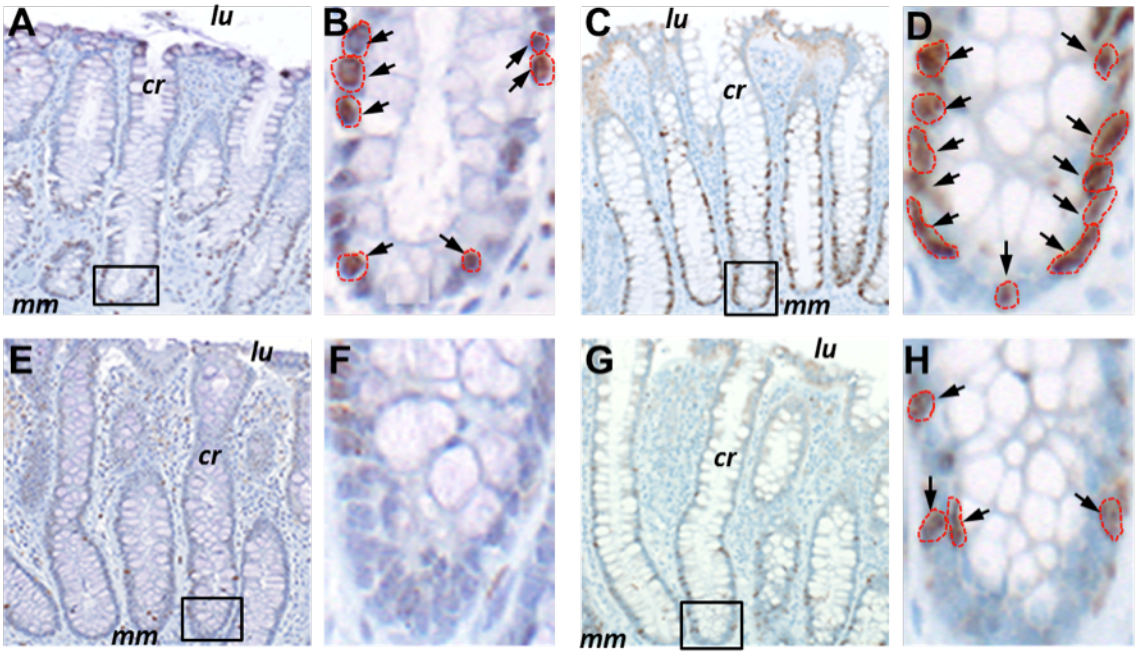

Figure S6. Comparison of pHH3 and Ki67 labeling of normal colon resections. Shown are examples of pHH3 (**A, B, E, F**) and Ki67 (**C, D, G, H**) labeling in a 26 year old individual (**A-D**) compared to an 80 year old individual (**E-H**). In each image, lu indicates the location of the colonic lumen, cr indicates a properly oriented crypt, and mm indicates the muscularis mucosae. The crypts outlined in (**A**) and (**C**) are magnified in (**B**) and (**D**), while the crypts outlined in (**E**) and (**G**) are magnified in (**F**) and (**H**). Arrows indicate antibody- labeled nuclei, which are outlined with red dashes; there were no labeled cells in (F). Images in (**A, C, E, G**) taken at 100x while those in (**B, D, F, H**) were taken at 400x.

Table S1. Measurements of Ki67 labeling of normal colon resections in Experiment 1A. The columns indicate, respectively, the case number, the age of the patient, the crypt number, the crypt height, the number of Ki67 positive cells within that crypt, and the positions of the positive cells within that crypt. Pathologist: Justin Poling.

Table S2. Measurements of Ki67 labeling of normal colon resections in Experiment 1A. The columns indicate, respectively, the area measured, the area occupied by Ki67, the case image, the age group of the patient, and the estimated fraction of Ki67 positive cells. Pathologist: Alex Baras.

Table S3. Measurements of Ki67 labeling of normal colon resections in Experiment 2. The columns indicate, respectively, the patient case number, the slide case number, the tissue type, the age of the patient, and the number of Ki67 positive cells within each crypt (one value per crypt; first eight positions on both sides of the crypt). Pathologist: Christine A. Iacobuzio-Donahue.

Table S4. Measurements of Ki67 labeling of normal colon resections in Experiment 3. The columns indicate, respectively, the set number, the patient case number, the region analyzed, the Ki67 counts (one value per crypt; first eight positions on both sides of the crypt), the average count per case, notes, and the age of the patient. Pathologist: Christine A. Iacobuzio-Donahue.

Table S5. Measurements of Ki67 labeling of normal esophagus resections in Experiment 1A. The columns indicate, respectively, the number of positive nuclei, the number of negative nuclei, the total number of nuclei, the case number, and the age of the patient. Pathologist: Meredith A. Pittman.

Table S6. Measurements of Ki67 labeling of normal esophagus resections in Experiment 1B. The columns indicate, respectively, the area measured, the area occupied by Ki67, the file name, the number of the case image, the estimated fraction of Ki67 positive cells, and the age of the patient. Pathologist: Alex Baras.

Table S7. Measurements of Ki67 labeling of normal esophagus resections in Experiment 2. The columns indicate, respectively, the image file name, the number of positive nuclei, the number of negative nuclei, the total number of nuclei, the proportion of Ki67 positive cells, and the age of the patient. Pathologist: Meredith A. Pittman.

Table S8. Measurements of Ki67 labeling of normal esophagus resections in Experiment 3. The columns indicate, respectively, the batch number, the patient case number, the age of the patient, and for each field analyzed the total number of nuclei and the total number of Ki67 positive nuclei. Pathologist: Christine A. Iacobuzio-Donahue.

Table S9. Measurements of Ki67 labeling of normal duodenum resections. The columns indicate, respectively, the patient case number, the age of the patient, the villus number, the villus height, the number of Ki67 positive cells within that villus, and the positions of the positive cells within that villus. Pathologist: Justin Poling.

Table S10. Measurements of Ki67 labeling of normal posterior ethmoid resections. The columns indicate, respectively, the patient case number, the age group of the patient, and for each region analyzed the number of Ki67 positive cells and the total number of cells counted. Pathologist: Michael C Haffner.

Table S11. Measurements of Ki67 labeling of normal colon, small intestine, and esophagus resections in mice. The columns indicate, respectively, the tissue type and the age of each cohort (5 mice per group), with the counting of the Ki67 positive cells, total, and percentages reported in one column per mouse. Pathologists: Justin Poling and Meredith A. Pittman.

References

1. Albertini RJ, Nicklas JA, O'Neill JP, & Robison SH (1990) In vivo somatic mutations in humans: measurement and analysis. *Annu Rev Genet* 24:305-326.
2. Albertini RJ*, et al.* (1988) Mutagenicity monitoring in humans by autoradiographic assay for mutant T lymphocytes. *Mutat Res* 204(3):481-492.
3. Ammenheuser MM, Au WW, Whorton EB, Jr., Belli JA, & Ward JB, Jr. (1991) Comparison of hprt variant frequencies and chromosome aberration frequencies in lymphocytes from radiotherapy and chemotherapy patients: a prospective study. *Environ Mol Mutagen* 18(2):126-135.
4. Ammenheuser MM, Berenson AB, Stiglich NJ, Whorton EB, Jr., & Ward JB, Jr. (1994) Elevated frequencies of hprt mutant lymphocytes in cigarette-smoking mothers and their newborns. *Mutat Res* 304(2):285-294.
5. Ammenheuser MM, Ward JB, Jr., Whorton EB, Jr., Killian JM, & Legator MS (1988) Elevated frequencies of 6-thioguanine-resistant lymphocytes in multiple sclerosis patients treated with cyclophosphamide: a prospective study. *Mutat Res* 204(3):509-520.
6. Araten DJ*, et al.* (2005) A quantitative measurement of the human somatic mutation rate. *Cancer research* 65(18):8111-8117.
7. Branda RF*, et al.* (1993) Measurement of HPRT mutant frequencies in T-lymphocytes from healthy human populations. *Mutat Res* 285(2):267-279.
8. Cole J, Green MH, James SE, Henderson L, & Cole H (1988) A further assessment of factors influencing measurements of thioguanine-resistant mutant frequency in circulating T-lymphocytes. *Mutat Res* 204(3):493-507.
9. Cole J & Skopek TR (1994) International Commission for Protection Against Environmental Mutagens and Carcinogens. Working paper no. 3. Somatic mutant frequency, mutation rates and mutational spectra in the human population in vivo. *Mutat Res* 304(1):33-105.
10. Davies MJ, Lovell DP, & Anderson D (1992) Thioguanine-resistant mutant frequency in T-lymphocytes from a healthy human population. *Mutat Res* 265(2):165-171.
11. DeMars R & Held KR (1972) The spontaneous azaguanine-resistant mutants of diploid human fibroblasts. *Humangenetik* 16(1):87-110.
12. Drake JW, Charlesworth B, Charlesworth D, & Crow JF (1998) Rates of spontaneous mutation. *Genetics* 148(4):1667-1686.
13. Jones S*, et al.* (2008) Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci U S A* 105(11):4283-4288.
14. Langlois RG, Bigbee WL, & Jensen RH (1986) Measurements of the frequency of human erythrocytes with gene expression loss phenotypes at the glycophorin A locus. *Hum Genet* 74(4):353-362.
15. Langlois RG, Nisbet BA, Bigbee WL, Ridinger DN, & Jensen RH (1990) An improved flow cytometric assay for somatic mutations at the glycophorin A locus in humans. *Cytometry* 11(4):513-521.

16. Tomasetti C, Vogelstein B, & Parmigiani G (2013) Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci U S A* 110(6):1999-2004.
17. Trainor KJ*, et al.* (1984) Mutation frequency in human lymphocytes increases with age. *Mechanisms of ageing and development* 27(1):83-86.
18. Vijayalaxmi & Evans HJ (1984) Measurement of spontaneous and X-irradiation-induced 6-thioguanine-resistant human blood lymphocytes using a T-cell cloning technique. *Mutat Res* 125(1):87-94.
19. Welch JS*, et al.* (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150(2):264-278.
20. Holland EC, Hively WP, DePinho RA, & Varmus HE (1998) A constitutively active epidermal growth factor receptor cooperates with disruption of G1 cell-cycle arrest pathways to induce glioma-like lesions in mice. *Genes Dev* 12(23):3675-3685.
21. O'Brien CA, Pollett A, Gallinger S, & Dick JE (2007) A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* 445(7123):106-110.
22. Visvader JE (2011) Cells of origin in cancer. *Nature* 469(7330):314-322.
23. Garraway LA & Lander ES (2013) Lessons from the cancer genome. *Cell* 153(1):17-37.
24. Kandoth C*, et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature* 502(7471):333-339.
25. Stratton MR, Campbell PJ, & Futreal PA (2009) The cancer genome. *Nature* 458(7239):719-724.
26. Vogelstein B*, et al.* (2013) Cancer genome landscapes. *Science* 339(6127):1546-1558.
27. Cancer Genome Atlas N (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487(7407):330-337.
28. Liu B*, et al.* (1995) Genetic instability occurs in the majority of young patients with colorectal cancer. *Nat Med* 1(4):348-352.