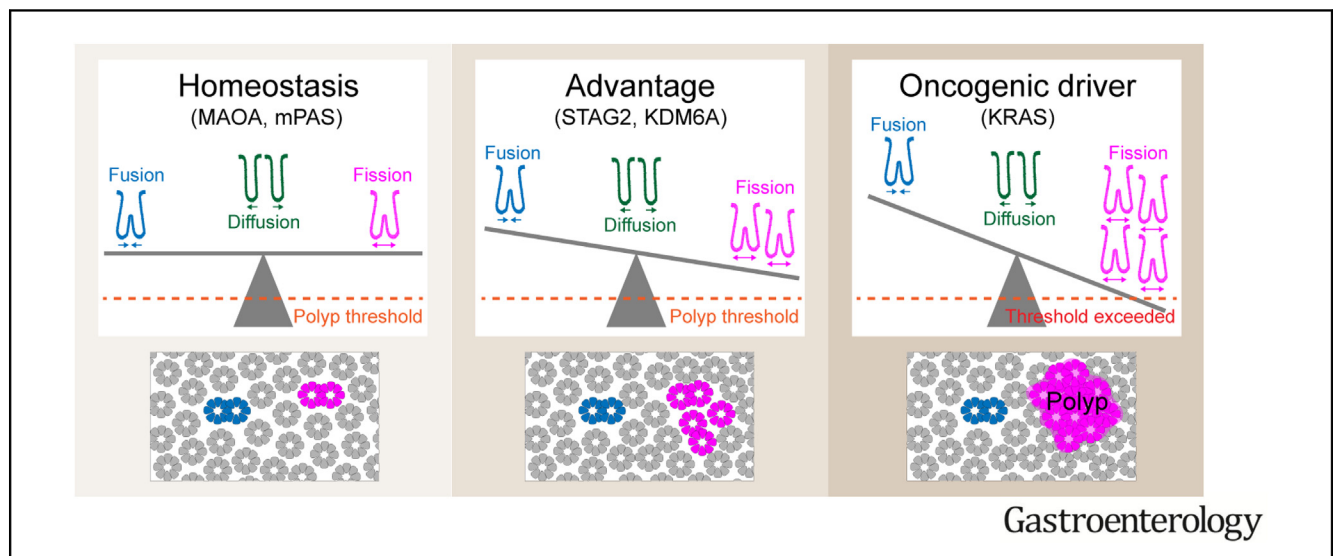




A Diffusion-like Process Accommodates New Crypts During Clonal Expansion in Human Colonic Epithelium

Cora Olpe,^{1,2,*} Doran Khamis,^{3,*} Maria Chukanova,¹ Nefeli Skoufou-Papoutsaki,^{1,2} Richard Kemp,¹ Kate Marks,⁴ Cerys Tatton,¹ Cecilia Lindskog,⁵ Anna Nicholson,¹ Roxanne Brunton-Sim,⁶ Shalini Malhotra,⁷ Rogier ten Hoopen,⁷ Rachael Stanley,⁶ Douglas J. Winton,^{1,2,§} and Edward Morrissey^{3,§}

¹Cancer Research-UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge, United Kingdom; ²Wellcome Trust-Medical Research Council, Cambridge Stem Cell Institute, Cambridge, UK; ³MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, United Kingdom; ⁴Pathology and Data Analytics, St James's University Hospital, Leeds, United Kingdom; ⁵Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden; ⁶Norwich Research Park BioRepository, Norwich, United Kingdom; and ⁷Department of Oncology, University of Cambridge, Cambridge, United Kingdom



BACKGROUND & AIMS: Colorectal cancer (CRC) is thought to arise when the cumulative mutational burden within colonic crypts exceeds a certain threshold that leads to clonal expansion and ultimately neoplastic transformation. Therefore, quantification of the fixation and subsequent expansion of somatic mutations in normal epithelium is key to understanding colorectal cancer initiation. The aim of the present study was to determine how advantaged expansions can be accommodated in the human colon. **METHODS:** Immunohistochemistry was used to visualize loss of the cancer driver KDM6A in formalin-fixed paraffin-embedded (FFPE) normal human colonic epithelium. Combining microscopy with neural network-based image analysis, we determined the frequencies of KDM6A-mutant crypts and fission/fusion intermediates as well as the spatial distribution of clones. Mathematical modeling then defined the dynamics of their fixation and expansion. **RESULTS:** Interpretation of the age-related behavior of KDM6A-negative clones revealed significant competitive advantage in intracrypt dynamics as well as a 5-fold increase in crypt fission rate. This was not accompanied by an increase in crypt fusion. Mathematical modeling of crypt spacing identifies evidence for a crypt diffusion process. We define the threshold fission rate at which diffusion fails to accommodate new crypts, which can be exceeded by KRAS activating mutations. **CONCLUSIONS:** Advantaged gene mutations in KDM6A expand dramatically by crypt fission but not fusion. The crypt

diffusion process enables accommodation of the additional crypts up to a threshold value, beyond which polyp growth may occur. The fission rate associated with KRAS mutations offers a potential explanation for KRAS-initiated polyps.

Keywords: Colorectal Cancer Initiation; Intestinal Crypt; Crypt Fission; Crypt Fusion.

*Authors share co-first authorship; §Authors share co-senior authorship.

Abbreviations used in this paper: bp, base pair; CI, confidence interval; CRC, colorectal cancer; FFPE, formalin-fixed paraffin-embedded; FUF, crypt fusion or fission form; χ , contribution of mutant FUFs to fusions; MAF, mutant allele frequency; M/M, mutant/mutant; M/W, mutant/wild-type; PCR, polymerase chain reaction; PPC, partially populated crypt; WPC, wholly populated crypt.

Most current article

© 2021 by the AGA Institute. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).
0016-5085

<https://doi.org/10.1053/j.gastro.2021.04.035>

It is widely recognized that many renewing epithelia acquire a substantial burden of cancer driver mutations while remaining apparently normal.¹⁻³ In the human colon, development of neoplastic disease is thought to be driven by elevated rates of gland replication or fission. Most notably, loss of the tumor suppressor gene APC generates adenomas in this way.⁴⁻⁷ Yet throughout life, normal crypts also undergo crypt replication at a low rate, which can be elevated by advantaged mutations.⁸⁻¹⁷

There does not appear to be an increase in the net density of crypts, or of colonic epithelial area, with age.¹⁸ This raises the question: how are local clonal expansions arising from elevated fission rates accommodated? One explanation might lie in crypt fusion. This process has recently been described and could counteract the consequences of fission.^{19,20} However, it remains unclear if fusion is an independent stochastic process or if it is locally coregulated with fission. The latter possibility may be particularly relevant to pro-oncogenic mutations as fusions at the edge of mutant patches could enable effective local invasion of wild-type crypts with a high probability of subsequent displacement of wild-type cells.

KDM6A (UTX) is an X-linked gene encoding a histone demethylase that specifically targets di- and tri-methyl groups on lysine 27 of histone H3. Inactivating mutations and deletions of *KDM6A* have been identified in a variety of human cancers, including colon, bladder, prostate, and esophageal cancer.²¹⁻²⁴ *KDM6A* featured among the 127 significantly mutated genes in The Cancer Genome Atlas study that analyzed 3281 tumors derived from 12 cancer types.²⁵ *KDM6A* mutations are infrequent in CRCs (<4% of all tumors, COSMIC database).

Here, in seeking additional cancer driver events that can be visualized as somatic clones, we identify loss of *KDM6A* as possessing advantage in both intracrypt fixation and subsequent expansion. The large multicrypt clones resulting from elevated rates of crypt fission are investigated to study the impact of expansion on crypt packing and the role of crypt fusion in relieving overcrowding. The increased fission rate within *KDM6A*⁻ clones is not accompanied by an increase in crypt fusion, suggesting the 2 processes are driven by independent mechanisms and that fusion does not act to relieve local overcrowding. Instead, it is demonstrated that new crypts generated by fission can be accommodated by localized crypt diffusion up to a threshold beyond which hyperplastic and neoplastic lesions may form.

Materials and Methods

Human Tissue

Normal colon tissue samples were obtained from Addenbrooke's Hospital Cambridge and Norfolk and Norwich University Hospital under full local research ethical committee approval (Documents 15/WA/0131 and 17/EE/0265, and 06/Q0108/307 and 08/H0304/85, respectively) according to UK Home Office regulations. The study included 273 individuals

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

As colorectal cancer is thought to arise from outgrowth of crypts harboring excess genomic alterations, quantification of mutation accumulation and spread within this tissue is key to understanding disease initiation.

NEW FINDINGS

Exploiting loss of the cancer driver *KDM6A*, we demonstrate that newly generated crypts resulting from increased crypt fission are accommodated by mass movement of surrounding crypts in a diffusion-type process.

LIMITATIONS

Only 2 cancer driver genes, *KDM6A* and *KRAS*, are modeled here, which does not preclude the existence of alternative expansion mechanisms.

IMPACT

The threshold fission rate beyond which diffusion cannot accommodate newly generated crypts is calculable defining when identifiable pathologies may form.

aged 13 to 93 years. Colectomy specimens were fixed in 10% neutral buffered formalin. From areas without macroscopically visible disease, mucosal sheets were removed and embedded en face in paraffin blocks. Sections were cut at 5- μ m thickness onto charged glass slides.

Histochemistry

mPAS staining. This was performed as previously published.²⁶

Immunohistochemistry. Antibodies are listed in Supplementary Table 1. For standard sections, immunohistochemistry was performed as previously published.²⁶ For laser capture microdissection, tissue was cut at 10- μ m thickness onto UV-irradiated PEN membrane slides (Zeiss, Oberkochen, Germany). Heat-induced epitope retrieval was performed in citrate buffer in a water bath at 76 °C for 16 hours. Counterstaining with Mayer's Hematoxylin was performed manually for 15 seconds followed by blueing in tap water for 1 minute.

Experimental Pathology

Data acquisition. Clones were scored by brightfield microscopy, followed by scanning of sections using a Leica (Wetzlar, Germany) Aperio AT2 scanner. The DeCryptICS image analysis tool developed by Edward Morrissey and Doran Khamis (<https://github.com/MorrisseyLab/DeCryptICS>) was used to count total number of crypts as well as Fusion or Fission (FUF) forms per section. This was followed by manual quality control using QuPath²⁷ including classification of FUFs into mutant/mutant (M/M), mutant/wild-type (M/W), or entirely wild-type.

Quality control. Within the dataset, 2 individuals aged 37 years with extreme average patch sizes were identified as outliers with respect to that measure and not included in

subsequent analyses of patch sizes, fusion rates, and newly generated crypts.

DNA Extraction From FFPE Tissue

Laser capture microdissection. Crypts were harvested into lids of 0.2-mm radius polymerase chain reaction (PCR) tubes using a Leica LMD7000 Laser Microdissection System; 10 μ L of Proteinase K solution from the Arcturus PicoPure DNA Extraction Kit (ThermoFisher, Waltham, MA) was added followed by lysis at 65 °C for 3 hours and inactivation at 95 °C for 10 minutes.

Extraction from sections for KRAS sequencing. The QIAmp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) was used according to the supplier's protocol.

Library Preparation and Sequencing

Primers, PCR reaction components, cycling conditions, and processing for amplification are described in [Supplementary Tables 2-6](#). Samples were barcoded using the Fast Start High Fidelity PCR System (Roche, Basel, Switzerland) according to the supplier's protocol. After pooling and purification by Clean & Concentrator Kit (Zymo Research, Irvine, CA) and size selection by PippinBlue (Sage Science, Beverly, MA), samples were sequenced using 150-base pair (bp) paired-end sequencing with 10% PhiX in-house on the Illumina (San Diego, CA) MiSeq platform.

Sequencing Data Analysis

Scripts are available under the following: <https://github.com/kemp05/>.

KDM6A. Amplicons were extracted by starting, finishing, and containing the expected sequence in the middle. Then, at every nucleotide position excluding the primer, the number of reads corresponding to the reference genome as well as those containing a base change were recorded. This enabled calculation of the noise at every position ([Supplementary Table 7](#)). Candidate mutations were identified when the mutant allele frequency (MAF) was either >4 times the mean of the noise at that position or >3.29 times the standard deviation at that position ($P \leq .001$). True mutations were called if present in all samples originating from the same patch in serial sections but absent in all wild-type samples from the same sections.

KRAS. Corresponding forward and reverse reads were combined using PANDAsq 2.11 with default options.²⁸ Amplicons were extracted by beginning and ending with the expected primer sequence and correct overall length (± 3 bp tolerance).

For codon 12 and 13 mutation calling, reads containing the sequences corresponding to wild-type as well as all possible mutations ([Supplementary Table 8](#)) were extracted. This yielded an MAF for every possible mutation in all 4 amplicons for every sample and revealed the noise. KRAS mutations were called if (1) >1000 reads were obtained for both KRAS amplicons and at least 1 mimic amplicon, and (2) the MAF in both KRAS amplicons was > 0.1% (corresponding to at least 10 mutant reads) but found at background levels in the mimic amplicons. These criteria correspond to ≥ 1.96 standard deviations or a P value of < .025 for the noisiest nucleotide position (G12D). The actual MAF for a particular mutation was calculated by subtracting the mean allele frequency (noise).

Mathematical Modeling

Stem cell dynamics and crypt fission. The stem cell dynamics and fission rate associated with loss of KDM6A were mathematically modeled as previously described.²⁶

Crypt fusion and diffusion. Crypt fusion was modeled as a process parallel to fission, with the same duration. Therefore, the following applied:

$$\frac{\text{number of fission events}}{\text{number of fusion events}} \sim \frac{\text{fission rate}}{\text{fusion rate}}$$

All M/W FUFIs were considered fusions, but M/M FUFIs could be fissions or fusions. Therefore, the number of fission events and fusion events required for the preceding equation were not directly measurable. However, they were calculable by sampling FUFIs from the edge of mutant patches, which enabled calculation of the proportion of M/M FUFIs that are fusions (termed χ) by accounting for the distribution of the mutational state of neighboring crypts (see supplemental materials for mathematical details).

To infer a diffusion coefficient, growth of a patch through initial mutation and subsequent fissions was modeled as a stochastically firing point source of mass at the clone centroid. Potential trajectories from mutation hit to patch of size 10 (i.e, stochastic event times) were simulated, and an ensemble diffusion coefficient was inferred by randomly drawing paths from the set of simulated potential trajectories. The inferred diffusion coefficient was then used in a theoretical study of patch expansion (see [supplemental materials](#) for mathematical details).

KRAS expansion. The data obtained here were combined with our previously published dataset and analyzed as described there.²⁶

Results

KDM6A-negative Clones Are Advantaged in Stem Cell Competition

We have previously used visualization of loss of X-linked genes as clonal marks to quantify human colonic stem cell dynamics.²⁶ In attempting to expand this methodology to X-linked genes with cancer association, clonal loss of KDM6A was identified by immunohistochemistry with 2 independent antibodies on normal human colonic epithelia ([Supplementary Figure 1A and B](#)). Intracrypt dynamics that describe the accumulation of clones wholly populating entire crypts (WPC) from partly populated (PPC) transition forms were determined for KDM6A⁻ clones using colonic FFPE sections from 120 patients aged 21 to 93 as previously described²⁶ ([Figure 1A and B](#)). This revealed that loss of KDM6A confers a competitive advantage to affected stem cells shown as a decrease in the fraction of PPC supporting the accumulation of WPC ([Figure 1C](#)).

KDM6A-negative Clones Expand by 5-Fold Increased Crypt Fission

Expansion of individual KDM6A⁻ clones was recognizable as large patches that frequently exceeded 10 crypts ([Figure 1D](#)). To confirm the clonal origin of such patches, we used laser capture microdissection followed by targeted

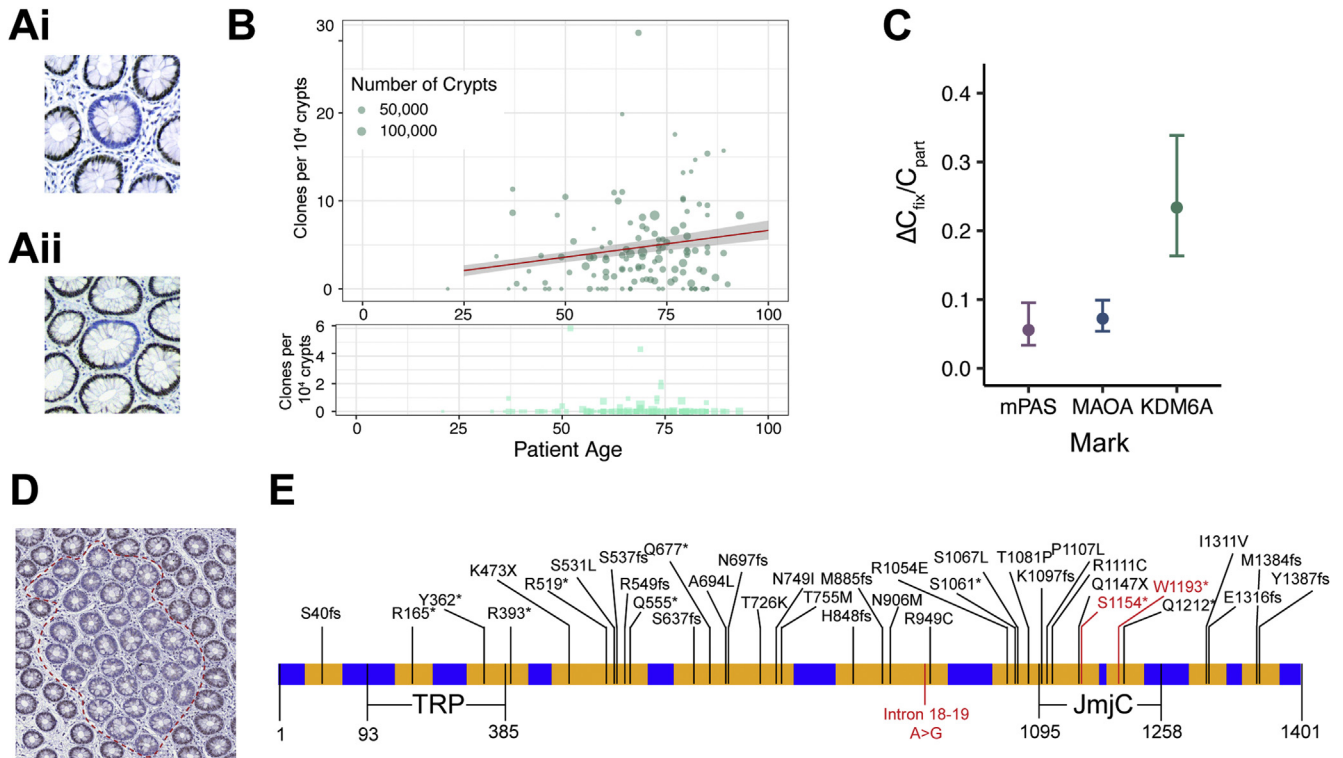


Figure 1. Detection of clonal loss of KDM6A in normal human colon. (A) Representative images of KDM6A⁻ WPC (i) and PPC (ii). (B) Top: Frequency plot showing age-related behaviors of KDM6A⁻ WPC (circles, darker) and PPC (squares, lighter). Red line: regression analysis showing age-related accumulation of WPC (slope of accumulation of fixed clones [ΔC_{fix}]: 6.04×10^{-6} per year) and 95% CI in gray. Bottom: PPC only on expanded y-axis. (C) Plot showing increased ratio of $\Delta C_{fix}/C_{part}$ for KDM6A (0.23, 95% CI 0.16–0.34) as compared with the neutral marks mPAS and MAOA (replotted from Nicholson et al.²⁶). Error bars = 95% CI. (D) Representative image of large KDM6A⁻ multicrypt patch highlighted by dashed line. (E) KDM6A cDNA structure annotated with sequenced areas (yellow), mutations found in COSMIC and mutations identified in KDM6A⁻ patches (red).

sequencing that covered 3.6 kb of exonic and flanking intronic sequence of *KDM6A* (24 amplicons), including sites frequently mutated in human cancers. No patch was found to carry more than 1 *KDM6A* mutation and mutant allele frequencies were in line with predictions from patient sex and stromal content, supporting clonality (Figure 1E, Supplementary Figure 1C–E).

The age-related size distribution of multicrypt clones was analyzed to infer the fission rate associated with *KDM6A* loss and compare it with those previously described for neutral (MAOA and mPAS) and advantaged clonal marks (STAG2).²⁶ This revealed an age-related increase in the frequency of large clones (≥ 10 crypts/patch) for STAG2⁻ and KDM6A⁻ but not for neutral marks (Figure 2A). Loss of *KDM6A* generates a higher proportion of large patches than the other clonal marks, whereas STAG2 loss generates more clones due to a higher event rate (Figure 2B and C). From these patch size distributions, the crypt fission rate associated with loss of *KDM6A* was calculated to be to 3.6% per year (95% confidence interval [CI] 3.2–4.1), approximately 5-fold higher than the background homeostatic rate previously derived from neutral clonal marks (Figure 2D). Consequently, in individuals older than 80 years, 13.5% of KDM6A⁻ clones are found as patches comprising more than 5 crypts

compared with 4.8% for STAG2, 1.7% for mPAS, and 0.8% for MAOA.

KDM6A⁻ and *STAG2*⁻ Patches Lack Significant Local Overcrowding

Expanding clones are more likely to undergo additional fissions as the probability scales with the number of crypts present. Consequently, the interval between fissions decreases rapidly with increasing patch size. For example, mathematical modeling of the expansion of *KDM6A*⁻ crypts indicates the median time taken to grow from 1 to 2 crypts is 19 years, but only 2 years to grow from 10 to 11 (Figure 2E). Therefore, recently formed larger patches might be expected to demonstrate overcrowding.

To test if larger patches are more densely packed, the area occupied by crypts and their surrounding stroma was determined for 24 STAG2⁻ and 20 KDM6A⁻ clones containing 10 crypts, the largest size for which sufficient data could be obtained (Figure 2F and G). The fraction of each patch occupied by stroma was then calculated. Adjacent to each mutant clone, 3 random groups of 10 crypts were defined as control “patches” and similarly analyzed (totaling 72 control patches for STAG2 and 60 control patches for *KDM6A*). Comparing the fraction of each patch occupied by

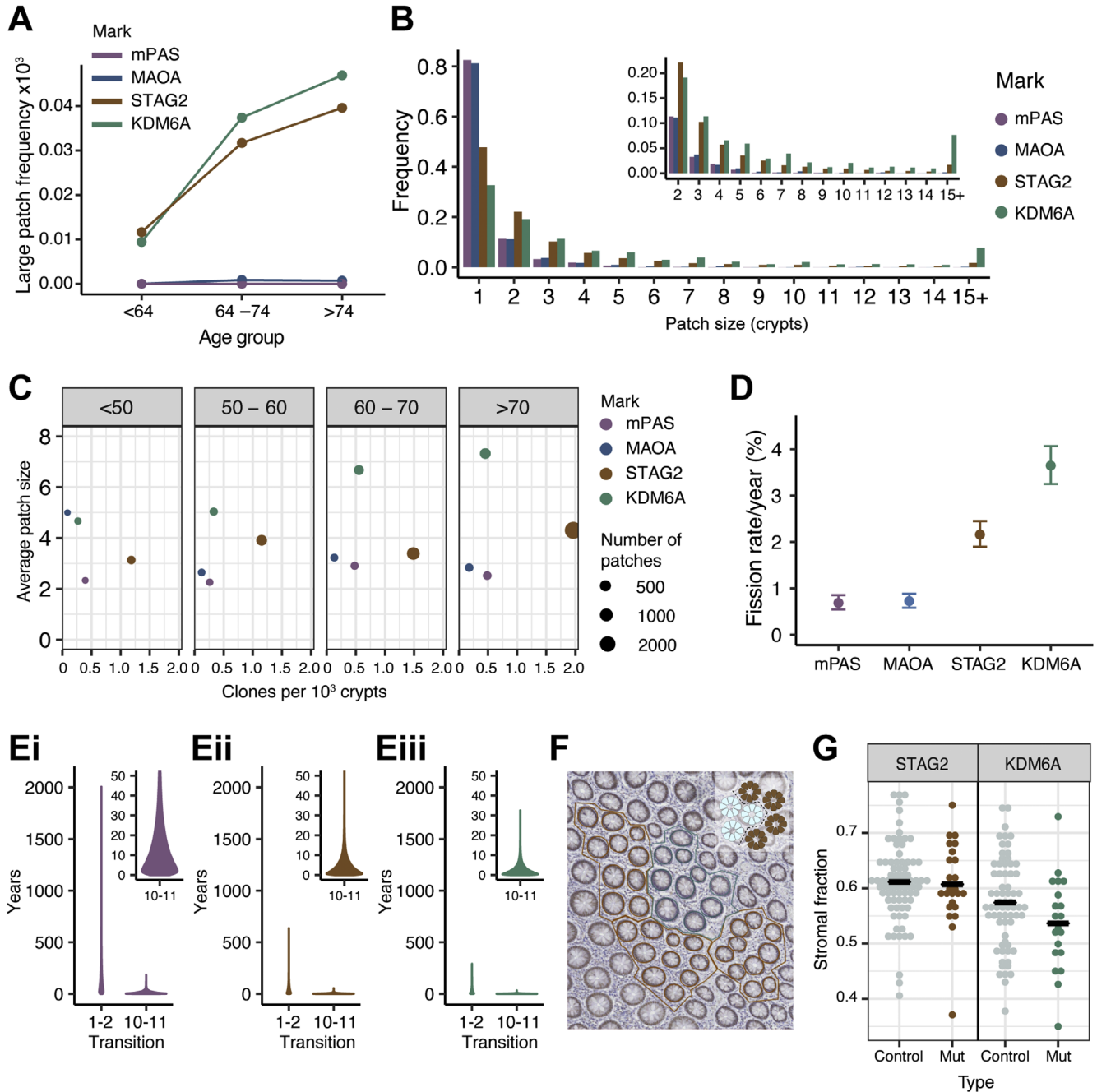


Figure 2. KDM6A⁻ patches lack significant overcrowding despite increased fission. (A) Plot showing mean frequency of large (≥ 10 crypts) patches for age groups shown. (B) Histogram showing the frequency of different patch sizes for mPAS, MAOA, KDM6A, and STAG2 across all ages. Inset shows patch size ≥ 2 crypts on expanded y-axis. (C) Dot plot of mean clone frequency plotted against mean average patch size for multicrypt clones in age groups shown for mPAS, MAOA, STAG2, and KDM6A. (D) Plot showing inferred fission rate/crypt/year for KDM6A compared with mPAS, MAOA, and STAG2 (data replotted from Nicholson et al²⁶). Error bars = 95% CI. (E) Simulation data showing the time in years taken for transitions between patch sizes 1–2 and 10–11 for (i) mPAS, (ii) STAG2, and (iii) KDM6A. Insets show 10–11 transition on expanded y-axis. (F) Image showing patch selection for assessment of crypt packing density within KDM6A⁻ (blue) and adjacent wild-type crypts (brown). Inset indicates placement of borders between crypts. (G) Dot plots showing stromal fraction for STAG2⁻ and KDM6A⁻ and adjacent wild-type patches comprising 10 crypts.

stroma to adjacent wild-type groupings indicated a slight trend toward increased packing density for STAG2⁻ as well as KDM6A⁻ crypts that failed to reach significance (Figure 2G). Considering that a lack of overcrowding may stem from a decrease in crypt size, the areas of crypts were

measured. This revealed that KDM6A⁻ crypt sections are approximately 1.3 times the size of adjacent wild-type crypts ($P < .001$) (Supplementary Figure 2). No difference was found between STAG2⁻ crypts and their wild-type neighbors (Supplementary Figure 2). Therefore, lack of

overcrowding cannot be attributed to reduced crypt size for either STAG2 or KDM6A loss. We also considered the possibility of accommodation of clonal expansions by “squashing” of neighboring crypts. However, analysis of crypt eccentricities provided no evidence for the predicted flattening of crypt architecture that would result (Supplementary Figures 3–5).

Together these observations suggest that crypts even within relatively recent clonal expansions avoid overcrowding to largely achieve ambient density.

Evidence for Crypt Fusion

The lack of overcrowding in KDM6A⁻ clones suggests a mechanism counteracting the localized increase in fission. An opposing process of crypt fusion has been recognized in mouse intestine.¹⁹ A homeostatic human fusion rate has been estimated by assuming equivalence in the rate of both fission and fusion.²⁰ On a tissue-wide basis such a balance of rates could act to maintain constant crypt density. However, local advantaged expansions can be balanced only if fission and fusion are locally coordinated.

We first sought confirmation that fusion occurs. The evidence in human epithelium is based on identification of branched crypts within which clonal loss of mitochondrial CCO activity is restricted to one branch. These are interpreted as transition intermediates in an active fusion process.²⁰ Analysis of en face tissue sections stained to visualize mPAS positivity confirmed the existence of rare heterotypic branched forms in normal human colonic epithelium. Analysis of more than 2×10^6 crypts in sections from 80 individuals containing mPAS⁺ clones identified 32 candidate mPAS⁺ branched forms that were either mixed (mutant and wild-type, M/W) or fully mutant (M/M) (Figure 3A). Of the 13 M/W forms, the positive epithelium was always restricted to 1 branch.

An alternative interpretation is that branched crypts are intermediate fission forms whereby heterotypic staining arises due to mutations occurring or segregating into a single branch (Figure 3B). We formally considered this possibility using the fusion duration estimate derived by Baker and colleagues²⁰ as well as our previous estimates of de novo mutation probability and clone fixation rates that together determine the frequency of monoclonal crypts present in individuals of different age.²⁶ Within the relatively small number of branched crypts present, none are predicted to contain monoclonal crypt branches by either mechanism (Figure 3C). This suggests that heterotypic forms represent genuine intermediates in an active fusion process. Because the bulk of branched crypts are unstained and can represent intermediates in either fusion or fission we propose the agnostic term FUFIs to describe these transition forms (Supplementary Figure 6).

Crypt Fission and Fusion Are Regulated Independently

To calculate crypt fusion rates, heterotypic and homotypic FUFIs were also evaluated for STAG2 and KDM6A loss by scoring approximately 3.9×10^6 and 1.8×10^6 crypts

from 53 and 102 individuals, respectively. In total, more than 28,000 FUFIs were evaluated. This identified 151 and 63 FUFIs with STAG2 (18,928 clones analyzed) and KDM6A loss (5,353 clones analyzed), respectively (Figure 3D). These could be found as single events or within multicrypt clones (Figure 3E).

Assuming equal duration for fission and fusion (ie, the time window during which FUFIs are detectable), the fusion rate is accessible by proportionality. Specifically, the ratio of the frequency of observed fission FUFIs to the fission rate (independently calculated from the patch size distribution) would equal the ratio of the frequency of observed fusion FUFIs to the fusion rate. However, although all M/W FUFIs are considered fusions, M/M FUFIs can be either fissions or fusions. Therefore, to exploit this proportionality, the relative contribution of M/M FUFIs events to the total number of fusions (termed χ) needs to be determined. The value for χ is calculable on the basis that M/M FUFIs have a probability of being a fusion event that depends on the number of M and W neighbors present at the onset of the fusion process. Therefore, the status of crypts neighboring FUFIs at the patch border of multicrypt clones were scored as W or M (totaling 4, 121, and 49 for mPAS, STAG2, and KDM6A, respectively) (Figure 3F and G). Single FUFIs had W neighbors only. Averaging the M and W neighbors of M/M FUFIs revealed χ to be 0.03 for mPAS. As neutral marks such as mPAS generate mostly small clones, most fusion events are W/M leading to low values of χ . At the edge of larger expansions generated by advantaged marks, M/M fusions occur more readily. Correspondingly, the χ values for STAG2 and KDM6A are 0.26 and 0.35, respectively (Figure 3H). Rates of fusion using these values of χ were then estimated using the proportionality described previously (see [supplemental information](#) for details).

This analysis indicates similar crypt fusion rates for mPAS, STAG2 and KDM6A of 0.3% per year (95% CI 0.1–0.6), 0.4% (95% CI 0.3–0.7), and 0.7% (95% CI 0.3–1.4) (Figure 3I). Comparison of mPAS fission (0.7% per year; 95% CI 0.5–0.9) and fusion rates show that these closely correspond. This suggests that in homeostasis the rates of both processes are balanced and will act together to maintain constant crypt numbers across the tissue, as has been suggested previously.²⁰ However, for mutations causing elevated fission rates there appears to be no evidence for a compensatory increase of the fusion rate. These analyses suggest that fission and fusion are independent processes and not coordinately regulated.

Crypt Diffusion Accommodates New Crypts Throughout Life

A striking feature of larger patches is that mutant crypts have over decades populated the territory initially occupied by multiple independent crypts without a significant increase in crypt density. In the absence of appreciable crypt fusion this suggests local adjustments to disperse crypts from the growing focus. With this rationale, we considered the possibility of random crypt movement in the form of a diffusion process.

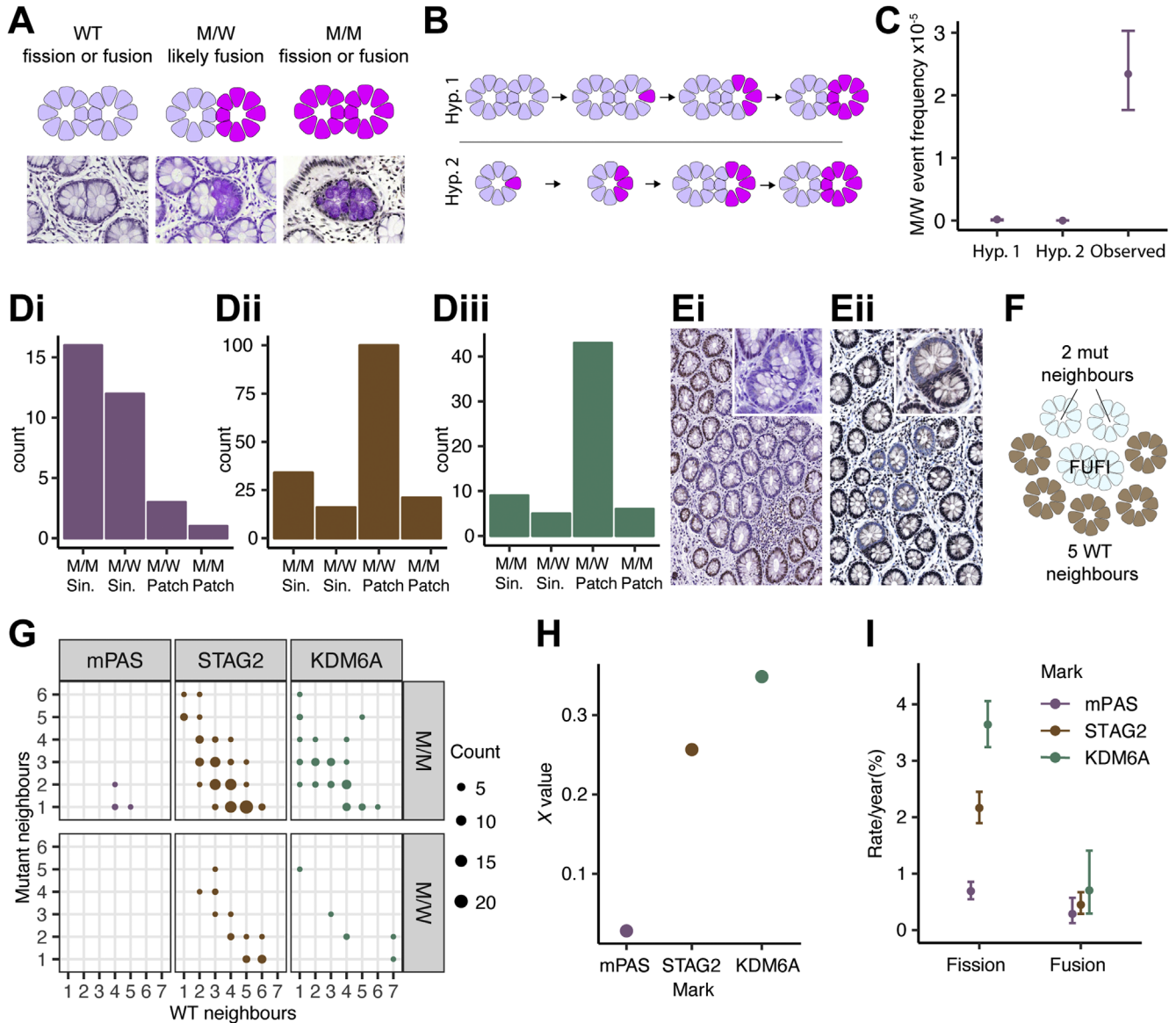


Figure 3. Crypt fission and fusion are independently regulated processes. (A) Schematic and representative mPAS-stained images of 3 types of fusion or fission forms. (B) Schematic representation of alternative origins of M/W forms. Hypothesis 1 (Hyp1): stem cell mutation in one branch of intermediate fission form followed by monoclonal conversion. Hypothesis 2 (Hyp 2): fission of a preexisting partially populated crypt with segregation of mutant and wild-type epithelium into each branch, followed by monoclonal conversion. (C) Comparison of M/W event frequencies simulated for hypotheses described in (B) and the observed frequency. Error bars = 95% CI. (D) Bar graphs showing numbers of FUFIs types scored for (i) mPAS, (ii) STAG2, and (iii) KDM6A. (E) Representative images of FUFIs at patch borders: (i) KDM6A M/M FUFIs, (ii) STAG2 M/W FUFIs. Insets show enlarged FUFIs. (F) Schematic showing scoring of FUFIs neighbors used for calculation of χ . (G) Dot plot showing neighboring crypt status of patch border FUFIs (M/M or M/W) for mPAS, STAG2, and KDM6A. Each dot represents 1 or more FUFIs for which neighbors were scored. (H) Plot comparing the value of χ for mPAS, STAG2, and KDM6A. (I) Plot comparing the derived fission and fusion rates for mPAS, STAG2, and KDM6A. mPAS and STAG2 fission rates correspond to data from Nicholson et al,²⁶ replotted. Error bars = 95% CI.

In the colonic epithelium it is the crypts that are being diffused in the “space” of the surrounding stroma (Figure 4A). The diffusion coefficient (change in area per unit time) can be estimated based on crypt packing (measured in terms of crypt area per unit area of mucosa) and consideration of all possible sequences of fission events, constrained by the patient age and calculated using the mutation and fission rates.

To find evidence supporting a diffusion-type process we revisited the STAG2⁻ and KDM6A⁻ patches of 10 crypts and surrounding control patches. For each mutant clone (24 for STAG2 and 20 for KDM6A) and the corresponding 3 arbitrary control groups of 10 adjacent wild-type crypts, each crypt was spatially mapped in X/Y coordinates. Total patch areas were divided in crypt domains that were defined as the area occupied by a crypt and its share of surrounding stroma

(Figure 4B). Areas of individual crypts as well as total patch area were measured and the distance between mutant clones and wild-type patch centroids, r , was determined (Figure 4C).

Using known mutation and fission rates, the potential trajectories from initial mutation to 10-crypt mutant clones were simulated and the most likely trajectories were used to calculate the age of the clone (Supplementary Figures 7 and 8). The diffusion coefficient was inferred to define an overall tissue process that best fits the differences in stromal densities between mutant and control patches. It describes how the burden of decreased stromal fraction resulting from a mutant clone is dispersed into the surrounding tissue over time. For older clones, the expectation is for the system to be close to the ambient density, whereas for younger clones, the perturbation to the local stromal fraction may still be evident. Examples of both presumptive young and old clones were readily detectable (Figure 4D, Supplementary Figures 9 and 10).

For a subset of 7 patches (4 for KDM6A, 3 for STAG2), a more detailed rolling window analysis was performed in which the preceding approach was applied but where fields of 10 crypts were moved outward from the mutant clone. Again, evidence of a reduction in stromal fraction consistent with perturbation in younger clones was observed (Figure 4E, Supplementary Figures 9 and 10). The diffusion coefficient was found to be 1.05 crypt domain areas per year (95% CI 0.339–9.70) (Figure 4F). Reassuringly, testing a null hypothesis, that there is no diffusion-type process and therefore no radial dependence in crypt packing, by considering neighborhood ambient densities of crypts across all patches (mutant, wild-type, or mixed), generated a significantly worse fit than the experimental comparisons (Supplementary Figure 11). Of note, this confirmation of radial dependence in packing argues against other possible alleviators of crypt density such as changes in the size or shape of crypts within mutant clones.

The inferred diffusion process can be used to define the number of crypt domains impacted to accommodate a new clonal expansion. For example, the model suggests that patches of 10 KDM6A⁻ crypts would require 264 crypt domains to undergo a 1% reduction in their spacing, whereas a 5% reduction would only require 53 crypt domains (Figure 4G and H).

Defining a Homeostatic Threshold

Limited evidence suggests that there are no significant age-related changes in colon length and crypt density.¹⁸ With respect just to STAG2 and KDM6A mutations, the relatively few new crypts arising during life could be easily accommodated by crypt movement. By the time individuals exceed 75 years of age, for every 10⁵ crypts, fission has added only approximately 200 and 290 new STAG2⁻ and KDM6A⁻ crypts, respectively (Figure 4J). However, it seems highly probable that additional genetic variants will also promote fission to different degrees. The potential for diffusion to locally balance this process as fission rate increases was investigated.

Simulations were performed escalating the homeostatic fission rate of 0.7% per year (Supplementary Figure 12).

When fission rates remain below approximately 12-fold that of homeostasis, diffusion can generate enough space to accommodate newly generated crypts (Figure 4J). Higher fission rates result in a proportion of clones reaching a threshold of maximum packing density within which crypts are directly touching. This suggests a potential boundary for polyp growth that is dependent on the physical processes of fission and diffusion. We have previously used targeted sequencing of FFPE sections to infer the effect of KRAS activating mutations on crypt fission.²⁶ Here, in expanding on that initial dataset (see Methods) KRAS activating mutations were found in 35 (22 new) of 256 individuals (130 new) in the age range 20 to 91 years, corresponding to 13.7% of the cohort (Supplementary Figure 13). Mutant allele frequencies in the range of 0.12% to 2.35% combined with total crypt numbers per section enabled estimation of clone size (Supplementary Figure 13). Subsequent mathematical inference indicates that a 17-fold increase in crypt fission rate to 12% (95% CI 10.8–13.7) per year best fits with the data. Approximately 1% of KRAS activating mutations are predicted to breach the threshold for lesion growth after 50 years (Figure 4J). A therapeutic intervention inhibiting crypt fission for any 10 years could reduce this to approximately 0.01% (Figure 4K).

Discussion

Several studies have identified large mutant expansions in seemingly normal epithelia.^{1,2,26} In the adult colon, this occurs by increased crypt fission rate, whereby biased mutations can generate large clones that are appropriately distributed within the tissue.²⁶ In contrast, elevated glandular fission rates are also known to drive the overgrowth of adenomas and CRCs, suggesting that differences in the rate of fission or the response to it must differ between normal and neoplastic tissues.^{4–7}

In considering the epithelial responses that compensate for elevated fission rates, we first validated a new advantaged clonal mark, KDM6A, that together with STAG2 provided gene-specific assays with 5- and 3-fold increased fission rates, respectively. Comparing the configuration of the size and frequency of clones for both genes demonstrates the different strategies by which age-related mutational burden can be achieved. STAG2 has the higher mutation rate and generates many relatively small clones, whereas KDM6A generates fewer but larger expansions. A corollary of the exponential growth of patches as their size increases is that larger patches will tend to be the most recent and therefore most likely to contain evidence of local adaptation to accommodate new crypts.

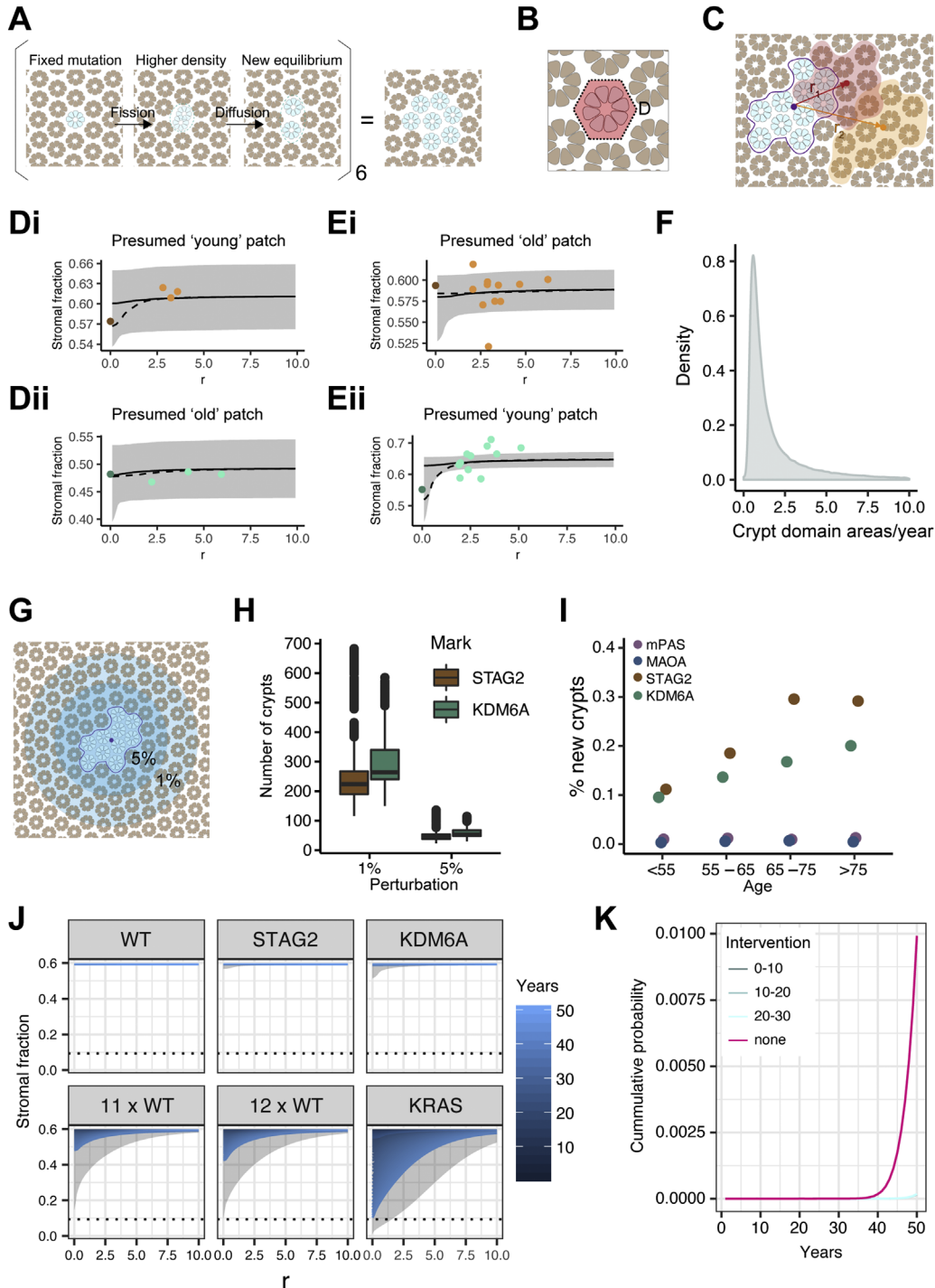
The recently recognized process of crypt fusion offers a potential mechanism to compensate for fission.¹⁹ Occurring at equal rates in homeostasis they could effectively balance crypt numbers on a population basis.²⁰ In considering fusion as a mechanism to accommodate new crypts, a baseline estimate was first established here and found to approximate that for fission; however, no upregulation of fusion accompanying the local expansions resulting from

STAG2 and KDM6A mutation was identified. Conceivably other mutations may impact fusion to ease local packing but it does not appear necessary to do so.

Multicrypt clones that form over decades populate the territory previously occupied by multiple independent crypts. Aiming to understand this dispersal, we sought and found evidence of a diffusion-type process in a subset of clones. These are consistent with a recent expansion “caught in the act” of being restored to an ambient crypt density. The behavior captured probably reflects a passive dispersal

mechanism rather than actual diffusion and must be accompanied by some level of stromal turnover.

The diffusion coefficient defines the rate of movement of crypt domains and the size of the larger impacted zone that is required to absorb new crypts. Parameterizing the process allows testing of the robustness of the tissue to deal with localized accelerated growth conferred by biased mutations. From this analysis, the homeostatic dispersal mechanism seems able to accommodate increased fission rates of more than 10-fold above baseline. Even for



mutations that generate higher crypt fission rates, only the fastest growing clones would overgrow the available space. For example, approximately 5% of clones carrying a gene mutation that confers a 19-fold increase in fission rate would reach a threshold where they lack a stromal domain between crypts and overgrow the available space after 50 years.

The actual threshold at which clonal expansions become recognizable as pathologies may be lower than the extreme one applied here. However, the implication remains that the distinction between phenotypically normal clones and those forming overt pathologies may be determined solely by a probabilistic process in which a recent succession of fission events overwhelms homeostatic dispersal mechanisms.

Activating mutations of KRAS have been described in normal colonic epithelium. The revised estimate of a 17-fold increase for the fission rate conferred by KRAS activating mutations is higher than that previously inferred (10-fold) and is based on analysis of many more patients.²⁶ It is intriguing that activating mutations of KRAS breach the extreme threshold defined here. KRAS is commonly mutated in a broad spectrum of benign and premalignant pathologies such as serrated lesions and adenomas that may arise at least in part due to the dispersal threshold being reached.^{29–33}

These findings have clinical significance with respect to bowel cancer screening programs with the implication that clonal expansions with a high malignant potential are not all contained within visible lesions such as sessile serrated adenomas despite having a comparable tissue footprint. Further, it is known that a proportion of adenomas spontaneously regress when observed in longitudinal studies.^{34–36} One plausible explanation for such phenomena is that these lesions first develop due to reaching the threshold resulting in local overcrowding but are transient because of ongoing crypt dispersal.

Loss of function mutations affecting the APC tumor suppressor gene are also initiated and expanded by glandular fission.^{4–7} Mutation of both APC and KRAS is frequent in CRCs. The 2 pathways are known to interact at the

molecular level.³⁷ It is likely their combined activation will also synergize to further elevate gland fission rate and promote overgrowth as fully neoplastic CRCs develop.

Obesity, a known risk factor for CRC, is known to be accompanied by increased crypt fission rate.³⁸ Furthermore, diets deficient for methyl donors are known to reduce crypt fission rates in the mouse.³⁹ An implication of the colon having the capacity to absorb many more new crypts is that modest time limited reductions in fission rates may not only slow the growth of lesions but prevent them from forming at all.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at <https://doi.org/10.1053/j.gastro.2021.04.035>.

References

1. Martincorena I, Roshan A, Gerstung M, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 2015;348:880–886.
2. Martincorena I, Fowler JC, Wabik A, et al. Somatic mutant clones colonize the human esophagus with age. *Science* 2018;362:911–917.
3. Lee-Six H, Olafsson S, Ellis P, et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* 2019;574:532–537.
4. Wasan HS, Park H-S, Liu KC, et al. APC in the regulation of intestinal crypt fission. *J Pathol* 1998;185:246–255.
5. Wong W-M, Mandir N, Goodlad RA, et al. Histogenesis of human colorectal adenomas and hyperplastic polyps: the role of cell proliferation and crypt fission. *Gut* 2002; 50:212LP–217.
6. Preston SL, Wong W-M, Chan AO-O, et al. Bottom-up histogenesis of colorectal adenomas. *Cancer Res* 2003; 63:3819–3825.
7. van den Brink GR, Offerhaus GJ. The morphogenetic code and colon cancer development. *Cancer Cell* 2007; 11:109–117.

Figure 4. Evidence for a crypt diffusion process. (A) Representation of the proposed crypt diffusion process. Fission generates higher local density which is relieved by diffusion. (B) Schematic showing crypt cross section with area of surrounding stroma defining a crypt domain. (C) Representation of areas measured to assess packing of mutant patch and surrounding crypts in rolling windows placed at different distances (r) from centroid of mutant patch. (D) Examples of radial variation in crypt packing in patches of 10 crypts when moving from mutant clone to 3 adjacent control groupings for (i) STAG2 and (ii) KDM6A. Points are data derived, the black line (gray ribbon) is the median (95% CI) theoretical stromal fraction as fitted from the diffusion model. Dashed line shows the average of the 25 most likely trajectories from initial mutation to clone of size 10, based on population average diffusion and neighborhood ambient stromal fraction. (E) As (D) but with rolling window data. (F) Density plot of values obtained for the diffusion coefficient in human colonic epithelium. (G) Representation of area of crypts affected if space is decreased by 1% or 5% surrounding a mutant patch of 10. (Crypt numbers not to scale.) (H) Boxplot showing simulated numbers of crypts affected if spacing is decreased by 1% or 5% respectively by addition of a patch of 10 mutant crypts. (I) Dot plot showing median frequency of newly generated crypts for mPAS, MAOA, STAG2, and KDM6A across 4 age bins. (J) Line graphs showing the stromal fraction resulting from crypt diffusion and crypt fission rates corresponding to the homeostatic (WT) rate as well as those associated with STAG2, KDM6A, and KRAS and multiples of 11 and 12 of the WT rate. Dotted line = whitespace fraction calculated from optimal hexagonal packing of circles. r = distance from centroid of patch in crypt domains. (K) Line graph showing results from simulations to calculate the cumulative probability of a clone developing into a lesion, defined as reaching the whitespace fraction for hexagonal packing shown in (J), over time. Plots using the fission rate associated with KRAS (17 times WT, none) are shown alongside those including therapeutic crypt fission inhibition applied for a decade either immediately after mutation acquisition (0–10), after 10 years (10–20) or 20 years (20–30).

8. Kim K-M, Shibata D. Methylation reveals a niche: stem cell succession in human colon crypts. *Oncogene* 2002;21:5441–5449.
9. Kim K-M, Shibata D. Tracing ancestry with methylation patterns: most crypts appear distantly related in normal adult human colon. *BMC Gastroenterol* 2004;4:8.
10. Greaves LC, Preston SL, Tadrous PJ, et al. Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. *Proc Natl Acad Sci U S A* 2006;103:714–719.
11. Clarke RM. The effect of growth and of fasting on the number of villi and crypts in the small intestine of the albino rat. *J Anat* 1972;112:27–33.
12. Cairnie AB, Millen BH. Fission of crypts in the small intestine of the irradiated mouse. *Cell Tissue Kinet* 1975;8:189–196.
13. Maskens A. Histogenesis of colon glands during post-natal growth. *Acta Anat (Basel)* 1978;100:17–26.
14. Maskens AP, Duhardin-Loits RM. Kinetics of tissue proliferation in colorectal mucosa during post-natal growth. *Cell Tissue Kinet* 1981;14:467–477.
15. Wright NA, Al-Nafussi A. The kinetics of villus cell populations in the mouse small intestine. II. Studies on growth control after death of proliferative cells induced by cytosine arabinoside, with special reference to negative feedback mechanisms. *Cell Tissue Kinet* 1982;15:611–621.
16. Cheng H, McCulloch C, Bjerknes M. Effects of 30% intestinal resection on whole population cell kinetics of mouse intestinal epithelium. *Anat Rec* 1986;215:35–41.
17. Clair WHS, Osborne JW. Crypt fission in the small intestine of the rat. *Br J Cancer* 1986;(Suppl 7):39–41.
18. Hounnou G, Destrieux C, Desme J, et al. Anatomical study of the length of the human intestine. *Surg Radiol Anat* 2002;24:290–294.
19. Bruens L, Ellenbroek SIJ, van Rheenen J, et al. In vivo imaging reveals existence of crypt fission and fusion in adult mouse intestine. *Gastroenterology* 2017;153:674–677.e3.
20. Baker A-M, Gabbutt C, Williams MJ, et al. Crypt fusion as a homeostatic mechanism in the human colon. *Gut* 2019;68:1986–1993.
21. Zhou Z, Zhang H-S, Liu Y, et al. Loss of TET1 facilitates DLD1 colon cancer cell migration via H3K27me3-mediated down-regulation of E-cadherin. *J Cell Physiol* 2018;233:1359–1369.
22. Nickerson ML, Dancik GM, Im KM, et al. Concurrent alterations in TERT, KDM6A, and the BRCA pathway in bladder cancer. *Clin Cancer Res* 2014;20:4935–4948.
23. Suvà ML, Riggi N, Bernstein BE. Epigenetic reprogramming in cancer. *Science* 2013;339:1567–1570.
24. van Haaften G, Dalgliesh GL, Davies H, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat Genet* 2009;41:521–523.
25. Kandath C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature* 2013;502:333–339.
26. Nicholson AM, Olpe C, Hoyle A, et al. Fixation and spread of somatic mutations in adult human colonic epithelium. *Cell Stem Cell* 2018;22:909–918.e8.
27. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep* 2017;7:16878.
28. Masella AP, Bartram AK, Truszkowski JM, et al. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 2012;13:31.
29. Zauber P, Marotta S, Sabbath-Solitare M. KRAS gene mutations are more common in colorectal villous adenomas and in situ carcinomas than in carcinomas. *Int J Mol Epidemiol Genet* 2013;4:1–10.
30. Juárez M, Egoavil C, Rodríguez-Soler M, et al. KRAS and BRAF somatic mutations in colonic polyps and the risk of metachronous neoplasia. *PLoS One* 2017;12:e0184937.
31. Hashimoto T, Tanaka Y, Ogawa R, et al. Superficially serrated adenoma: a proposal for a novel subtype of colorectal serrated lesion. *Mod Pathol* 2018;31:1588–1598.
32. McCarthy AJ, Serra S, Chetty R. Traditional serrated adenoma: an overview of pathology and emphasis on molecular pathogenesis. *BMJ Open Gastroenterol* 2019;6:e000317.
33. De Palma FDE, D'Argenio V, Pol J, et al. The molecular hallmarks of the serrated pathway in colorectal cancer. *Cancers (Basel)* 2019;11:1017.
34. Knoernschild H. Growth rate and malignant potential of colonic polyps: early results. *Surg Forum* 1963;14:137–138.
35. Hoff G, Foerster A, Vatn MH, et al. Epidemiology of polyps in the rectum and colon: recovery and evaluation of unresected polyps 2 years after detection. *Scand J Gastroenterol* 1986;21:853–862.
36. Loeve F, Boer R, Zauber AG, et al. National Polyp Study data: evidence for regression of adenomas. *Int J Cancer* 2004;111:633–639.
37. Jeong W-J, Ro EJ, Choi K-Y. Interaction between Wnt/ β -catenin and RAS-ERK pathways and an anti-cancer strategy via degradations of β -catenin and RAS by targeting the Wnt/ β -catenin pathway. *npj Precis Oncol* 2018;2:5.
38. Sainsbury A, Goodlad RA, Perry SL, et al. Increased colorectal epithelial cell proliferation and crypt fission associated with obesity and roux-en-y gastric bypass. *Cancer Epidemiol Biomarkers Prev* 2008;17:1401–1410.
39. Hanley MP, Kadaveru K, Perret C, et al. Dietary methyl donor depletion suppresses intestinal adenoma development. *Cancer Prev Res* 2016;9:812–820.

Received August 28, 2020. Accepted April 13, 2021.

Correspondence

Address correspondence to: Douglas J. Winton, PhD, Cancer Research-UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, United Kingdom. e-mail: doug.winton@cruc.cam.ac.uk; or Edward Morrissey, PhD, MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DS, United Kingdom. e-mail: edward.morrissey@imm.ox.ac.uk.

Acknowledgments

The authors thank the Biobanking, Histology and Genomics cores at the CRUK Cambridge Institute for technical support. The authors acknowledge the contribution and support to this project provided by the Norwich Research Park BioRepository (Human Tissue Authority licence number 11208), a facility supported by the BBSRC, the Norfolk and Norwich University Hospitals NHS Foundation Trust, Not University of East Anglia. Equally, the

authors acknowledge contribution and support provided by the Addenbrooke's Human Research Tissue Bank that is supported by the NIHR Cambridge Biomedical Research Centre.

Sequencing data: NCBI BioProject PRJNA449344 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA449344>)

Preprint server DOI: <https://doi.org/10.1101/2020.07.10.193748>

CRediT Authorship Contributions

Cora Olpe, PhD (Formal analysis: Supporting; Investigation: Lead; Methodology: Equal; Writing – original draft: Lead; Writing – review & editing: Equal).

Doran Khamis, PhD (Formal analysis: Lead; Investigation: Supporting; Methodology: Equal; Writing – original draft: Supporting; Writing – review & editing: Supporting).

Maria Chukanova, MSci (Investigation: Supporting). Nefeli Skoufou-Papoutsaki, MRes (Investigation: Supporting; Writing – review & editing: Supporting).

Richard Kemp, PhD (Formal analysis: Supporting; Methodology: Supporting).

Kate Marks, MBChB, PhD (Methodology: Supporting; Resources: Supporting). Cerys Tatton, N/A (Investigation: Supporting).

Cecilia Lindskog, PhD (Resources: Equal).

Anna M. Nicholson, PhD (Writing – review & editing: Supporting).

Roxanne Brunton-Sim, BS (Resources: Equal).

Shalini Malhotra, PhD (Resources: Equal).

Rogier ten Hoopen, PhD (Resources: Supporting; Writing – review & editing: Supporting).

Rachael Stanley, PhD (Resources: Equal).

Edward Morrissey, PhD (Conceptualization: Equal; Formal analysis: Supporting; Writing – original draft: Supporting; Writing – review & editing: Supporting).

Douglas J. Winton, PhD (Conceptualization: Equal; Writing – original draft: Supporting; Writing – review & editing: Equal).

Conflict of interest

The authors disclose no conflicts.

Funding

This study was funded by a Wellcome Trust Grant (103805), Wellcome Trust PhD studentship (109141/Z/15/Z), MRC Computational Biology Fellowship (MC_UU_12025), and CRUK Cambridge Institute core funding (A24456).

Mathematical Background

Calculating the Crypt Fusion Rate

For individual crypts, fusion/fission events occur at a rate ρ and have a duration $\Delta\tau$. If we take a snapshot of a piece of tissue at a time t_0 we see all fusion/fission events that occurred in the window $[t_0 - \Delta\tau, t_0]$. Calculating the average number of events per crypt, X , in a time $\Delta\tau$ over many snapshots is the same as calculating the probability of an event for a single crypt in the window $\Delta\tau$ (as we can only have a single event in any time window equal to the event duration). The number of events for a single crypt follows a Poisson distribution,

$$X_1 \sim \text{Poi}(\rho\Delta\tau). \tag{1}$$

We want to observe events on the edge of mutant patches such that we can differentiate fission events from fusion events. For a patch with edge length N (that is, N crypts define the patch perimeter, each with at least one wild-type crypt as a neighbor), the number of crypts undergoing fusion or fission is distributed as

$$X_N \sim \text{Poi}(N\rho\Delta\tau). \tag{2}$$

For a given patch with edge length N , then, the probability of zero events in a window $\Delta\tau$ is

$$p\{X_N = 0\} = e^{-N\rho\Delta\tau} \sim 1 - N\rho\Delta\tau + o(N^2\rho^2\Delta\tau^2), \tag{3}$$

where $N\rho$ is considered small compared with $\Delta\tau$ such that we may define a parameter $\varepsilon = N\rho\Delta\tau$ where $\varepsilon \ll 1$. Correspondingly, the probability of seeing at least 1 event in a window $\Delta\tau$ is

$$p\{X_N \geq 1\} = 1 - p_0 \sim N\rho\Delta\tau + o(\varepsilon^2). \tag{4}$$

This equation can be applied to either fusion events or fission events separately by changing the event rate ρ to ρ_{fu} or ρ_{fi} , the event rates of fusion and fission, respectively, assuming that the event duration $\Delta\tau$ is approximately equal for fission and fusion.

Calculating the fusion rate. Let the number of partially mutant (partial) and fully mutant (monoclonal) FUFU events observed on the edge of mutant patches be n_p (termed M/W in main text) and n_m (termed M/M in main text), respectively. These numbers are combined over many different mutant patches and tissue samples, with a total patch edge length of N . To calculate the fusion rate ρ_{fu} given the fission rate ρ_{fi} , we use the following observations and assumptions:

1. All fission events are monoclonal.
2. Not all monoclonal events are fissions.
3. All partial events are fusions.
4. Not all fusions are partials.

The first and third assumptions stem from the belief that the timescale over which monoclonal conversion occurs in a crypt is short compared to the time between fusion/fission events. The second and fourth observations are alternative statements of the fact that fusion at the patch edge can happen inwards: a mutant crypt on the patch edge can fuse with a mutant crypt within the patch, hence creating a monoclonal event. We define the parameter χ to be the proportion of fusion events that are M/M. Then the ratio n_p/n_m may be written as

$$\frac{n_p}{n_m} = \frac{2n_{fu}(1-x)}{n_{fi} + 2n_{fu}x}, \tag{5}$$

where n_{fu} and n_{fi} are the number of fusion and fission events, respectively. The factor of 2 accompanying each instance of n_{fu} in (5) comes from the hypothesis that fusion can be initiated by either of the participating crypts, so the number of events associated with the fusion rate of any individual crypt should be half the number observed. We do not know n_{fu} and n_{fi} a priori, however.

We may recast the probability $p\{X_N \geq 1\}$ from (4) as the number of observed events over sample size, n_{events}/N , for fusion and fission, we get

$$\frac{n_{fu}}{N} \sim N\rho_{fu}\Delta\tau, \tag{6}$$

and

$$\frac{n_{fi}}{N} \sim N\rho_{fi}\Delta\tau, \tag{7}$$

where we have assumed the duration of a fusion event is approximately equal to that of a fission event. Thus, we find the approximate equivalence

$$\frac{n_{fi}}{n_{fu}} \sim \frac{\rho_{fi}}{\rho_{fu}} \tag{8}$$

between the ratios of numbers and rates of events. Using this, we may rewrite (5) as

$$\frac{n_p}{n_m} = \frac{2(1-x)}{\frac{\rho_{fi}}{\rho_{fu}} + 2x}, \tag{9}$$

and subsequently find an expression for the fusion rate.

$$\rho_{fu} = \frac{\rho_{fi}}{2\frac{n_m}{n_p}(1-x) - 2x}. \tag{10}$$

The simplest model for the parameter χ is to assume (isotropic) fusion, such that the proportion of fusions that are monoclonal is simply

$$x = \frac{N_t - N_w}{N_t}, \tag{11}$$

where N_t and N_w are the total number of neighbors and number of wild-type neighbors of a given fusion event,

respectively. To get a bulk measure of χ for each clonal mark, N_t and N_w were averaged over all observed M/M events.

Diffusion Model of Tissue Reorganization

In the following we lay out the theoretical framework and statistical methods used for understanding tissue rearrangement due to clonal expansion in the gut as a diffusion process.

Parameterization and derivation. We approach the idea of crypt packing by defining a quantity γ that represents the local stromal fraction of the tissue (we also refer to this as the “white space” fraction); then, the crypt area per unit area of mucosa is given by $1-\gamma$. If you look at a small region of tissue in cross-section, γ is the fraction of that region that is taken up by stroma rather than epithelial cells (ie, that fraction that is not part of a crypt). It is useful to think of the crypts as a density that is moving in the “free space” of the stroma. Then, we can define another quantity ψ that represents this density in such a way that $\psi \equiv 1/\gamma$. This density ψ tells us the number of units of area we would need to observe to see 1 unit of area of white space. This is a convenient definition, as it allows the density to be expressed as a function unbounded in the positive real numbers, whereas the local crypt fraction $1-\gamma$ is bounded in $[0,1]$.

In intestinal tissue, we assume there is a patient-specific homeostatic degree of crypt packing that can be represented by “ambient” values γ_a, ψ_a for the stromal fraction and crypt density, respectively. Near to a region of clonal expansion (a fission-driven mass source), the tissue is perturbed and the crypt density is altered such that

$$\psi(r, t) = \psi_a + \tilde{\psi}(r, t), \tag{12}$$

where the spatiotemporal variation in the crypt density is contained in the perturbation term $\tilde{\psi}(r, t)$. By centering our polar coordinate system $r = (r, \theta)$ at the initiation point of the clonal expansion we can state our far-field condition: $\tilde{\psi} \rightarrow 0$ as $r \rightarrow \infty$, meaning that we expect the tissue to remain in its ambient structure far from any perturbation. We choose to model the density perturbation $\tilde{\psi}$ as undergoing diffusive dynamics governed by the 2D diffusion equation.

$$\frac{\partial \tilde{\psi}}{\partial t} = \nabla \cdot (D \nabla \tilde{\psi}), \tag{13}$$

where the coefficient of diffusion D quantifies the speed with which the tissue can react to new mass being created by fission by rearranging to accommodate it. We assume D is isotropic and homogeneous such that (13) simplifies to

$$\frac{\partial \tilde{\psi}}{\partial t} = \nabla \cdot (D \nabla \tilde{\psi}). \tag{14}$$

We want to use (14) to understand the dynamical process underlying observed patches of clonally expanded mutant tissue; we note that although we know the initial size (a single crypt) and the current size (n mutant crypts) of the of the clonal expansion, we do not know the total age of the mutant patch or when the individual fission events driving the expansion occurred. We first show how to solve the system in the case of an instantaneous injection of mass at time $t=0$ before showing how to use this solution to build a more realistic model of dynamic fission events.

We approach (14) by taking a spatial Fourier transform such that the transformed density

$$\hat{\psi}(k_x, k_y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{\psi}(x, y, t) e^{-ik_x x - ik_y y} dx dy, \tag{17}$$

is governed by

$$\frac{\partial \hat{\psi}}{\partial t} = -D(k_x^2 + k_y^2) \hat{\psi}, \tag{18}$$

where $k = (k_x, k_y)$ is the wave vector in a Cartesian coordinate system (x, y) with its origin at the initiation point of clonal expansion. Equation (18) can be solved by integrating with respect to time and applying the initial condition $\hat{\psi}(0) = F[M \delta(r)]$, where M is the extra crypt area introduced by the injection of mass. We find

$$\hat{\psi}(k, t) = M e^{-D|k|^2 t}. \tag{19}$$

We proceed by inverting the Fourier transform to find the crypt density in terms of the spatial coordinates x and y . First, note that the inversion can be separated as follows:

$$\psi(x, y, t) = M \left(\frac{1}{2\pi} \right)^2 \left(\int_{-\infty}^{\infty} e^{-Dk_x^2 t + ik_x x} dk_x \right) \left(\int_{-\infty}^{\infty} e^{-Dk_y^2 t + ik_y y} dk_y \right). \tag{20}$$

Completing the square in the exponent of the x integrand we can recast the integral as

$$\int_{-\infty}^{\infty} e^{-Dk_x^2 t + ik_x x} dk_x = e^{-\frac{x^2}{4Dt}} \int_{u(-\infty)}^{u(\infty)} e^{-Dtu^2} du, \tag{21}$$

where $u = k_x + ix/2Dt$ and the integral on the right hand side of (21) can be evaluated:

$$\int_{u(-\infty)}^{u(\infty)} e^{-Dtu^2} du = \sqrt{\frac{\pi}{Dt}}. \tag{22}$$

By symmetry, the full solution is found to be

$$\psi(r, t) = \frac{M}{4\pi Dt} e^{-\frac{r^2}{4Dt}}, \tag{23}$$

where $r^2 = x^2 + y^2$ is the radial distance from the center of the clone. The stromal fraction near to a mutant patch may be expressed using (23) as

$$\gamma(r, t) = \frac{1}{\psi_a + \psi(r, t)} = \frac{\gamma_a}{1 + \frac{M\gamma_a}{4\pi Dt} E_d(r)}, \quad (24)$$

where for notational ease we have defined the exponential function.

$$E_d(x) = e^{-\frac{x^2}{4Dt}} \quad (25)$$

Diffusion with a stochastically firing point source. To investigate the temporal aspect of clonal expansion, the diffusion model was extended to accommodate growth over time due to stochastic crypt fission at the patch center. One can intuitively think of this as overlaying identical diffusion processes in space with different initiation times. Mathematically, this can be achieved by breaking the solution space into chunks separated by each fission event time. If the Fourier-space density $\psi \wedge_1(t)$ is the solution of (18) in a time interval $[t_0, t_1]$, achieving a value $\psi \wedge_1(t_1)$ at the end of this period, then the solution $\psi \wedge_2(t)$ for the next period $[t_1, t_2]$, where a fission event occurs at t_1 , is found by solving (18) with the initial condition $\psi \wedge_2(t_1) = \psi \wedge_1(t_1) + F[a_m \delta(r)]$ where a_m is the area of the new mutant crypt. Performing this process iteratively produces, with initial area perturbation $A_0 = a_m - a_w$ generated by the mutational hit, the compound diffusion solution

$$\psi(k, t) = A_0 e^{-D|k|^2 t} + \sum_{i=1}^{n_f} a_m e^{-D|k|^2 \tau_i}, \quad (26)$$

where $\tau_i = t - \sum_{j=1}^i t_j$ for $i \in [1, n_f]$ with t_k the event times of n_f crypt fissions. Inverting the Fourier transform gives the full spatial solution for the density perturbation:

$$\psi(r, t) = \frac{A_0}{4\pi Dt} e^{-\frac{r^2}{4Dt}} + \sum_{i=1}^{n_f} \frac{a_m}{4\pi D\tau_i} e^{-\frac{r^2}{4D\tau_i}}, \quad (27)$$

The formulation (27) supposes that we do not know how large a mutant patch will grow in a given amount of time t , but that we can generate fission times t_k and take those events with $t_k < t$ to define our solution up to the desired time t . We can so generate fission times given a fission rate ρ by assuming exponential waiting times (between events generated by a Poisson process) and using

$$t_n = -\frac{\ln(1 - u_{01})}{\rho n}, \quad (28)$$

where n is the current patch size, t_n is the time of the n th fission event, and u_{01} is a random draw from the unit uniform distribution $U(0, 1)$. This method can also be used to generate mutation times using the mutation rate α (9.3×10^{-7} for KDM6A; 1.7×10^{-6} for STAG2), the monoclonal accumulation rate ΔC (6.0×10^{-6} for KDM6A; 2.1×10^{-5} for

STAG2), the number of stem cells per crypt undergoing neutral drift N_{stem} (7) and the total number of crypts in the gut N_{crypts} ($\sim 10 \times 10^6$). So given a mutation that confers an expansion bias in terms of an increased fission rate over wild type epithelium, we can simulate the distribution of theoretical clonal expansion dynamics by generating a mutation time and a sequence of fission event times using (28) and calculate the evolution of the packing density over time using (27). The full theoretical white space fraction is given by inserting (28) into

$$\gamma(r, t) = \frac{1}{1/\gamma_a + \psi(r, t)}. \quad (29)$$

Comparing theory to experiment. Although (29) fully describes the theoretical diffusion process we posit as an explanation for the alleviation of crypt packing in the gut in lieu of mass crypt fusion, we must do more work to form a quantity suitable for comparison with experimental measurements. The data we have are measurements of the total area of patches of n crypts (including their “share” of the stromal space bordering the patch) and the area of the individual crypts making up each patch. Thus we can calculate the total white space Γ in a patch by subtracting the summed crypt areas from the total patch area. We can get a comparable theoretical quantity Γ_{th} by integrating the theoretical white space fraction (29) over the area A occupied by a patch P of n crypts:

$$\Gamma_{th} = \iint_{PA} \gamma(r, t) dA. \quad (30)$$

The integral (30) is nontrivial for an arbitrary patch geometry. To simplify the problem, we transform the geometry of the patch into a subsector of a circle centered at the coordinate origin while conserving the patch area.

To do this, 3 quantities must be found: the angle θ_s subtended by the subsector, its inner radius R_{in} , and its outer radius R_{out} . To find θ_s the distance d between the centroid of the patch and the centroid of the mutant source patch (the coordinate origin) is first calculated. Then,

$$\theta_8 = \begin{cases} 2a \tan \frac{R_w}{d}, & \text{if } d > 0, \\ 2\pi, & \text{if } d = 0, \end{cases} \quad (31)$$

where R_w is the radius of the patch being transformed (and where “radius” means the radius that would produce the area of the patch if the patch were a circle). Now, the area of a subsector is given by

$$A_{sec} = \frac{\theta_8}{2} (R_{out}^2 - R_{in}^2), \quad (32)$$

and area conservation enforces the equality $A_{sec} = \pi R_w^2$. If $d > 0$ we may use the approximation $R_{in} = d - cR_w$ and

$R_{out} = d + cR_w$ for some small $c > 0$ and substitute into (32) to find

$$c = \frac{\pi R_w}{2d\theta_s}. \tag{33}$$

Using (31) to fix the value for R in as

$$R_{in} = \begin{cases} d - \frac{\pi R_w^2}{2d\theta_s}, & \text{if } d > 0, \\ 0, & \text{if } d = 0, \end{cases} \tag{34}$$

we may then invoke area conservation once more to find

$$R_{out} = \sqrt{R_{in}^2 + \frac{2\pi R_w^2}{\theta_s}}. \tag{35}$$

We can now approximate the integral (30) with

$$\Gamma_{th} \approx \int_0^\theta \int_{R_{in}}^{R_{out}} \gamma(r, t) r dr d\theta, \tag{36}$$

which we evaluate numerically. The theoretical value for the total patch white space defined in (36) can then be used to fit the diffusion model given a data set of patch measurements.

Statistical model for inferring the diffusion coefficient. To infer the parameters defining the tissue-intrinsic diffusion process, we define the likelihood as

$$\Gamma^{(pq)} \sim N(\Gamma_{th}^{(pq)}(D, r_a^{(p)}, t^{(p)}, m^{(p)}, L^{(pq)}), \sigma\Gamma), \tag{37}$$

where $\Gamma^{(pq)}$ and Γ_{th} are respectively the observed and theoretical total white space in patch q of the neighborhood of the p th mutant patch (as defined in [36]). The parameters defining $\Gamma_{th}^{(pq)}$ are the single tissue-intrinsic diffusion coefficient D , the ambient stromal fraction γ_a , sequence of event times (mutation and subsequent fissions) $t^{(p)}$, and input areas $M^{(p)}$ (mutant and wild type crypt areas a_m, a_w) of the p th mutant patch, and the location parameters $L^{(pq)}$ which define the transformed inner and outer radii and angle subtended by the patch q of the neighborhood of the p th mutant patch. We simultaneously infer the coefficient of diffusion D and the local ambient white spaces γ_a .

We use a hierarchical model to constrain γ_a such that each value is drawn from a population distribution

$$\gamma_a^{(p)} \sim \text{Beta}(\alpha_{pop}, \beta_{pop}), \tag{38}$$

with hyperpriors

$$\alpha_{pop} \sim \text{Gamma}(10, 1), \beta_{pop} \sim \text{Gamma}(10, 1), \tag{39}$$

chosen to be uninformative around an unbiased mean value $\gamma_a^{pop} = 1/2$. Standard normal distributions bounded to the positive real line are used as prior distributions for

the diffusion coefficient and the standard deviation σ_Γ about the theoretical total white space,

$$D \sim N^{(+)}(0, 1), \sigma\Gamma \sim N^{(+)}(0, 1). \tag{40}$$

Inference for the full model defined by equations (37) to (40) was performed by MCMC sampling using the Stan probabilistic programming language. The sequence of event times $t^{(p)}$ an input to the model, generated using (28). For each clone 1000 potential sequences (trajectories) were generated and the inference was run 250 times with a random sample of the potential trajectories used each time. This allowed us to be sure that the resulting model parameters accounted for the large variation in the temporal development of the patches due to the stochastic nature of the mutational hit and the fissions themselves. From the resulting distributions, “most likely” trajectories were selected by fixing the diffusion coefficient D and ambient stromal fraction of each neighborhood γ_a to their median values and evaluating the likelihood of the data under the model for each of the 1000 potential trajectories per clone. For each clone, the 25 most likely trajectories were selected and averaged to give an approximation of the likely age of each patch.

To test whether we are justified in imposing the form of a diffusion process on the data we performed the inference above for a null model wherein each neighborhood around a mutant clone would be explained by a constant stromal fraction. This involved inferring γ_a for each patch and setting $\gamma(r) = \gamma_a$, and then calculating $\Gamma_{null}^{(pq)}$ and using it in (37) to evaluate the likelihood of the data under the null model. To compare the diffusion model to the null model, the full set of 1000 potential trajectories for each clone were used as input to calculate a distribution of model likelihoods given the inferred population median diffusion coefficient and the per-clone median ambient stromal fraction. The results show that there is more evidence to support the hypothesis of identifying the mutant patch as the source of clonal expansion causing a diffusion-like radial dependence in the local stromal fraction.

Polyp initiation. Here we quantify when the biological process might break down due to physical constraints on crypt density. In the theoretical scheme defined previously, as crypt density $\psi \rightarrow \infty$ the stromal fraction $\gamma \rightarrow 0$. Below some local value of γ , there will not be enough space for a crypt to fission. We posit that this may be a mechanism for outward growth, or polyp formation. A useful threshold to set on the available white space can be borrowed from the mathematics of optimal packing; identical circles may be optimally hexagonally packed to fill a fraction $\pi/\sqrt{12}$ of the space. Thus, we set the lower bound on the stromal fraction $\gamma_b = 1 - \pi/\sqrt{12}$. A mutant patch that creates mass through fission at such a rate that the tissue-intrinsic diffusion cannot act to fully accommodate new crypts will have a decrease over time of the available stromal fraction as the patch grows exponentially. We claim that if the available

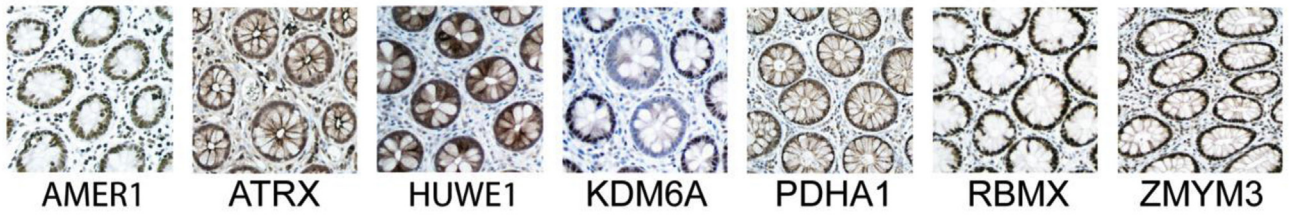
white space dips below γ_b then there the patch has a finite chance of initiating outgrowth.

We can quantify this more concretely by simulating an ensemble of mutant patches with a given fission rate. At each time point, we can find the radial extent r_i of each instance i of the mutant patch by constraining the integral over the nonstromal fraction to equal the total area of the n_i mutant crypts in the patch,

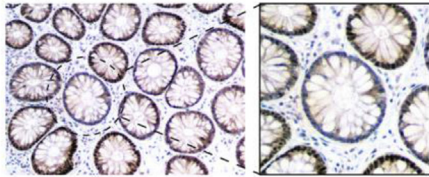
$$\int_0^{2\pi} \int_0^{r_i} (1 - \gamma(r)) r dr d\theta = n_i a_m, \quad (41)$$

where the mutant crypts have area a_m . To approximate the effective local stromal fraction throughout the patch, the cumulative average of $\gamma(r)$, $\bar{\gamma}(r)$ is calculated starting from the patch center $r = 0$. For each patch area πr_i^2 we may then calculate the fraction p_f of this area for which $\gamma < \gamma_b$. This gives us an idea of the portion of the patch that is at risk of initiating a polyp with the next fission event. Averaging p_f over the whole ensemble, we find an estimate for the probability of a clonal expansion developing into a polyp for a given time t after the initial mutation.

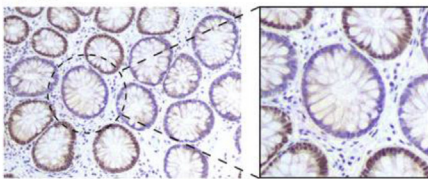
A



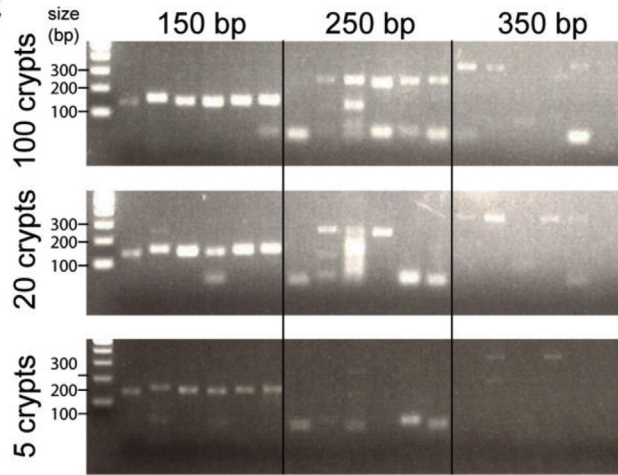
Bi



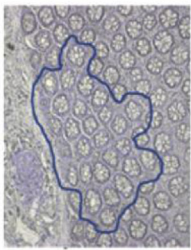
Bii



C



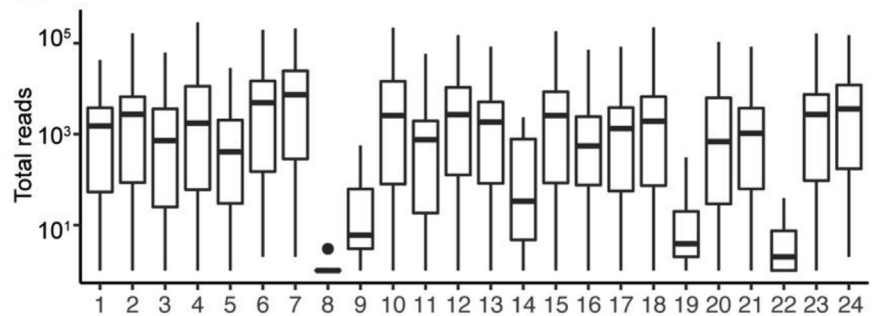
Di



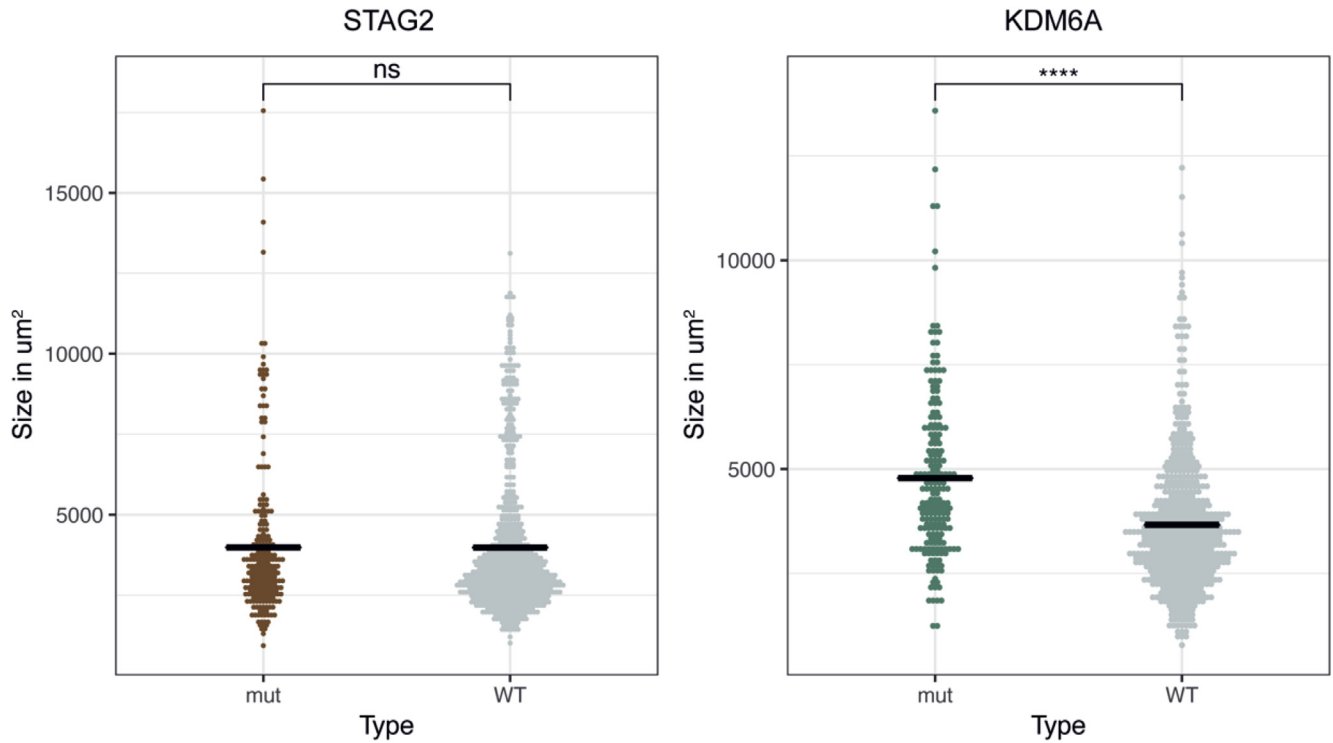
Dii



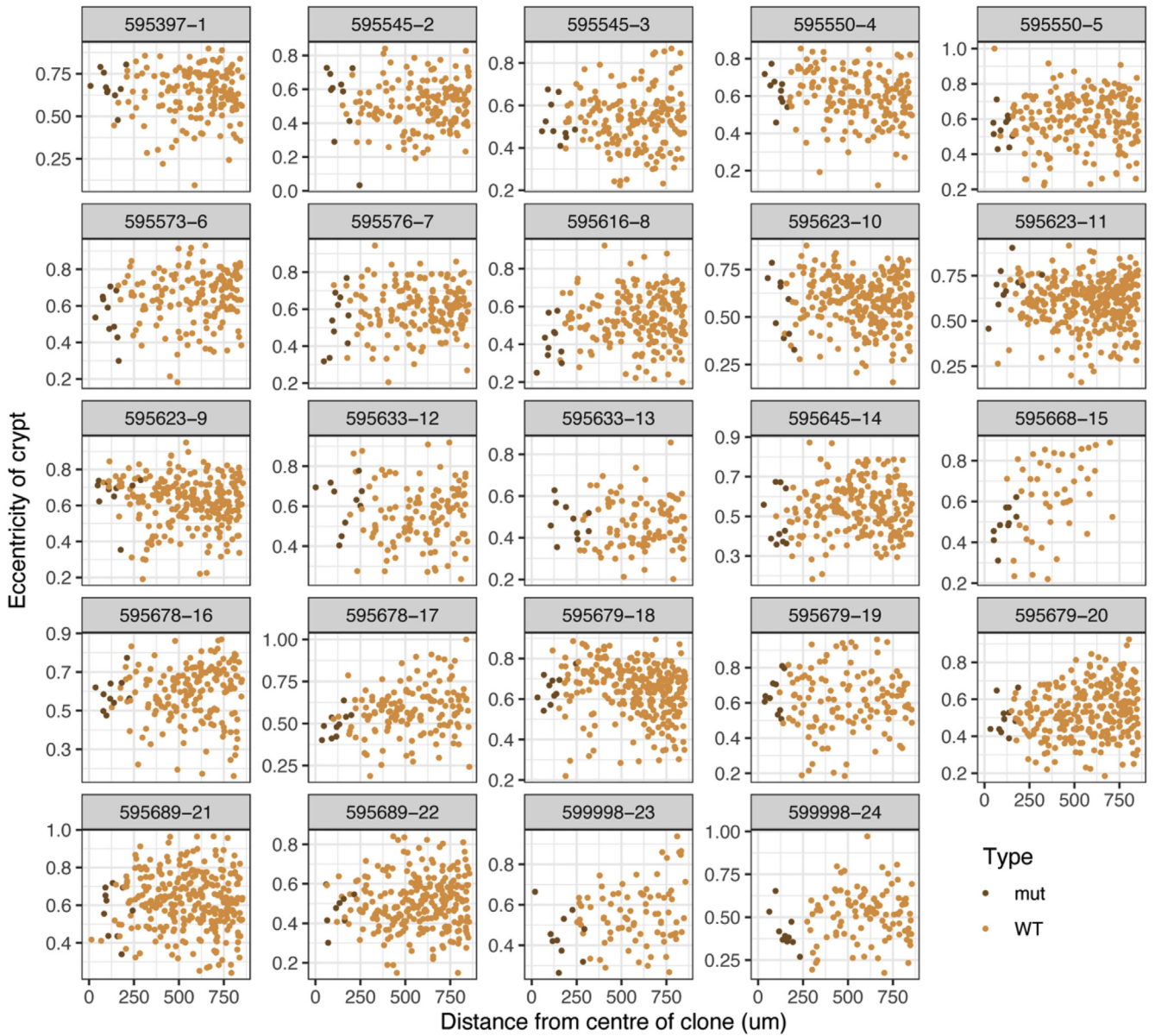
E



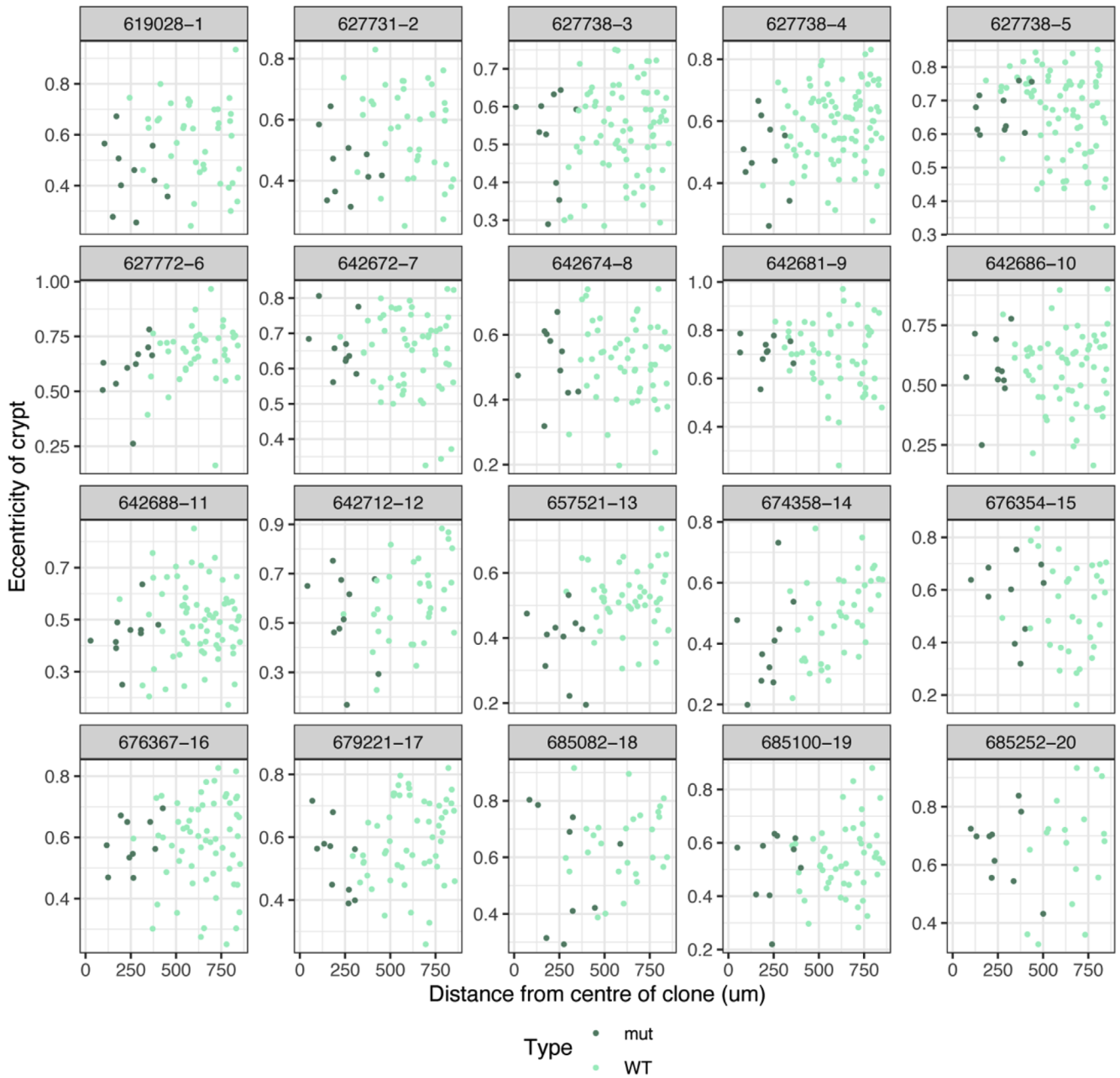
Supplementary Figure 1. Detection and laser capture of KDM6A⁻ clones. (A) Representative images of immunohistochemistry (IHC) for cancer-associated X-linked genes on human colonic sections. Of the 7 assessed genes, a staining pattern indicative of loss-of-function mutation was only found for KDM6A (*middle*). (B) Serial human colonic sections stained with 2 independent antibodies against KDM6A: (*i*) human protein atlas, (*ii*) Cell Signaling Technology. KDM6A⁻ crypts are circled and enlarged. (C) For assessment of optimal amplicon size, DNA extracted from FFPE sections was diluted to equivalents of 5, 20, and 100 crypts and used for PCR reactions (primers in [Supplementary Table 2](#)). Image of gel electrophoresis of PCR products on FFPE DNA diluted to equivalents of varying amounts of crypts. This revealed that an amplicon size of 150 bp ensures reliable amplification. (D) Image of IHC for KDM6A on laser capture slide (*i*) before and (*ii*) after laser capture microdissection. (E) Box plot showing reads obtained for each of the 24 amplicons used for sequencing of KDM6A⁻ and control patches. Amplicon libraries from 4 patches failed quality control. Of the remaining 7 patches, mutations were identified in 4 at mutant allele frequencies consistent with patient sex and estimated stromal fraction within the captured material ([Supplementary Table 4](#)). These results both confirm antibody specificity for KDM6A and the clonality of multicrypt patches that share only a single mutation.



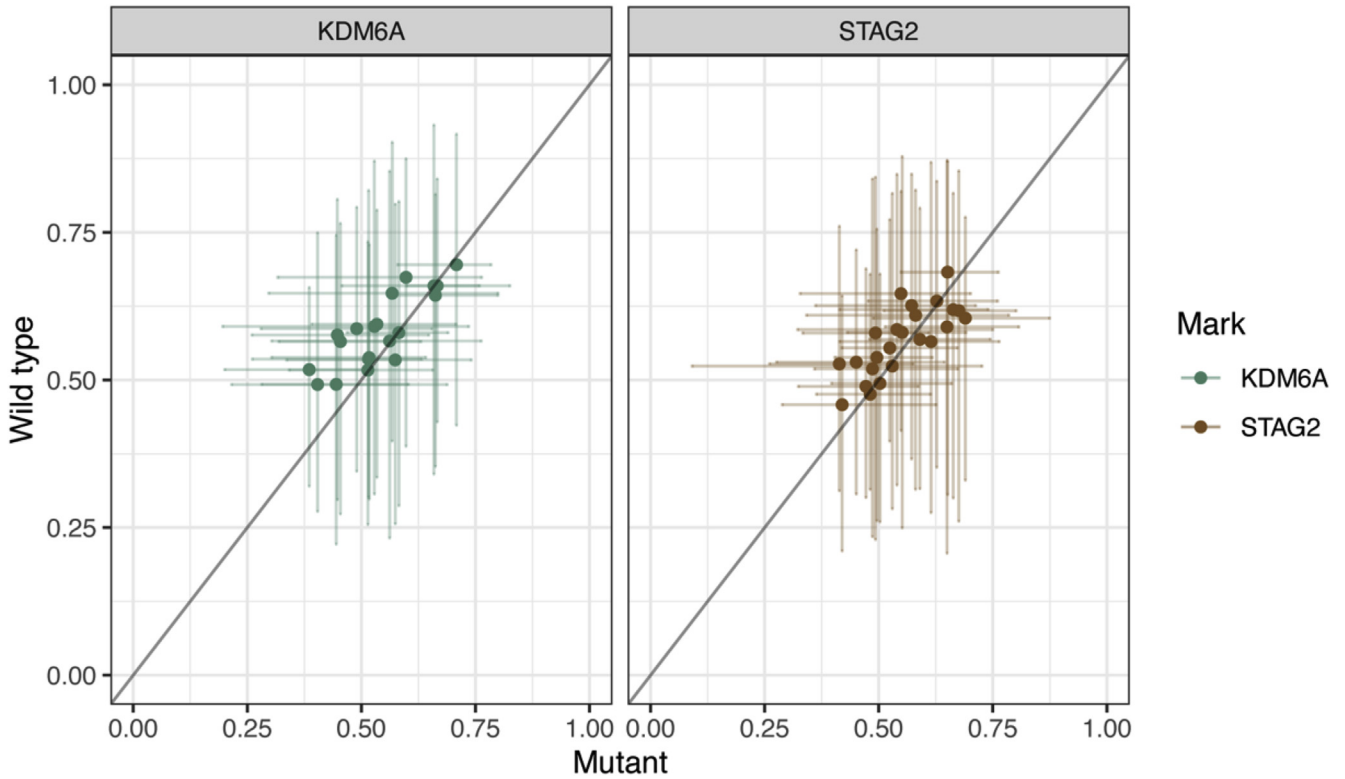
Supplementary Figure 2. Size measurement of STAG2⁻ and KDM6A⁻ crypts. Crypt areas in multicrypt clones containing 10 crypts that are STAG2⁻ or KDM6A⁻ (24 and 20, respectively). The area of individual crypts was determined using QuPath software. As controls, adjacent or nearby wild-type groups of 10 crypts were analyzed in the same way. Jitter plots show areas of crypts. Black line = median, **** $P < .001$.



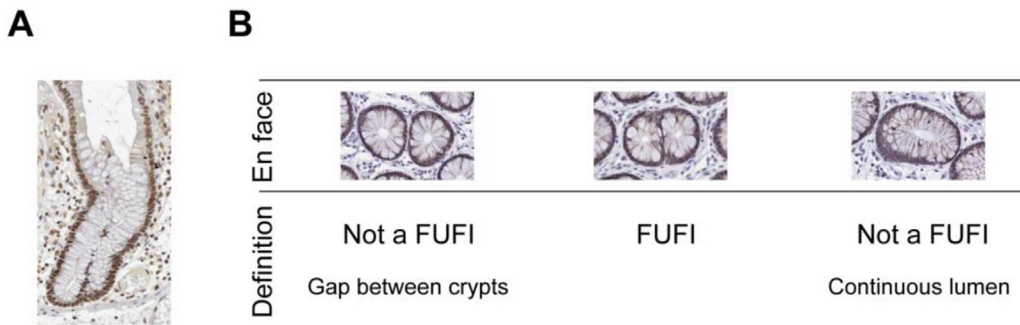
Supplementary Figure 3. Eccentricities of crypts around STAG2⁻ clones of size 10 as a function of radial distance from crypt centroid. Crypt squashing due to clone growth would impact wild-type (WT) (*orange*) crypts closest to the expanding patch, predicting a right to left decline in the crypt eccentricities in the plots above with distance from mutant (*brown*) crypts. This pattern is not observed.



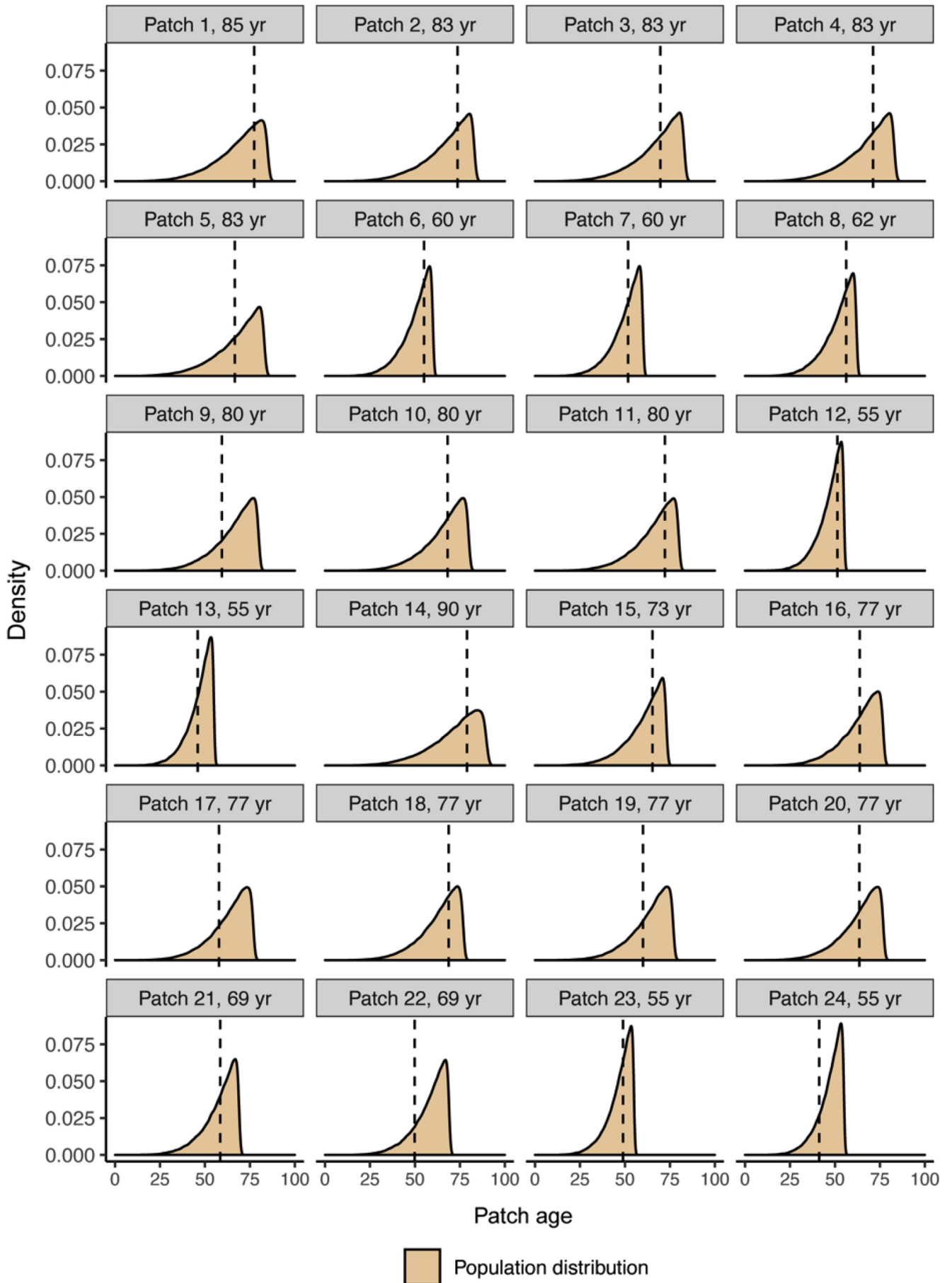
Supplementary Figure 4. Eccentricities of crypts around KDM6A^- clones of size 10 as a function of radial distance from crypt centroid. Crypt squashing due to clone growth would impact wild-type (WT) (*light turquoise*) crypts closest to the expanding patch, predicting a right to left decline in the crypt eccentricities in the plots above with distance from mutant (*teal*) crypts. This pattern is not observed.

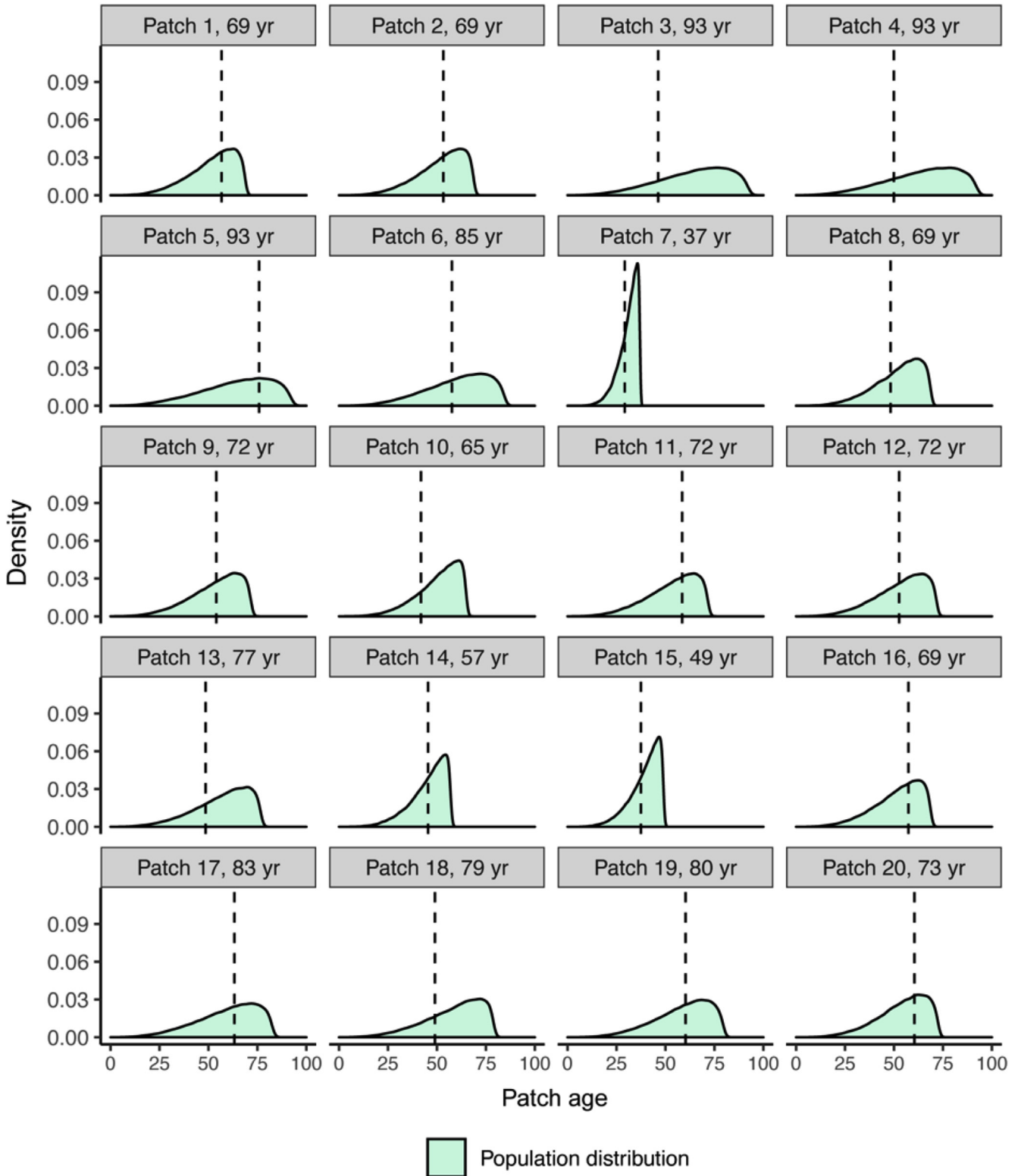


Supplementary Figure 5. Comparison of mean eccentricities of mutant and adjacent WT crypts in clones. For each assessed patch of 10 KDM6A⁻ or STAG2⁻ crypts, the mean eccentricity of the mutant crypts is plotted on the x-axis (0 = perfect circle) against the mean eccentricity of the surrounding wild-type crypts on the y-axis. The diagonal line represents x = y, where the mean eccentricities of mutant and surrounding WT crypts correspond. Error bars = 95% CI around the mean.



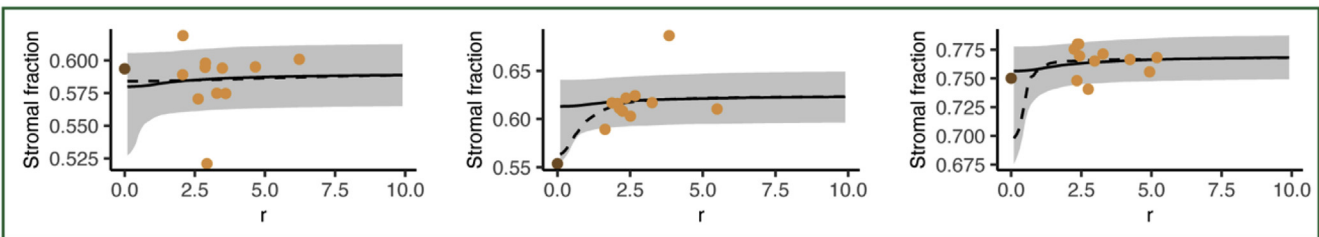
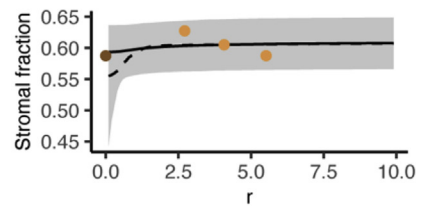
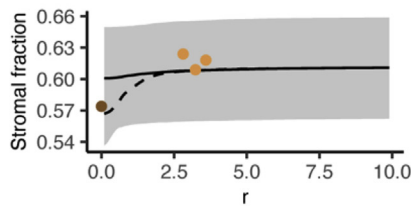
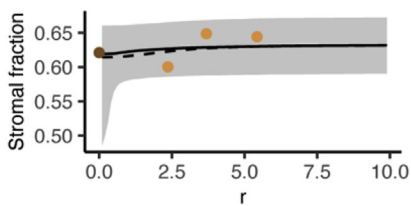
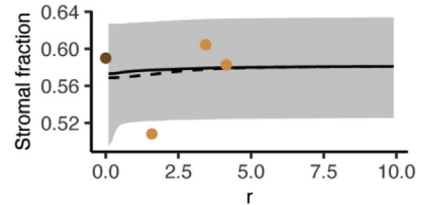
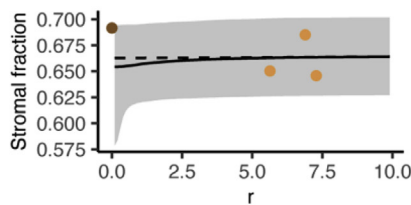
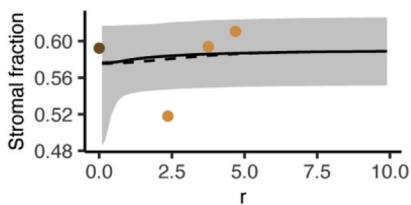
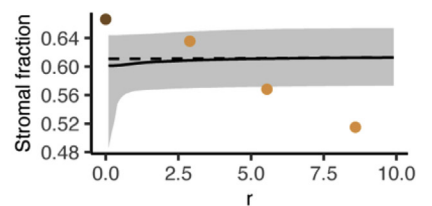
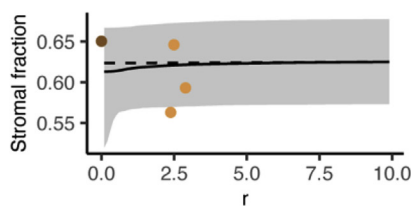
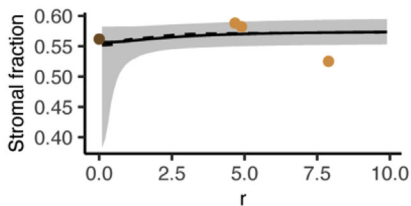
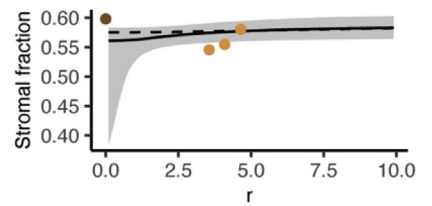
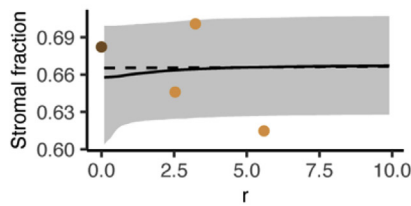
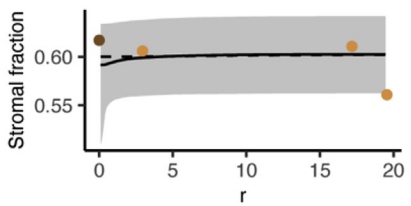
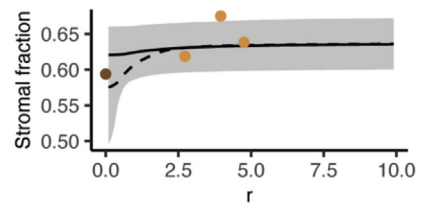
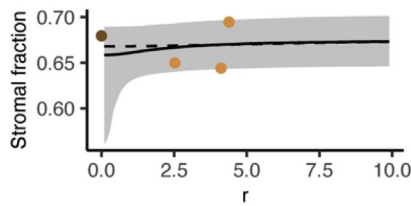
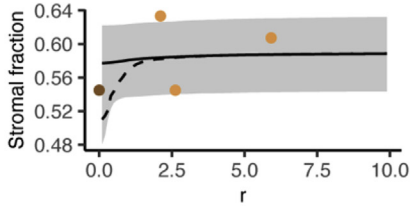
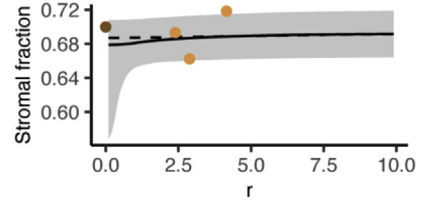
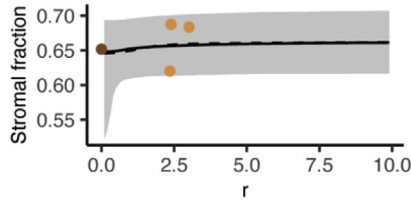
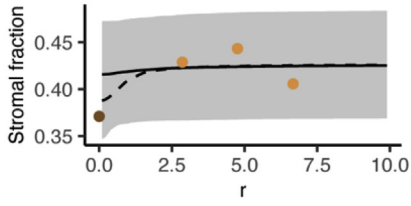
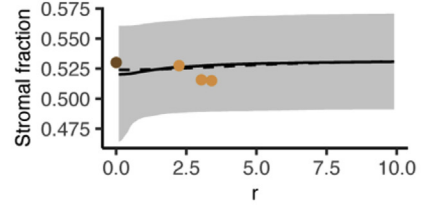
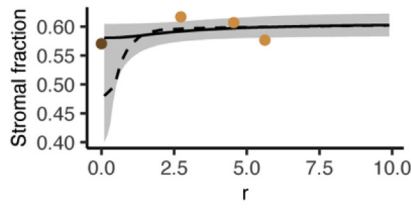
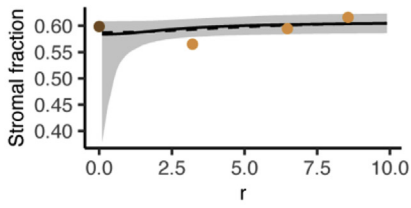
Supplementary Figure 6. Defining the FUF1. (A) Representative image of KDM6A-positive bifurcating crypts sectioned longitudinally. (B) A FUF1 is defined as 2 adjoining crypts viewed in a transverse section with 2 clearly discernible lumina lacking any separating gap.



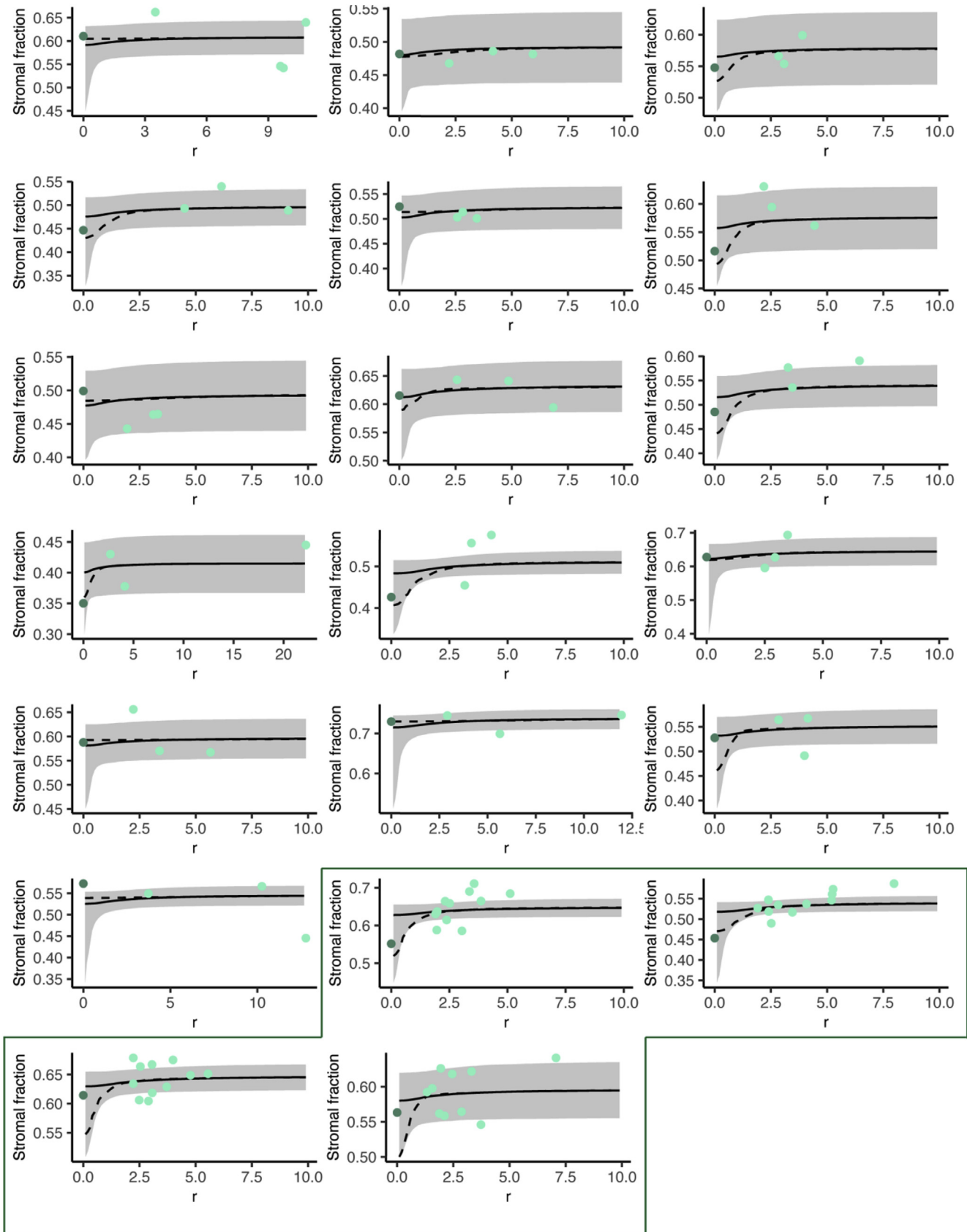


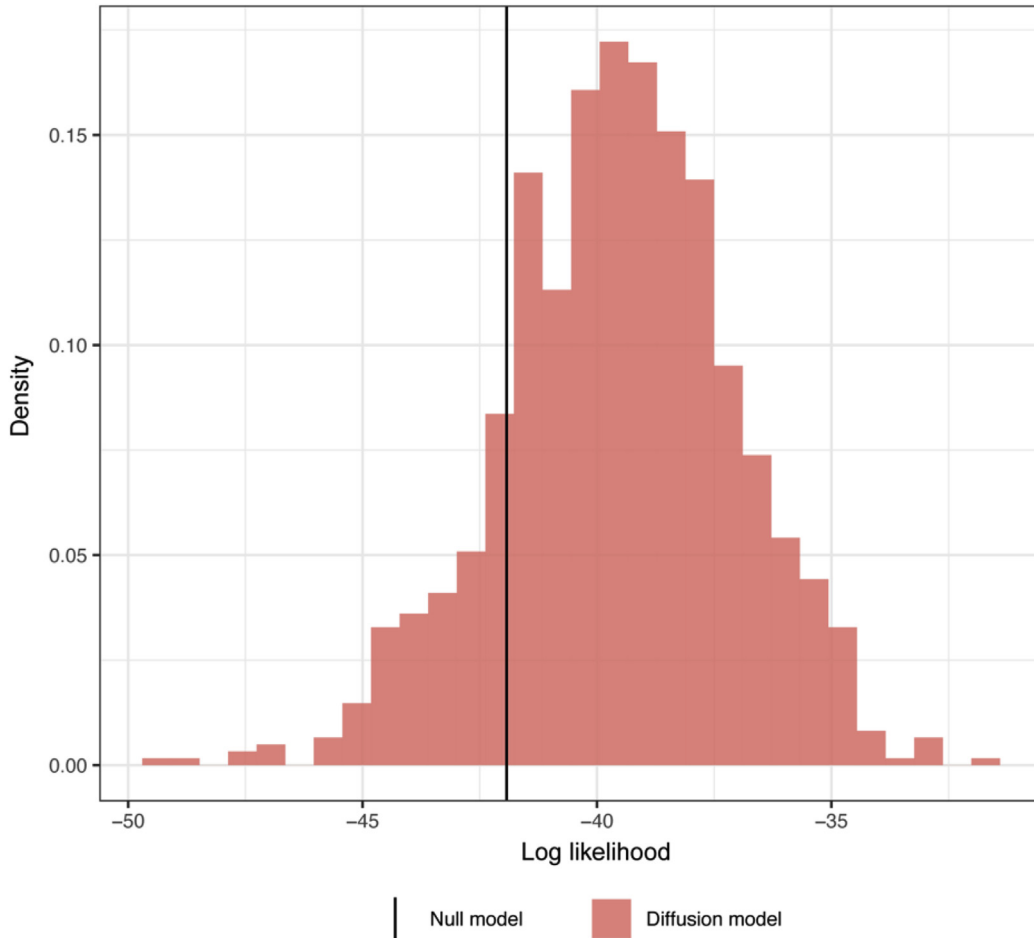
Supplementary Figure 8. Inferred ages of KDM6A⁻ clones. Theoretical distributions for the age of KDM6A⁻ clones in years covering the period from single crypt to a clone comprising 10 crypts. The theoretical density of patch age is calculated using the mutation rate, the monoclonal accumulation rate, and the fission rate (modeling fission as a stochastic birth process). *Dashed lines* show the most likely patch age for each clone derived from the average of the 25 most likely trajectories.

Supplementary Figure 7. Inferred ages of STAG2⁻ clones. Theoretical distributions for the age of STAG2⁻ clones in years covering the period from single crypt to a clone comprising 10 crypts. The theoretical density of patch age is calculated using the mutation rate, the monoclonal accumulation rate, and the fission rate (modeling fission as a stochastic birth process). *Dashed lines* show the most likely patch age for each clone derived from the average of the 25 most likely trajectories.



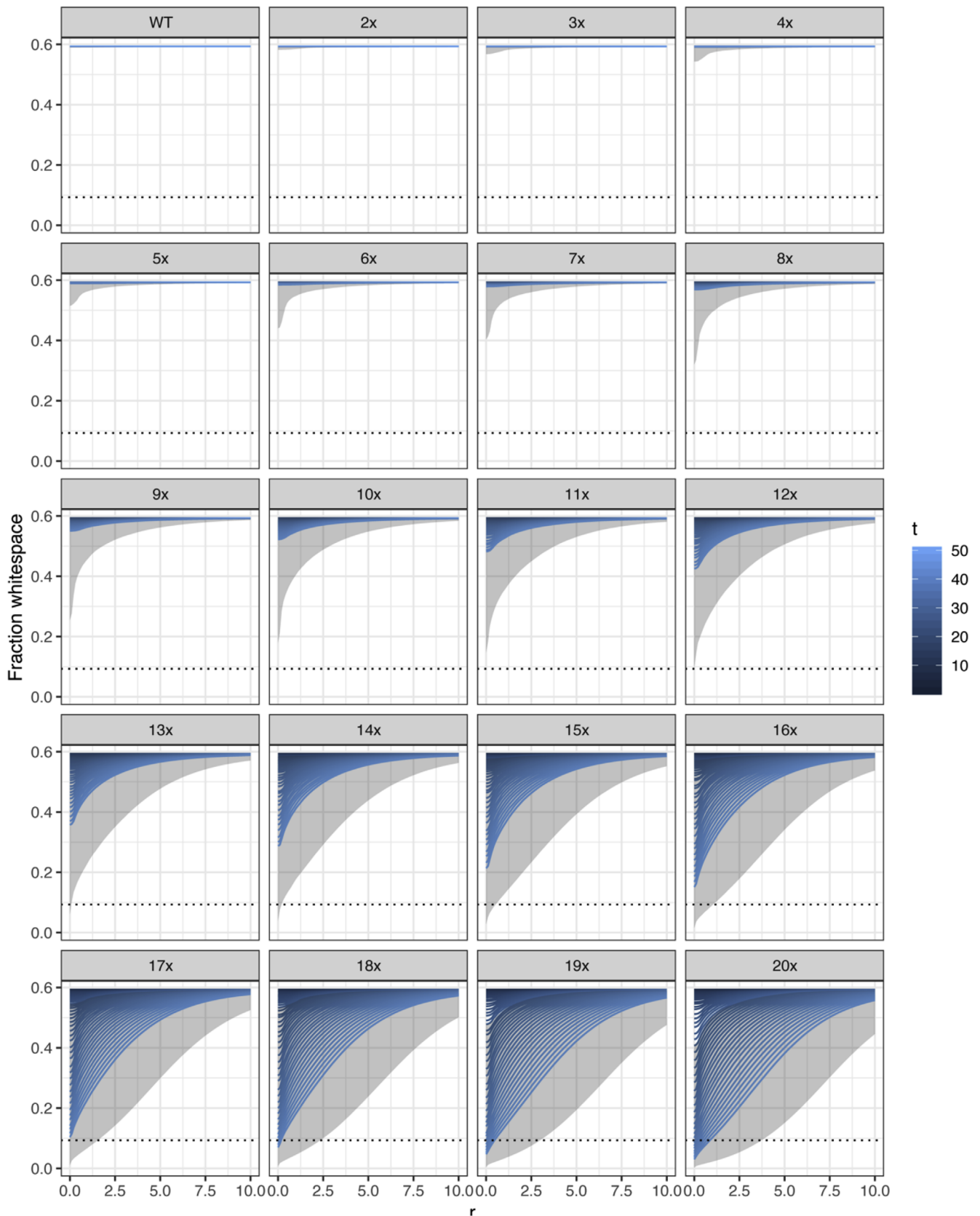
Supplementary Figure 9. Diffusion model prediction of stromal fraction changes for STAG2⁻ clones. Plots show the stromal fraction measured for STAG2⁻ as well as surrounding wild-type (WT) patches of 10 crypts. The radial distance r (in crypt domains) is measured from adjacent patch centroid to mutant patch centroid. The *black line* and *gray ribbon* represent the median and 95% CI theoretical stromal fraction as fitted from the diffusion model. *Dashed line* shows the solution from averaging the 25 most likely trajectories from initial mutation to clone of size 10, assuming population average diffusion and neighborhood ambient stromal fraction. *Green box*: patches for which a “rolling window” was applied. For each mutant patch, surrounding measurements included 2 areas comprising 3 mutant and 7 WT, 2 mutant and 8 WT, and 1 mutant and 9 WT crypts as well as 5 surrounding WT patches at varying distances (combined: 1 data point for the mutant patch and 11 for surrounding patches).



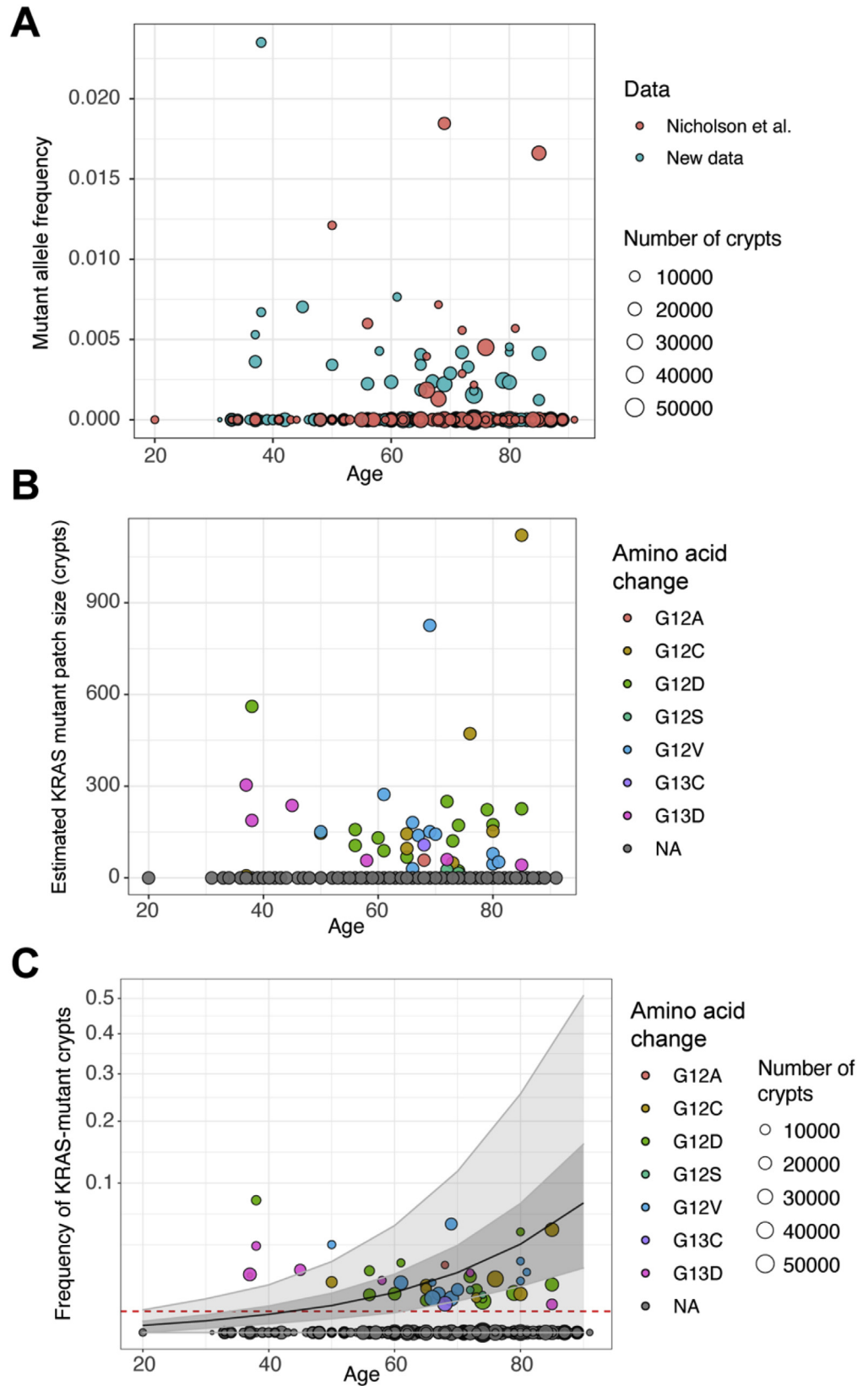


Supplementary Figure 11. Comparison of diffusion model to null model. Results for model fit comparison where the log likelihood of the data under a null model with constant per-clone stromal fraction is compared with the log likelihood of the data under the diffusion model for 1000 potential trajectories (from wild-type to 10-crypt mutant clone) per clone. The *black line* lies to the left of the density mass, meaning there is evidence for a diffusion-like radial dependence in the crypt packing data.

Supplementary Figure 10. Diffusion model prediction of stromal fraction changes for KDM6A⁻ clones. Plots show the stromal fraction measured for KDM6A⁻ as well as surrounding wild-type (WT) patches of 10 crypts. The radial distance *r* (in crypt domains) is measured from adjacent patch centroid to mutant patch centroid. The *black line* and *gray ribbon* represent the median and 95% CI theoretical stromal fraction as fitted from the diffusion model. *Dashed line* shows the solution from averaging the 25 most likely trajectories from initial mutation to clone of size 10, assuming population average diffusion and neighborhood ambient stromal fraction. *Green box*: patches for which a “rolling window” was applied. For each mutant patch, surrounding measurements included 2 areas comprising 3 mutant and 7 WT, 2 mutant and 8 WT, and 1 mutant and 9 WT crypts as well as 5 surrounding WT patches at varying distances (combined: 1 data point for the mutant patch and 11 for surrounding patches).



Supplementary Figure 12. Simulating the “breaking point.” Simulations were performed to find the fission rate at which clones may generate new crypts more quickly than can be accommodated by crypt diffusion. Line graphs show the stromal fraction resulting from crypt diffusion and different crypt fission rates in multiples of the homeostatic (wild-type) rate (0.7% per year). *Dotted line* = stromal fraction calculated from optimal hexagonal packing of circles. *r* = distance from centroid of patch in crypt domains. *Gray area* = 95% CI.



Supplementary Figure 13. Evidence for expansion of *KRAS*-mutant clones in human colon. Targeted amplicon sequencing focusing on *KRAS* codons 12 and 13 was performed on DNA from FFPE tissue sections from 256 patients of the age range 20 to 91 years. In total, 35 individuals displayed detectable mutations. (A) Mutant allele frequency data of *KRAS* codons 12 and 13 mutations plotted against age. Data previously published separated from new data obtained as part of this study. (B) *KRAS*-mutant patch sizes inferred from the mutant allele frequencies and known crypt numbers for 256 individuals plotted against age. (C) Frequency of *KRAS* mutant crypts for 256 individuals plotted against age. The mean accumulation of mutant clones using the model is plotted in *black* as well as the 95% CI in *gray*. *Red dotted line* shows detection threshold.

Supplementary Table 1. Primary Antibodies Used for Immunohistochemistry

Antigen	Antibody	Supplier	Titer
KDM6A	HPA002111	HPA	1:100
KDM6A	#33510	CST	1:200
MAOA	SC-271123	Santa Cruz	1:200
STAG2	LS-B11284	LSBio	1:1000

Supplementary Table 2. Primers Used for Amplification of FFPE to Assess Amplifiability

Primer	Sequence (5' - > 3')
MAOA_150_F1	ACCCATCAGTTACTCCTTCCC
MAOA_150_R1	GGGATTAAGCTGGGAGTTTCT
MAOA_150_F2	TAGCAGGGCCTTGAATCTGT
MAOA_150_R2	GATAGTGCCCAGAGTCACCA
STAG2_150_F1	GGAGAAGAAGACACAGTTGGATG
STAG2_150_R1	TTCTGTGAGGCATTTAGGGAAAA
STAG2_150_F2	CCTATGCTCGCACAACCTATGAG
STAG2_150_R2	GGAAGCCACACATCCTCTCT
CASD1_150_F1	ACCTGGAACCCTATGCTCAA
CASD1_150_R1	TGCAGCTATACATGCCAACC
LOC_150_F1	TGGTCTGCTTCATCCTCCTC
LOC_150_R1	GCCTAACATGCTTGGACCAC
MAOA_250_F1	TGCAAGTCTTAGGTTGGTTGC
MAOA_250_R1	TCAGTAATGGGTCATGTGCAAA
MAOA_250_F2	AAGACATGTAGGTTGGGGC
MAOA_250_R2	CAGAACACCCTGCTCTAACCT
STAG2_250_F1	GACTCTAAGGCCAGGTCAGG
STAG2_250_R1	GGAGGTGAGTTGTGGTGTCT
STAG2_250_F2	GCCTAATCATTCTCCCTGACCT
STAG2_250_R2	TGGTGTCAAAATCCATTCCCTC
CASD1_250_F1	GGTTAGAGGAAGACAAAAGTGG
CASD1_250_R1	CCTCAGTCCACACTTTGATACAC
LOC_250_F1	AGCTTACCTCTTTGTCTTCTCT
LOC_250_R1	CAACCTCAAAGTATCACGTGGA
MAOA_350_F1	TTCTTCAGAAATTGAATCCTTG
MAOA_350_R1	CCTGGGAGAAAGCAAAATCA
MAOA_350_F2	TCCCGGAGTATCAGCAAAAG
MAOA_350_R2	CATGAGAGACCCCAACAC
STAG2_350_F1	TCCGAATATTTTTGGTGCATT
STAG2_350_R1	CAGAGCCTTGATGAGTGCTG
STAG2_350_F2	TCTGAAGGAATGCTATGGTATGAA
STAG2_350_R2	TTGTCAAGGGTCATAGACACAA
CASD1_350_F1	CTTTGGGAAGCTTTGCGTAAAA
CASD1_350_R1	CGATTGAGGAGATGTAAGCCA
LOC_350_F1	TCAGGAATGATGGTCTACGTGA
LOC_350_R1	TCTCAGCTCTATTCCGTGAGT

Supplementary Table 3. Primers Used for Amplification of KDM6A

Primer	Sequence (5' - > 3')
1_F	ACACTGACGACATGGTTCTACACGCTTTCGGTGATGAGGAAA
1_R	TACGGTAGCAGAGACTTGGTCTCCGTACCTGTCCAGTCCG
2_F	ACACTGACGACATGGTTCTACATCTTTCAGGGCAATTAAGCATT
2_R	TACGGTAGCAGAGACTTGGTCTACAACCTACCTTTAAACTAGACTCA
3_F	ACACTGACGACATGGTTCTACAGTACAATTGGACCATGGCCA
3_R	TACGGTAGCAGAGACTTGGTCTAGTGACAGAGGTATTACTACAACCT
4_F	ACACTGACGACATGGTTCTACACAGGATGCCATTAATGCTACTT
4_R	TACGGTAGCAGAGACTTGGTCTTCTGGGGAAATATGTGGCTTT
5_F	ACACTGACGACATGGTTCTACAATGCTGTGTACATCCTCCA
5_R	TACGGTAGCAGAGACTTGGTCTACTTGTGTTGCTACCTCTACTCCT
6_F	ACACTGACGACATGGTTCTACATGACAGATGAGACCAACAGGA
6_R	TACGGTAGCAGAGACTTGGTCTCAGGCTGAGAGACGCTAGG
7_F	ACACTGACGACATGGTTCTACACTGCCTACAACTCAGTCTCTG
7_R	TACGGTAGCAGAGACTTGGTCTCAGAAAAGGGTCCATTGGCC
8_F	ACACTGACGACATGGTTCTACATAACCGCACAAAACCTGACCA
8_R	TACGGTAGCAGAGACTTGGTCTTCTCTCAAAGTGATAAAACCCAGT
9_F	ACACTGACGACATGGTTCTACACGACCTCTCTCTTCCACTGG
9_R	TACGGTAGCAGAGACTTGGTCTAATGCCTTGTGTCCACCTG
10_F	ACACTGACGACATGGTTCTACAGGCTGCTCTCAATCACCTCT
10_R	TACGGTAGCAGAGACTTGGTCTGCAGTGTCTGTTAGGTGTCTC
11_F	ACACTGACGACATGGTTCTACAGAGACACCTAACAGCACTGC
11_R	TACGGTAGCAGAGACTTGGTCTTCCCATCAACAAGGCAGAGA
12_F	ACACTGACGACATGGTTCTACAGCCATTCAACAGCAACACC
12_R	TACGGTAGCAGAGACTTGGTCTGGGGCTCTGAGATTCTTCCA
13_F	ACACTGACGACATGGTTCTACAGGAAGAATCTCAGAGCCCCA
13_R	TACGGTAGCAGAGACTTGGTCTCACACTAACCTGCATGCCTT
14_F	ACACTGACGACATGGTTCTACAATGGACTTGTGCAAATGCCTAGTAA
14_R	TACGGTAGCAGAGACTTGGTCTTGAGGTGGACATTTATCCAACAA
15_F	ACACTGACGACATGGTTCTACATGTTTTCTGAGATCTAACCACA
15_R	TACGGTAGCAGAGACTTGGTCTCAAGGCCACGTATTACTGTAACA
16_F	ACACTGACGACATGGTTCTACATGTAGAACACTAACTAGACTGCT
16_R	TACGGTAGCAGAGACTTGGTCTACACAGTATTAGAAACATGCCTTTT
17_F	ACACTGACGACATGGTTCTACAGTTCTGGGAGGAGGAGGAAA
17_R	TACGGTAGCAGAGACTTGGTCTAGCACAGGATAACTCTTTGCA
18_F	ACACTGACGACATGGTTCTACAAAACCTCCACAGGTATTTGTAGC
18_R	TACGGTAGCAGAGACTTGGTCTCCAACATGGCTTAGAAGATTTCC
19_F	ACACTGACGACATGGTTCTACAGTGAAGTTGCAGCTACATGA
19_R	TACGGTAGCAGAGACTTGGTCTTGCTCCCTGGAACCTTCATG
20_F	ACACTGACGACATGGTTCTACAACCGTGTGCTAACCAATTGC

Supplementary Table 3.Continued

Primer	Sequence (5' – > 3')
20_R	TACGGTAGCAGAGACTTGGTCTACAAACCATTACAGTCACCT
21_F	ACACTGACGACATGGTTCTACAGGAGCTTCTTAATGTAGTTGATCC
21_R	TACGGTAGCAGAGACTTGGTCTGCTGAATAAACCTATACACTGGAAC
22_F	ACACTGACGACATGGTTCTACACTAATGGGTTCTTGGTGGCC
22_R	TACGGTAGCAGAGACTTGGTCTTGAACCCAATGAACAGTGCC
23_F	ACACTGACGACATGGTTCTACAGCTGGTCACAAATAATTTCTCCC
23_R	TACGGTAGCAGAGACTTGGTCTTGGAGCTGGTTCTTCTTTGTCC
24_F	ACACTGACGACATGGTTCTACAACTTGGAAAACCTTTGTGGTGCT
24_R	TACGGTAGCAGAGACTTGGTCTACTGCTGCTTCATAACCCA

F, forward primer; Number, amplicon number; R, reverse primer.
NOTE. Sequence includes Fluidigm CS adapters.

Supplementary Table 4. Primers Used for Amplification of Area Around KRAS Codons 12 and 13 as Well as the Mimic Gene PITPNM2

Primer	Sequence (5' – > 3')
KRAS_1_F	ACACTGACGACATGGTTCTACATAAGGCCTGCTGAAAATGACT
KRAS_1_R	TACGGTAGCAGAGACTTGGTCTATGGTCCTGCACCAGTAATATG
MIMIC_1_F	ACACTGACGACATGGTTCTACATAGCACCCAGCCAGCTTG
MIMIC_1_R	TACGGTAGCAGAGACTTGGTCTATGACCACCCATGAAATATGAGCT
KRAS_2_F	ACACTGACGACATGGTTCTACAGGTGGAGTATTTGATAGTGTATTAACC
KRAS_2_R	TACGGTAGCAGAGACTTGGTCTTAGCTGTATCGTCAAGGCAC
MIMIC_2_F	ACACTGACGACATGGTTCTACACCTGCTGTTCTACAAAGCTG
MIMIC_2_R	TACGGTAGCAGAGACTTGGTCTAGGCTTCTCCCGTCTAAGGA

Note. Mimic amplicons contained the sequence context of KRAS codons 12 and 13, including 5 preceding and 3 subsequent nucleotides (total synchronous sequence: GAGCT GGTGGC GTA) at the same position within the amplicon. Sequence includes Fluidigm CS adapters.

F, forward primer; Number, amplicon number; R, reverse primer.

Supplementary Table 5. Polymerase Chain Reaction Components

Gene	Primers	dNTPs	5X Phusion HF Reaction Buffer (New England BioLabs),	Phusion High-Fidelity DNA Polymerase	Water
KDM6A	1 μ M	0.5 mM	5 μ L	1 U	To 25 μ L
KRAS	1.2 μ M pair 1 0.8 μ M pair 2	0.5 mM	5 μ L	1 U	To 25 μ L

Note. For KDM6A, pre-amplification was performed in multiplex groups (group 1: primers 1, 4, 7...group 2: 2, 5, 8... group 3: 3, 6, 9...). For KRAS, each sample was amplified in 2 independent duplex polymerase chain reactions containing a KRAS and corresponding mimic primer pair.

Supplementary Table 6. Polymerase Chain Reaction Cycling Conditions Used for Amplification of KDM6A and KRAS

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	2 min	1
Denaturation	95 °C	10 s	35
Annealing	60 °C	10 s	
Elongation	72 °C	15 s	
Final Elongation	72 °C	5 min	1

Note. Polymerase chain reaction was followed by ExoSAP-IT enzyme treatment (2 μ L of enzyme for 5 μ L of sample, ThermoFisher) at 37 °C for 15 minutes and 15-minute inactivation at 80 °C and 1:10 dilution in DNA Suspension Buffer (Teknova). For KDM6A, samples were further amplified using the Fluidigm Access Array according to the supplier's protocol.

Supplementary Table 7. Summary of Mutations Identified in KDM6A⁻ Patches

Gender	Sample	Genomic coordinate	Exon	Ref	Alt	Alt freq	Maximum alt freq	Change
Male	1A	X:45078391	intron 18	A	G	85.8	100	Intronic
Male	1B	X:45078391	intron 18	A	G	72.9	100	Intronic
Male	2A	X:45078391	intron 18	A	G	82.9	100	Intronic
Male	2B	X:45078391	intron 18	A	G	85.3	100	Intronic
Male	3A	X:45085892	24	C	A	51	100	S1154>STOP
Male	3B	X:45085892	24	C	A	0	100	N/A
female	4A	X: 45089773	25	G	A	15.8	50	W1193>STOP
female	4B	X: 45089773	25	G	A	10.4	50	W1193>STOP

Note. Frequencies derived from frequency of mutant reads in samples. Of note, the mutation found in intron 18 was present in 2 independent patches. Furthermore, the mutations S1154* and W1193* have previously been found in cancer samples (COSMIC database).

Alt, alternative nucleotide -> this is the mutation; Change, predicted consequence of the mutation; freq., frequency; Ref, nucleotide in reference genome.

Supplementary Table 8. Sequence Context Used to Extract Mutant Reads From Sequencing Data

Context	Nucleotide change	Amino acid change
GAGCTGGTGGCGTA	WT	WT
GAGCTGATGGCGTA	G>A	G12D
GAGCTGCTGGCGTA	G>C	G12A
GAGCTGTTGGCGTA	G>T	G12V
GAGCTAGTGGCGTA	G>A	G12S
GAGCTCGTGGCGTA	G>C	G12R
GAGCTTGTGGCGTA	G>T	G12C
GAGCTGGTGACGTA	G>A	G13D
GAGCTGGTGCCGTA	G>C	G13A
GAGCTGGTGTCGTA	G>T	G13V
GAGCTGGTAGCGTA	G>A	G13S
GAGCTGGTCGCGTA	G>C	G13R
GAGCTGGTTGCGTA	G>T	G13C
GAGCTGGAGGCGTA	T>A	WT
GAGCTGGCGGCGTA	T>C	WT
GAGCTGGGGGCGTA	T>G	WT
GAGCTGGTGGAGTA	G>T	WT
GAGCTGGTGGGGTA	G>C	WT
GAGCTGGTGGTGTA	T>C	WT

Note. For each variant, the mutant allele frequency was calculated by dividing by the total read number.