Published in final edited form as:
Nature. 2021 October 01; 598(7881): 510–514. doi:10.1038/s41586-021-03965-7.

Mutant clones in normal epithelium outcompete and eliminate emerging tumors

B. Colom[1], **A. Herms**[1], **M.W.J. Hall**[1,2], **S.C. Dentro**[1,3], **C. King**[1], **R.K. Sood**[1], **M.P. Alcolea**[4,5], **G. Piedrafita**[1,6], **D. Fernandez-Antoran**[1,§], **S.H. Ong**[1], **J.C. Fowler**[1], **K.T. Mahbubani**[7], **K. Saeb-Parsy**[7], **M. Gerstung**[3,8], **B.A. Hall**[9], **P.H. Jones**[1,2,*]

[1]Wellcome Sanger Institute, Hinxton CB10 1SA, UK

[2]MRC Cancer Unit, University of Cambridge, Hutchison-MRC Research Centre, Cambridge Biomedical Campus, Cambridge CB2 0XZ, UK

[3]European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge CB10 1SD, UK

[4]Wellcome-MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge CB2 0AW, UK

[5]Department of Oncology, University of Cambridge, Hutchison/MRC Research Centre, Hills Road, Cambridge Biomedical Campus, Cambridge CB2 0XZ, UK

[6]Spanish National Cancer Research Centre (CNIO), Madrid, Spain

[7]Department of Surgery and Cambridge NIHR Biomedical Research Centre, Biomedical Campus, Cambridge CB2 0QQ, UK

[8]European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany

[9]Department of Medical Physics and Biomedical Engineering, University College London, London, United Kingdom

Abstract

Human epithelial tissues accumulate cancer-driver mutations with age[1–9], yet tumor formation remains rare. The positive selection of these mutations argues they alter the behavior and fitness of proliferating cells[10–12]. Hence, normal adult tissues become a patchwork of mutant clones competing for space and survival, with the fittest clones expanding by eliminating their less-competitive neighbors[11–14]. However, little is known about how such dynamic competition in normal epithelia influences early tumorigenesis. Here we show that the majority of newly

[*]**Correspondence and requests for materials** should be addressed to P.H.J. pj3@sanger.ac.uk.
[§]Current address: Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK.

Author contributions: B.C. and M.P.A. designed experiments. B.C. performed the majority of experiments and data analysis. A.H. and M.P.A. contributed to mouse sample collection and performed preliminary experiments. M.W.J.H. conducted permutation analysis of leukocyte location. G.P. performed initial analysis of immune cell location in tumors and normal areas. D.F.A. analyzed human esophageal samples. B.C., M.W.J.H., C.K., R.K.S. and S.H.O. analyzed sequencing data. S.C.D. analyzed chromosomal copy number alterations under M.G. supervision. M.W.J.H. developed the mathematical modelling, under B.A.H. supervision. K.T.M., K.S.P., and J.C.F. collected human samples. B.C., M.W.J.H., and P.H.J. wrote the paper. P.H.J. supervised the research.

Competing interests: Authors declare no competing interests.

Europe PMC Funders Author Manuscripts

formed esophageal tumors are eliminated through competition with mutant clones in the adjacent normal epithelium. We followed the fate of nascent, microscopic, pre-malignant tumors in a mouse model of esophageal carcinogenesis, finding most are rapidly lost with no indication of tumor cell death, decreased proliferation, or an anti-tumor immune response. However, deep-sequencing of 10-day and 1-year-old tumors showed evidence of selection on the surviving neoplasms. Induction of highly competitive clones in transgenic mice increased early tumor removal, while pharmacologically inhibiting clonal competition reduced tumor loss. These results support a model where survival of early neoplasms depends on their competitive fitness relative to that of mutant clones in the surrounding normal tissue. Mutant clones in normal epithelium have an unexpected anti-tumorigenic role in purging early tumors through cell competition, thereby preserving tissue integrity.

Somatic mutations in cancer-driver genes are frequent in phenotypically normal adult human epithelia such as that of the esophagus, skin, endometrium, lung, bladder and colon[1–9]. Some of these mutations are more prevalent in normal tissue than in neoplastic lesions, arguing that mutant clones in normal epithelium might have an unanticipated protective role against cancer[2,5]. However, it is unclear whether the development of emerging neoplasms is impacted by competition with the strongly selected mutant clones in the surrounding normal epithelium during early tumorigenesis. Here we investigated this issue in a mouse model of esophageal carcinogenesis.

The mouse esophageal epithelium consists of a single basal layer of progenitor cells that divide continually (Extended Data Fig. 1a). Dividing cells have an equal probability of generating progenitor and differentiating daughters, which then leave the basal layer, migrate to the tissue surface and are shed. The balanced production of progenitors and differentiating cells maintains tissue homeostasis (Extended Data Fig. 1b, Supplementary Note)[15,16]. Some mutations result in an increase in the likelihood of generating progenitor cells, thus conferring a competitive advantage on mutant cells over their less fit neighbors (Extended Data Fig. 1c)[11,14].

Administration of diethylnitrosamine (DEN), a mutagen present in tobacco smoke, in drinking water induces diverse mutations in the mouse esophageal epithelium[13]. Cells with competitive “winner” mutations form clones that expand laterally in the proliferative basal cell layer until they collide with other mutants of similar fitness, at which point they return towards neutral drift (Extended Data Fig. 1d)[13]. With time, mutant clones in the normal epithelium undergo selection as they compete for the limited space in the basal layer, and generate a mutational landscape similar to that found in the aging human normal esophagus[13]. Most of the tissue remains phenotypically normal apart from a slight increase in cell density in the basal layer[13]. However, DEN treatment also leads to the formation of discrete esophageal premalignant tumors scattered across the tissue[11,17]. DEN administration thus constitutes an ideal model to study interactions between nascent tumors and mutant clones in the adjacent normal epithelium during early carcinogenesis.

Most esophageal tumors are lost over time

Tumors or neoplasms are defined as abnormal growths of tissue, and can be benign or malignant (Fig. 1a). To study the dynamics of emerging tumors interacting with mutant clones in the surrounding epithelium we first characterized the development of early tumors from 10 days post-DEN exposure, following their progression over the next 18 months (Fig. 1b). At multiple time points, the entire esophageal epithelium was collected, stained for nuclei as well as the stress marker protein Keratin 6 (KRT6), and 3D-imaged at single cell resolution by confocal microscopy (Fig. 1c, Supplementary Video 1). Tumors were identified from their abnormal morphology and strong KRT6 staining as compared to normal epithelium (Fig. 1c, Extended Data Figs. 2a-d)[17]. High levels of KRT17, a keratin induced in squamous tumors in humans and mice which regulates signaling and transcription were also observed in DEN-generated neoplasms (Extended Data Fig. 2e)[17–20]. This technique allowed the detection of hundreds of 'micro-tumors' per esophagus at 10 days post-DEN withdrawal (Fig. 1d). These were as small as ~30 μm in diameter, containing as few as ~20 cells, and exhibited mild histological features of dysplasia (Fig. 1c, Extended Data Figs. 2b, f). Despite their size, ~35% of these lesions displayed early signs of angiogenesis (Extended Data Figs. 2g-h; and Supplementary Video 2). Remarkably, the number of tumors fell rapidly during the following months and then stabilized (Fig. 1d, Supplementary Table 1). Persisting lesions continued growing in size and from three months onwards all tumors analyzed exhibited an intricate vascular network (Figs. 1d-e, Extended Data Figs. 2g, i; Supplementary Table 1, Supplementary Video 3). At 1 year post-DEN exposure, tumors showed signs of dysplasia characterized by marked hyperkeratosis and mild to moderate inflammation, and by 18 months squamous cell carcinoma was observed (Extended data Fig. 2f)[17,21,22]. We conclude that, while the persisting neoplasms acquire features of transformation, most microscopic tumors are resolved soon after they are formed (Figs. 1a, f).

The evolving tumor mutational landscape

To gain insight into how the early neoplasms are lost we compared the mutational landscapes of tumors collected 10 days and 1 year post-DEN exposure (Fig. 2a; Extended Data Figs. 3 and 4a). Deep-targeted sequencing (median on-target coverage 236x) of 192 genes implicated in epithelial carcinogenesis identified a total of 11,149 mutations from 141 tumors (Fig. 2b, Extended Data Fig. 4b, Supplementary Tables 2-3). The mutational spectrum comprised mainly T>A/A>T, T>C/A>G and C>T/G>A alterations, consistent with DEN mutagenesis in mouse esophagus (Extended Data Figs. 4c-d)[13]. Functionally, most mutations were protein altering, with missense SNVs being the most common (Extended Data Figs. 4e-f). Mutation types and the mutational spectrum were similar between tumors at both time points (Extended Data Figs. 4d-e).

Analysis of the maximum variant allele frequency (VAF) indicated that 10 day-old tumors were highly polyclonal, whereas the majority of surviving tumors at 1 year were monoclonal (Extended Data Fig. 4g), a result consistent with the selection of mutant clones by intra-tumoral competition[23,24]. To test whether selection of mutations was a feature of surviving tumors we calculated dN/dS (non-synonymous-to synonymous mutation) ratios, a

robust measure of selection in both tumors and normal tissues[2,25–27], across all sequenced genes, with values over 1 indicating positive selection[2,3,26]. Mutant *Notch1* and *Trp53* were positively selected in tumors at 10 days, whereas *Atp2a2, Notch1, Notch2, Chuk* and *Adam10* were selected at 1 year (Fig. 2c, Supplementary Tables 4-5). Whole exome sequencing of 1year tumors did not detect any additional selected genes (Extended Data Figs. 5a-g, Supplementary Tables 6-7). The number of positively selected mutant genes per tumor increased from 10 days to 1 year (Extended Data Fig. 4h). This was mainly due to mutations in *Atp2a2* and *Notch1,* which dominated clonal selection within tumors over time, exhibiting the highest gain in dN/dS values, number of mutations and estimated proportion of tumor area occupied by the mutant gene (Figs. 2c-f and Extended Data Figs. 4i, j). The selection of *Atp2a2* is notable as mice heterozygous for this gene develop spontaneous esophageal squamous cell carcinoma[28,29]. Whole exome and whole genome sequencing analysis found that chromosomal alterations were negligible in the persisting tumors collected at 9, 12 or 18 months, arguing that genome instability does not make a substantial contribution to the survival of DEN-generated tumors (Extended Data Figs. 5h-j).

Together, these results indicate that nascent tumors differ genetically from those persisting in the tissue at 1 year, reflecting intra-tumor clonal evolution and/or competitive selection of entire lesions, which may determine their survival or elimination from the tissue.

Potential mechanisms of early tumor loss

We next sought to identify the mechanism(s) responsible for early tumor loss. In line with our previous observations we found negligible activated Caspase-3 positive apoptotic cells in the DEN-treated normal epithelium (~0.04% of cells)[13]. Notably, no Caspase-3 positive cells were detected within the tumors, indicating that early neoplasms are not lost through apoptosis. (Extended Data Figure. 6a). We then assessed whether tumor loss was caused by a decrease in tumor cell proliferation. An EdU (5-ethynyl-2′-deoxyuridine) assay for cells in S phase of the cell cycle labeled a similar proportion of cells in the tumor and the surrounding normal epithelium (Extended Data Figs. 6b-c). In addition, the rate of cell division was measured in transgenic *Rosa26*^{M2rtTA}*/TetO-HGFP* mice which, upon treatment with the drug Doxycycline, express the stable Histone-Green Fluorescent Protein (HGFP)[30]. Once Doxycycline is withdrawn, HGFP expression falls at a rate dependent on the rate of cell division. Measurement of HGFP expression at different time points following Doxycycline withdrawal demonstrated that normal and tumor cells divide at a similar rate (Extended Data Figs. 6d-g, Supplementary Table 8). From these results we conclude that early neoplasms are not lost by tumor cell apoptosis or loss of proliferative cells.

Another potential mechanism of tumor elimination is by the immune system. However, although immune (CD45^{+}) cells were occasionally observed in close contact with the emerging tumors, their density and distribution pattern was no different than in the surrounding normal areas (Extended Data Figs. 6h-n). This suggests that early tumors are not detected by the immune system. Furthermore, immune-deficient *NOD.Cg-Prkdc*^{scid}*-Il2rg*^{tm1Wjl}/SzJ mice, which lack the typical anti-tumor immune response, showed similar tumor formation and loss rates as wild-type controls (Extended Data Figs. 6o-p). These

results argue that the immune system is implicated neither in the initiation nor the elimination of early tumors.

Epithelial mutant clones displace tumors

Within DEN-treated normal epithelium, highly competitive mutant clones expand and eliminate their less-fit neighboring clones[13]. This led us to hypothesize that emerging microscopic tumors may also be outcompeted by highly fit mutant clones expanding in the surrounding normal tissue, leading to tumor loss (Fig. 3a, Supplementary Note). For such a mechanism to be credible, competitive mutant clones should be present in a substantial proportion of normal epithelium adjacent to the early tumors. We therefore performed ultra-deep targeted exome sequencing of normal epithelium surrounding the 10-day tumors (median on-target coverage 220x), identifying 12,789 mutations in 81 micro-biopsies covering ~0.4mm^2 (Extended Data Figs. 7a-b, Supplementary table 9). The distribution of mutation effects and mutational spectrum were similar to those found in the 10-day tumors (Extended Data Figs. 7c-d). Hundreds of small mutant clones were detected per mm^2 of normal epithelium (Fig. 2g, Extended Data Fig. 7e). dN/dS analysis identified 3 genes, *Notch1, Trp53* and *Fat1* as positively selected, with *Notch1* mutations being the most common, estimated to be present in ~24 to 44% of normal cells (Extended Data Figs. 7f-h, Supplementary Table 10). This pattern of selection is highly reminiscent of that in aging human normal esophagus[2]. A comparative analysis of neoplastic and normal epithelium revealed that whereas nonsynonymous *Notch1* mutants were preferentially enriched in the normal tissue, *Atp2a2* mutants were more prevalent in tumors, a divergence that was exacerbated after a year post-DEN (Figs. 2h-i). Thus, consisting with the hypothesis, our results exposed strong competition and selection of highly fit mutant clones already taking place in the normal esophageal epithelium at 10 days post-DEN withdrawal. The data also suggest that different selective pressures may act on normal and neoplastic epithelium, influencing the resulting mutational landscape.

We next investigated whether the induction of highly fit mutant clones in the normal epithelium altered early tumor survival (Fig. 3a, Extended Data Figs. 8a-b, Supplementary Note). For this we used *Ahcre^{ERT}/R26$^{DNM\text{-}GFP/wt}$ (MAML-Cre)* mice, which carry an inducible highly competitive dominant negative mutant allele of *Maml-1 (DN-Maml1)* that inhibits Notch signaling (Extended Data Fig. 8c)[11]. *DN-Maml1* is fused to GFP, allowing the mutant expressing clones to be imaged. Induction of *MAML-Cre* mice 10 days post-DEN treatment accelerated tumor removal (Figs. 3b-c, Extended Data Figs. 8d-f). The loss of tumors was proportional to the area of epithelium occupied by *DN-Maml1* mutant clones (Fig. 3d, Extended Data Fig. 8g). Images of tumors being encircled by *DN-Maml1* or confetti clones identified a potential mechanism of tumor loss (Extended Data Fig. 9). Mutant clones in the normal epithelium may outcompete tumor cells at the basal layer. When the footprint of the tumor within this layer is displaced by the expanding normal clones the lesion will be shed (Fig. 3e, Extended Data Fig. 9c, g), a mechanism paralleling the extrusion of mutant cells described in other epithelia[31,32].

The results above are compatible with a model where tumors are eliminated by competition with highly fit clones in the adjacent normal epithelium (Extended Data Figs. 10a-b,

Supplementary Note). To challenge this hypothesis, we explored the effect of neutralizing clonal competition following tumor formation. We have shown that *Notch1* inactivating mutations confer the strongest competitive advantage of all mutant genes in DEN-treated esophageal epithelium[13]. Therefore, inactivating Notch signaling, by treating mice with Dibenzazepine (DBZ), would be expected to 'level up' the competitive fitness across the tissue and decrease tumor loss, as mutant epithelial clones would have no advantage over tumors (Extended Data Figs. 10c-e, Supplementary Note)[11]. The results show that DBZ treatment does indeed decrease tumor loss (Extended Data Figs. 10f-g). We conclude that expansion of mutant clones in the normal epithelium removes early tumors through competition, acting as a selective pressure on tumor survival (Extended Data Figs. 10h-i, Supplementary Note).

Discussion

Humans are protected by highly efficient innate mechanisms to battle tumors, exemplified by the immune system. In this study, we have identified an additional tumor-protective mechanism, independent of the immune system, by which expanding highly fit mutant clones present in the normal epithelium outcompete and eliminate early neoplastic lesions. These studies in the context of carcinogenesis complement transgenic experiments that reveal a role for wild type cells in the elimination of less fit mutant cells and epithelial lesions[12,31,32]. Importantly, the results imply that the survival of early tumors does not depend solely on the mutations they carry, but also on the mutational landscape of the neighboring normal tissue. These findings help reconcile the seemingly conflicting observations of a relatively low cancer incidence despite the large burden of cancer-associated mutations observed in healthy tissues[1–9], and suggest that some mutations found in normal tissues may contribute to the maintenance of homeostasis, with clear implications in cancer and aging.

Whether an analogous mechanism of micro-tumor elimination exists in humans is unknown. However, there are parallels with the mouse model described here. As in the mouse, strongly selected Notch1-mutant clones colonize the majority of human normal esophageal epithelium[2,5,13]. Furthermore, rare, small neoplasms phenotypically similar to those described in the mutagenized mouse model are seen in otherwise healthy human esophagus (Extended Data Figs. 10j-k). These findings hint that the micro-tumor elimination by mutant clones in normal epithelium may be conserved in humans. Whether this mechanism acts in other human epithelia carrying mutant clones under competitive selection such as skin, endometrium, lung and colon[1,2,6–8] will require further investigations.

Our results show that rather than being viewed solely as the potential origin of cancer, mutant clones in normal epithelium also have a tumor suppressive role by eliminating emerging micro-tumors. Treatments that manipulate the mutant clonal architecture of normal epithelium might thus have a role in early cancer prevention[12,14].

Methods

Mice

All animal experiments were conducted at the MRC ARES facility, Babraham UK or the RSF facility at the Wellcome Sanger Institute according to the UK Home Office Project Licenses 70/7543, P14FED054 or PF4639B40. The endpoint limits permitted by these licenses were not exceeded in any of the experiments. Endpoints for mouse experiments were predetermined time points based on previous experiments. Mice were housed in individually ventilated cages, fed on standard chow and maintained at SOPF health status. Male and female adult mice were used for experiments. C57BL/6N wild-type mice and *NOD.Cg-Prkdc*^{scid} *Il2rg*^{tm1Wjl}*/SzJ*[33] mice, which lack typical anti-tumor immune responses including a complete absence of B and T lymphocytes, NK cells, and defective dendritic cells and macrophages, were obtained from core colonies bred at the RSF facility. Double mutant *R26*^{M2rtTA}*/TetO–HGFP* and *Ahcre*^{ERT}*/R26*$^{DNM–GFP/wt}$ strains on a C57BL/6N background were generated as described previously[11,15,17]. R26Cre^{ERT2} mice (from core colonies bred at the RSF facility) were crossed with R26^{flConfetti} mice[17] to generate the double mutant R26Cre^{ERT2}/R26^{flConfetti}.

Chemical model of mouse esophageal neoplasia

Adult (>10 week old) mice were treated with Diethylnitrosamine (DEN) in sweetened drinking water (40 mg per 1,000 ml) for twenty-four hours, three days a week (Monday, Wednesday and Fridays) for eight weeks[11,17]. Animals received sweetened water in between DEN dosages and normal water after DEN treatment was completed. Esophageal samples were collected from 10 days to 18 months after DEN treatment.

Mouse whole-esophagus sampling and processing

To prepare the esophagus for whole-tissue staining and imaging, tissues were dissected and placed in cold PBS. The esophageal tube was cut longitudinally and the muscle layer removed by pulling gently with micro-forceps. To separate the epithelium from the underlying submucosa the entire tissue was incubated for 2–3 hours in 5 mM EDTA at 37 °C and both layers were pulled away under a dissecting microscope with fine forceps. The whole esophageal epithelium was then flattened, fixed in 4% paraformaldehyde for 30 min at room temperature, washed in PHEM buffer (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, and 4mM MgSO4·7H20) and finally stored at 4 °C until further processing.

Mouse whole-esophagus immunostaining

For immunostaining, whole-tissue esophageal epitheliums were blocked for one hour in 800 μl of staining buffer (0.5% bovine serum albumin, 0.25% fish skin gelatin, 0.5% Triton X-100 in PHEM buffer and 10% donkey serum) followed by an overnight incubation at room temperature with primary antibodies against KRT6, KRT17, GFP, CD31, CD45 or Caspase 3, in staining buffer. Samples were then washed four times (x20 minutes) with 0.2% Tween-20 in PHEM buffer and, when necessary, incubated with secondary antibodies in staining buffer for three hours at room temperature. Samples were finally washed as above, incubated overnight at room temperature with 1 μg/ml DAPI or 0.4 μM TO-PRO™-3 Iodide

Europe PMC Funders Author Manuscripts

solution to stain cell nuclei, and mounted using VECTASHIELD Mounting Media before image acquisition.

Human samples

Esophagi were obtained from deceased organ donors. Informed consent for the use of tissue was obtained from the donor's family (REC reference: 15/EE/0152 NRES Committee East of England - Cambridge South). A full thickness segment of mid-esophagus was excised within 60 minutes of circulatory arrest and preserved in University of Wisconsin (UW) organ preservation solution (Belzer UW® Cold Storage Solution, Bridge to Life, USA) at 4 °C until further processing.

Immunostaining of human esophagus

Donor esophageal samples were cut longitudinally, flattened and the muscle layer removed. Samples were blocked for one hour in 1mL of staining buffer (0.5% bovine serum albumin, 0.25% fish skin gelatin, 1% Triton X-100 and 10% donkey serum in PHEM buffer), followed by two days incubation at room temperature with primary antibody against KRT6 in staining buffer. Samples were then washed for 2 days in PHEM buffer before incubation with secondary antibody in staining buffer for two days at room temperature. Samples were washed as above and incubated overnight at room temperature with 0.4 μM TO-PRO™-3 Iodide solution to stain cell nuclei. Finally, samples were incubated for one hour at 37 °C and then mounted in RapiClear 1.52 (SUNJin lab) solution, for image acquisition.

Confocal microscopy

To obtain high quality confocal images of the entire mouse esophagus, whole-mounted tissues were imaged in a Leica TCS SP8 confocal microscope coupled to a high precision motorized stage. Contiguous 3D images of all epithelial layers (basal + suprabasal) were obtained and merged computationally using the mosaic function of the Leica Software. Typical settings for acquisition of multiple z stacks were 2.5μm z-step size, zoom x1, optimal pinhole, line average 4, scan speed 400 Hz and a resolution of 1,024 * 1,024 pixels using a 10X HC PL Apo CS Dry objective with a 0.4NA. Visualization and image analysis was performed using IMARIS (bitplane), ImageJ or Volocity 3D Image Analysis Software (Perkin Elmer). Otherwise, close up images were obtained with ×10, ×20 or ×40 objectives with typical settings for acquisition of z stacks being optimal pinhole, line average 3–4, scan speed 400-600 Hz and a resolution of 512 x 512 or 1,024 × 1,024 pixels.

KRT6 staining as an early marker of mouse esophageal epithelial tumors

KRT6 is a keratinocyte stress-induced protein upregulated at both RNA and protein levels in mouse esophageal tumors[17]. We used KRT6 staining of whole-tissues to help identify DEN-generated neoplasms in the esophageal epithelium from as early as 10 days post-DEN treatment. To characterize the protein expression levels in the tumors relative to the surrounding normal epithelium, the KRT6 staining profile in and around tumors was analyzed from esophagi collected 10 days after DEN treatment. The intensity of KRT6 and the nuclear marker DAPI were measured with ImageJ from defined ROIs containing tumors

or their normal surrounding epithelia. The percentage of KRT6 positive or negative tumors was measured at different time points post-DEN withdrawal.

Tumor identification and quantification of number and size of tumors

Whole-tissues were stained with antibodies against KRT6 and a nuclear marker such as DAPI or TO-PRO™-3. High definition 3D confocal images of the whole esophagus were obtained and visualized with IMARIS software or ImageJ. Tumors were identified on the basis of altered topology of cell nuclei and positive KRT6 staining. The number of tumors quantified in the entire esophagus was divided by the total area of the tissue to obtain the number of tumors per mm^2 of esophageal epithelium. The size of individual neoplasms was calculated by assuming an ellipse shape using the following formulae: $A = R_1 R_2 \pi$, where R_1 and R_2 are the length of the radius for the major and minor axis as measured with IMARIS software.

Tissue histology

Esophagus from DEN-treated mice were dissected, fixed in 10% formalin for at least 24 hours and stored at 4°C. Tissues were then embedded in paraffin and cut at 5 μm thickness. Sections were stained with hematoxylin and eosin and scanned.

Analysis of tumor angiogenesis

Whole mount esophageal epithelia (including the submucosal layer) were stained with Dapi and antibodies against CD31 and KRT6. Confocal images were obtained with a x40 objective and analyzed with IMARIS software. Tumors were classified as angiogenic when they presented a complex vasculature surrounding them or show small vessel sprouts pointing towards them.

Targeted sequencing of 10 days and 1 year post-DEN esophageal tumors

Tissue preparation—Mice were treated with DEN for eight weeks and culled 10 days or 1 year after treatment. Mouse esophagi were dissected, cut longitudinally and the muscle layer removed. Tissues were incubated for 2–3 hours in 5 mM EDTA at 37 °C to separate the epithelium from the underlying submucosa. The whole epithelium was flattened, fixed in 4% paraformaldehyde for 30 min at room temperature and washed in PHEM buffer. Tissues were then immuno-stained for KRT6 and the nuclei labelled with DAPI or TO-PRO™-3 to identify the tumors before being imaged by confocal microscopy as described above.

Excision and digestion of tumors—Esophageal tumors from animals culled 10 days or 1 year post-DEN treatment were identified from the confocal images and manually cut using micro-scalpels under a fluorescent dissecting microscope. Tumors were placed in 0.2 ml collector tubes containing 20 μl lysis buffer (30 mM Tris-HCl pH = 8.0, 0.5% Tween-20, 0.5% NP-40 and 0.056 μAU Proteinase K) and kept on ice until digested at 56 °C for 60 minutes followed by 76 °C for 30 minutes, before being finally stored at -80 °C. For each animal, one or more pieces of ear skin were cut and processed as above to be used as germline controls.

Library preparation—Samples were processed using a recently developed enzymatic fragmentation-based library preparation method specifically designed to detect small amounts of DNA (low-input) without the need for whole genome amplification[1,34]. Briefly, genomic DNA was purified by mixing 20 μl of the lysed samples with 50 μl Ampure XP beads (Beckman Coulter), and left for 5 minutes at room temperature. After magnetic separation, bead-DNA samples were washed twice with 75% ethanol and re-suspended in 26 μl TE buffer (10 mM Tris-HCl, 1 mM EDTA). For DNA fragmentation, samples were mixed with 7 μl 5x Ultra II FS buffer and 2 μl Ultra II FS enzyme (New England BioLabs), and incubated 12 minutes at 37 °C followed by 30 minutes at 65 °C. Samples were then incubated with a 30 μl ligation mix, 1 μl ligation enhancer (New England BioLabs), 0.9 μl nuclease-free water and 0.1 μl duplexed adapter (100 μM; 5′-ACACTCTTTCCCTACACGACGCTCTTCCGATC*T-3′, 5′-phos-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG-3′). Adapter-ligated libraries were then purified by mixing with 65 μl Ampure XP beads (Beckman Coulter) followed by magnetic separation and elution with 65 μl TE buffer. For the PCR reaction, 21.5 μl of the DNA samples were mixed with 25 μl KAPA HiFi HotStart ReadyMix (KAPA Biosystems), 1 μl PE1.0 primer (100 μM; 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T-3′) and 2.5 μl iPCR-Tag (40 μM; 5′-CAAGCAGAAGACGGCATACGAGATXGAGATCGGTCTCGGCATTCCTGCTGAACCGCTC TTCCGATC-3′), in which ‘X’ represents one of 96 unique 8-base indexes. PCR conditions were: 5 min at 98 °C, followed by 12 cycles of 30 s at 98 °C, 30 s at 65 °C and 1 min at 72 °C, and finally 5 min at 72 °C. Amplified libraries were purified by mixing with Ampure beads at a ratio 0.7:1, adjusted to 2.4nM with nuclease-free water.

DNA sequencing—A custom bait capture to target the exonic sequences of 192 genes frequently mutated in cancer was designed using Agilent SureDesign. The list of genes selected for ultra-deep targeted sequencing was: *abcb11, abcc2, adam10, adam29, adcy10, aff3, ajuba, akap9, akt1, akt2, apob, arid1a, arid2, arid5b, asxl1, atm, atp2a2, atrx, aurka, b2m, bbs9, bcas3, bcl11b, bcr, braf, brca2, c1s, cacna1d, card11, casp8, ccnd1, cdc16, cdh1, cdkn2a, chuk, clgn, cnot1, cntnap4, cobll1, col12a1, copb2, cr2, crebbp, csmd2, ctcf, cul3, cyld, cyp2b13, dclk1, dclre1a, ddr2, dicer1, dmxl2, dnm2, dnmt3a, dst, egfr, eif2d, ep300, epha2, erbb2, erbb3, erbb4, ezh2, fat1, fat4, fbn2, fbxo21, fbxw7, fgfr1, fgfr2, fgfr3,* f*lg2,* f*lt3, fn1, gcn1l1, grin2a, grm3, gtf3c5, hist1h2bm, hmcn1, hras, huwe1, hydin, igsf1, insrr, iqgap1, irf6, kcnh5, kdm5b, kdm6a, kdr, keap1, kit, klrc3, kmt2c, kmt2d, kras, krt5, krtap4-9, loxhd1, lrp1, lrp1b, lrp2, ltf, maml1, mcm7, met, mrgprb4, mtor, myof, nf1, nf2, nfe2l2, nfkb1, nlrp12, notch1, notch2, notch3, notch4, nras, nsd1, nup214, opn3, pard3, pcdha5, pced1b, pde4dip, peg10, pign, pik3ca, pkhd1, plcb1, prex2, psme4, ptch1, pten, ptprt, rac1, rasa1, rb1, rbm46, rhbdf2, ripk2, ripk4, ros1, rpgrip1, rpl10, ryr2, sall1, scn10a, scn11a, scn1a, scn3a, setd2, setx, sgk3, sis, slc13a1, smad4, smarca4, smo, snx25, soat2, sox2, spen, st18, sufu, synm, taf2, tas2r102, tet2, tnr, trp53, trp63, trrap, tsc1, ttc27, usp24, usp26, usp9x, vhl, vmn2r81, vps13b, wnk1, zan, zfhx3, zfp39, zfp457, zfp521, zfp644* and *zfp750*.

Samples were multiplexed and sequenced on the HiSeq 2500 platform (Illumina) according to the manufacturer's instructions, with the exception that we used iPCR-Tag (5′-AAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTC-3′) to read the library index. Paired-end reads were aligned with BWA-MEM (v0.7.17, https://github.com/lh3/bwa)[35] with optical and PCR duplicates marked using Biobambam2 (v2.0.86, https://gitlab.com/german.tischler/biobambam2, https://www.sanger.ac.uk/science/tools/biobambam).

Mutation calling with deepSNV—In order to identify somatic mutations present in a small fraction of cells within the samples, we used the ShearwaterML algorithm from the deepSNV package (v1.21.3, https://github.com/gerstung-lab/deepSNV) to call for mutational events on ultra-deep targeted data[2,3,36]. Instead of using a single-matched normal sample, deepSNV uses a collection of deeply-sequenced normal samples as the reference for variant calling that enables the identification of mutations at very low allele frequencies. For this study we used a total of 19 normal samples from 12 animals, providing a normal panel with a combined coverage of 7439x. A depth filter of 50x was used in the final variant call results.

Mutation burden—The average number of mutations per Mb in a given sample was estimated from the VAFs and the number of bases within the bait set with synonymous mutations (dubbed as synonymous footprint), as described before[2,3], and assuming no alterations in copy number as indicated by either the whole exome and genome sequencing data (Extended Data Figs. 5h to j). Only the synonymous mutations were used for this calculation as the mutation burden can be inflated by the alterations in strongly selected genes when using targeted sequenced data.

Percentage of mutant epithelium—The size of mutant clones within each sample can be calculated by taking into account the area of the biopsy and the fraction of cells carrying a mutation within the sample, as described previously[2,3]. We have determined that DEN treatment does not alter the copy number variation (Extended Data Fig. 5h to j), therefore, the fraction of cells carrying a particular mutation can be estimated as twice the VAF of that mutation. The lower (=VAF) and upper (=2xVAF) bound estimates of the percentage of tissue covered by clones carrying non-synonymous mutations in a given gene was calculated for each sample[2,3]. The fraction of epithelium covered by the mutant genes was then calculated from the mean of summed VAF (capped at 1.0) of all the samples.

Targeted sequencing of normal esophageal epithelium at 10 days post-DEN

Tissue preparation—Mice were treated with DEN for eight weeks and culled 10 days after treatment. Tissues were fixed, stained and imaged by confocal microscopy as described above. Samples of normal epithelium, matching the 10 day tumor sizes, were collected, processed for sequencing and analyzed as above.

Generation of the patchwork summary plot—A graphical summary representation of the mutant clones detected in the normal esophagus is shown in Figure 2g. The figure summarizes the number, density and estimated sizes of mutant clones (each shown as a circle). Only non-synonymous mutations in the 11 genes under positive selection found in

10 day or 1 year tumors or normal epithelium[13] are represented (shown in different colors). The generation of the plot was adapted from previous publications[2,3]. Briefly, the number, density and estimated sizes of the clones are based on the sequencing data. All mutations were assumed to be heterozygous. The nesting of clones and subclones is informed by the sequencing data whenever possible or randomly simulated when the data is uninformative, particularly when involving small subclones. The placement of clones in space is randomly simulated.

Mutational spectra and signature analysis

Mutational spectra for single base substitutions were plotted and compared to 65 known mutational signatures present in human cancers[37] using linear decomposition with the deconstructSigs R package (https://github.com/raerose01/deconstructSigs)[38]. The mutational spectra across DEN-treated samples were highly consistent, precluding deconvolution into separate signatures (either known or *de novo).*

Gene selection (dN/dS)

We used the maximum-likelihood implementation of the dNdScv algorithm (v0.0.1.0, https://github.com/im3sanger/dndscv) to identify genes under positive selection[26]. dNdScv estimates the ratio of non-synonymous to synonymous mutations across genes, controlling for the sequence composition of the gene and the mutational signatures, using trinucleotide context-dependent substitution matrices to avoid common mutation biases affecting dN/dS. Values of dN/dS significantly higher than 1 indicate an excess of non-synonymous mutations in that particular gene and therefore imply positive selection, whereas dN/dS values significantly lower than 1 suggest negative selection. In our experimental set up, only mutations that reach a minimum size are detected by deepSNV. Therefore, a significant value of dN/dS>1 indicates that a clone that acquires a non-synonymous mutation in that particular gene will have a higher probability to reach a detectable size as compared to a synonymous mutation in the same gene. Hence, genes with dN/dS>1 are considered drivers of clonal expansion.

Whole exome sequencing of surviving esophageal tumors

Sample preparation and imaging—Mice were treated with DEN in drinking water 3 times a week for 8 weeks as described above. 9 or 18 months after DEN removal mice were culled and the esophagus harvested. Tissues were incubated for 2–3 h in 5 mM EDTA at 37 °C before removing the submucosa from the epithelium as described above. The esophageal epitheliums were stained with Dapi and antibodies against KRT6, and then imaged by confocal microscopy to identify tumors. The projected area of the tumors was measured using Volocity 3D Image Analysis Software.

Tumor isolation and sequencing—49 tumors from 16 mice were manually cut under a fluorescent micro-dissecting scope (Leica Microsystems) using ultra fine forceps and micro-scalpels. Individual tumors were collected in low binding DNA tubes and digested in 3 μl RLT buffer (Qiagen Cat# 1048449) for 30min at room temperature. Digested samples were diluted 1:10 in water, separated in triplicates, transferred to 96-well plates and incubated 15 min at room temperature with Agencourt AMPure XP magnetic beads

(Beckman Coulter Cat# A63881) at a 1:1 ratio. Beads with bound DNA were separated with a magnet and washed 3 times with 70% ethanol. DNA was re-suspended in 10 µl elution buffer and transferred to a new plate. Whole genome DNA was amplified using 1 µl polymerase enzyme from the illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Cat# 25-6600-32) and 9 µl of sample with the following conditions: 95 °C for 3 min, 4 °C for 5 min, 30 °C for 1.5 hours and 65 °C for 10min. DNA was then purified by mixing with beads at a 1:0.6 DNA/beads ratio followed by 3 washes with 70% ethanol and eluted with 30 µl of elution buffer (Qiagen Cat# 19086). Whole-exome sequencing was performed using the Mouse_Exome_Targets baitset from the Wellcome Sanger Institute pipeline. Captured material was sequenced on Illumina HiSeq 2500 sequencers using paired-end 75bp reads, to an average cumulative coverage of 1130x (Extended Data Fig. 5b).

Mutation calling and sequencing analysis—Substitutions were called using the CaVEMan (Cancer Variants through Expectation Maximization) variant caller (version v1.11.2, http://cancerit.github.io/CaVEMan)[39]. Insertions and deletions were called using cgpPindel (version v2.2.4, http://cancerit.github.io/cgpPindel)[40]. Mutations were annotated using VAGrENT (v3.3.3, https://github.com/cancerit/VAGrENT)[40]. Mutations occurring outside of the exome were removed by retaining only mutations annotated with the effects 'missense', 'nonsense', 'ess_splice', 'frameshift' or 'silent'. The mutational spectra and signature of single clones was analyzed as described above.

Copy number analysis of whole exome sequenced tumors

The whole exome sequencing results were used to analyze the copy number alterations of surviving tumors collected 9 or 18 months post-DEN (see above). We used a modified version of QDNAseq (https://github.com/ccagc/QDNAseq/) to call somatic copy number alterations[41]. QDNAseq was adapted to use only so-called “off-target” reads from exome sequencing, which yield an even, shallow coverage across the genome[42]. To obtain off-target reads, we first counted all reads per 1Kb bins along the genome, then removed any bin that overlaps with the start/end coordinates of a protein-coding gene (Ensembl version GRCm38.p6) and mapped the remaining onto bins of 1Mb. Read counts from the tumor sample were subsequently adjusted by those from the control to obtain relative tumor coverage (commonly referred to as "logr"). Here we followed the approach implemented in the Battenberg algorithm[43]. The regular QDNAseq pipeline was then applied to adjust the logr for mappability and GC-content, and the genome segmented into regions of constant signal. Each segment was then subjected to a significance test to obtain copy number alterations. A t-test was applied between the logr of the segment and a logr of 0 (the expected logr if the segment is not altered), with the standard deviation of both distributions set to that of the observed data. A segment was not tested if its logr fell within [-0.05, 0.05] to conservatively account for remaining noise. The t-test had different power between segments, as these differed in size and therefore in the number of bins from which logr data was drawn. To account for this difference, we sampled up to 500 bins without replacement from the segment to perform the test. A one-sided t-test was then applied for gains and losses separately and p-values were adjusted for multiple testing using Bonferroni correction.

Europe PMC Funders Author Manuscripts

As every tumor was sequenced three times we applied a pipeline that used these triplicates to obtain confident copy number calls for each tumor. First, the method described above was applied to each sample and matched control, and we only kept alterations if they were called in all three samples that represent a tumor. This yielded 5 calls across 5 tumors. A tumor, however, had only a single matched control. This single control was used to normalize the data for all three triplicates of a tumor and could introduce false positive calls. We reasoned that a *true* alteration would remain when a sample is run against another control. We therefore randomly matched samples with a different control that was matched for sex, requiring each sample in a triplicate to be matched with a different control. The copy number pipeline was applied to each sample part of the 5 tumors that were found to contain an alteration in the first step, and again kept calls obtained in all three triplicates. This reduced the number of calls to 2, spanning 2 tumors.

Whole genome sequencing of 1 year post-DEN esophageal tumors

Sample preparation and sequencing—Wild type mice were treated with DEN and the esophagus collected 1 year later. Esophageal epitheliums were stained with Dapi and KRT6 to label the tumors and imaged by confocal microscopy. 64 tumors from 9 mice were manually cut out under a fluorescent dissecting microscope. Samples were digested and DNA isolated as above (see WES methods section) and subjected to whole genome sequencing (median coverage 3.13x), using 150-base-pair clipped reads, on a HiSeq 4000 platform (Illumina).

Copy number analysis—We applied a modified version of QDNAseq (https://github.com/ccagc/QDNAseq/) that takes into account the control that is available for each sample for an additional normalization step, to obtain calls where the coverage log ratio is significantly different from a normal diploid state. Inspection of the results revealed a number of calls in a distinct region within chromosome 12 and on chromosome X in samples where the control was sample MD5921b. A run of control samples against each other also revealed frequent calls on chromosome 12, spanning a region 5Mb (bp 18,200,001 to 23,300,000), which was subsequently blacklisted. We further used the largest call (1.15Mb) in this run of control samples to establish a threshold on the minimum genomic region to be covered, and kept calls only if they exceed it (3 calls filtered). Inspection of control sample MD5921b revealed unusually low coverage, which suggests the gains called in samples where this control was used during analysis are likely false positives. We therefore removed calls on chromosome X in these samples (14 calls filtered). After filtering, the final set of alterations contains a gain that covers the whole of chromosome 19 in sample MD5924e and a focal loss on chromosome 14 in sample MD5928e (Extended Data Fig. 5h).

Detection of apoptosis by activated caspase-3 staining

Wild-type mice received DEN in drinking water for eight weeks as described above, and culled 10 days after the treatment stopped. Esophagus were collected, fixed, and immuno-stained to label active caspase-3 and KRT6 followed by Dapi staining. Tissues were imaged by confocal microscopy using a x40 objective and the presence of caspase-3 positive cells within the tumors or in the surrounding normal epithelium was analyzed with Image J.

Europe PMC Funders Author Manuscripts

EdU lineage tracing assay

EdU (5-ethynyl-2′-deoxyuridine) incorporates into dividing cells and therefore quantification of EdU positive cells can be used as a direct measure of cell proliferation in the tissue. Mice received DEN in drinking water for 8 weeks as described above. 10 days after finishing DEN treatment mice were administered 10 μg of EdU (i.p.) and the esophagus collected 1h later. Tissues were peeled, fixed and the incorporated EdU detected using a Click-iT EdU imaging kit (Life technologies Cat# C10086) according to the manufacturer's instructions. Tissues were imaged by confocal microscopy using a x40 objective and the presence of EdU positive cells in the tumors and their adjacent normal epithelium analyzed using IMARIS software.

In vivo transgenic label-retaining cell assay

Rosa26^{M2rtTA}*/TetO-HGFP* mice were used to measure the rate of cell division in DEN generated tumors as compared to their adjacent normal epithelium. These animals are double-transgenic for a reverse tetracycline-controlled transactivator (rtTA-M2) targeted to the Rosa 26 locus and a *HIST1H2BJ/EGFP* fusion protein (Histone-Green Fluorescent Protein, HGFP) expressed from a tetracycline promoter element[30]. Administration of doxycycline (Doxy, Sigma Aldrich Cat# D9891) induces the transient expression of HGFP, resulting in nuclear fluorescent labelling throughout the entire epithelium. When Doxy is withdrawn, HGFP is no longer expressed and is diluted lineally by half after every cell division cycle (Extended Data Fig. 6d). As a result, the decline in fluorescence intensity can be used to calculate the cell division rate. *Rosa26*^{M2rtTA}*/TetO-HGFP* mice received DEN for eight weeks as described above. During the last 4 weeks of DEN treatment, Doxy (2mg/ml) was also supplied in the drinking water. After DEN-Doxy treatment, mice were culled and esophagus collected at times t= 0 (immediately after the treatments stopped), t = 10 days or t= 30 days. Tissues were peeled, fixed and stained (nuclei and KRT6) as detailed above, and imaged on a confocal microscope using a 40x objective. The intensity of HGFP in the tumors and the adjacent normal epithelia was analyzed using ImageJ.

Leukocyte density in tumors and normal epithelium

Wild-type mice were treated with DEN for 2 months and the esophagus collected 10- or 30-days after treatment. The esophageal epitheliums were processed and immuno-stained with anti-CD45 and anti-KRT6 antibodies to visualize leukocytes and tumors, respectively, and imaged on a confocal microscope. Typical settings for acquisition of multiple z stacks were 2.5μm z-step size, zoom x1, optimal pinhole, line average 4, scan speed 400 Hz and a resolution of 1,024 × 1,024 pixels using a 10X HC PL Apo CS Dry objective with a 0.4NA. Visualization and image analysis were performed using IMARIS (bitplane) and ImageJ software. Images, with the tumors placed at the center of a 300 x 300 microns field of view, were used to quantify the number of leukocytes within and outside the tumors. This number was normalized by the area of the tumor or the surrounding normal tissue (calculated by subtracting the area of the tumors to the total area).

Permutation analysis of leukocyte location within tumors and normal tissue

To investigate whether the number of immune cells was enriched or excluded within the tumors, or instead followed a random distribution, we used a permutation analysis based on the experimental measurements of leukocyte density in tumors and normal epithelium obtained as detailed above. For this, tumors were assigned a circular shape, with an area matching that measured experimentally. Immune cells ($CD45^+$ cells) were also assumed to be circular with a diameter of 8 microns. For each field of view, the location of the immune cells was left intact while the location of the tumor was randomly shuffled (ensuring the entirety of the tumor remained within the boundaries of the field of view), and the number of immune cells in contact with the "phantom" tumor was counted (Extended Data Fig. 6l). This was repeated 1000 times to produce the expected distribution assuming no association between tumor and immune cell locations. The proportion of the expected distribution at least as extreme as the observed value was calculated. This proportion was then doubled to produce the two-tailed p-value.

Induction of DN-Maml1 clones

Ahcre^{ERT}/*R26*$^{DNM–GFP/wt}$ mice were used for induction of the dominant negative mutant of *Maml1* (DN-Maml1)[11]. This mutant inhibits Notch intracellular domain induced transcription, therefore disrupting the Notch signaling pathway. It is also fused to GFP, which allows for clonal labelling of the mutant. Expression of DN-Maml1 can be genetically induced following treatment with ß-napthoflavone (BNF, MP Biomedicals Cat# 156738) and tamoxifen (TAM, Sigma Aldrich Cat# N3633). Specifically, transcription of the *Cre* mutant estrogen receptor fusion protein (CreERT) is induced following intraperitoneal (i.p) BNF injection. A subsequent i.p injection of TAM is necessary in order for the CreERT protein to gain access to the nucleus and excise the loxP flanked "STOP" cassette resulting in the expression of DN-Maml1 tagged to GFP. As the switch occurs at the gene level, the descendants of the originally labelled cell (clones) will also constitutively express DN-Maml1, and can be visualized by fluorescent microscopy.

Mice received DEN for 8 weeks as described above. 10 days after DEN withdrawal, mice were given a single injection of BNF (80 mg kg) and TAM (1 mg), or left un-induced. Esophagus were collected 30 days post-DEN treatment (20 days after induction of DN-Maml1 clones). Whole tissues were processed, stained with anti-GFP and KRT6 antibodies and imaged on a confocal microscope as described above. The percentage of esophageal epithelium occupied by DN-Maml1 mutant clones and the number of tumors within and outside DN-Maml1$^+$ areas were measured using Volocity 3D Image Analysis Software (Perkin Elmer) or IMARIS software.

Percentage of tumors eliminated by DN-Maml1 clones

The percentage of tumors removed by expanding DN-Maml1 clones was calculated as follow: % of tumors eliminated = (1-T/(DNneg)) 100. Where T is the total number of tumors per mm^2 of esophageal epithelium and DNneg is the number of tumors per mm^2 of DN-Maml1 negative epithelium within the same tissue.

Quantification of DN-Maml1 clone-tumor interactions

Interactions of tumors and DN-Maml1 clones were classified in four different categories as follow: Category 1 (Distant), where at least 1 cell separates the clone and the tumor, Category 2 (Adjacent), where DN-Maml1 clones are in direct contact with the tumor, but without clear signs of the clone encircling the tumor, Category 3 (Encircling), with DN-Maml1 clones partially surrounding the tumor and Category 4 (Enclosed), where tumors are completely surrounded by DN-Maml1 clones.

Induction of Confetti clones in mutagen treated mice

R26Cre^{ERT2}*/R26*^{flConfetti} mice were dosed with a single i.p injection of TAM (3mg per 20g of body weight) to clonally label cells with one of four different fluorescent proteins (YFP, GFP, RFP or CFP). Animals were then treated with DEN in drinking water for 2 months and esophagus collected 10 or 30 days post-DEN withdrawal. Whole mount tissues were processed as described above. Fluorescent clones and tumors were imaged on a confocal microscope and analyzed with IMARIS software (bitplane).

In vivo treatment with the γ-secretase inhibitor DBZ

Wild-type mice received DEN in drinking water for 8 weeks as described above. 10 days after DEN withdrawal, mice were injected i.p. with the γ-secretase inhibitor DBZ (S2711; Selleckchem) at 30μmol/Kg of bodyweight, using a stock solution of 2.8mg/ml in 0.5% hydroxypropyl methylcellulose (Methocel 65HG, 64670-100G-F, Sigma) and 0.1% Tween-80 in water. Control animals were injected with vehicle solution. Mice received DBZ or vehicle every three days, a total of 5 injections, and were culled 2 weeks after initiating the treatment (24 days post-DEN). Esophagus were collected and epitheliums from control and DBZ-treated mice processed and stained with KRT6 antibodies. Whole-mounted esophageal epitheliums were imaged by confocal microscopy and the number of tumors per mm^2 of epithelium measured using IMARIS software.

Fitting of model to data

See Supplementary Note. The parameter sets which best fit the data were found using Approximate Bayesian Computation based on sequential Monte Carlo sampling (ABC-SMC)[44] implemented in the python package pyABC[45]. This method randomly generates parameter combinations, then uses summary statistics to compare the model's output using those parameters to the observed data, and rejects parameter combinations for which the model's output is too dissimilar to the data. Multiple generations are run with increasingly stringent thresholds for acceptance, with each new generation of parameters based on the accepted parameter combinations from the previous generation. In this manner, the method identifies the regions of parameter space that produce model outputs most similar to the observed data. We ran 25 generations with a population of 10000 (the number of accepted parameter combinations in each generation). For the summary statistic, we used the total squared distance of the mean model prediction from the mean observed tumor density. For the tumor drift model with no interaction with mutant clones in surrounding normal tissue, the fit was to the tumor density following DEN-treatment. For the model including clone interaction, the fit was to the tumor density following DEN-treatment and the tumor density

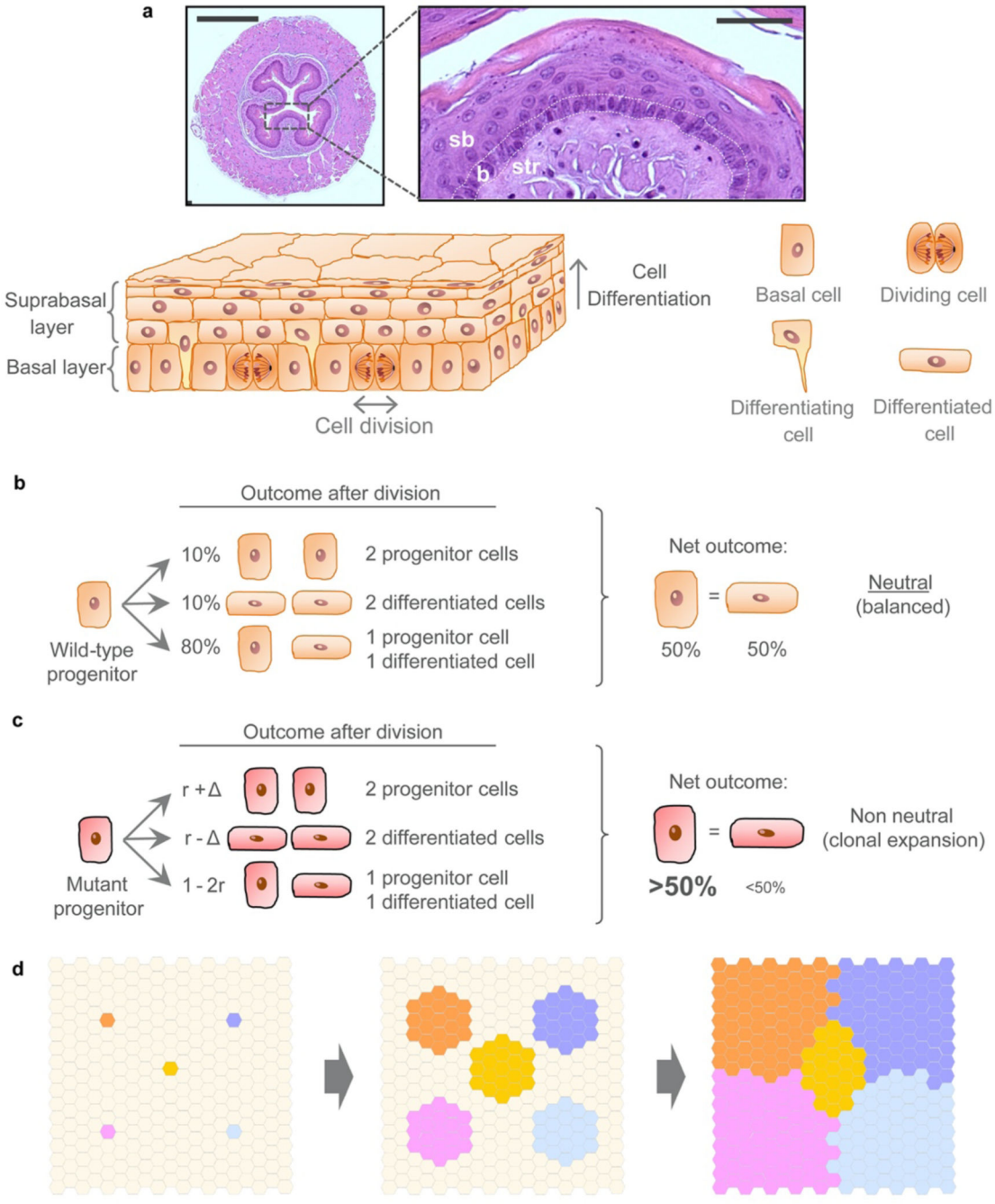
in the DBZ experiment. Uniform distributions using the bounds in Supplementary Table 11 were used as the initial prior distributions of the parameters.

Statistics and Reproducibility

Data are expressed as mean values ± s.e.m. unless otherwise indicated. No statistical method was used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. Fig. 1c: 10 days and 18 months whole tissue images and insets of normal epithelium are representative of 4 and 6 mice, respectively. Tumor images in insets are representative of 588 (10 days), 408 (1 month), 73 (3 months), 92 (6 months), 58 (12 months) and 42 (18 months) tumors. Fig. 2a: Tumor images are representative of 79 (10 days) and 62 (12 months) tumors. Fig. 3e: Images are representative of 14 tumors. Extended Data Figs. 2a-b: Images are representative of 588 tumors. Extended Data Fig. 9a: Images are representative of 156, 142, 74 and 14 events for categories 1, 2, 3 and 4, respectively. Extended Data Fig. 9c: Images are representative of 14 tumors. Extended Data Figs. 10j-k: Images are representative of 5 lesions from 2 donors.

Europe PMC Funders Author Manuscripts

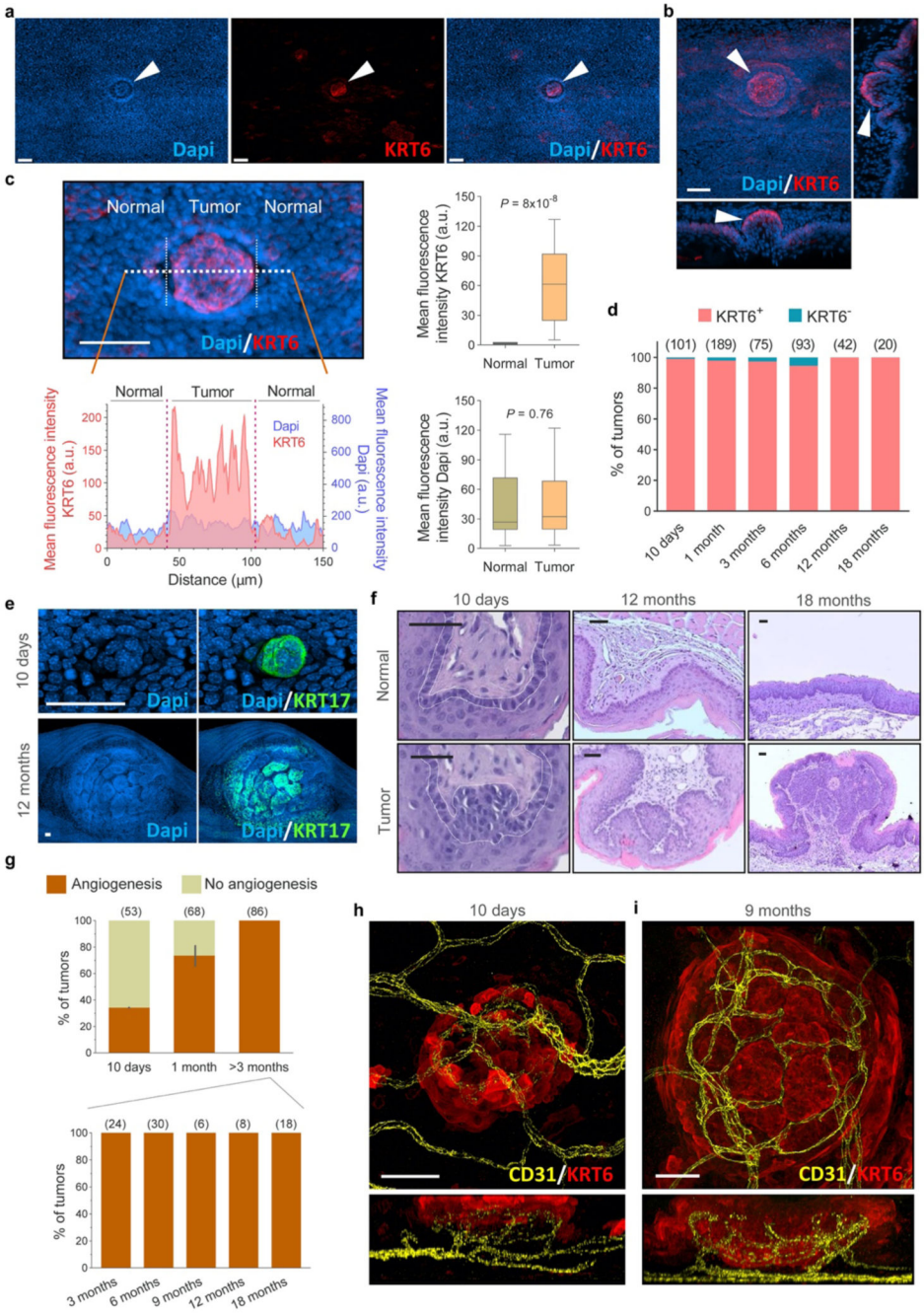
Extended Data



Extended Data Figure 1. Architecture and dynamics of murine esophageal epithelium.
(**a**) H&E images and schematic of the mouse esophageal epithelium architecture and dynamics; b = basal cell layer (delineated by white dots), sb = suprabasal layers, str = stroma. Scale-bars: 500μm (left image), 50μm (inset). Dividing (progenitor) cells are confined to the basal layer. Differentiating cells exit the cell cycle, migrate out of the basal layer, through the suprabasal layers, and are finally shed into the lumen. (**b-c**) The

single progenitor model (see Supplementary Note). All progenitor cells in the basal layer are functionally equivalent and following division produce either: two progenitors that will persist in the tissue, two differentiating cells that will cease division, stratify and be lost, or one cell of each type. In homeostasis (**b),** the likelihood of the two progenitor and two differentiating cell outcomes is equal. Mutations (**c**) may tip the balance towards a non-neutral behavior, resulting in clonal growth if they favor proliferation of daughter cells. (**d**) Expansion of mutant clones is defined by their relative competitive fitness to adjacent clones. Initially, a fit “winner” mutant progenitor (colored) shows a fate bias towards proliferation and outcompetes its less fit “loser” surrounding cells, resulting in clonal expansion. Eventually, mutant clones begin to collide with each other until surrounded by similarly competitive mutants, at which point their cell fate reverts towards balance and their expansion slows (see Supplementary Note).

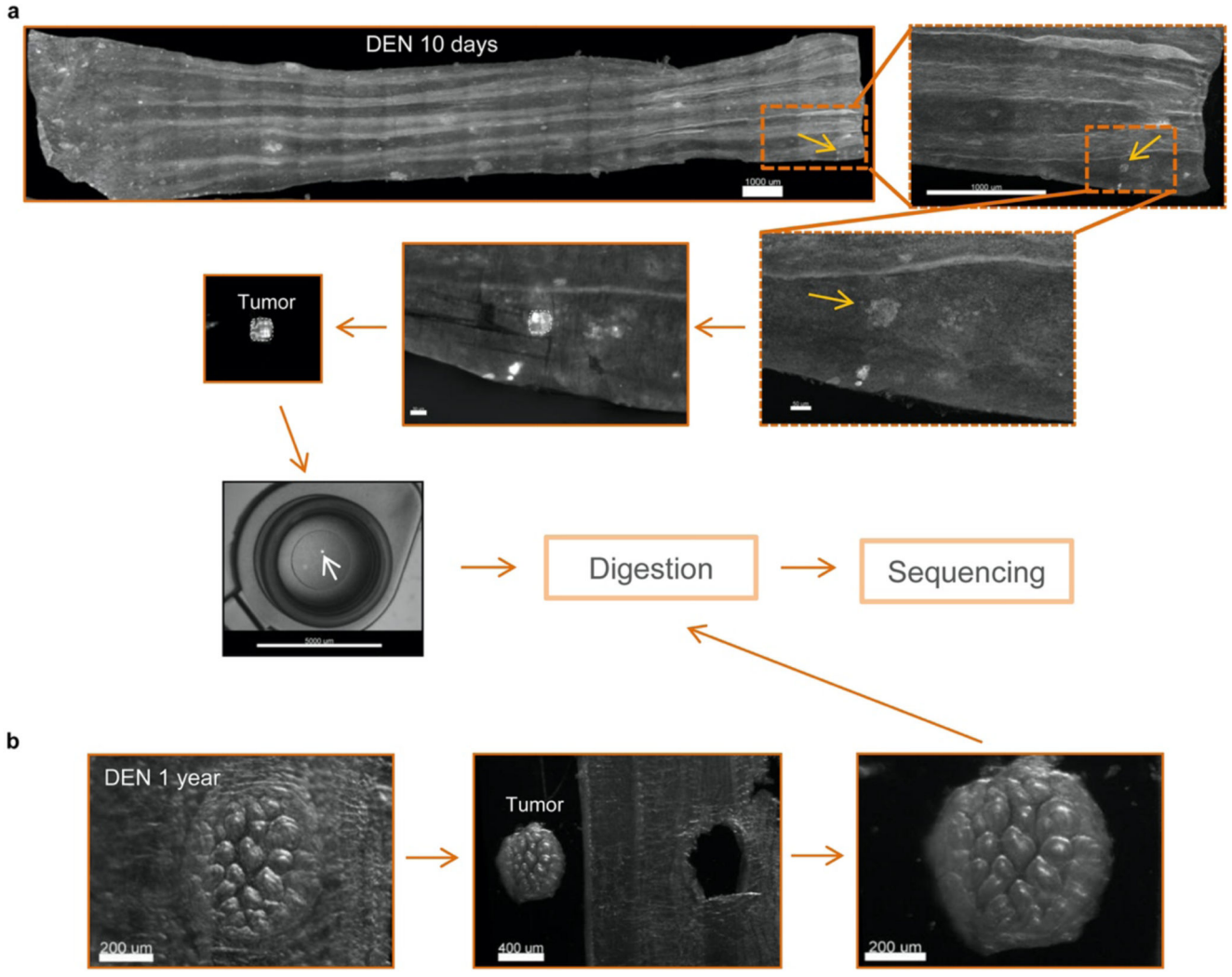
Europe PMC Funders Author Manuscripts



Extended Data Figure 2. Characterization of Diethyl Nitrosamine (DEN) generated tumors.

(**a, b**) Confocal images of esophagus collected 10 days post-DEN treatment and stained with Dapi and KRT6. Images show top-down (**a, b**) and lateral (**b**) projections and illustrate typical morphological features and expression of KRT6 in tumors (arrowheads) and normal surrounding epithelium. (**c**) Quantification of the mean fluorescence intensity of KRT6 and Dapi in 10 days post-DEN tumors and the adjacent normal epithelium, measured from the images as shown. Two-sided Mann-Whitney test, n=48 tumors from 3 mice. (**d**) Percentage of tumors positive for KRT6 staining. The number of tumors analyzed for each time point

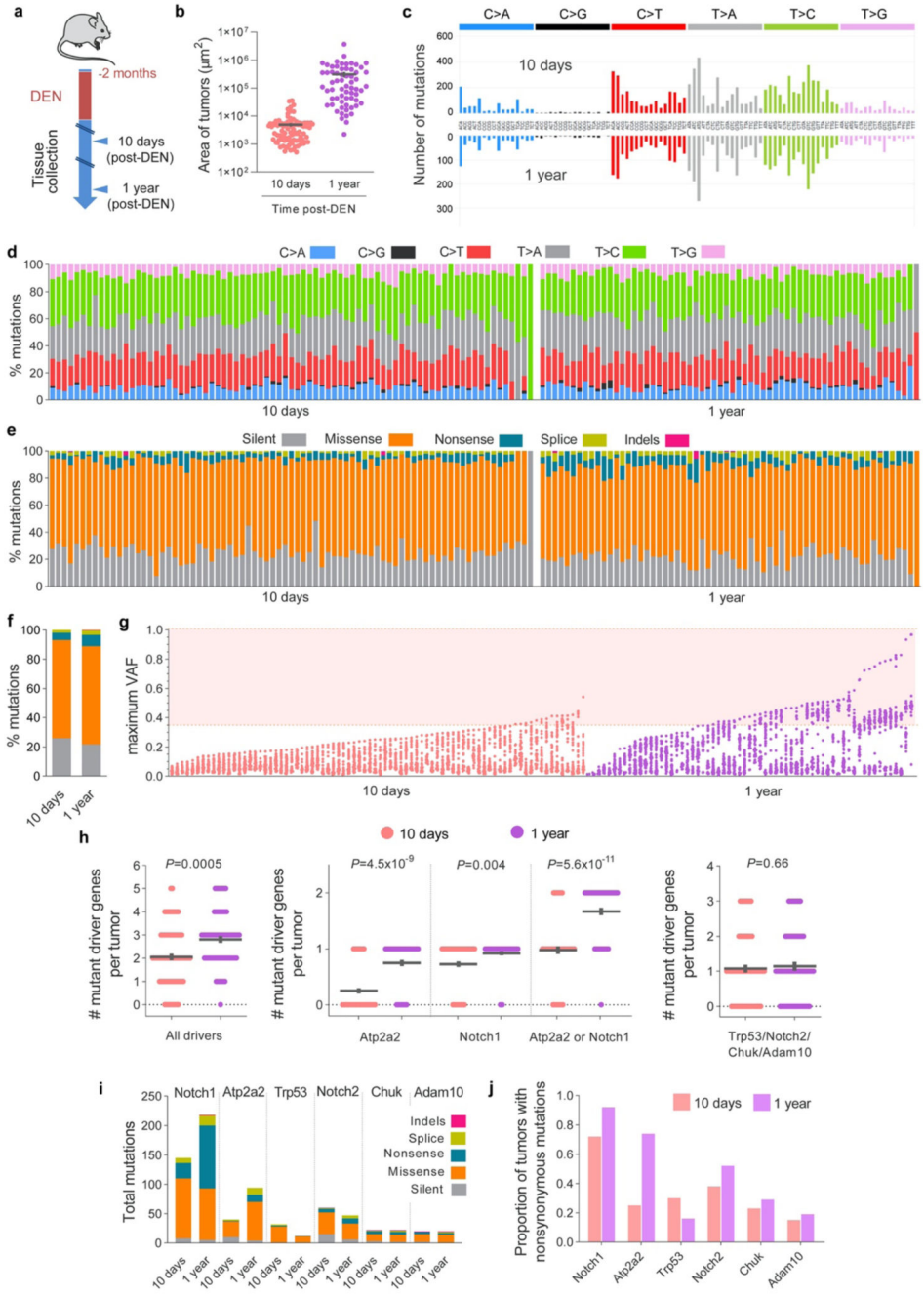
is shown between brackets. (**e**) Confocal images of 10-days and 12-months post-DEN tumors showing increased expression of KRT17 (green). Images are representative of 31 and 29 tumors, from 3 and 7 mice, respectively. (**f**) H&E images of esophageal tumors (bottom images) and normal epithelium (top images) from tissues collected at the indicated time points post-DEN treatment. (**g**) Percentage of angiogenic tumors at the indicated time points post-DEN treatment. The number of tumors analyzed at each time point is shown in brackets. (**h, i**) Projected top-down and lateral confocal images of 10 day (**h**) and 9 months (**i**) post-DEN tumors immuno-stained for the endothelial cell marker CD31 and KRT6. Images are representative of 18 and 6 tumors, respectively. Scale-bars for panels (**a-c), (e-f**) and (**h-i**) are 50μm.



Extended Data Figure 3. Collection of 10-day and 1-year esophageal tumors.

(**a-b**) Mouse esophagus was collected 10-days (**a**) or 1-year (**b**) post-DEN treatment. The esophagus was cut open longitudinally and the epithelium separated from the underlying muscle and stroma. The epithelium was flattened, fixed, stained with KRT6 (grey), mounted and 3D-imaged on a confocal microscope. Tumors were identified from the

processed images and manually dissected under a fluorescent microscope and sequenced (see Methods).



Extended Data Figure 4. Targeted sequencing of 10 days and 1 year tumors.

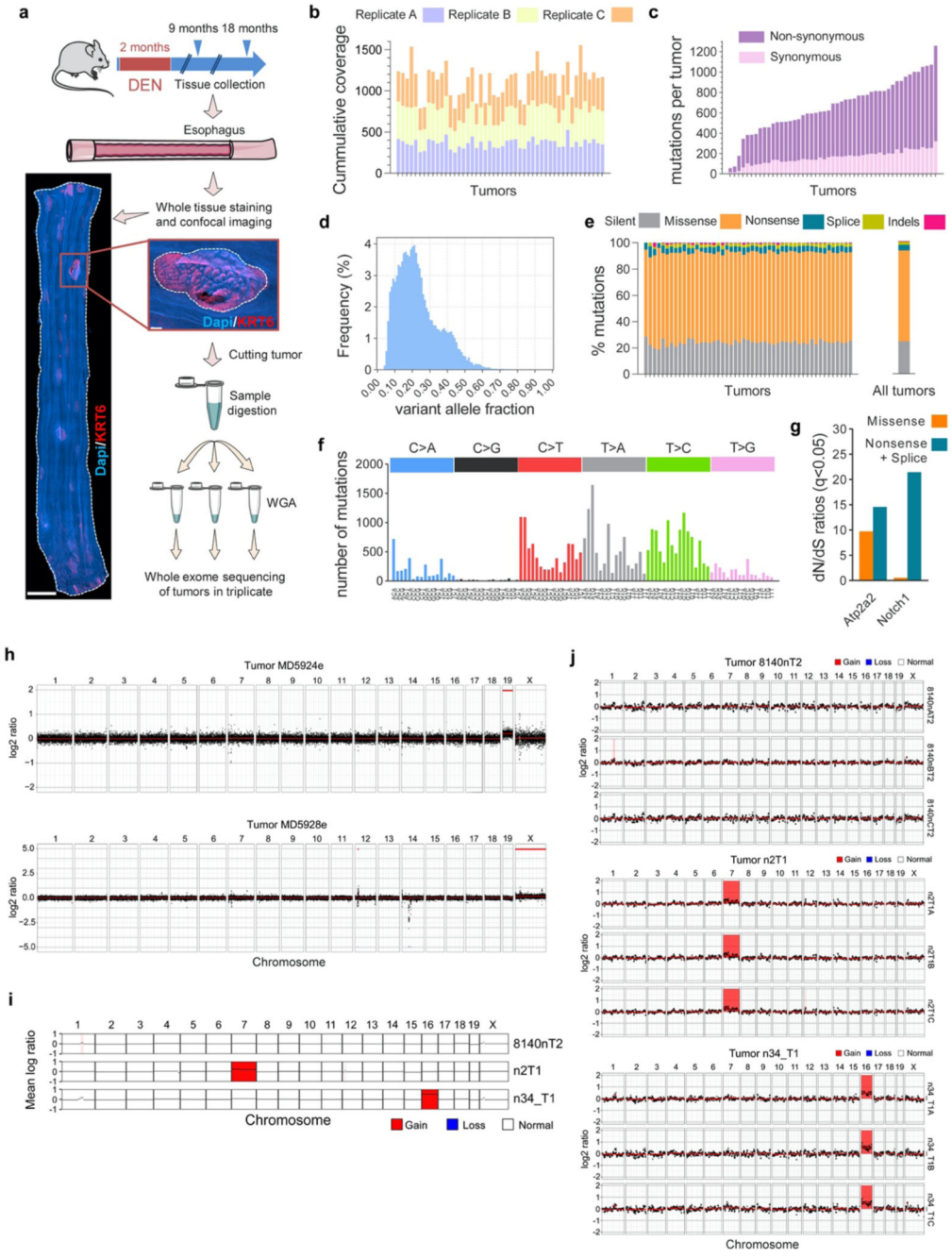
(**a**) Protocol: wild-type mice were treated with DEN for two months and the esophageal tumors collected 10-days or 1-year later. Tumors were then sequenced with a targeted approach (192 gene panel). (**b**) Area of 10-day and 1-year post-DEN tumors (n=89 and 64 tumors from 2 and 9 mice, respectively). Lines show mean ± s.e.m. (**c**) Mutational

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

spectrum of 10-day and 1-year tumors. The bar plots illustrate the number of mutations in each of the 96 possible trinucleotides. The mutational spectrum of individual tumors is shown in (**d). (e-f**) Percentage of silent, missense, nonsense and splice mutations and indels identified in 10-day or 1-year tumors. Graphs show the values for individual tumors (**e,** each column is a tumor), or the average for all tumors at each time point (**f). (g**) Maximum VAF for mutations identified in each tumor at 10 days or 1 year post-DEN (n=80 and 63 tumors, respectively). Colored shadow illustrates VAF > 0.35, as an estimation for clonality. (**h**) Number of positively selected mutant genes in tumors at 10 days or 1 year post-DEN treatment (n=80 and 63 tumors, respectively). Lines show mean±s.e.m (two-tailed Mann-Whitney test). (**i**) Number and type of mutations in the positively selected genes identified by dN/dS analysis from 10-day or 1-year tumors. (**j**) Proportion of 10-day and 1-year tumors carrying nonsynonymous mutations in the indicated genes.

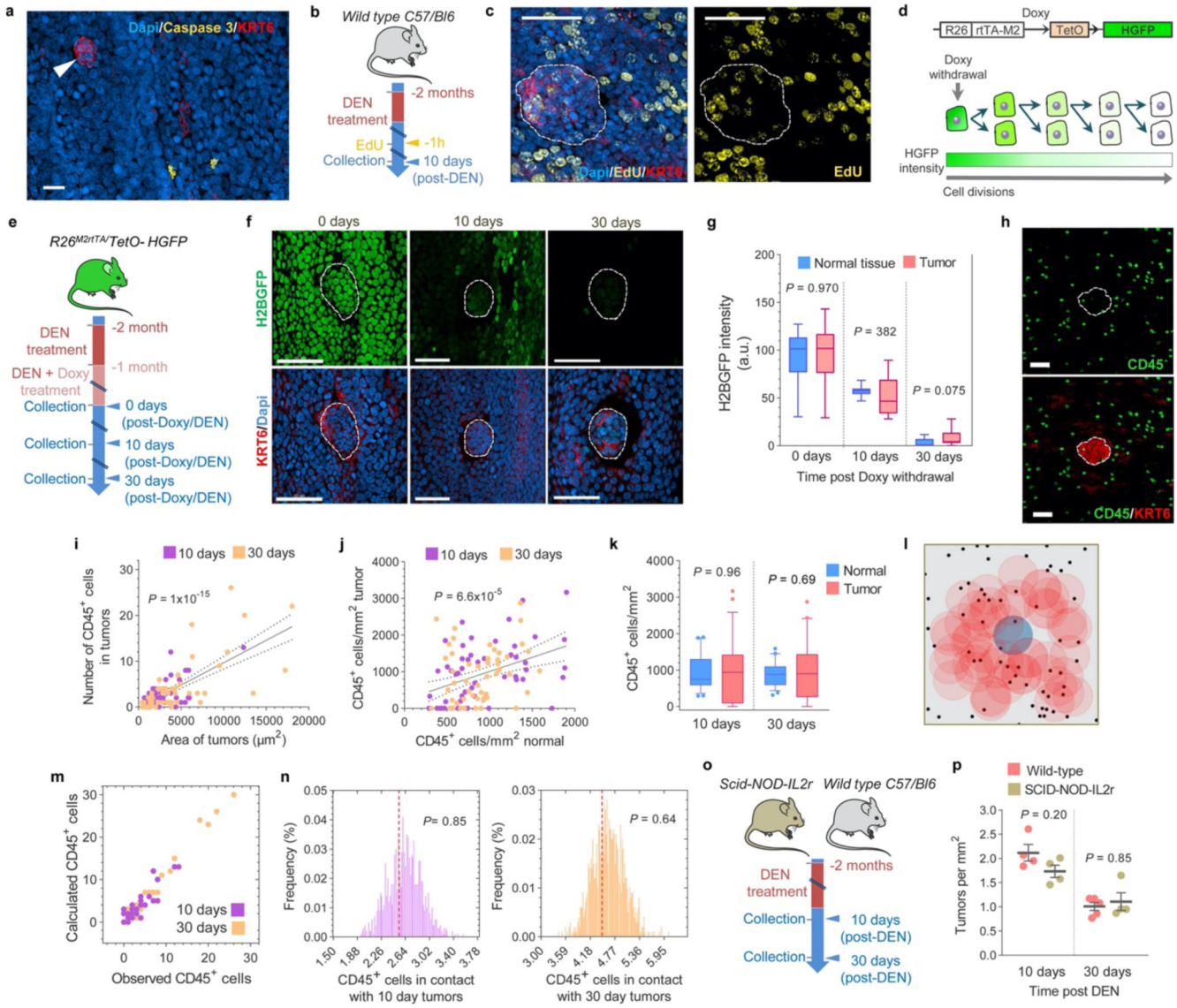
Europe PMC Funders Author Manuscripts



Extended Data Figure 5. Whole exome sequencing, whole genome sequencing, and chromosomal alteration of surviving tumors.

(**a**) Wild-type mice received DEN for two months and the esophagus was collected 9 or 18 months after treatment (n=49 tumors from 16 mice). Tissues were stained for Dapi (blue) and KRT6 (red) and confocal imaged to identify tumors. Scale-bars=2mm (main), 150μm (inset). Individual tumors were manually cut under a fluorescent microscope, digested and separated in triplicates. Each triplicate was whole genome amplified (WGA) and whole exome sequenced. To exclude artefactual SNVs generated during WGA, only mutations

shared by all three amplified triplicates were considered for further analysis. A total of 32,736 mutations, including silent, missense, nonsense and splice mutations and indels were identified. (**b**) Cumulative sequencing coverage of the whole exome triplicate samples. (**c**) Number of synonymous and non-synonymous mutations per tumor, ranked by mutation burden. (**d**) Distribution of the variant allele fraction (VAF) for the mutations common within triplicates in each tumor. (**e**) Percentage of silent, missense, nonsense and splice mutations and indels for individual tumors and for all tumors combined. (**f**) Mutational spectrum of tumors. The bar plots illustrate the percentage of mutations in each of the 96 possible trinucleotides. (**g**) dN/dS ratios for missense and truncating (nonsense + essential splice site) substitutions indicating genes under significant positive selection in the tumors ($q<0.05$, calculated with R package dNdScv[26]). (**h**) Analysis of chromosomal copy number alterations (CNAs) by whole genome sequencing of 1-year post-DEN tumors (n=64 tumors from 9 mice). Only 2 tumors, MD5924e and MD5928e, exhibited CNAs. (**i**) Summary of chromosomal alterations found by whole exome sequencing data of 9 or 18 months post-DEN tumors (n=49 tumors from 16 mice). Only alterations present in all 3 whole genome amplified triplicates (**j**) were considered valid calls. 8140nT2 (top) is a representative example of a tumor without chromosomal alterations. 2 tumors, n2T1 and n34_T1, showed small alterations.



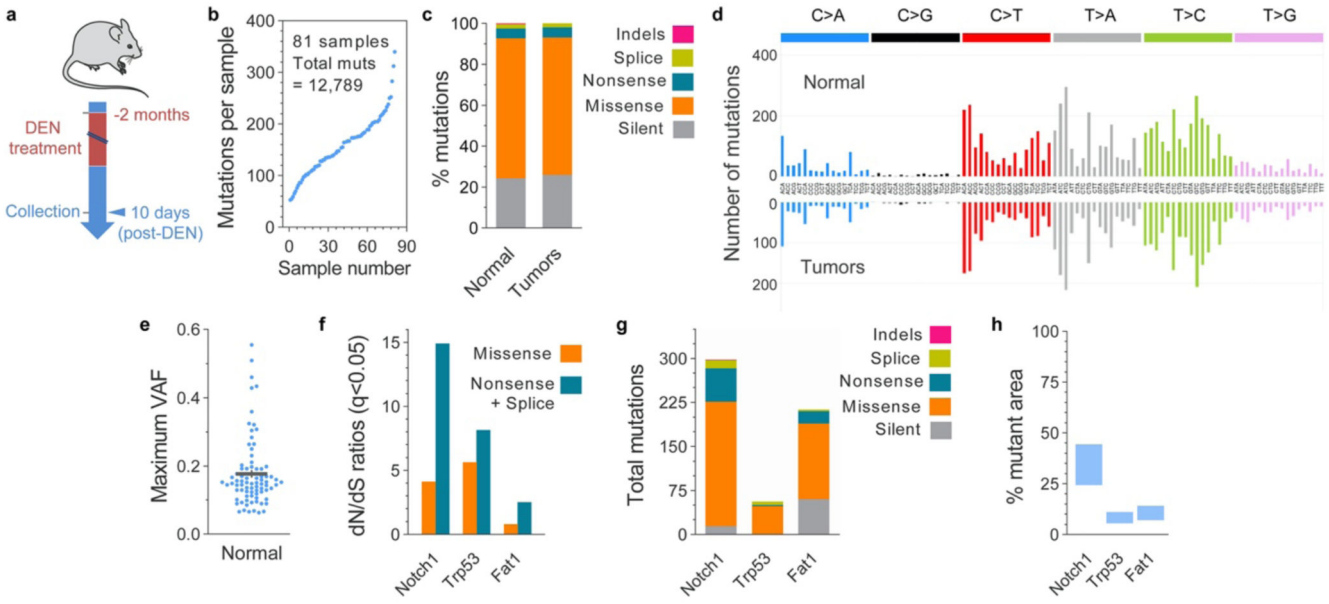
Extended Data Figure 6. Early tumors are not eliminated by tumor cell apoptosis, abnormal proliferation or the immune system.

(**a**) Representative confocal image of a 10-day post-DEN esophageal epithelium immuno-stained for activated Caspase 3+ (Yellow) and KRT6 (red). No apoptotic cells were detected in the tumors (arrowhead) (n=23 tumors, 2 mice). Scale-bar: 20μm. (**b**) Protocol: wild-type mice were treated with DEN for two months and the tissues collected ten days later. Mice received EdU 1h before tissue collection. (**c**) Representative confocal images showing EdU incorporation (1h pulse) in tumors (dotted line) and the surrounding normal epithelium (n=22 tumors, 3 mice). Scale-bars: 20μm. (**d**) *In vivo* label-retaining assay using *Rosa26*^{M2rtTA}*/TetO-HGFP* transgenic mice to measure the rate of progenitor cell division (see Methods). (**e**) Protocol: *Rosa26*^{M2rtTA}*/TetO-HGFP* mice received DEN for one month followed by DEN+Doxy for another month. Tissues were collected at times 0, 10 and 30 days after Doxy/DEN withdrawal. (**f-g**) Representative confocal images (**f**) and

Europe PMC Funders Author Manuscripts

quantification (**g**) of the histone-green fluorescent protein (H2BGFP) intensity in esophageal tumors (dotted lines) and the surrounding normal epithelium at the indicated time points post-Doxy withdrawal. Scale bars: 50μm. Graph shows median (central box line), 25^{th}-75^{th} percentiles (box) and 5^{th}-95^{th} percentiles (whiskers). Two-tailed Mann-Whitney test (n=10 images per group). (**h**) Confocal images depicting $CD45^+$ (immune) cells (green) within a 10-days post-DEN tumor (white dotted line) and its adjacent normal epithelium. (**i-j**) Correlation between the size of tumors and the number of $CD45^+$ cells within them (**i**) and the number of $CD45^+$ cells in the tumor and the normal epithelium (**j).** Lines show the two-tailed Pearson correlations: R^2=0.5039 (**i),** and R^2=0.1464 (**j),** with 95% confidence interval (dotted lines). (**k**) Number of $CD45^+$ cells per area of tumor or normal esophageal epithelium at 10- or 30-days post-DEN treatment. Graph shows median (central box line), 25^{th}-75^{th} percentiles (box), 5^{th}-95^{th} percentiles (whiskers) and outliers (dots). Two tailed Wilcoxon matched-pairs test, n=53 and 50 images, respectively. (**l**) Permutation analysis of leukocyte location within tumors, based on the experimental measurements obtained from (**h).** For each image, the location of $CD45^+$ cells (black dots) was left intact while the location of the tumor (colored circles) was randomly shuffled (blue shows original location, red shows shuffled “phantom” tumors), and the number of immune cells in contact with the tumor was counted. This was repeated 1000 times to produce the expected distribution. (**m**) Calculated (within the original location) vs. experimentally observed number of $CD45^+$ cells in 10- or 30-days post-DEN tumors. (**n**) Distribution of the average number of $CD45^+$ cells in 10- or 30-days post-DEN tumors, obtained from the permutation analysis in (**l).** Red dotted line shows the experimentally observed average number of $CD45^+$ cells within tumors at the indicated time-points. Statistics are two-tailed permutation tests (Methods). (**o**) Protocol: Wild-type and immunocompromised *Scid-NOD-IL2r* mice were treated with DEN for two months and the tissues collected 10 or 30 days post-DEN withdrawal. (**p**) Tumor density in mice collected 10-days (n=4) or 30-days (n=5 and 4, respectively) post-DEN treatment. Mean±s.e.m (two-tailed Mann-Whitney test).

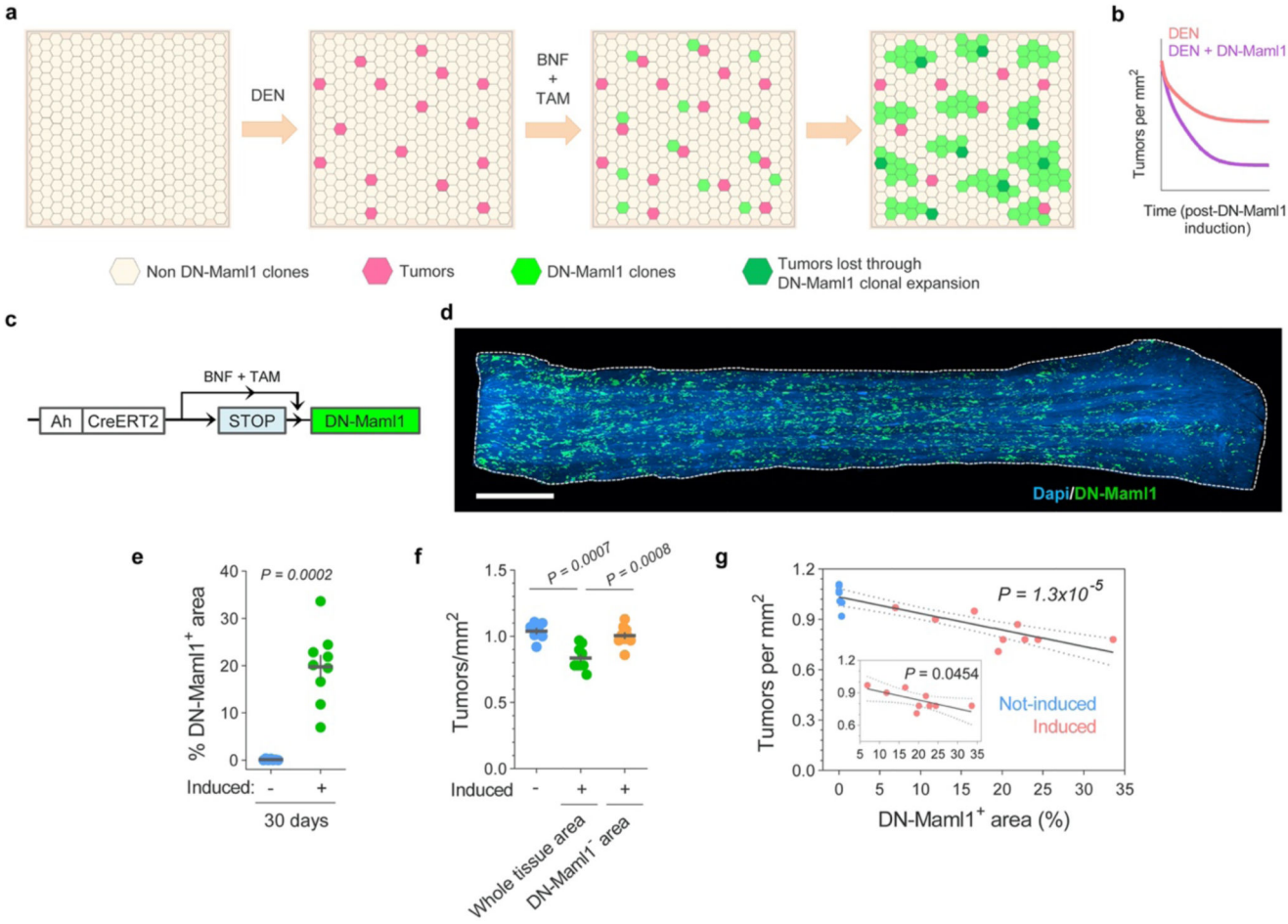
Europe PMC Funders Author Manuscripts



Extended Data Figure 7. Mutational landscape of normal esophageal epithelium at 10 days post-DEN treatment.

(**a**) Protocol: wild-type mice were treated for two months with DEN. 10 days after DEN withdrawal, normal epithelial samples matching the size of the tumors (Extended data Fig. 4b) were collected and sequenced with a targeted approach (192 gene panel). (**b**) Number of mutations per sample, including essential splice, frameshift, missense, nonsense and silent mutations (n=81 samples from 2 mice). (**c**) Percentage of mutation types identified in normal epithelium and tumors collected 10 days post-DEN treatment. (**d**) Mutational spectrum of 10-day post-DEN tumors and normal epithelium. The bar plots illustrate the number of mutations in each of the 96 possible trinucleotides. (**e**) Dots show the value of the mutation with maximum VAF identified in each sample (n=81 samples). Lines are mean ± s.e.m. (**f**) dN/dS ratios for missense and truncating (nonsense + essential splice site) mutations of positively selected genes (dN/dS >1) (only significant genes are shown, $q<0.05$ calculated with R package dNdScv[26]). (**g**) Number and type of mutations in the positively selected genes identified by dN/dS. (**h**) Estimated percentage of 10 days post-DEN normal epithelium carrying non-synonymous mutations in the positively selected genes. Range indicates upper and lower bound estimates.

Europe PMC Funders Author Manuscripts



Extended Data Figure 8. Expansion of highly competitive clones in normal tissue eliminates early tumors.

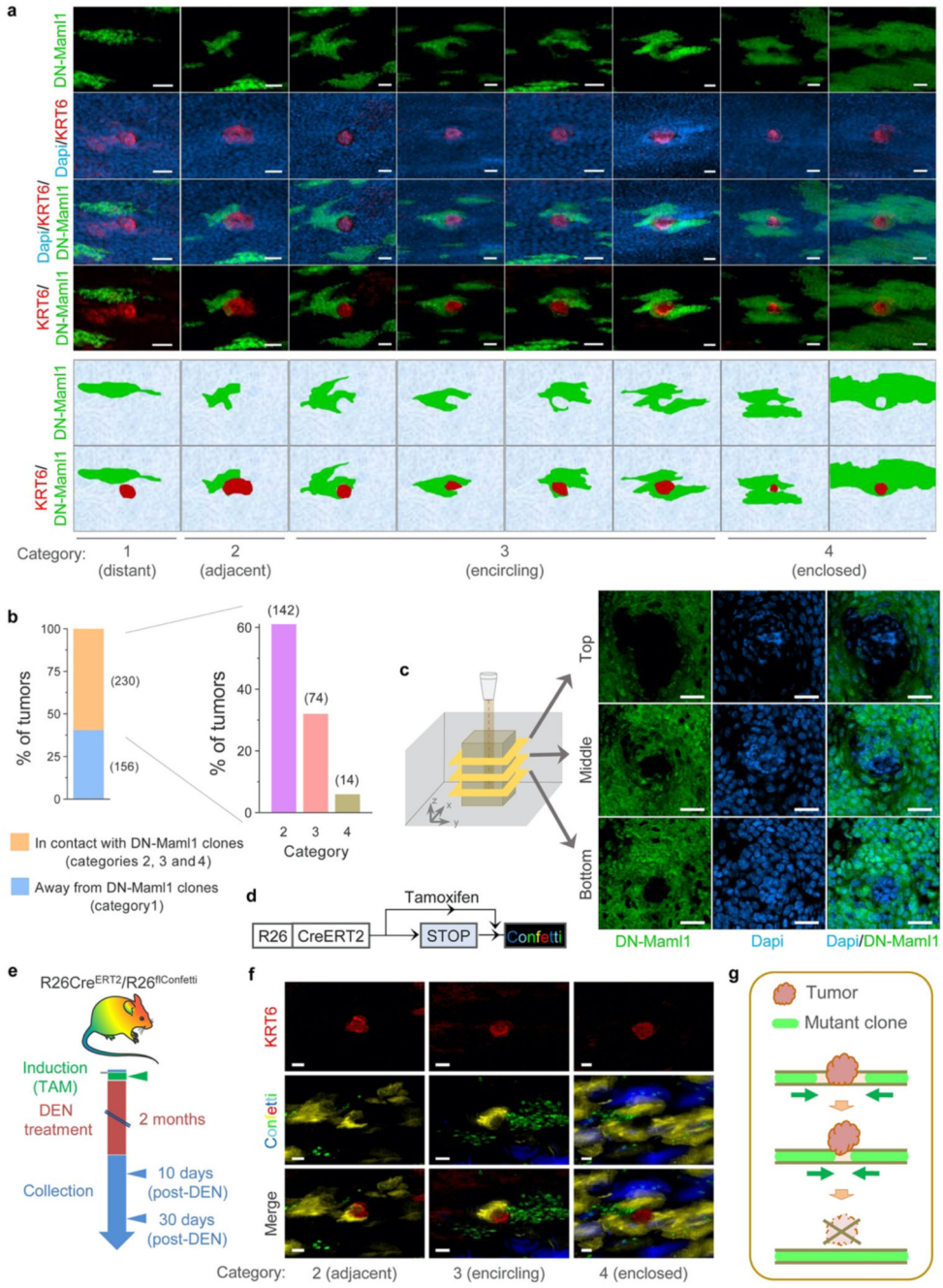
(**a**) Cartoon illustrating the elimination of tumors due to competition with the highly competitive *DN-Maml1* clones. Following the generation of tumors by DEN treatment, *DN-Maml1* clones are induced by BNF and TAM injections. As *DN-Maml1* clones in the normal epithelium expand they eliminate less fit tumors. (**b**) Qualitative representation of tumor dynamics following *DN-Maml1* induction as compared to non-induced controls (see Supplementary Note). (**c**) *In vivo* genetic lineage tracing using $Ahcre^{ERT}/R26^{DNM-GFP/wt}$ reporter mice. Upon injection of the drugs tamoxifen (TAM) and ß-napthoflavone (BNF), Cre-mediated recombination results in the heritable expression of the highly competitive dominant negative allele of *Maml-1* fused to GFP fluorescent protein *(DN-Maml1),* which will then be expressed in all the progeny of the single marked cells, generating clusters of labelled mutant clones. (**d**) Confocal image (representative of 9 mice) of a DEN-treated $Ahcre^{ERT}/R26^{DNM-GFP/wt}$ mouse whole esophageal epithelium, depicting *DN-Maml1* clones (green). Mice were induced ten days after DEN withdrawal and esophagus collected twenty days later (as in Fig. 3b). Scale-bar: 2mm. (**e**) Percentage of epithelium covered by *DN-Maml1* clones in induced and non-induced (control) mice (n=9 and 7 mice, respectively). Error-bars are mean±s.e.m (two-tailed Mann-Whitney test). (**f**) Number of

Europe PMC Funders Author Manuscripts

tumors per mm^2 of esophageal epithelium in control non-induced (n=7) and induced (n=9) *Ahcre*^{ERT}/*R26*$^{DNM–GFP/wt}$ mice. Data for the induced mice is show as the number of tumors per mm^2 of whole tissue area and per mm^2 of area not covered by *DN-Maml1* clones (DN-Maml1$^-$ area). Error-bars are mean±s.e.m. (two-sided Mann-Whitney test). (**g**) Correlation between the area covered by *DN-Mam1* clones and the number of tumors in induced and non-induced *Ahcre*^{ERT}/*R26*$^{DNM–GFP/wt}$ mice. Line shows the Pearson correlation (two-tailed, R^2=0.7526) with 95% confidence interval (dotted lines). Inset shows the correlation of the induced mice only (R^2=0.4577).

Europe PMC Funders Author Manuscripts

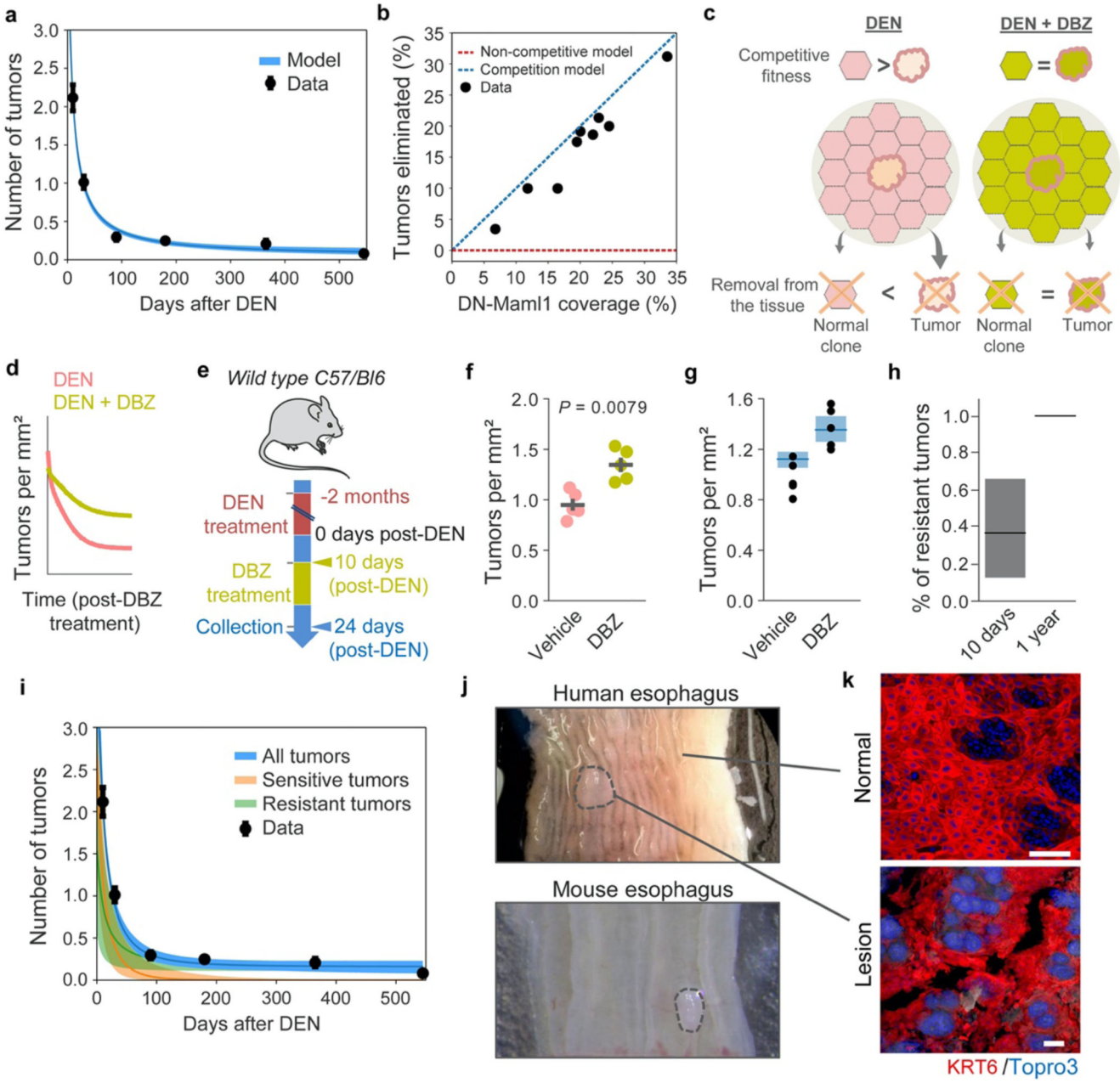
Europe PMC Funders Author Manuscripts



Extended Data Figure 9. Spatial interaction between mutant clones in mutagenized normal epithelium and tumors.

(**a**) Confocal images and cartoons of esophageal epithelium from *Ahcre*^{ERT}/*R26*$^{DNM–GFP/wt}$ mice treated with DEN for 2 months, induced 10-days post-DEN withdrawal and tissues collected 20-days later (as in Fig. 3b). The images show different categories of *DN-Maml1* clones (green) interacting with early tumors (red). (**b**) Percentage of tumors in each category. Number of tumors in each group is shown between brackets (n=386 tumors from 9 mice). (**c**) Top-down confocal images depicting a *DN-Maml1* clone surrounding a tumor (see

also Fig. 3e). A smaller tumor surface area at the base is consistent with a process of extrusion by the expanding *DN-Maml1* clone as illustrated in (**g**). **(d**) In vivo genetic lineage tracing using the multicolor confetti reporter allele in $R26cre^{ERT2}R26^{flConfetti}$ mice. Injection of the drug tamoxifen (TAM), results in Cre-mediated inversion and excision recombination events in scattered single cells. This confers heritable expression of one of four fluorescent proteins (YFP, GFP, RFP or CFP) resulting in labelled clones. (**e**) Protocol: $R26cre^{ERT2}R26^{flConfetti}$ mice were induced with TAM to label Confetti clones, followed by 2-months DEN treatment. Esophagi were collected 10 days or 30 days after DEN withdrawal. (**f**) Confocal images of Confetti clones surrounding early tumors (KRT6, shown in red) in the categories 2-4 as in (**a**), suggesting that tumor extrusion events may also take place in mutagenized epithelium in the absence of *DN-Maml1* induced clones. Images are representative of 13 tumors from 8 mice. (**g**) Cartoon illustrating a potential mechanism of tumor extrusion by expanding mutant clones in the normal epithelium. Once the tumor footprint in the basal layer is displaced by expanding mutant clones in the adjacent normal tissue, the tumor will be shed. Scale-bars in panels (**a), (c**) and (**f**) are 25μm.



Extended Data Figure 10. Increasing competitive fitness throughout the epithelium slows down tumor loss. Human esophageal lesions.

A stochastic drift model (Supplementary Note) fit to the observed number (mean±s.e.m) of tumors following DEN treatment. (**b**) Experimentally observed (black dots) and model predicted (dashed lines) % of tumors eliminated when highly fit mutant clones (such as DN-Maml1) are induced in esophageal epithelium following DEN-treatment. Dashed lines represent models where the induced mutant clones remove tumors they encounter (blue) or where tumor loss is independent of clones in the surrounding tissue (red) (Supplementary Note). (**c**) Cartoon illustrating the predicted effects of DBZ administration. When clones in the normal surrounding epithelium have a higher competitive fitness than the tumors,

Europe PMC Funders Author Manuscripts

the probability that tumors will be eliminated from the tissue may be higher (left) than when administration of DBZ levels the competition between the clones and tumors (right). (**d**) Qualitative representation of tumor dynamics following DBZ vs. control treatment. (**e**) Protocol: wild-type mice were treated with DEN for two months. Ten days post-DEN withdrawal mice received DBZ or vehicle control. Tissues were harvested two weeks later. (**f-g**) Experimental (**f**) and simulated (**g**, Supplementary Note) tumors per mm^2 of esophageal epithelium in DBZ and vehicle control treated mice (n=5 mice/group). Lines in (**f**) are mean±s.e.m (two-tailed Mann-Whitney test). Black dots in (**g**) show experimental data. (**h-i**) Numerical example (see Supplementary Note) of the model showing increased proportion of tumors resistant to displacement by mutant clones (**h**) and the decrease in tumor density following DEN-treatment (**i).** Experimental data depicts mean±s.e.m. (**j**) Images of human (top) and mouse (bottom) esophagus. Dotted lines delineate lesions. (**k**) Confocal images of human normal (top) and neoplastic (bottom) esophageal epithelium stained with KRT6 (red) and Topro3 (nuclei, blue). Scale-bars: 100μm. Simulations in (**a, g, h and i**) show the mean and range between the minimum-maximum outputs of the model run with the accepted parameters from Approximate Bayesian Computation (Methods).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Esther Choolun, Tom Metcalf and staff at the MRC ARES and Sanger RSF animal facilities for technical support. This work was supported by grants from the Wellcome Trust to the Wellcome Sanger Institute (098051 and 296194) and Cancer Research UK Programme Grants to P.H.J. (C609/A17257 and C609/A27326). B.A.H. and M.W.J.H. are supported by the Medical Research Council (Grant-in-Aid to the MRC Cancer unit grant number MC_UU_12022/9 and NIRG to B.A.H. grant number MR/S000216/1). A.H. benefited from the award of an EMBO long term fellowship. M.W.J.H. acknowledges support from the Harrison Watson Fund at Clare College, Cambridge. B.A.H. acknowledges support from the Royal Society (grant no. UF130039). S.D. benefited from the award of an ESPOD fellowship, 2018-21, from the Wellcome Sanger Institute and the European Bioinformatics Institute EMBL-EBI. K.T.M. benefited from the support of the Chan Zuckerberg Initiative. We thank Betania Mahler-Araujo and the MRC Metabolic Diseases Unit [MC_UU_00014/5] for the histological analysis of tumor samples. We are grateful to the Cambridge Biorepository for Translational Medicine for access to human tissue. The authors acknowledge the use and/or adaption of diagrams from Servier Medical Art (https://smart.servier.com).

Data availability

Individual data sets are available in Supplementary Tables 1–11. The accession number for the DNA targeted sequencing of 10 day normal tissue and 10 days and 1 year tumors is ENA: ERP022921. Accession numbers for WES and WGS of tumors are ENA: ERP015469 and ENA: ERP122780, respectively.

Code availability

Code is available at https://github.com/michaelhall28/Colom_lesions.

References

1. Lee-Six H, et al. The landscape of somatic mutation in normal colorectal epithelial cells. Nature. 2019; 574: 532–537. [PubMed: 31645730]

Europe PMC Funders Author Manuscripts

2. Martincorena I, et al. Somatic mutant clones colonize the human esophagus with age. Science. 2018; 362: 911–917. [PubMed: 30337457]
3. Martincorena I, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015; 348: 880–886. [PubMed: 25999502]
4. Suda K, et al. Clonal Expansion and Diversification of Cancer-Associated Mutations in Endometriosis and Normal Endometrium. Cell reports. 2018; 24: 1777–1789. [PubMed: 30110635]
5. Yokoyama A, et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. Nature. 2019; 565: 312–317. [PubMed: 30602793]
6. Moore L, et al. The mutational landscape of normal human endometrial epithelium. Nature. 2020; 580: 640–646. [PubMed: 32350471]
7. Yoshida K, et al. Tobacco smoking and somatic mutations in human bronchial epithelium. Nature. 2020; 578: 266–272. [PubMed: 31996850]
8. Fowler JC, et al. Selection of Oncogenic Mutant Clones in Normal Human Skin Varies with Body Site. Cancer Discov. 2021; 11: 340–361. [PubMed: 33087317]
9. Lawson ARJ, et al. Extensive heterogeneity in somatic mutation and selection in the human bladder. Science. 2020; 370: 75–82. [PubMed: 33004514]
10. Vermeulen L, et al. Defining stem cell dynamics in models of intestinal tumor initiation. Science. 2013; 342: 995–998. [PubMed: 24264992]
11. Alcolea MP, et al. Differentiation imbalance in single oesophageal progenitor cells causes clonal immortalization and field change. Nat Cell Biol. 2014; 16: 615–622. [PubMed: 24814514]
12. Fernandez-Antoran D, et al. Outcompeting p53-Mutant Cells in the Normal Esophagus by Redox Manipulation. Cell stem cell. 2019; 25: 329–341. [PubMed: 31327664]
13. Colom B, et al. Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. Nature genetics. 2020; 52: 604–614. [PubMed: 32424351]
14. Murai K, et al. Epidermal Tissue Adapts to Restrain Progenitors Carrying Clonal p53 Mutations. Cell stem cell. 2018; 23: 687–699. e688 [PubMed: 30269904]
15. Doupe DP, et al. A single progenitor population switches behavior to maintain and repair esophageal epithelium. Science. 2012; 337: 1091–1093. [PubMed: 22821983]
16. Piedrafita G, et al. A single-progenitor model as the unifying paradigm of epidermal and esophageal epithelial maintenance in mice. Nat Commun. 2020; 11 1429 [PubMed: 32188860]
17. Frede J, Greulich P, Nagy T, Simons BD, Jones PH. A single dividing cell population with imbalanced fate drives oesophageal tumour growth. Nat Cell Biol. 2016; 18: 967–978. [PubMed: 27548914]
18. Nair RR, et al. A role for keratin 17 during DNA damage response and tumor initiation. Proceedings of the National Academy of Sciences. 2021; 118 e2020150118
19. Hobbs RP, et al. Keratin-dependent regulation of Aire and gene expression in skin tumor keratinocytes. Nat Genet. 2015; 47: 933–938. [PubMed: 26168014]
20. Liu Z, et al. Keratin 17 activates AKT signalling and induces epithelial-mesenchymal transition in oesophageal squamous cell carcinoma. Journal of Proteomics. 2020; 211 103557 [PubMed: 31669361]
21. Narushima K, et al. Establishment of a DEN-induced mouse model of esophageal squamous cell carcinoma metastasis. Esophagus. 2017; 14: 131–137.
22. Rubio CA, Liu FS, Chejfec G, Sveander M. The induction of esophageal tumors in mice: dose and time dependency. In Vivo. 1987; 1: 35–38. [PubMed: 2979761]
23. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017; 168: 613–628. [PubMed: 28187284]
24. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012; 481: 306–313. [PubMed: 22258609]
25. Hall MWJ, Jones PH, Hall BA. Relating evolutionary selection and mutant clonal dynamics in normal epithelia. Journal of the Royal Society Interface. 2019; 16 20190230
26. Martincorena I, et al. Universal Patterns of Selection in Cancer and Somatic Tissues. Cell. 2017; 171: 1029–1041. e1021 [PubMed: 29056346]

27. Martincorena I, Jones PH, Campbell PJ. Constrained positive selection on cancer mutations in normal skin. Proceedings of the National Academy of Sciences. 2016; 113: E1128–E1129.
28. Prasad V, et al. Haploinsufficiency of Atp2a2, Encoding the Sarco(endo)plasmic Reticulum Ca^{2+}-ATPase Isoform 2 Ca^{2+} Pump, Predisposes Mice to Squamous Cell Tumors via a Novel Mode of Cancer Susceptibility. Cancer Research. 2005; 65: 8655–8661. [PubMed: 16204033]
29. Liu LH, Boivin GP, Prasad V, Periasamy M, Shull GE. Squamous Cell Tumors in Mice Heterozygous for a Null Allele of Atp2a2, Encoding the Sarco(endo)plasmic Reticulum Ca2+-ATPase Isoform 2 Ca2+Pump. Journal of Biological Chemistry. 2001; 276: 26737–26740.
30. Tumbar T, et al. Defining the epithelial stem cell niche in skin. Science. 2004; 303: 359–363. [PubMed: 14671312]
31. Brown S, et al. Correction of aberrant growth preserves tissue homeostasis. Nature. 2017; 548: 334–337. [PubMed: 28783732]
32. Kon S, et al. Cell competition with normal epithelial cells promotes apical extrusion of transformed cells through metabolic changes. Nature Cell Biology. 2017; 19: 530. [PubMed: 28414314]
33. Shultz LD, et al. Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. Journal of immunology (Baltimore, Md.: 1950). 2005; 174: 6477–6489.
34. Ellis P, et al. Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing. Nature Protocols. 2021; 16: 841–871. [PubMed: 33318691]
35. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv. 2013. 1303.3997
36. Gerstung M, Papaemmanuil E, Campbell PJ. Subclonal variant calling with multiple samples and prior knowledge. Bioinformatics. 2014; 30: 1198–1204. [PubMed: 24443148]
37. Alexandrov LB, et al. The repertoire of mutational signatures in human cancer. Nature. 2020; 578: 94–101. [PubMed: 32025018]
38. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome biology. 2016; 17: 31. [PubMed: 26899170]
39. Jones D, et al. cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. Current Protocols in Bioinformatics. 2016; 56: 15.10.11–15.10.18.
40. Menzies A, et al. VAGrENT: Variation Annotation Generator. Current Protocols in Bioinformatics. 2015; 52 15.18.11
41. Scheinin I, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome research. 2014; 24: 2022–2032. [PubMed: 25236618]
42. Kuilman T, et al. CopywriteR: DNA copy number detection from off-target sequence data. Genome biology. 2015; 16: 49. [PubMed: 25887352]
43. Nik-Zainal S, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012; 149: 979–993. [PubMed: 22608084]
44. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface. 2009; 6: 187–202. [PubMed: 19205079]
45. Klinger E, Rickert D, Hasenauer J. pyABC: distributed, likelihood-free inference. Bioinformatics. 2018; 34: 3591–3593. [PubMed: 29762723]
46. Consortium, T. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017; 541: 169–175. [PubMed: 28052061]
47. Klein AM, Brash DE, Jones PH, Simons BD. Stochastic fate of p53-mutant epidermal progenitor cells is tilted toward proliferation by UV B during preneoplasia. Proceedings of the National Academy of Sciences. 2010; 107: 270–275.
48. Clayton E, et al. A single type of progenitor cell maintains normal epidermis. Nature. 2007; 446: 185–189. [PubMed: 17330052]

49. Kendall DG. On the Generalized "Birth-and-Death" Process. The Annals of Mathematical Statistics. 1948; 19: 1–15.

50. Orr HA. THE POPULATION GENETICS OF ADAPTATION: THE DISTRIBUTION OF FACTORS FIXED DURING ADAPTIVE EVOLUTION. Evolution. 1998; 52: 935–949. [PubMed: 28565213]

Europe PMC Funders Author Manuscripts

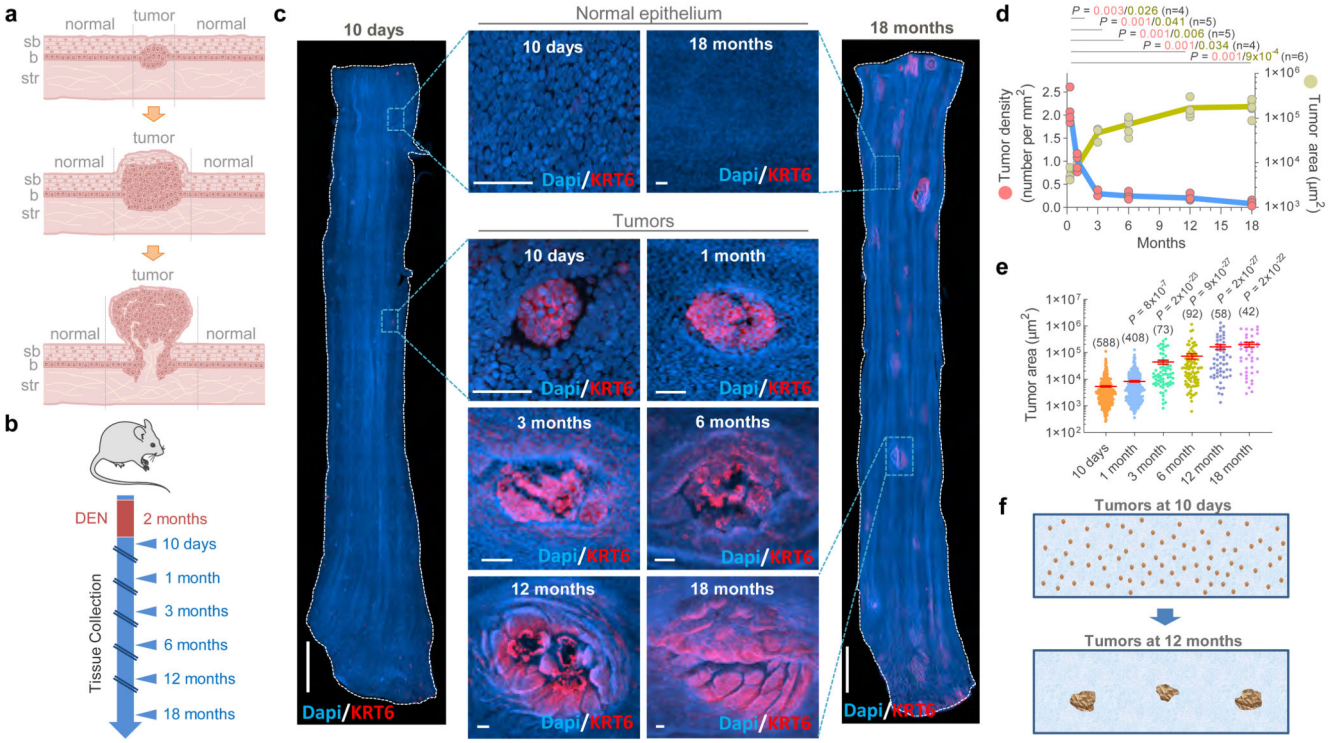


Figure 1. Most esophageal tumors are lost over time.

(**a**) Cartoon depicting epithelial tumor formation and growth over time. Sb, subrabasal; b, basal; str, stroma. (**b**) Protocol: wild-type mice received the carcinogen Diethylnitrosamine (DEN) in drinking water. Esophagi were collected at the indicated time-points. (**c**) Examples of confocal images of the entire mouse esophageal epithelium collected at 10 days (left) or 18 months (right) post-DEN treatment, stained for Dapi (blue) and Keratin 6 (KRT6, red). Insets show esophageal normal epithelium (top images) or tumors (bottom images) from the indicated time points post-DEN treatment. Scale-bars: 2mm, insets: 50µm. (**d**) Tumor density (left axis) and size (right axis) at the indicated time-points post-DEN treatment. Statistics are two-tailed Welch's t test vs 10-days (n=number of mice, shown in brackets). Colored lines connect the mean values. (**e**) Area of individual tumors at the indicated time points post-DEN treatment. Number of tumors per time point is indicated in brackets. Lines indicate mean±s.e.m. Statistics are two-tailed Mann Whitney test vs 10-days. (**f**) Cartoon illustrating tissue tumor clearance over time.

Europe PMC Funders Author Manuscripts

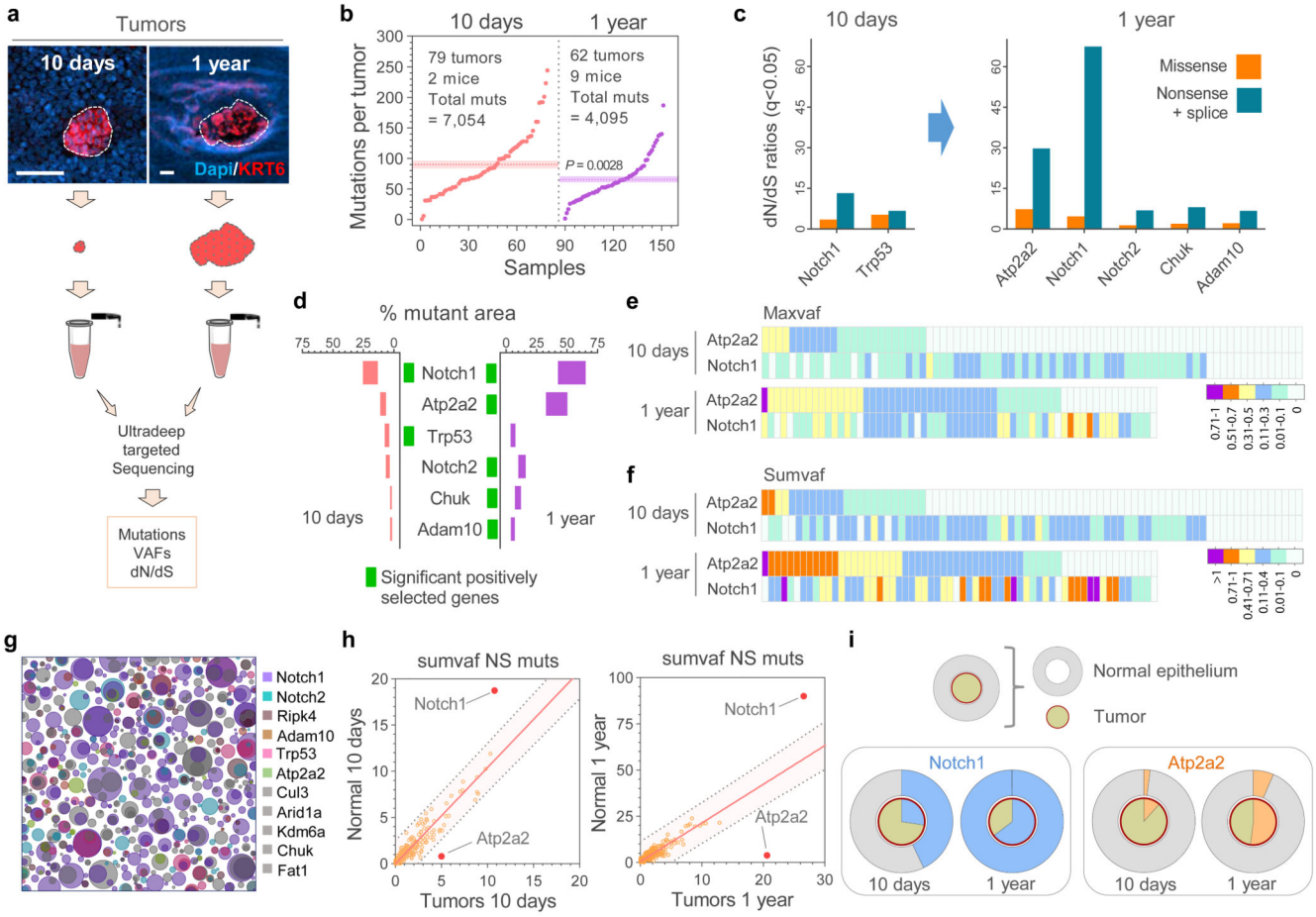


Figure 2. Sequencing of tumors and adjacent normal tissue.

(**a**) Esophageal tumors were collected 10-days or 1-year post-Diethylnitrosamine (DEN) treatment and ultra-deep targeted sequencing with a 192 gene bait-set performed. Scale-bars: 50μm. (**b**) Mutations per tumor, including essential splice, frameshift, missense, nonsense and silent mutations. Dotted lines and colored shadows indicate mean±s.e.m (two-tailed Mann-Whitney test). (**c**) dN/dS ratios for missense and truncating (nonsense + essential splice site) mutations of positively selected genes (dN/dS>1) in the 10 days and 1 year post-DEN tumors (only significant genes for each time point are shown, q<0.05 calculated with R package dNdScv[26]). (**d**) Estimated percentage of tumor area carrying non-synonymous mutations in the positively selected genes (identified with green rectangles for each time point). Range indicates upper and lower bound estimates. (**e, f**) Heatmaps indicating the maximum (**e**) and summed VAF (**f**) for non-synonymous mutations in *Atp2a2* and *Notch1* in 10-day and 1-year post-DEN tumors. (**g**) Schematic representation of the mutant clones in normal esophageal epithelium collected 10 days post-DEN treatment. Only mutations of the positively selected genes (dN/dS>1) in normal or tumor samples at 10 days and 1 year[13] are shown. The density and size of the clones are inferred from the sequencing data, and the nesting of clones and subclones is inferred from the data when possible and randomly allocated otherwise. (**h**) Correlations between the summed VAF of nonsynonymous mutations (NS muts) in the 192 genes sequenced from normal epithelium

Europe PMC Funders Author Manuscripts

and tumors at 10-days (left) and 1-year (right) post-DEN treatment. Centre lines show two-tailed Pearson correlations (R^2=0.8329; P=1.6x10^{-76}) and R^2=0.7081; P=3.2x10^{-53}, respectively). Red dots indicate genes outside the 99% predicted band (colored area within grey dotted lines). (**i**) Percentage of area covered by *Notch1* (blue, left graphs) and *Atp2a2* (orange, right graphs) mutations in the normal epithelium (outer circles) versus the tumors (inner circles) collected at 10-days or 1-year post-DEN treatment.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

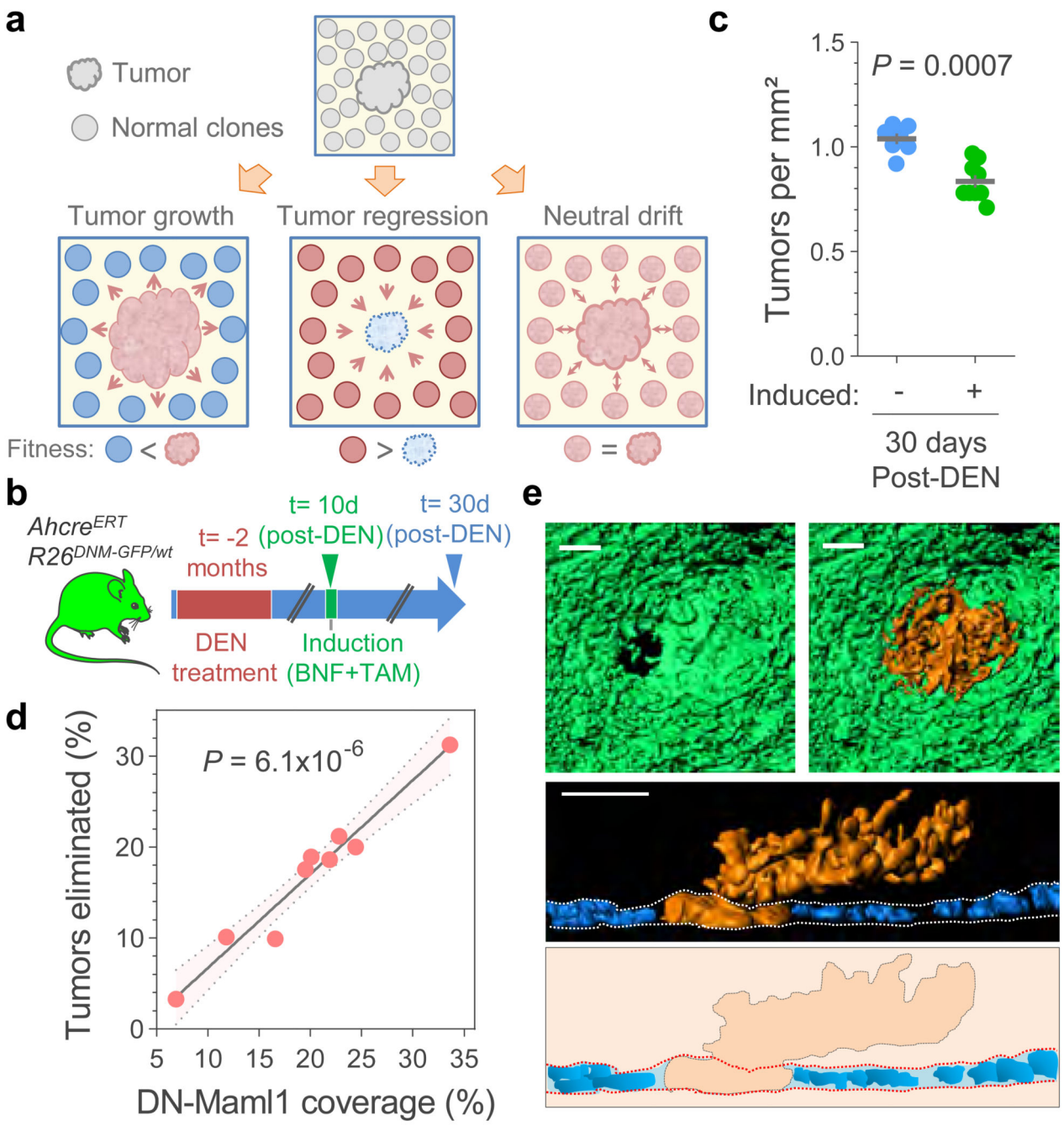


Figure 3. Elimination of early tumors by expansion of mutant clones in the adjacent normal epithelium.

(**a**) Hypothesis: Tumor survival depends on their competitive fitness relative to that of the surrounding normal tissue, predicting 3 possible scenarios: 1) tumor cells are fitter than clones in the normal adjacent epithelium (left), tilting the balance towards tumor growth; 2) Tumor cells are less fit than the adjacent normal epithelium (middle), resulting in tumor regression and elimination, and 3) cells in tumor and adjacent normal epithelium are equally fit (right), resulting in neutral drift. (**b**) Protocol: *Ahcre*^{ERT}/*R26*$^{DNM–GFP/wt}$ mice received

DEN for two months. 10 days after DEN withdrawal DN-Maml1 clones were induced and tissues harvested twenty days later. Control mice received DEN but were not induced. (**c**) Number of tumors per mm^2 of esophageal epithelium in induced and non-induced $Ahcre^{ERT}/R26^{DNM-GFP/wt}$ mice (n= 9 and 7 mice, respectively). Lines indicate mean±s.e.m (two-tailed Mann-Whitney test). (**d**) Correlation between the area covered by DN-Mam1 clones and tumors being eliminated (see Methods). n=9 mice. Line shows the two-tailed Pearson correlation (R^2=0.9541) with 95% confidence interval (colored area). (**e**) Rendered confocal images showing top views (top images), lateral view (middle image) and schematic (bottom image) of a tumor (orange) surrounded by a DN-Maml1 clone (green). Dapi is shown in blue. Scale-bars: 20μm.

Europe PMC Funders Author Manuscripts