Somatic mutant selection is altered by prior *NOTCH1* mutation in aging esophageal epithelium

Charlotte King[1], Emilie Abby[1], Joanna C. Fowler[1], Irina Abnizova[1], Roshan K. Sood[1], Swee Hoe Ong[1], Michael W. J. Hall[1,2], Faye Lynch-Williams[3], Benjamin A. Hall[4], Philip H. Jones[1,2,5]

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

[2]Department of Oncology, University of Cambridge, Hutchinson Research Centre, Cambridge Biomedical Campus, Cambridge, CB2 0XZ, UK.

[3]School of Clinical Medicine, Addenbrooke's Hospital, Hills Rd, Cambridge CB2 0SP, UK

[4]Department of Medical Physics and Biomedical Engineering, University College London, London, WC1E 6BT, UK.

[5]Correspondence to: pj3@sanger.ac.uk

Abstract

In cancer evolution, genome alterations often occur in a specific order, implying selection depends on the prior clonal genotype[1-3]. It is unknown if similar constraints operate in normal epithelia. Here, we mapped mutations in normal mid-esophagus of aged UK subjects. Mutant *NOTCH1* clones colonized most of the epithelium by age 60 and the tissue became saturated with mutants under strong competitive selection by age 70. Compared to samples that were mostly *NOTCH1* wildtype, samples predominantly mutant for *NOTCH1* showed weaker selection of mutant *TP53* and increased selection of *NOTCH2* mutants. In mouse esophagus lacking *Notch1* we observed strong selection of mutant *Notch2,* not seen in wild type epithelium. In aging esophagus, the first driver mutation may change the trajectory of subsequent somatic evolution by altering mutational selection and by restricting space available for the expansion of other mutant clones.

Main Text

There is strong evidence that the order in which somatic genome alterations occur is critical in carcinogenesis[1]. In the squamous esophagus, for example, loss of both alleles of *TP53* is an initiating event, being found from the earliest lesions to esophageal squamous cell carcinoma (ESCC)[4,5]. Conversely, *NOTCH1* mutations are more common in normal epithelium than ESCC, and in mice *Notch1* mutation reduces tumor growth[6].

The esophagus is an ideal tissue in which to investigate if 'order matters' for somatic mutations in a normal epithelial context. Esophageal epithelium is progressively colonized over decades by mutant clones under strong positive selection, which compete for the limited space available[7-9]. We began our investigation by mapping the mutational landscape of a large area of epithelium in older subjects.

We employed three DNA sequencing methods to map the mutational landscape in normal mid-esophageal epithelium. We analysed 2,856 samples by combining results from 22 transplant donors aged up to 79 years with 9 previously sequenced subjects (**Fig. 1**, **Supplementary Table 1**)[7]. All donors had no history of cancer or oesophageal disease. Analysis of germline single-nucleotide polymorphisms altering the risk of developing ESCC found the donors were genetically representative of the UK population (**Methods**, **Supplementary Table 2**)[10-14].

The epithelium was dissected into a gridded array of rectangular samples each 2 mm^2 in area (**Fig. 1**). A total 2,405 two mm^2 samples were analysed (total area 48.1cm^2). In addition, a further 451 samples were collected using a circular micro biopsy punch (area=0.05 mm^2) in a spaced array (**Fig. 1**)[8].

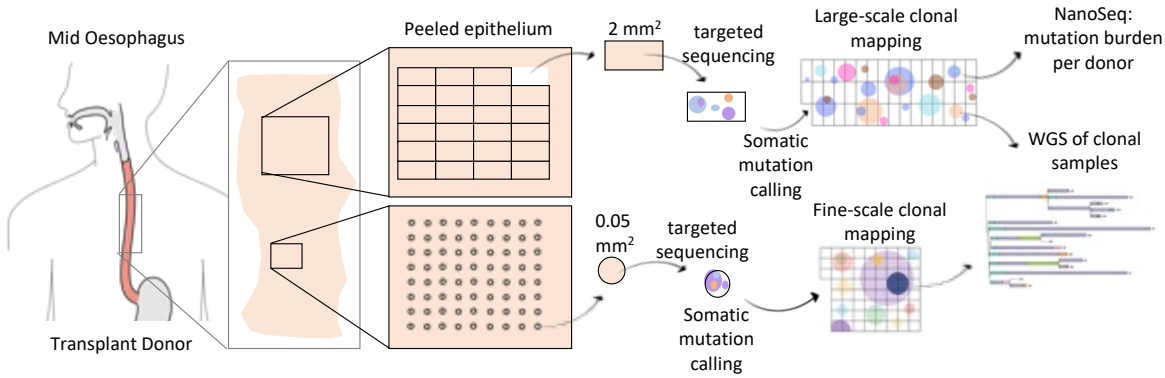
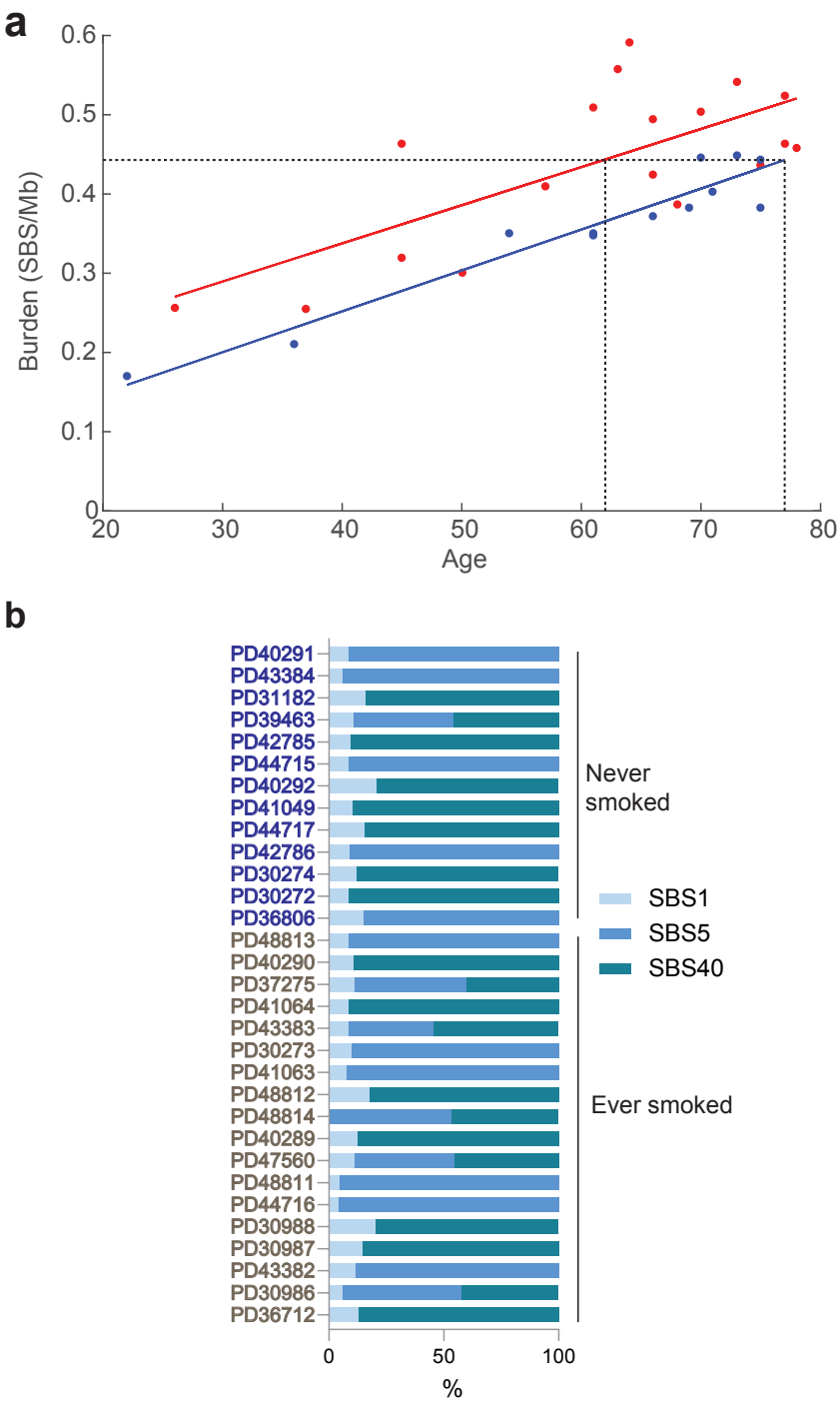


Figure 1: Protocol. Normal epithelium of the middle third of the esophagus from transplant donors was sampled to enable large- and fine-scale mapping of somatic mutant clones by taking arrayed samples of 2 mm^2 and 0.05 mm^2 respectively for targeted sequencing. A subset of clonal samples was analysed by whole genome sequencing (WGS). Mutational burden was determined from NanoSeq analysis of a 2 mm^2 sample from each donor.

We first estimated the genome-wide mutation burden in all 31 donors using a modified form of duplex sequencing (NanoSeq), an accurate single molecule assay[15]. Substitution burdens ranged from 0.17-0.59 mutations/megabase (**Supplementary Table 3).** A general linear model (ANCOVA) suggested that substitution burden increased with both donor age (p=5.71e-08) and smoking status (p= 8.23e-04). Having smoked increased oesophageal mutational burden by the equivalent of about 14 years of age compared to never smoking (**Extended data Fig. 1a**).



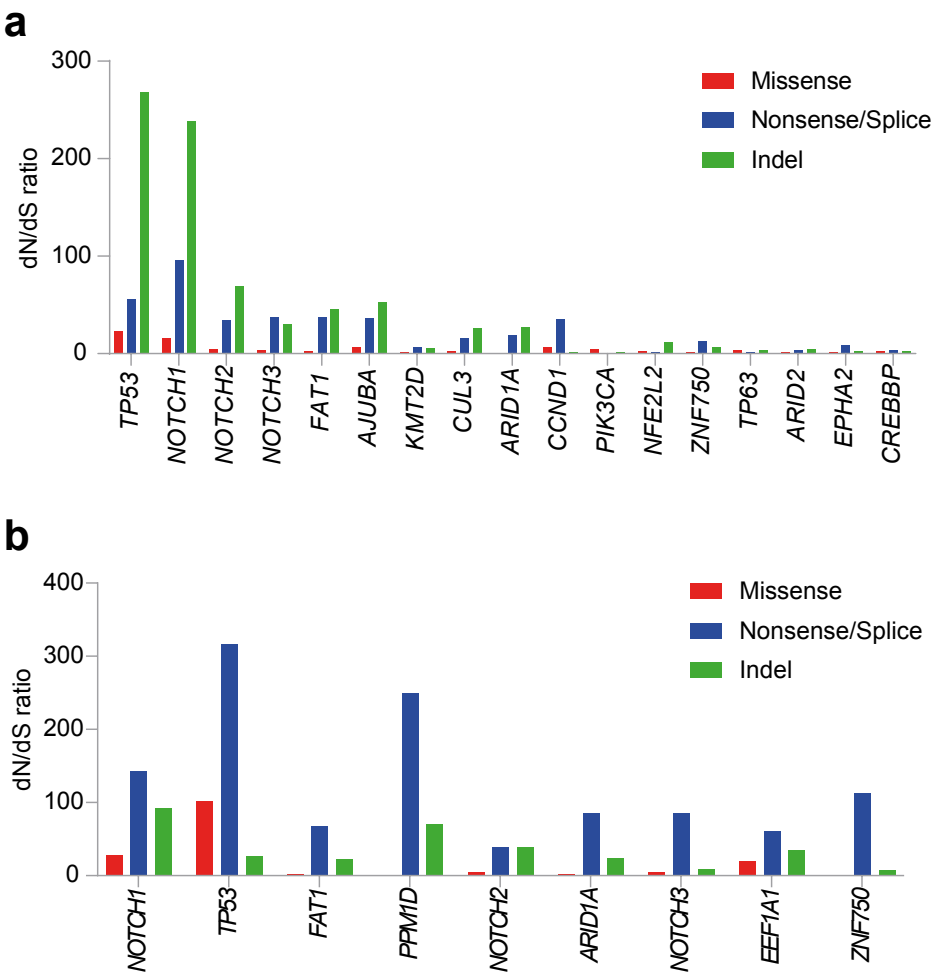
Extended data Figure 1: NanoSeq analysis. **a**. Mutational burden (single base substitutions/megabase) in ever smokers (magenta), and never smokers (blue). Lines show linear correlation of burden with age (smokers R=0.714, p=0.00128; non-smokers R=0.95 p =0.00001). Dashed lines projected to the x-axis show age difference between smokers and non-smokers to reach the same mutational burden, dash lined projected to the y-axis. **b** Mutational signatures of single base substitutions.

Mutational signature analysis revealed three age correlated single base signatures SBS1, 5 and 40 (https://cancer.sanger.ac.uk/signatures/sbs)[16] (**Extended Data Fig. 1b, Supplementary Table 3**). There was no variation in single base mutational signatures associated with smoking

status, consistent with a study of over 500 ESCC genomes, which identified a tobacco-associated SBS signature in only 8% of tumors in smokers[17].

Next, we mapped the mutations across 2 mm^2 samples using targeted sequencing (median coverage 855x, range 219-1454x) (**Methods, Supplementary Table 4**). After merging clones spanning adjacent samples 20,589 independent mutations were identified (**Supplementary Table 5**). In addition, 451 0.05 mm^2 samples were collected in regularly spaced arrays from the epithelium of 7 donors aged 61-75 years (**Fig. 1**) and subjected to targeted sequencing of 324 cancer associated genes (**Supplementary Table 1, Methods**)[18]. 1406 independent mutations were identified after merging mutants spanning adjacent samples (**Supplementary Tables 6,7**).

The ratio of protein altering to synonymous mutations (dN/dS) for each gene indicates whether mutants are under selection[7,19]. dN/dS analysis revealed a total of 17 mutant genes under positive selection for missense, protein truncating and indel mutations in 2 mm^2 and 0.05 mm^2 samples (*dNdScv*, $q<0.01$, **Extended data Fig.2a-b**, **Supplementary Table 8**). Of these, mutant *EPHA2* and *EEF1A1* have not been reported to drive clonal expansion in normal esophagus. *EPHA2* encodes a receptor for EPH ligands, and promotes keratinocyte differentiation, consistent with mutants driving clonal expansion[20]. Mutants of *EEF1A1,* which has functions beyond protein synthesis, drive clonal expansion in the epidermis [21,22].



Extended Data Figure 2: Positive selection of mutant genes. a. Ratio of missense (red), nonsense/splice (blue) and indels (green) for positively selected genes in 2 mm^2 samples. Neutral, unselected genes have dN/dS=1. dN/dS ratios calculated with dNdScv. (q<0.01, see **Supplementary Table 8**). **b** Ratio of missense (red), nonsense/splice (blue) and indels (green) for positively selected genes in 0.05 mm^2 samples. Neutral, unselected genes have dN/dS=1. dN/dS ratios calculated with dNdScv. (global q<0.01, **Supplementary Table 8**).

We used data from the 2 mm^2 samples to generate spatial maps summarising colonization by mutant clones in different age groups (**Fig. 2a, Methods**). The area of mutant epithelium rose progressively, and above age 70 the tissue became saturated with mutant clones under strong positive selection (**Fig. 2b**). The most prevalent selected mutant gene was *NOTCH1*, which occupied the majority of wild type epithelium by age 60 and colonized a median of 80% of epithelium in the over 70s (**Fig. 2c**)[6-8]. Notably, the median proportion of epithelium colonized by mutant *TP53* increased from under 5% in those under 60 to 19% for donors over 70 (**Fig. 2d, Supplementary Table 9**).

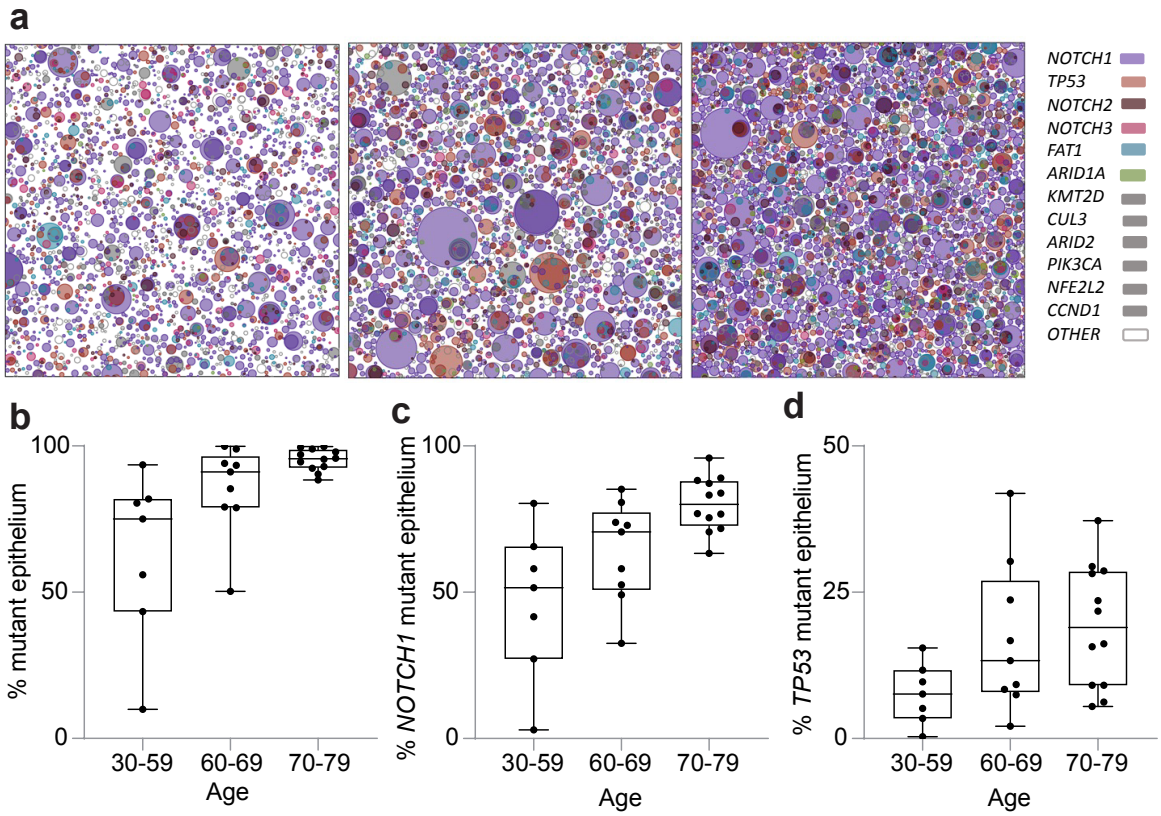


Figure 2: Mutant clones in aging normal esophageal epithelium. a Top-down views of nonsynonymous mutant clones for genes under positive selection in 9 cm^2 of epithelium in age groups. Samples were randomly selected, and mutations displayed as circles, randomly distributed in the space. Sequencing data, including copy number, was used to infer the size and number of clones and, where possible, the nesting of sub-clones. Otherwise, sub-clones are nested randomly. **b** Estimate of the percentage of mutant epithelium. **c, d** Proportion by area of epithelium mutant for *NOTCH1*, **c** and *TP53*, **d**. Boxes indicate median and quartiles, bars range, dots are individual donors (**Methods**, **Supplementary Table 9**).

To investigate clonal genome alterations, we performed whole genome sequencing (WGS) on forty 2 mm^2 and 173 0.05 mm^2 samples in which a mutation was present in 50% or more cells (**Supplementary Tables 1, 10–14**). Phylogenetic analysis revealed the 0.05 mm^2 samples derived from 63 independent clones **(Supplementary Table 11)**. The median mutational burden of 0.05 mm^2 samples was 1850 single base substitutions per genome (range 202-3011). Mutational signature analysis was consistent with NanoSeq, except in three donors in whom signature 16, associated with alcohol, was detected (**Supplementary Table 13**).

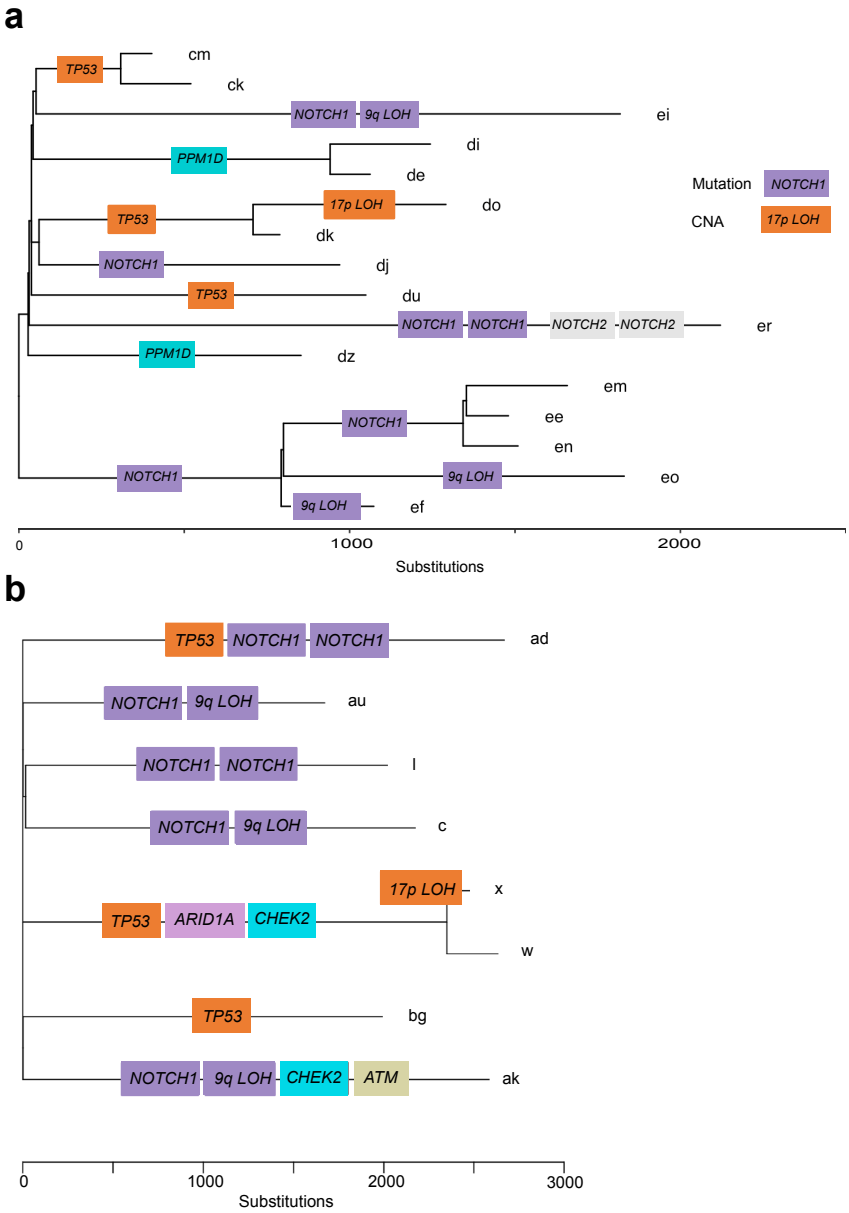


Figure 3: Whole genome sequencing (WGS) of samples. Phylogenies derived from whole genome sequencing of clonal 0.05 mm^2 (**a**) and 2 mm^2 (**b**) samples. Positively selected mutations and associated copy number changes are shown, colors indicate locus. **a,** 75-year-old non-smoking male (PD43384). **b**, 69-year-old non-smoking female (PD40292).

Figure 3a summarises data from WGS of sixteen 0.05 mm^2 samples from eight clones in a 75-year-old male never smoker, each carrying at least one positively selected mutation. Three clones had a *TP53* mutation and one (sample *do*) had undergone loss of heterozygosity (LOH) at the *TP53* locus on chromosome 17p (**Supplementary Table 11**). Conversely, out of eight *NOTCH1* mutant clones there were three occurrences of LOH at the *NOTCH1* locus on chromosome 9q (samples *ei, eo,* and *ef*). Across all 173 WGS samples we observed only one example of *TP53* LOH but 22 occurrences of *NOTCH1* LOH, four of which occurred in the absence of a *NOTCH1* mutation (**Supplementary Table 11**). Additional copy number alterations (CNA) were found in only two clones (18p deletion with 20p amplification and Chromosome 7 amplification, **Supplementary Table 11**).

Genomes of the larger clones in 2 mm^2 samples carried between one and three positively selected drivers (**Fig. 3b**, **Supplementary Table 12**). The most commonly mutated genes were *NOTCH1* (25 samples) and *TP53* (11 samples). Biallelic disruption of these genes was observed in 23/25 and 3/11 samples respectively.

We used targeted sequencing data to extend our analysis of CNA at the *NOTCH1* and *TP53* loci to 393 0.05 mm^2 samples (**Methods**). In samples sequenced by both WGS and targeted sequencing the calls at both loci agreed. 25% (97 of 393) of samples carried CNA at the *NOTCH1* locus, but only 0.3% (1 of 393) CNA at the *TP53* locus. Of the samples with *NOTCH1* CNA, 70% (68 of 97) carried one or more *NOTCH1* mutation(s) (**Supplementary Table 6**). Overall genomes in aging normal esophagus remain genome stable apart from LOH at driver loci.

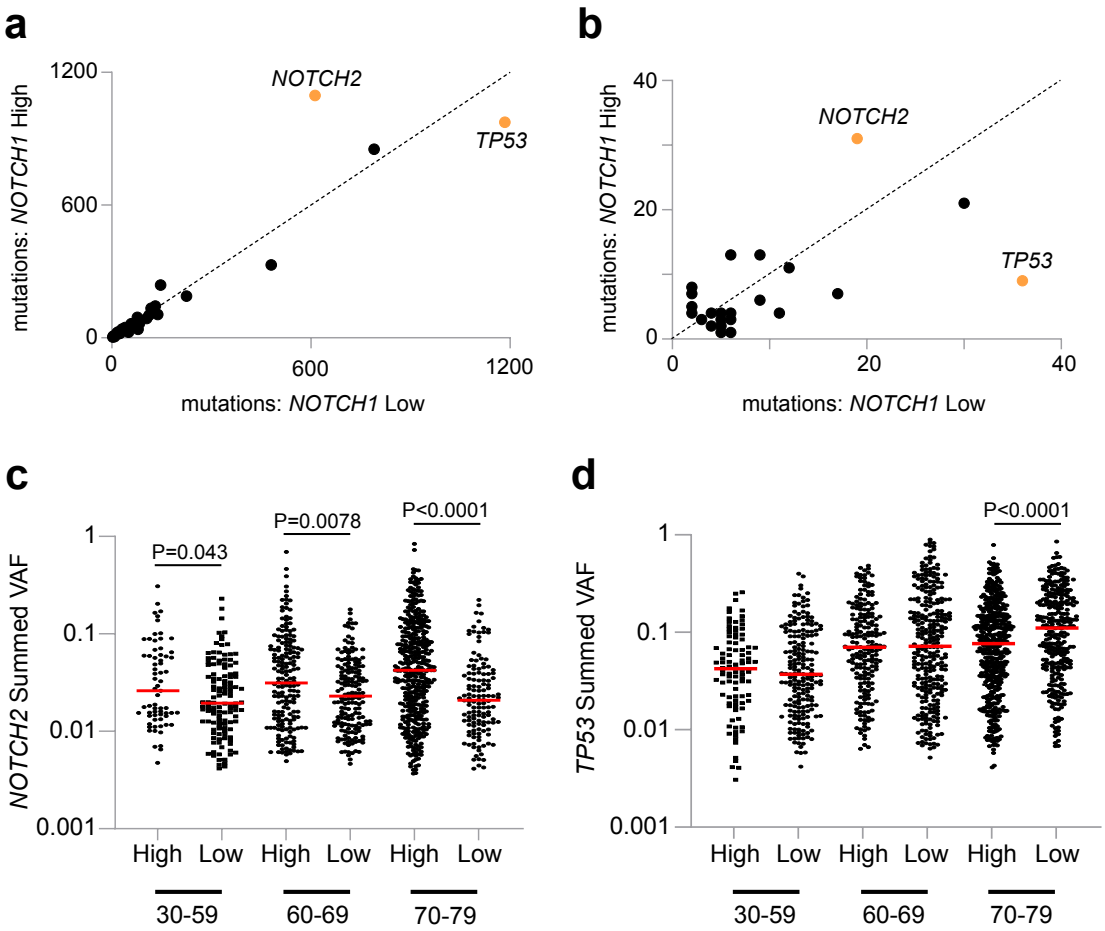


Figure 4: Mutant selection in samples with different levels of *NOTCH1* mutation. a,b 2 mm^2 **(a)** and 0.05 mm^2 **(b)** samples were sorted by the summed VAF of all NOTCH1 mutants they carried into mutant *NOTCH1*High (*NOTCH1* summed VAF >0.4) and *NOTCH1*Low (*NOTCH1* summed VAF <0.4). n samples **a**, *NOTCH1High* 1237, *NOTCH1*Low 1073 **b** *NOTCH1*High 188, *NOTCH1*Low 181. The number of protein-altering mutations in other genes is plotted for each group, orange indicates significant differences. P values **a**, mutant *NOTCH2* <0.0001, *TP53* 0.02, **b,** mutant *NOTCH2*, 0.04 *TP53,* 0.0002, 2 tailed Z test. **c,d,** summed VAF for mutant *NOTCH2* (**c**) and *TP53* (**d**) in mutant *NOTCH1*High and *NOTCH1*Low samples in different age groups, 2 tailed Mann Whitney test.

The most striking observation was the disruption of *NOTCH1* across most of the tissue by middle age. We speculated that this may alter the selection of mutations that occur in the subsequent decades of life. To test this, we first compared the number of non-synonymous mutants in samples which were predominantly *NOTCH1* mutant (*NOTCH1*High), i.e. the summed VAF for all *NOTCH1* mutant clones exceeded 0.4 (**Methods**) with the remaining *NOTCH1*Low samples. In both 0.5 mm^2 and 2 mm^2 *NOTCH1*High samples, selection of mutant *NOTCH2* was increased and selection of mutant *TP53* mutants decreased compared with the corresponding *NOTCH1*Low samples, while other positively selected mutants were unchanged (**Fig. 4a-b**). Furthermore, the VAF of *NOTCH2* mutations was increased in 2 mm^2 *NOTCH1*High samples across all age groups (**Fig. 4c**). *TP53* mutant clones had a smaller

VAF in *NOTCH1*High samples in donors aged over 70 (**Fig 4d**). These results suggests that selection may be altered following *NOTCH1* mutation.

To challenge this hypothesis, we turned to a mouse model. Mouse esophagus differs from human in having a single, basal, proliferative cell layer, and fewer suprabasal cell layers which undergo keratinized differentiation[23,24]. However, exposure to mutagens generates a mutational landscape resembling that of aging human esophagus[9,25], in which mutant *Notch1* and *p53* drive clonal expansion[6,26].

Notch1 null esophageal epithelium was generated in *YFPCreNotch1*$^{flox/flox}$ mice, a triple transgenic strain carrying a drug inducible *Cre* allele, and two conditional alleles, a neutral EYFP reporter in the *Rosa26* locus and a conditional *Notch1* allele (**Fig. 5a, Methods**)[6]. *Cre* was induced and animals aged, after which the epithelium was divided into a gridded array of 2 mm^2 samples and subjected to targeted sequencing for *Notch* pathway and cancer genes (**Supplementary Table 16**)[6]. Uninduced littermates with a *Notch1* wild type esophagus were used as controls. The average sequencing depth was 365x in control samples and 339x in *Notch1*$^{-/-}$ samples. The sequencing depth of exon1 of *Notch1*, which is deleted by *Cre* recombination, was used to check for recombination in each sample (**Extended Data Fig. 3a**). Sparse spontaneous mutations were detected in the mouse aged epithelium (44 mutations in 51 *Notch1*$^{-/-}$ samples and 39 mutations in 50 *Notch1*$^{+/+}$ control samples) (**Supplementary Tables 17-20**). Most nonsynonymous mutations were in *Notch1* in *Notch1*$^{+/+}$ epithelium (25/35) and in *Notch2* in *Notch1*$^{-/-}$ tissue (35/41) (**Fig. 5b**). dNdS analysis revealed mutant *Notch1* was positively selected in wild type esophagus whereas *Notch2* mutants were selected in *Notch1* null esophagus, with the selected mutant genes colonizing similar areas of epithelium in their respective strains (**Fig 5b-c**).

To confirm this result, we aged a second cohort of induced and control mice and immunostained the esophageal epithelium for NOTCH2, finding foci of altered NOTCH2 staining in *Notch1*$^{-/-}$ but not control esophagi (**Fig. 5d-f**). Targeted sequencing showed the areas of altered NOTCH2 staining carried clonal nonsynonymous *Notch2* mutations (**Extended Data Fig 3b-e, Supplementary Tables 20,21**). We conclude that prior *Notch1* deletion increases selection of *Notch2* in aging mouse esophagus, a finding consistent with altered selection seen in mutant *NOTCH1*High samples in humans.

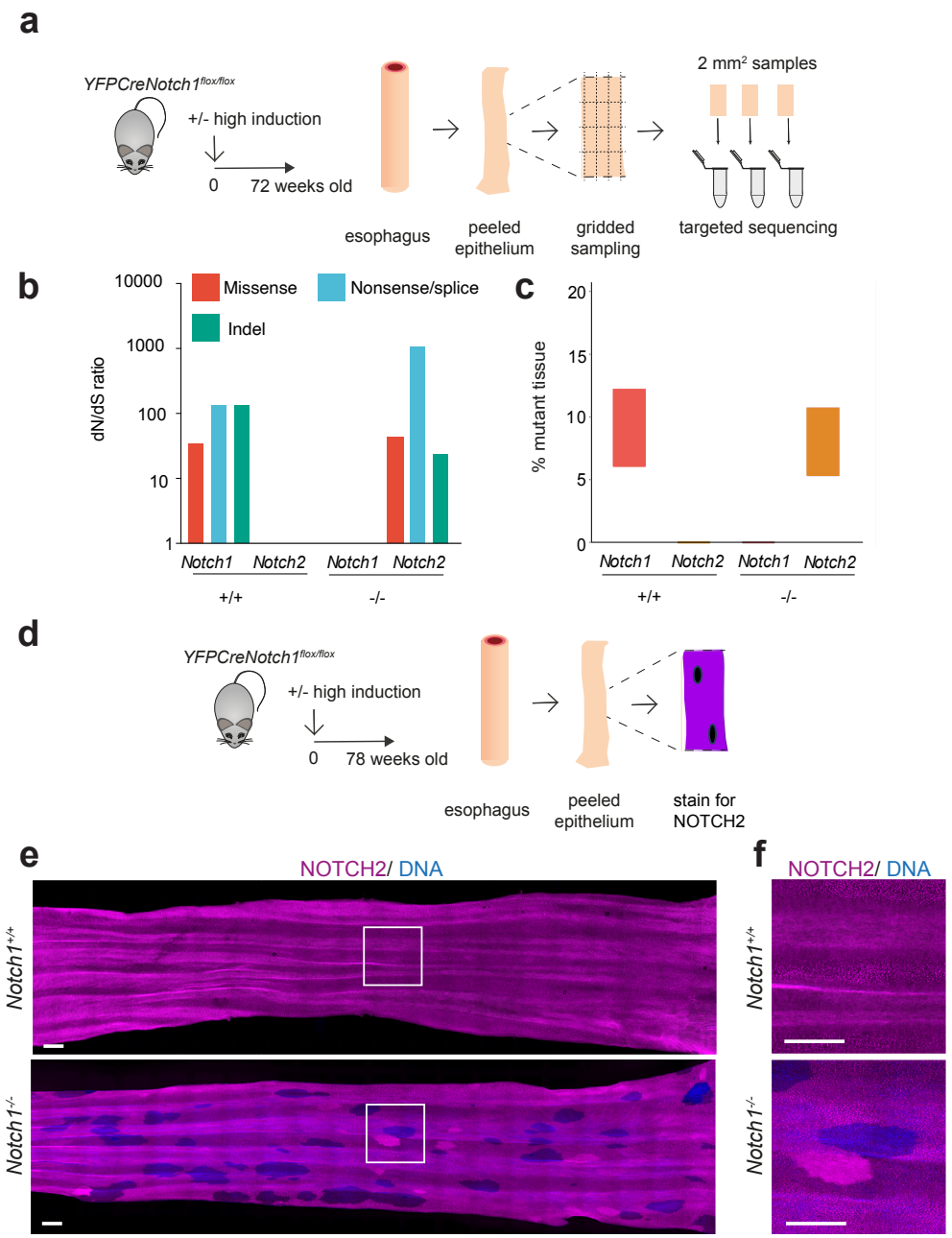
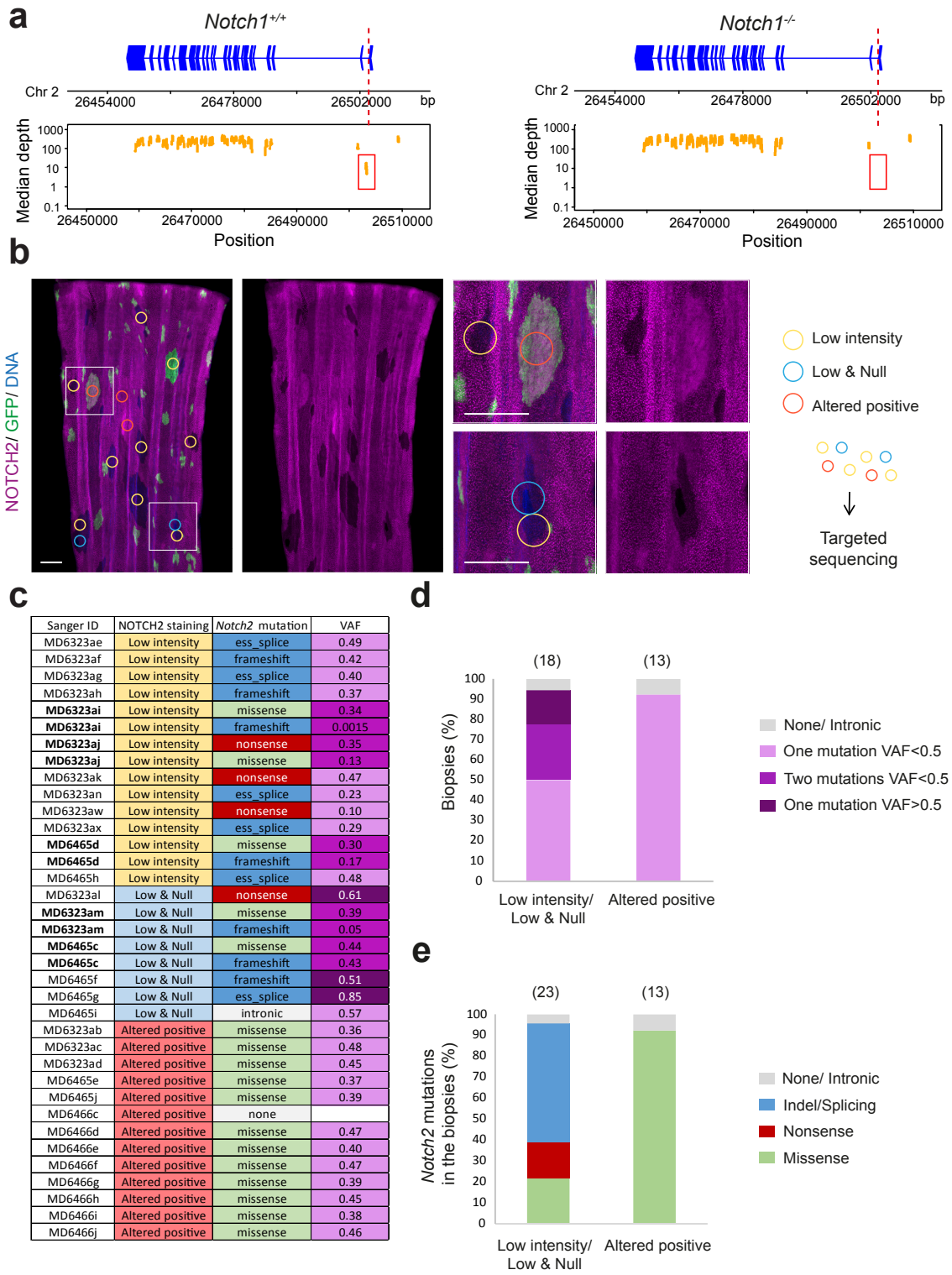


Fig 5. Selection of *Notch2* mutations in aging *Notch1*$^{-/-}$ mouse esophagus

a. Protocol. Highly induced or uninduced control *YFPCreNotch1*$^{flox/flox}$ mice were aged. The esophagus was peeled and gridded into 2 mm^2 samples that were processed for targeted sequencing (n= 50-51 total biopsies from 2 mice for each genotype, mice were 72 weeks old ± 4 weeks). **b.** dN/dS ratios for mutant *Notch1* and *Notch2* in control (+/+) and induced *Notch1*$^{-/-}$ (-/-) esophagus. **c.** Proportion by area of epithelium mutant for *Notch1* and *Notch2* in in control (+/+) and induced *Notch1-/-* (-/-) esophagus. **d:** protocol: NOTCH2 immunostaining in *YFPCreNotch1*$^{flox/flox}$ mice of up to 78 weeks of age. **e, f.** Representative immunofluorescence staining for NOTCH2 (magenta) and DNA (blue) in epithelium from aged wild type and*Notch1*$^{-/-}$ mice. Images are representative of entire esophagus from 5 *Notch1*$^{+/+}$ and 11 *Notch1*$^{-/-}$ animals. White frames in **e** indicate magnified images shown in insets (**f**). Scale bars, 500µm.



Sanger ID	NOTCH2 staining	*Notch2* mutation	VAF
MD6323ae	Low intensity	ess_splice	0.49
MD6323af	Low intensity	frameshift	0.42
MD6323ag	Low intensity	ess_splice	0.40
MD6323ah	Low intensity	frameshift	0.37
MD6323ai	Low intensity	missense	0.34
MD6323ai	Low intensity	frameshift	0.0015
MD6323aj	Low intensity	nonsense	0.35
MD6323aj	Low intensity	missense	0.13
MD6323ak	Low intensity	nonsense	0.47
MD6323an	Low intensity	ess_splice	0.23
MD6323aw	Low intensity	nonsense	0.10
MD6323ax	Low intensity	ess_splice	0.29
MD6465d	Low intensity	missense	0.30
MD6465d	Low intensity	frameshift	0.17
MD6465h	Low intensity	ess_splice	0.48
MD6323al	Low & Null	nonsense	0.61
MD6323am	Low & Null	missense	0.39
MD6323am	Low & Null	frameshift	0.05
MD6465c	Low & Null	missense	0.44
MD6465c	Low & Null	frameshift	0.43
MD6465f	Low & Null	frameshift	0.51
MD6465g	Low & Null	ess_splice	0.85
MD6465i	Low & Null	intronic	0.57
MD6323ab	Altered positive	missense	0.36
MD6323ac	Altered positive	missense	0.48
MD6323ad	Altered positive	missense	0.45
MD6465e	Altered positive	missense	0.37
MD6465j	Altered positive	missense	0.39
MD6466c	Altered positive	none	
MD6466d	Altered positive	missense	0.47
MD6466e	Altered positive	missense	0.40
MD6466f	Altered positive	missense	0.47
MD6466g	Altered positive	missense	0.39
MD6466h	Altered positive	missense	0.45
MD6466i	Altered positive	missense	0.38
MD6466j	Altered positive	missense	0.46

Extended Data Fig. 3: Selection of *Notch2* mutations in aging *Notch1*$^{-/-}$ mouse esophagus
a Median depth of sequencing along *Notch1* sequence in *Notch1*$^{+/+}$ and *Notch1*$^{-/-}$ biopsies from aging mice shown in **Fig. 5a-c**. Dashed red line and red box indicate site of *Cre* induced excision of the first exon of *Notch1*. **b.** Whole mounts of highly induced *YFPCreNotch1*$^{flox/flox}$ mouse epithelium after aging up to 78 weeks old and staining for NOTCH2 (magenta), GFP

(green) and DNA (blue). Large ovoid areas of altered NOTCH2 staining (low intensity, low and null or altered positive staining classified as shown) were sampled for targeted sequencing (n=31 samples from 3 mice). White frames on the left images indicate magnified areas shown in the right hand images. Scale bars, 500µm. **c.** Summary table. *Notch2* mutations in samples from altered staining areas. Sample ID, sample identification, samples in bold carry multiple *Notch2* mutations, None, no mutation detected. **d.** Proportion of samples from areas of altered staining with no detected *Notch2* mutations or intronic mutations, one mutation with VAF<0.5, two mutations with VAF<0.5 or one mutation with VAF>0.5. Number of biopsies is shown in brackets. **e.** Proportion of *Notch2* missense, nonsense, indel/splicing or intronic/ undetected mutations in samples. Number of mutations is shown in brackets.

Overall, our study found the normal esophageal epithelium of the UK becomes saturated with mutants above age 70, the most prevalent of which is mutant *NOTCH1*. Analysis of clone genomes found only rare (<2%) clones with one or more hallmarks of the ESCC genome such as biallelic *TP53* loss, multiple driver mutations and widespread copy number alterations[5] . The APOBEC mutational signature, prominent in over 90% of ESCC, was not detected[17]. The absence of multiple ESCC genome changes in normal esophagus, together with the dominance of the anti-oncogenic mutant *NOTCH1* seems consistent with the comparatively low incidence of ESCC in the UK.

What is the significance of increased *NOTCH2* selection in *NOTCH1* mutant epithelium? The proportion of epithelium carrying nonsynonymous *NOTCH2* mutants in human esophagus reaches 5.7% in 70–79-year-olds (**Fig. 6a**). However, *NOTCH2* mutations are found in only 2.8% of ESCC (38/1345, https://cancer.sanger.ac.uk/cosmic). The similar prevalence of mutants in normal tissue and cancer argues against *NOTCH2* mutants having a major role in promoting transformation and is consistent with the proposition that mutant *NOTCH1* protects against transformation.

The observation of decreased selection of mutant *TP53* and lower VAF of mutant *TP53* clones in *NOTCH1*High samples above the age of 70 is intriguing. By age 70 the tissue is saturated with mutants (**Fig. 2a, b**), so in *NOTCH1*High samples *TP53* mutant clones are likely to compete for space with *NOTCH1* mutants. The competitive fitness of *TP53* mutants is less than mutant *NOTCH1* as evidenced by their ability to colonize normal esophagus (**Fig. 2 c,d**). Furthermore, in transgenic mouse esophagus, it appears that mutagen induced *Notch1* mutants may restrict the expansion of *p53* mutant clones (**Fig 4d**)[26]. Such spatial constraint of mutant *TP53* by the fitter mutant *NOTCH1* seems a plausible mechanism for our observations.

Mutant selection in the esophagus is a dynamic process in which the competitive fitness required for clone survival increases with age. Colonization by mutant *NOTCH1* acts to alter selection in two ways, first by promoting the selection of a seemingly non-oncogenic mutant and secondly by constraining expansion of an oncogenic mutant (**Fig. 6b**). More generally, future mutant selection may be altered by the current genotype in epithelia, shaping clonal evolution and the likelihood of transformation.

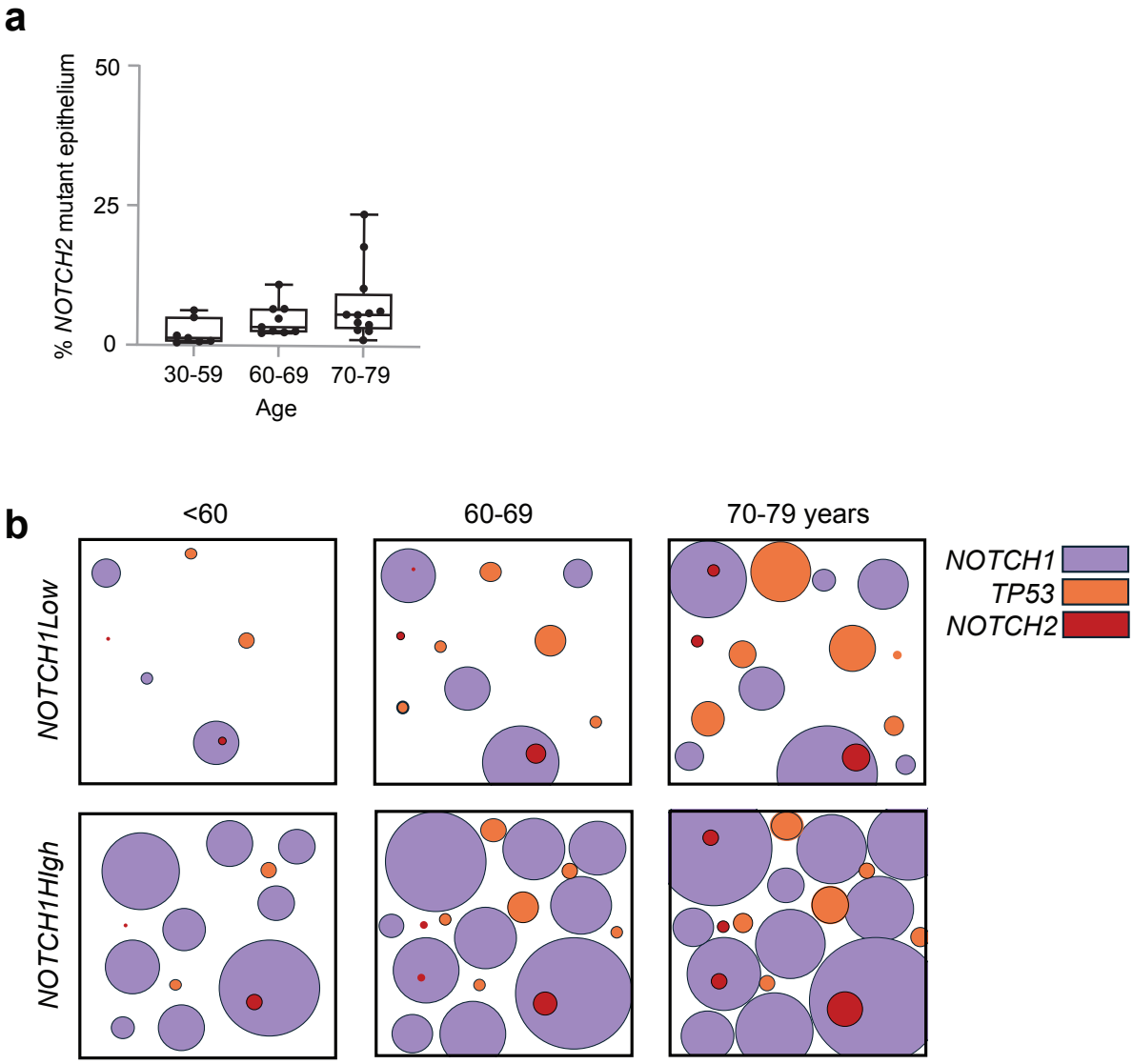


Figure 6: *NOTCH1* mutation alters selection and constrains mutant *TP53*
a. Proportion of *NOTCH2* mutant epithelium with age. Each dot is a donor, boxes indicate median and quartiles, bars range. **b.** Summary. New mutations accumulate at a constant rate. The area of mutant epithelium is determined by fitness of mutant genes under positive selection. Mutant *NOTCH1*Low samples accumulate new mutations at the same rate as mutant *NOTCH1*High samples. The area of mutant epithelium is predominantly determined by competitive fitness of expanding mutant clones. In *NOTCH1*Low samples, *TP53* mutant expansion is unconstrained, whereas by age 70, the fitter *NOTCH1* mutant clones constrict *TP53* mutant clone growth. *NOTCH2* mutants are selected preferentially within *NOTCH1* mutant epithelium.

Acknowledgements

We thank Esther Choolun and Tom Metcalf for their assistance in *in vivo* experiments, and Wellcome Sanger Institute RSF facilities for technical support.

This work was supported by grants from the Wellcome Trust to the Wellcome Sanger Institute (098051, 296194, and 108413/A/15/D) and Cancer Research UK Programme Grants to P.H.J. (C609/A17257 and C609/A27326). BH is supported by a Royal Society URF grant (UF130039).

Author Contributions

J.C.F., C.K. and E.A. performed the experiments. C.K., R.S., S.H.O., F. W-L, and J.C.F. processed and analyzed sequencing data. I.A. and M.W.J.H performed statistical analysis supervised by B.H.. P.H.J. supervised the project and wrote the manuscript with input from C.K., E.A. R.S., and J.C.F.

Competing Interests Statement

The authors declare no competing interests.

Methods

Ethical Approval for animal experiments

Approval was given by the local ethical review committees at the Wellcome Sanger Institute and conducted according to Home Office project licenses PPL70/7543, P14FED054 and PF4639B40.

Ethical approval for human study

The study protocol was ethically reviewed and approved by the UK National Research Ethics Service Committee East of England - Cambridge South, Research Ethics Committee protocol reference 15/EE/0152 NRES. Esophageal tissue was obtained from deceased organ donors. Written Informed consent was obtained from the donor's next of kin.

Human samples

A sample of mid-esophagus was removed within 60 minutes of circulatory arrest from 22 deceased organ donors at Addenbrooke's Hospital (Cambridge, UK) and placed in University of Wisconsin (UW) organ preservation solution (Belzer UW® Cold Storage Solution, Bridge to Life, USA). Human esophagus epithelium was isolated and cut into a gridded array of 2 mm^2 samples as described [7]. DNA was extracted using QIAMP DNA microkit (Qiagen, 56304).

Germline samples were collected from flash frozen oesophageal muscle and DNA was extracted as for the epithelial samples.

DNA from a 2 mm^2 sample was analyzed by nanorate sequencing (NanoSeq)[15]. As well as the 22 donors collected in this study, DNA from 9 donors reported in a previous study was also analyzed by NanoSeq.

NanoSeq

NanoSeq is a method of single DNA molecule sequencing with an error rate less than five per billion base pairs, therefore providing an accurate estimation of genome-wide mutation burden[15]. Calling mutations from a single molecule of DNA also means that mutation burden estimates are not affected by the clonality of the sample. Nonetheless, to maximise consistency in sampling across donors, 2 mm^2 samples were selected for NanoSeq if mutated for *NOTCH1* (as determined by targeted sequencing, see below), since this is the most common somatic cell genotype in aged oesophageal epithelium. Substitution counts were corrected for genome-wide trinucleotide contexts and genome-wide burden estimated by dividing by the sequence footprint (**Supplementary Table 3**).

50ng of DNA from a 2 mm^2 grid sample was used for library preparation according to[15]. Briefly, DNA was digested with mung bean nuclease, A-tailed, repaired and tagged. 0.3fmol of indexed tagged library were sequenced with 14 PCR cycles before quantifying and sequencing on a NovaSeq6000 (Illumina) with 150bp PE reads to give a median 0.9x duplex coverage. 30x coverage whole genome sequence of oesophageal muscle from the same patient was used as the germline control for calling of SNPs and indels. Only samples with verifyBAMID alpha freemix <0.005 were used in downstream analysis. For both SNV and indels, only calls which passed all the defined filters were used https://github.com/cancerit/NanoSeq.

Donor Genotyping

Donor genotypes were ascertained by whole-genome sequencing of donor muscle. A list of 34 ESCC risk alleles were obtained from the National Human Genome Research Institute-European Bioinformatics Institute GWAS (https://www.ebi.ac.uk/gwas/) and Open Targets (https://genetics.opentargets.org) databases. Associated genes are suggested for each SNP using PheWAS and eQTL data from the Open Targets Genetics database. Risk allele frequencies for Great Britain (GBR), East Asia (EAS) and Africa (AFR) are quoted according to the 1000 Genomes Project Phase 3 to show donors in this study are representative of the UK population in terms of ESCC risk (**Supplementary Table 2**).

2 mm^2 sample sequencing

For targeted sequencing of 2 mm^2 grid samples 200ng of genomic DNA was fragmented (average size distribution ~ 150bp, LE220, Covaris Inc), purified, libraries prepared (NEBNext

Ultra II DNA Library prep Kit, New England Biolabs), and index tags applied (Sanger 168 tag set). Index tagged samples were amplified (6 cycles of PCR, KAPA HiFi kit, KAPA Biosystems), quantified (Accuclear dsDNA Quantitation Solution, Biotium), then pooled together in an equimolar fashion. 9 500ng of pooled material was taken forward for hybridization, capture and enrichment (SureSelect Target enrichment system, Agilent technologies), normalised to ~6nM, and sequenced with 75bp PE reads on HiSeq2500. Samples were sequenced with one of two bait panels. The 74 gene bait panel covered the following genes (*ADAM29, ADAMTS18, AJUBA, AKT1, AKT2, APOB, ARID1A, ARID2, AURKA, BAI3, BRAF, CASP8, CCND1, CDH1, CDKN2A, CR2, CREBBP, CUL3, DICER1, EGFR, EPHA2, ERBB2, ERBB3, ERBB4, EZH2, FAT1, FAT4, FBXW7, FGFR1, FGFR2, FGFR3, FLG2, GRIN2A, GRM3, HRAS, IRF6, KCNH5, KEAP1, KMT2A, KMT2C, KMT2D, KRAS, MET, MUC17, NF1, NFE2L2, NOTCH1, NOTCH2, NOTCH3, NOTCH4, NRAS, NSD1, PCED1B, PIK3CA, PLCB1, PPP1R3A, PREX2, PTCH1, PTEN, PTPRT, RB1, RBM10, SALL1, SCN11A, SCN1A, SETD2, SMAD4, SMO, SOX2, SPHKAP, SUFU, TP53, TP63* and *TRIOBP*.) while the 32 gene bait panel covered a subset of these (*ADAM10, AJUBA, ARID1A, ARID2, CCND1, CUL3, DICER1, ERBB4, FAT1, FAT4, FBXW7, HRAS, KMT2C, KMT2D, KRAS, NF1, NFE2L2, NOTCH1, NOTCH2, NOTCH3, NRAS, NSD1, PIK3CA, PREX2, PTCH1, PTEN, RB1, SMAD4, SMO, TP53, TP63*, ZNF750)

BAM files were mapped to the human reference genome (GRCh37d5) using BWA-mem (version 0.7.17)[27]. Duplicate reads were marked using SAMtools (version 1.11)[28]. Depth of coverage (ca. 900x) was calculated using SAMtools to exclude reads which were: unmapped, not in the primary alignment, failing platform/vendor quality checks or were PCR/Optical duplicates. BEDTools (version 2.23.0) coverage program was then used to calculate the depth of coverage per base across samples[29,30]. BAM files of samples for nine published donors were then incorporated for all downstream analysis (**Supplementary Table 1**)[7].

For targeted sequencing data, sub-clonal mutation variant calling was made using the deepSNV R package (also commonly referred to as ShearwaterML), version 1.26.1 commit d45c86c, available at https://github.com/gerstung-lab/deepSNV, used in conjunction with R version 3.5.2[7]. Variants were annotated using VAGrENT, https://github.com/cancerit/VAGrENT, version 3.3.3[31].

Mutations called by ShearwaterML were filtered using the following criteria:

- Germline variants called from the same individual were removed from the list of called variants.
- Adjustment for FDR and mutations demanded support from at least one read from both strands for the mutations identified.
- Pairs of SNVs on adjacent nucleotides within the same sample are merged into a dinucleotide variant if at least 90% of the mapped DNA reads containing at least one of the SNV pair, contained both SNVs.

- Identical mutations found in multiple contiguous tissue biopsies are merged and considered as a single clone to prevent duplicate clone counting.

ShearwaterML was run with a normal panel coverage of approximately 42,000x for the 74-gene panel and around 12,000x for the 32-gene panel.

0.05 mm^2 sample sequencing

0.05 mm^2 samples were collected with circular biopsy punch (0.25 mm diameter, Stoelting Europe), in a rectangular array, with approximately a punch diameter between each sample. DNA was extracted using the Arcturus picopure kit (ThermoFisher).

Shearing and library prep was performed using the NEBNext®Ultra™ II Fragmentase System. DNA was hybridised using a bait panel described previously[32] which contains many cancer driver genes. Material was subjected to 12 PCR cycles (Initial denaturation: 95C 5mins, 98C 30sec, 65C 30sec, 72C 2mins 12 cycles, 72C 10mins). Variants were called using the CaVEMan (version 1.15.1), http://cancerit.github.io/CaVEMan/, and cgpPindel (version 3.7.0), http://cancerit.github.io/cgpPindel, algorithms[33,34]. For SNVs CaVEMan was run with the major copy number set to 10 and the minor copy number set to 2. Only SNVs which passed all CaVEMan filters were kept.

Whole Genome Sequencing

2 mm^2 and 0.05 mm^2 samples were selected for WGS if targeted sequencing determined that at least half the tissue was dominated by a single clone. Sequences were aligned to the human reference genome (GRCh37d5) using BWA-MEM[27]. Single-base substitutions were called against a donor's matched germline sample using CaVEMan[33]. Mapping artefacts were removed, the median alignment score of reads supporting a variant had to be at least 140 and the number of clipped reads equal to zero. In addition, the proportion of mutant reads present in the matched sample also had to be zero. Small (<200 bp) insertions and deletions were called using cgpPindel. Only indels which passed all cgpPindel filters were kept. For 0.05 mm^2 samples, a large excess of single base pair insertions at homopolymers of length five or more were removed, an artefact caused by PCR amplification of low-input DNA concentrations. Telomere lengths were estimated using the Telomerecat software package (version 4.0.2, https://github.com/cancerit/telomerecat/)[35]. Telomere length is given as a median of all chromosomes for all cells in that sample.

Phylogeny Construction

A pile-up of quality reads was constructed, and the number of mutant and wildtype reads counted for every locus that had a mutation called in any sample of a donor. Reads with a mapping quality of at least 30 and bases with a base quality of at least 25 were used. Mutations with a VAF < 0.2 were assigned a genotype of 0 and those with a VAF >= 0.3 a genotype of 1. All other mutations were set a genotype of NA (not applicable) for the purposes of phylogeny

construction. Maximum parsimony trees were drawn for each donor after 1000 iterations with MPBoot[36]. Substitutions in mutant genes under selection were assigned to each branch of a phylogeny. Copy number aberrations were manually assigned to each branch where present.

Mutational signature analysis

Mutational spectra and signatures are described using the notation employed by the PCAWG Mutational Signatures working group[37]. COSMIC signature definitions (v3.2) (https://cancer.sanger.ac.uk/signatures/sbs/) were used for Single Base Substitutions (SBS), Double base Substitutions (DBS) and Indel signature classification. Using SigProfiler packages MatrixGenerator (v1.2.12), Extractor (v1.1.12), Assignment (v0.0.13), Plotting (1.2.2). The frequency of mutations within each trinucleotide context was calculated using SigProfiler within the SBS288[37].

The maximum-likelihood implementation dNdScv (v0.0.1.0, https://github.com/im3sanger/dndscv) was used to identify genes under positive selection. dNdScv estimates the ratio of non-synonymous to synonymous mutations across genes, adjusting for sequence composition and mutational spectrum. dN/dS values significantly greater than 1 indicate positive selection.

The proportion of mutant epithelium was calculated from mutant VAFs as described[7,38].

Copy number calling

Copy number calling on WGS was inferred with ASCATNGS (https://github.com/cancerit/ascatNgs), using oesophageal muscle from the same patient as a germline control[34] . For inferring copy number from the 0.05 mm^2 punch samples sequenced with a targeted bait panel ASCAT (v3.1.0) was used. Regions were restricted to the *NOTCH1* and *TP53* loci, only SNPS within these regions were identified in the normal samples. Allele counts were generated for SNP positions from these normal samples allowing calling for both 0.05 mm^2 punch and germline samples. These were used to derive LogR values and BAF values used to generate segment calls of allele-specific copy number. https://github.com/VanLoo-lab/ascat[39].

Mouse experiments

Animals and strains

Animals were 10 to 16 weeks old at the start of the experiments. Mice were housed in Specific and Opportunistic Pathogen Free unit with a 12-hour light/12-hour dark cycle. Lights were switched off at 19:30 and there was no twilight period. Ambient temperature was 21 ± 2 °C, and humidity 55 ± 10%. Mice were kept at 3–5 mice per cage, in cages measuring 365 mm × 207 mm × 140 mm (length × width × height), with a floor area, of 530 cm^2. Individually ventilated caging (Tecniplast, Sealsafe 1284L) with 60 air changes/ hour was used. As well as Aspen bedding substrate, environmental enrichment with Nestlets, a cardboard

tunnel and chew blocks was provided. Water and standard chow diet were provided ad libitum. Experiments were carried out with male and female animals and no sex specific differences were observed. Animals were randomly assigned to experimental groups.

YFPCreNotch1 triple mutant mice[6] were generated by crossing B6.129X1-*Notch1*$^{tm2Rko/GridJ}$ mice purchased from the Jackson Laboratory with *Rosa26*^{floxedYFP} and *AhCre*^{ERT} mice [23,40,41]. In these animals, *Cre* mediated recombination is induced by treatment with β-Naphthoflavone and Tamoxifen. An enhanced yellow fluorescent (EYFP) reporter is expressed from the Rosa 26 locus after *Cre* mediated excision of a *LoxP* flanked 'stop' cassette. Exon1 of *Notch1* is flanked by *LoxP* sequences so that *Cre* mediated recombination blocks *Notch1* expression.

Aging experiments

YFPCreNotch1$^{flox/flox}$ mice aged between 10-16 weeks-old were injected intraperitoneally (i.p.) with β-Naphthoflavone (MP Biomedicals, catalog no.156738) at 80 mg.kg^{-1} and tamoxifen (Sigma Aldrich, catalog no. N3633) at 0.25mg and collected at up to 78 weeks old. *Notch1*$^{+/+}$ or non-induced *YFPCreNotch1*$^{flox/flox}$ mice were used as wild type controls.

Whole-mount preparation and immunostaining of mouse esophagus

Mouse esophagus was opened longitudinally before removing the underlying muscle layer with forceps. Esophageal tissue, comprising mucosa and submucosa, was incubated for 2h15 to 3 hours in 5 mM Ethylenediaminetetraacetic acid (EDTA) at 37 °C before peeling away the epithelium. The flattened epithelium was fixed in 4% paraformaldehyde for 75 minutes at room temperature under agitation then washed and stored in PBS at 4 °C [9]. All subsequent steps were performed at room temperature. Tissues were incubated in staining buffer for a one-hour blockage (10% donkey serum, 0.5% bovine serum albumin, 0.5% Triton X-100, 0.25% fish skin gelatin in PHEM). This step was followed by overnight incubation with primary antibodies in staining buffer, three 30 min washes in 0.2% Tween-20 PHEM, and a 3 hour-long incubation with secondary antibodies in staining buffer (**Supplementary Table 23**). After washes in PHEM, tissues were incubated for 1h in 0.5 µM Sytox™ Blue nuclei counterstain (Biolegend, catalog no. 425305) and mounted on slides using Vectashield media (Vector Laboratories, catalog no. H-1000).

Sequencing of aged esophageal epithelial wholemounts

Whole mounts from aged, induced *YFPCreNotch1*$^{flox/flox}$ or control mice were flattened and dissected into a grid of 2 mm^2 contiguous biopsies. DNA extraction was performed using the QIAMP DNA microkit (Qiagen, catalog no. 56304) following the manufacturer's instructions. Each sample was then subject to targeted sequencing.

Identification and sequencing *Notch2* mutant clones by immunostaining

We used a similar approach to that described for identifying *Notch1* mutant clones in a previous study[6]. Whole mounted epithelium was stained for NOTCH2, EYFP (using an anti-GFP antibody) and Sytox™ Blue and imaged by confocal microscopy. Large, ovoid areas with either

low intensity or null NOTCH2 staining suggested the presence of *Notch2* mutant clones (**Extended data Fig. 3a**). Similarly, large and ovoid areas with non-nuclear localization of the NOTCH2 protein also suggested the presence of *Notch2* mutations that would alter the cleavage or nuclear translocation of NICD2, as the EDTA incubation at 37ºC used to prepare the wholemounts activates NOTCH2 cleavage[42]. Finally, YFP staining helped identifying putative mutant clones, appearing as large and ovoid areas devoid or fully stained with YFP (**Extended data Fig. 3b**).

Putative clones with altered NOTCH2 staining were sampled under a Fluorescent Stereo Microscope Leica M165 FC (Leica) using 0.25 mm diameter punch (Stoelting, catalog no. 57391). DNA extraction was performed using Arcturus® PicoPure® DNA extraction kit (Applied Biosystems, catalog no. 11815-00) following the manufacturer's instructions. Targeted sequencing was then performed on each sample.

Targeted Sequencing of mouse samples

After multiplexing, samples were sequenced on Illumina HiSeq 2000 sequencer using paired-end 75-base pair (bp) reads. An Agilent SureSelect custom bait capture kit comprising 73 cancer related and Notch pathway related genes (baitset_3181471_Covered.bed, **Supplementary Table 16**) was used[6].

For 2 mm^2 samples the average depth of coverage across all genes was: *Notch1*$^{+/+}$: 365x; *Notch1*$^{-/-}$: 339x.
For 0.05 mm^2 samples the average depth of coverage across all genes was 228x.

Paired-end reads were aligned to the GRCm38 reference mouse genome with BWA-MEM (v.0.7.17, https://github.com/lh3/bwa)[27] with optical and PCR duplicates marked using Biobambam2 (v.2.0.86, https://github.com/gt1/biobambam2).

Mutations in 0.05 mm^2 samples from aged *YFPCreNotch1*$^{-/-}$ mice were analysed with CaVEMan (version 1.13.14) and with cgpPindel (version 3.3.0)[34] for insertions and deletions. Sequence variants were annotated with VAGrENT (version 3.3.3)[33,43] .

Mutations in 2 mm^2 samples were also called with deepSNV (v1.26.1), https://github.com/gerstung-lab/deepSNV). The total coverage provided by the references used for the mouse bait kit was 22885x in a total of 34 germline samples.

In a gridded array of samples, some mutations can span between several adjacent biopsies. To prevent counting them multiple times and obtain accurate estimates of clone sizes and number, mutations extending over two or more biopsies were merged as individual calls[14,27].

Selection was determined using dNdScv (v0.0.1.0) commit c3437e.

The proportion of cells carrying a mutation within each sample was calculated from the fraction of cells carrying a mutation within the sample, as described previously[7,38]. The lower (=VAF) and upper (=2xVAF) bound estimates of the percentage of epithelium covered by clones carrying non-synonymous mutations in each gene was then calculated. The proportion of mutant epithelium was derived from mean of summed VAF (capped at 1.0) of all samples in the esophagus.

Statistical analysis

Data in graphs are shown as mean values ± standard error of the mean (SEM) unless otherwise stated. In general, P values <0.05 were considered statistically significant, but for dN/dS ratios calcuated by dNdScv, a global q<0.01 was considered statistically significant. All significance tests were 2 tailed. Each mouse experiment was performed using multiple biological replicates. The numbers of experimental replicates are given in figure legends and Supplementary tables. Statistical tests described in figure legends were performed using GraphPad Prism software 8.3.1 and Python package SciPy 1.7.3 (https://scipy.org/citing-scipy/). No statistical method was used to predetermine sample size. Animals of the relevant genotype were allocated to experimental groups at random.

Data availability statement

The sequencing datasets generated during study are available at the European Genome-phemone Archive (EGA) with the following dataset accession numbers:
EGAD00001015256 – 2 mm^2 samples, targeted
EGAD00001015257 – 2 mm^2 samples, WGS
EGAD00001015258 - 0.25 mm^2 samples targeted
EGAD00001015259 - 0.25 mm^2 samples, WGS
EGAD00001015260 - Nanoseq
EGAD00001015261 - Nanoseq germline

Targeted sequencing of mouse tissue is deposited in the European Nucleotide Archive (ENA) under the following accession code:
ERP164438- Targeted sequencing of aged *Notch1*$^{+/+}$ mouse esophageal epithelium, targeted sequencing of aged *Notch1*$^{-/-}$ mouse esophageal epithelium, and targeted sequencing of aged *Notch1*$^{-/-}$ mouse esophageal epithelial microbiopsies.

References

1 Gerstung, M. *et al.* The evolutionary history of 2,658 cancers. *Nature* **578**, 122-128 (2020). https://doi.org:10.1038/s41586-019-1907-7

2 Ortmann, C. A. *et al.* Effect of mutation order on myeloproliferative neoplasms. *N Engl J Med* **372**, 601-612 (2015). https://doi.org:10.1056/NEJMoa1412098

3 Nguyen, L. H., Goel, A. & Chung, D. C. Pathways of Colorectal Carcinogenesis. *Gastroenterology* **158**, 291-302 (2020). https://doi.org:10.1053/j.gastro.2019.08.059

4 Chang, J. *et al.* Genomic alterations driving precancerous to cancerous lesions in esophageal cancer development. *Cancer Cell* **41**, 2038-2050.e2035 (2023). https://doi.org:https://doi.org/10.1016/j.ccell.2023.11.003

5 Consortium, T. Integrated genomic characterization of oesophageal carcinoma. *Nature* **541**, 169-175 (2017). https://doi.org:10.1038/nature20805
6 Abby, E. *et al.* Notch1 mutations drive clonal expansion in normal esophageal epithelium but impair tumor growth. *Nat Genet* (2023). https://doi.org:10.1038/s41588-022-01280-z
7 Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science (New York, N.Y.)* **362**, 911-917 (2018). https://doi.org:10.1126/science.aau3879
8 Yokoyama, A. *et al.* Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312-317 (2019). https://doi.org:10.1038/s41586-018-0811-x
9 Colom, B. *et al.* Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. *Nature Genetics* **52**, 604–614 (2020).
10 Abnet, C. C. *et al.* A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat Genet* **42**, 764-767 (2010). https://doi.org:10.1038/ng.649
11 Wu, C. *et al.* Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. *Nat Genet* **43**, 679-684 (2011). https://doi.org:10.1038/ng.849
12 Wu, C. *et al.* Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat Genet* **44**, 1090-1097 (2012). https://doi.org:10.1038/ng.2411
13 Wu, C. *et al.* Joint analysis of three genome-wide association studies of esophageal squamous cell carcinoma in Chinese populations. *Nat Genet* **46**, 1001-1006 (2014). https://doi.org:10.1038/ng.3064
14 Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415-1424 (2021). https://doi.org:10.1038/s41588-021-00931-x
15 Abascal, F. *et al.* Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405-410 (2021). https://doi.org:10.1038/s41586-021-03477-4
16 Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020). https://doi.org:10.1038/s41586-020-1943-3
17 Moody, S. *et al.* Mutational signatures in esophageal squamous cell carcinoma from eight countries with varying incidence. *Nat Genet* **53**, 1553-1563 (2021). https://doi.org:10.1038/s41588-021-00928-6
18 Fowler, J. C. *et al.* Selection of Oncogenic Mutant Clones in Normal Human Skin Varies with Body Site. *Cancer Discov* **11**, 340-361 (2021). https://doi.org:10.1158/2159-8290.Cd-20-1092
19 Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029-1041.e1021 (2017). https://doi.org:10.1016/j.cell.2017.09.042
20 Lin, S., Gordon, K., Kaplan, N. & Getsios, S. Ligand targeting of EphA2 enhances keratinocyte adhesion and differentiation via desmoglein 1. *Mol Biol Cell* **21**, 3902-3914 (2010). https://doi.org:10.1091/mbc.E10-03-0242
21 Olafsson, S. *et al.* Effects of psoriasis and psoralen exposure on the somatic mutation landscape of the skin. *Nature Genetics* **55**, 1892-1900 (2023). https://doi.org:10.1038/s41588-023-01545-1
22 Wilson, R. B. *et al.* Elongation factor 1A1 regulates metabolic substrate preference in mammalian cells. *J Biol Chem*, 105684 (2024). https://doi.org:10.1016/j.jbc.2024.105684

23 Doupe, D. P. *et al.* A single progenitor population switches behavior to maintain and repair esophageal epithelium. *Science (New York, N.Y.)* **337**, 1091-1093 (2012). https://doi.org:10.1126/science.1218835
24 Piedrafita, G. *et al.* A single-progenitor model as the unifying paradigm of epidermal and esophageal epithelial maintenance in mice. *Nature Communications* **11**, 1429 (2020). https://doi.org:10.1038/s41467-020-15258-0
25 Colom, B. *et al.* Mutant clones in normal epithelium outcompete and eliminate emerging tumours. *Nature* **598**, 510-514 (2021). https://doi.org:10.1038/s41586-021-03965-7
26 Murai, K. *et al.* p53 mutation in normal esophagus promotes multiple stages of carcinogenesis but is constrained by clonal competition. *Nature Communications* **13**, 6206 (2022). https://doi.org:10.1038/s41467-022-33945-y
27 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv*, 1303.3997 (2013).
28 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10** (2021). https://doi.org:10.1093/gigascience/giab008
29 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010). https://doi.org:10.1093/bioinformatics/btq033
30 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078-2079 (2009). https://doi.org:10.1093/bioinformatics/btp352
31 Menzies, A. *et al.* VAGrENT: Variation Annotation Generator. *Current Protocols in Bioinformatics* **52**, 15.18.11-15.18.11 (2015). https://doi.org:10.1002/0471250953.bi1508s52
32 Lee-Six, H. *et al.* The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532-537 (2019). https://doi.org:10.1038/s41586-019-1672-7
33 Jones, D. *et al.* cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Current Protocols in Bioinformatics* **56**, 15.10.11-15.10.18 (2016). https://doi.org:10.1002/cpbi.20
34 Raine, K. M. *et al.* cgpPindel: Identifying Somatically Acquired Insertion and Deletion Events from Paired End Sequencing. *Current protocols in bioinformatics* **52**, 15.17.11-15.17.12 (2015).
35 Farmery, J. H. R., Smith, M. L. & Lynch, A. G. Telomerecat: A ploidy-agnostic method for estimating telomere length from whole genome sequencing data. *Sci Rep* **8**, 1300 (2018). https://doi.org:10.1038/s41598-017-14403-y
36 Hoang, D. T. *et al.* MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. *BMC Evolutionary Biology* **18**, 11 (2018). https://doi.org:10.1186/s12862-018-1131-3
37 Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020). https://doi.org:10.1038/s41586-020-1943-3
38 Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (New York, N.Y.)* **348**, 880-886 (2015). https://doi.org:10.1126/science.aaa6806
39 Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences* **107**, 16910-16915 (2010). https://doi.org:doi:10.1073/pnas.1009843107

40 Kemp, R. *et al.* Elimination of background recombination: somatic induction of Cre by combined transcriptional regulation and hormone binding affinity. *Nucleic Acids Res* **32**, e92 (2004). https://doi.org:10.1093/nar/gnh090

41 Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10**, 547-558 (1999).

42 Rand, M. D. *et al.* Calcium depletion dissociates and activates heterodimeric notch receptors. *Molecular and cellular biology* **20**, 1825-1835 (2000). https://doi.org:10.1128/mcb.20.5.1825-1835.2000

43 Menzies, A. *et al.* VAGrENT: Variation Annotation Generator. *Curr Protoc Bioinformatics* **52**, 15.18.11-15.18.11 (2015). https://doi.org:10.1002/0471250953.bi1508s52