Kakiuchi, N., Yoshida, K., Uchino, M. *et al*. Frequent mutations that converge on the NFKBIZ pathway in ulcerative colitis. *Nature* **577**, 260–265 (2020). https://doi.org/10.1038/s41586-019-1856-1

Frequent mutations converging into NFKBIZ signalling in ulcerative colitis

Nobuyuki Kakiuchi[1,2,3], Kenichi Yoshida[1], Motoi Uchino[4], Takako Kihara[5], Kotaro Akaki[6], Yoshikage Inoue[1,2,7], Kenji Kawada[7], Satoshi Nagayama[8], Akira Yokoyama[1,9], Shuji Yamamoto[3], Minoru Matsuura[3], Takahiro Horimatsu[9], Tomonori Hirano[1,2,3], Norihiro Goto[3], Yasuhide Takeuchi[1,2,10,11], Yotaro Ochi[1,2], Yusuke Shiozawa[1], Yasunori Kogure[1,12], Yosaku Watatani[1,2], Yoichi Fujii[1,2], Soo Ki Kim[1,3], Ayana Kon[1,2], Keisuke Kataoka[1,12], Tetsuichi Yoshizato[1], Masahiro M. Nakagawa[1], Akinori Yoda[1], Yasuhito Nanya[1,2], Hideki Makishima[1,2], Yuichi Shiraishi[13], Kenichi Chiba[13], Hiroko Tanaka[14], Masashi Sanada[1,15], Eiji Sugihara[16], Taka-aki Sato[16], Takashi Maruyama[17], Hiroyuki Miyoshi[18], Makoto Mark Taketo[18], Jun Oishi[19], Ryosaku Inagaki[19], Yutaka Ueda[19], Shinya Okamoto[20], Hideaki Okajima[20], Yoshiharu Sakai[7], Takaki Sakurai[10], Hironori Haga[11], Seiichi Hirota[5], Hiroki Ikeuchi[4], Hiroshi Nakase[3,21], Hiroyuki Marusawa[3], Tsutomu Chiba[3,22], Osamu Takeuchi[6], Satoru Miyano[13,14], Hiroshi Seno[3], Seishi Ogawa[*1,2,23]

[1]Department of Pathology and Tumour Biology, Kyoto University, Kyoto, Japan
[2]Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto University, Kyoto, Japan
[3]Department of Gastroenterology and Hepatology, Kyoto University, Kyoto, Japan
[4]Department of Inflammatory Bowel Disease, Division of Surgery, Hyogo College of Medicine, Nishinomiya, Japan
[5]Department of Surgical Pathology, Hyogo College of Medicine, Nishinomiya, Japan
[6]Department of Medical Chemistry, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[7]Department of Surgery, Kyoto University, Kyoto, Japan
[8]Gastroenterological Center, Department of Gastroenterological Surgery, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan
[9]Department of Clinical oncology, Kyoto University, Kyoto, Japan
[10]Department of Diagnostic Pathology, Kyoto University Hospital, Kyoto, Japan
[11]Department of Diagnostic Pathology, Kyoto University, Kyoto, Japan
[12]Division of Molecular Oncology, National Cancer Centre Research Institute, Tokyo, Japan
[13]Laboratory of DNA information Analysis, Human Genome Centre, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
[14]Laboratory of Sequence Analysis, Human Genome Centre, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
[15]Department of Advanced Diagnosis, Clinical Research Centre, Nagoya Medical Centre, Nagoya, Japan
[16]Research and Development Centre for Precision Medicine, University of Tsukuba, Tsukuba, Ibaraki, Japan

[17]Department of Immunology, Akita University Graduate School of Medicine, Akita, Japan
[18]Division of Experimental Therapeutics, Kyoto University, Kyoto, Japan
[19]DSP Cancer Institute, Sumitomo Dainippon Pharma Co., Ltd., Osaka, Japan.
[20]Division of Hepato-Biliary-Pancreatic Surgery and Transplantation, Department of Surgery, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[21]Department of Gastroenterology and Hepatology, Sapporo Medical University School of Medicine, Sapporo, Japan
[22]Kansai Electric Power Hospital, Osaka, Japan
[23]Department of Medicine, Centre for Haematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden

[*]Correspondence should be addressed to S.Ogawa. (sogawa-tky@umin.ac.jp).

SUMMARY PARAGRAPH:

Chronic inflammation is accompanied by recurring cycles of tissue destruction and repair and explains a major cancer risk[1-3]. However, it is poorly understood how such cycles affect clonal composition of tissues, particularly in terms of cancer development. Here we show that in patients with ulcerative colitis (UC), the inflamed intestine undergoes widespread remodelling by pervasive clones positively selected by acquiring mutations commonly involving *NFKBIZ, TRAF3IP2, ZC3H12A*, *PIGR,* and *HNRNPF*, many of which are implicated in downregulation of IL-17 and other proinflammatory signalling. Substantially different mutation profiles between UC-epithelia and cancer indicate distinct mechanisms of positive selection between both. Particularly, *NFKBIZ* mutations highly prevalent in UC epithelia were rarely found in both sporadic and colitis-associated cancer, suggesting negative selection of *NFKBIZ*-mutated cells during colorectal carcinogenesis, which was further supported by significantly attenuated colitis-induced tumour formation in *Nfkbiz*-deficient mice and compromised cell-competition of *NFKBIZ*-disrupted colorectal cancer cells. Our results highlight common and discrete mechanisms of clonal selection in inflammatory tissues, which unexpectedly reveal cancer vulnerability potentially utilized for therapeutics of colorectal cancer.

MAIN TEXT:

Chronic inflammation is a major cause of morbidity and mortality in the human population and explains a substantial cancer risk therein[1,2]. A common feature of chronic inflammation is recurring tissue injury and repair fuelled by long-lasting, unregulated immunity, often terminating in tissue remodelling and debilitating organ dysfunction[3]. Meanwhile, it has recently been demonstrated that extensive tissue remodelling can take place even in apparently normal tissues in an age-dependent manner and is mediated by expansion of numerous clones carrying common cancer-related mutations[4-6]. Given the inflammation-associated cancer risks, clonal expansion of somatically mutated clones may also play a role in inflammation-associated tissue remodelling, explaining cancer risk. However, it is largely unknown whether this actually happens, how frequently and extensively and with what mutations it occurs if ever, and how it modifies the inflammatory disease process and cancer development.

Ulcerative colitis (UC) is a common form of inflammatory bowel disease (IBD), affecting ~2,100,000 people in Europe alone[7]. It is characterized by persistent inflammation affecting the large intestine, causing a severe destruction of intestinal mucosa and intractable ulcer formation. Although the pathophysiology of UC is not fully understood, deregulated immune cells and cytokines, as well as intestinal dysbiosis, have been implicated in the persistent inflammation[8], which is associated with an increased risk (15-20%) of cancer development (colitis-associated colorectal cancer; CAC)[9]. In this study, to understand how chronic inflammation shapes tissue remodelling and how it correlates with cancer development, we investigated clonal expansion in UC and non-UC epithelia, using intensive multi-regional sampling followed by unbiased detection of somatic mutations.

Clonal history of normal and UC crypts

Colorectal epithelium is composed of single-layer columnar cells that are compacted into numerous small replication units called ‘crypts’. Each crypt comprises ~2,000 cells, which are replenished by the stem cells located at its bottom[10]. During life, physiological or pathological deficits of crypts can occur, which should be filled by crypt fission[11]. Thus, to understand the dynamics of UC and non-UC epithelia, we first evaluated somatic mutations in single crypts isolated from normal and UC specimens, using whole exome sequencing (WES) of whole genome-amplified (WGA) single crypt DNA **(EDFigs. 1, 2 and 3)**. Mutations in non-UC crypts increases with age at an annual mutation rate of 0.61 mutations/exome/year, which is significantly increased in UC-derived crypts **(Fig. 1a)**. When corrected for age at diagnosis and disease duration **(Methods)**, UC-derived crypts showed a >3 times higher mutation rate than non-UC crypts (1.9 mutations/exome/year). Whereas the mutation rate did not differ between right- and left-side colons in non-UC individuals, we observed increased mutation rates for the crypts from the left-side colon in UC patients, in which inflammation tends to be more severe than in the right-side colon **(Fig. 1b)**[8]. Although no significantly mutated genes among crypts from normal epithelia, three genes were significantly affected by nonsynonymous mutations (dN/dS>1.0, q<0.05) in UC epithelia, including *NFKBIZ, PIGR,* and *ARID1A* **(EDFig. 3c,d)**, suggesting their driver roles.

Next, we analysed history of crypt expansion in normal and inflamed epithelia, in which a cluster of crypts were isolated from a small area of colon epithelium (0.25-4 mm^2) and analysed by WES, keeping their geographical information **(EDFig. 2b)**. We analysed crypt clusters from 3 each of UC and non-UC subjects aged 38-56 and 78-88 years, respectively. As seen in solitary crypts, UC-derived crypt clusters had substantially higher mutation rates (2.86-15.5 mutations/exome/year), compared with non-UC crypt clusters (0.63-0.75 mutations/exome/year). Most of the crypts (99.4%) in both UC and non-UC subjects shared one or more mutations with other crypts, based on which phylogenetic trees were constructed and superimposed on their geographical maps **(Fig. 1c and EDFig. 4a-d)**. In non-UC epithelia, the majority (99/143, 69.2%) of the branchpoints in the phylogenetic trees were estimated to occur before 20 years of age (corresponding to 12.2 mutations) **(EDFig. 4e)**. Thus, it is suggested that the number of crypts steadily increases until 20 years old with an annual fission rate of ~0.136 to expand the intestinal size, while in the steady state throughout the remaining life, crypt fissions occur only occasionally to balance the physiological deficits of crypts (0.00413 fissions/year) **(Methods)**. By contrast, the crypts from 3 UC patients had very different phylogenetic structures **(Fig. 1c and EDFig. 4c,d)**. They were characterized by a prominently long trunk having known driver mutations, which terminated in many short branches. Thus, it is suggested that after repetitive destruction and regeneration, these crypts rapidly expanded from a positively selected common ancestor crypt very recently, during the past 1.7-4.7 years (3.3 on average).

Positive selection in UC epithelia

To further investigate clonal expansion in UC epithelia, we performed WES of bulk crypts isolated from 269 non-dysplasia specimens collected from 22 UC and 23 non-UC individuals with a median of 6 (4-8) specimens per individual **(EDFig. 1b)**. We also analysed an additional 294 bulk crypt samples, which were collected from surgically obtained rectums from 6 UC patients at regular intervals **(EDFig. 1b)**. Compared with non-UC samples, UC-derived samples had much higher numbers of mutations (51.2 vs. 1.9 mutations/sample on average) **(EDFig. 5a)**. Maximum mutated cell fractions (MCFs) was significantly larger for UC samples (mean=0.58 [0-1]) than those from non-UC individuals (mean=0.21 [0-1]) (P<0.01) **(Methods)** with 22.6% of UC samples showing maximum MCF of 1.0. The number of mutations and the maximum MCF in

UC samples weakly but significantly correlated with the interval between diagnosis and sampling **(EDFig. 5b,c)**. Samples from the left-side colon tended to show larger mutation numbers and MCFs, compared with those from the right-side colon **(EDFig. 5d,e)**.

To decipher the mechanism of positive selection in UC epithelia, we evaluated dN/dS for 30,062 mutations detected in 399 UC-derived non-dysplastic samples and identified 14 driver genes with evidence of positive selection (dN/dS>1.0, q<0.05) **(Fig. 2a,b and EDFig. 6) (Methods)**, while no positively selected mutations were detected in 179 non-UC samples. Mutations in most of these drivers in UC epithelium were predicted to cause a protein truncation and loss-of-function. Exceptions were mutations in *ZC3H12A, KRAS,* and *HNRNPF*, which exhibited prominent mutational hotspots indicative of some neomorphic functions (**Fig. 2a and EDFig. 6a)**. *NFKBIZ, TRAF3IP2,* and *RNF43* frequently underwent biallelic inactivation, which was less common in other drivers **(Fig. 2b)**.

A single bulk-crypt sample may contain multiple independent clones and subclones carrying discrete driver mutations. Thus, to precisely determine the frequency of mutations in UC epithelia on the basis of independent clones, we first clustered mutations detected in bulk-crypt samples using PyClone and based on the cell fraction predicted for each mutation cluster, identified a total of 183 independent clones according to the '*Pigeonhole principle'* **(EDFig. 1d) (Methods)**. Among these, 166 (90%) harboured one or more driver mutations, which most frequently affected *NFKBIZ*, followed by *ARID1A, PIGR, KRAS,* and *ZC3H12A* **(Fig. 2b)**. Conspicuously, the molecules mutated in UC epithelia displayed a surprising overlap to signalling pathways converging into NFKBIZ (also known as IκBζ) and related molecules (ZC3H12A and PIGR) **(Fig. 2c)**, where the mutations in these molecules occurred in a mutually exclusive manner **(Fig. 2b)**. Among these mutational targets most prominent was the molecules involved in IL-17 signalling, including IL-17 receptors (IL17RA/IL17RC) and its adaptor, TRAF3IP2, which were shown to be positively selected in UC epithelia **(Methods)**, suggesting a major role of IL-17 signalling **(Fig.2c)**. Moreover, other inflammation-related cytokine receptors and pattern recognition receptors (PRRs) implicated in NFKBIZ upregulation, including OSMR, LTBR, IL20RA, and TLR9[12] **(EDFig. 6b)**, also underwent loss-of-function mutations in occasional clones **(Methods)**.

Expansion of clones carrying these cytokine/PRR pathway mutations was intensively evaluated in surgically obtained rectum specimens from 6 UC patients using densely collected samples **(Fig. 2d and EDFig. 7a)**. We identified 41 independent driver-mutated clones that involved ≥2 samples in these specimens, of which 31 had the mutations in the above pathways. Conspicuously, in these patients, 53-83% of the entire rectum was replaced by clones having one or more driver mutations, most of which were explained by cytokine/PRR pathway mutations **(Fig. 2e)**. The expansion could be extensive, occupying an epithelial area as large as 19 cm^2 in size **(EDFig. 7b)**. In HCM22, almost entire rectal epithelium, containing ~300,000 crypts, was replaced by only two clones carrying mutations in *NFKBIZ* (S339fs) and *ZC3H12A* (S438L). In the analysis of phylogenetic structure, 37 clones had one or more driver mutations in their main trunk, of which 26 had IL-17 or other cytokine/PRR pathway mutations, suggesting that these mutations were acquired before or early in UC development **(Fig. 2d, EDFig. 7)**.

Effects of *NFKBIZ* and related mutations

Among these cytokine/PRR signalling genes, most commonly mutated were *NFKBIZ* (30%), *PIGR* (21%), and *ZC3H12A* (14%). Upregulated upon a number of primary stimuli, such as IL-17A, IL-1β, and LPS, NFKBIZ is induced via activated NFκB to regulate expression of secondary response genes, including many proinflammatory cytokine genes **(Fig. 2c and EDFig. 6b)**[13]. *PIGR,* encoding a key molecule for transcytosis of IgA[14], is one of such genes[15] and known to be downregulated in the colon epithelium in IL-17R-deficient mice**[15]** and in patients with IBD[16]. Thus, to understand the functional link between *NFKBIZ* and *PIGR* with regard to the IL-17 signalling, we established *Nfkbiz*$^{-/-}$ and *Nfkbiz*$^{+/+}$ organoids from colon epithelium from *Nfkbiz*$^{fl/fl}$*VilCre* and *Nfkbiz*$^{fl/fl}$ mice, in which *Pigr* expression upon IL-17A stimulation was examined **(Fig. 3a)**. In *Nfkbiz*$^{+/+}$ organoids, *Pigr* expression was induced by IL-17A stimulation, slightly after upregulation of *Nfkbiz*, and increased until 24 hours after stimulation, whereas in *Nfkbiz*$^{-/-}$ organoids, the *Pigr* induction was significantly attenuated with no *Nfkbiz* induction. These observations confirmed that NFKBIZ is a key regulator of *Pigr* expression upon IL-17A stimulation.

ZC3H12A encodes an RNase known as Regnase-1, which is implicated in the degradation of mRNAs of *NFKBIZ* and many downstream secondary response genes, including *PIGR*[17,18]. Mutations in UC epithelia were highly enriched in the DSGxxS motif and the C-terminal part, showing a prominent mutational peak at the S438 residue **(Fig. 2a)**, which is shown to be phosphorylated by the IKK complex to promote ubiquitination and subsequent degradation of the protein[19]. In fact, a mouse mutant at the corresponding site, S435A, is reported to prevent ubiquitin-mediated degradation and lead to an elevated ZC3H12A expression upon IL-1β stimulation[19]. We also demonstrated that another amino-acid substitution (S438L) showed an elevated protein expression level, compared with wild-type transfected cells **(Fig. 3b and EDFig. 6c)**. Of interest, we found a UC sample harbouring a nonsense mutation affecting the E3 ligase implicated in the ubiquitination of ZC3H12A, *BTRC*. In fact, we demonstrated that expression of a dominant-negative form of BTRC (BTRC(ΔF)) stabilized ZC3H12A upon stimulation with IL-17A and IL-1β **(EDFig. 6d)**[19], further supporting the role of DSGxxS mutants in positive selection. An elevated expression was also demonstrated for a C-terminal-truncating mutant (W543*). Of note, the W543* mutant showed much higher protein expression than the S438L mutant, even though their mRNA levels were similar **(Fig. 3b)**, suggesting that the C-terminus deleted protein has a higher stability regardless of IL-17A stimulation. Also implicated in mRNA degradation[20], *HNRNPF* showed mutational hotspots affecting highly conserved tyrosine-containing residues (aa201-210) **(EDFig. 6a)**, although its link to IL-17 or other cytokine signalling pathways is unknown.

We next evaluated the role of IL-17 signalling in the positive selection of *NFKBIZ*-deficient clones in UC epithelia. When an equal number of colon organoids from *Nfkbiz*$^{-/-}$ and *Nfkbiz*$^{fl/fl}$ mice were mixed and co-cultured with or without IL-17A, we observed an increasing allelic ratio of *Nfkbiz*$^{-}$ relative to *Nfkbiz*^{fl} in the presence of IL-17A, whereas the ratio was rather decreased in culture without IL-17A, supporting a selective advantage of *Nfkbiz*-null *cells* over *Nfkbiz*-intact cells under the presence of IL-17A **(Fig. 3c)**. We also investigated the driver role of *NFKBIZ* in UC *in vivo* by generating *R26*$^{LacZ/wt}$*Nfkbiz*$^{fl/fl}$*VilCre*^{ER} and *R26*$^{LacZ/wt}$*Nfkbiz*$^{wt/wt}$*VilCre*^{ER} mice, in which colitis was induced with dextran sulphate sodium (DSS), where aberrant IL-17 signalling has been implicated[21]. In *R26*$^{LacZ/wt}$*Nfkbiz*$^{fl/fl}$*VilCre*^{ER} mice, *Nfkbiz*-deleted cells randomly introduced in the colorectal epithelium were visualized by LacZ staining. After tamoxifen treatment, we induced chronic colitis in these mice by oral administration of DSS and the size of LacZ-positive areas was compared between *Nfkbiz*$^{-/-}$ and *Nfkbiz*$^{+/+}$ mice 11-12 weeks after tamoxifen treatment **(EDFig. 8)**. Overall, the total size of LacZ-positive area, as well as severity of inflammation as evaluated by body weight, colon length, and histology, did not significantly differ between these mice **(EDFig. 8f-h)**. However, the number of LacZ-positive areas sized 2^{8}-2^{14} µm^{2} was significantly increased in *Nfkbiz*$^{-/-}$ mice, compared with *Nfkbiz*$^{+/+}$ mice **(Fig. 3d)**, which was not observed without DSS treatment **(EDFig. 8d)**. RNA sequencing of DSS-treated colorectal epithelium revealed down-regulation of IL-17 signalling pathway genes and downstream targets of NFKBIZ in *Nfkbiz*$^{-/-}$ mice compared with *Nfkbiz*$^{+/+}$ mice **(EDFig. 8i)**, also supporting the role of *NFKBIZ* mutations in the positive selection in UC epithelium.

Negative role of *NFKBIZ* mutation in CAC

Next, we investigated driver mutations and copy number abnormalities (CNAs) in 99 CAC samples, including 65 with publicly available WES data[22-25] and 34 newly analysed with WES, together with 356 TCGA cases with sporadic colorectal cancer (sCRC) **(Fig. 4 and EDFig. 9)**. We also performed WES in 13 samples with low- and high-grade dysplasia samples from UC patients.

CAC and UC dysplasia samples had significantly higher numbers of mutations, compared with UC non-dysplasia samples **(EDFig. 9c,d)**. Analysis of mutational signature in UC-derived samples disclosed two predominant signatures, Signatures A and B **(Methods) (EDFig. 9c,e)**. Corresponding to COSMIC signature 1, Signature A mostly consisted of C>T transitions in the CpG contexts and is predominant across in all tissue types**,** while , corresponding to COSMIC signature

17, Signature B were characterized by T>G and T>C substitutions at C[T]T context and highly enriched in UC-derived samples **(EDFig. 9f)**, suggesting its possible link to chronic inflammation.

Driver mutation profile was substantially different between UC non-dysplasia and cancer samples. dN/dS analysis identified 6 genes significantly positively selected in CAC **(EDFig. 9a)**, which were also commonly mutated in sCRC **(Fig. 4a)** but were rarely affected in UC non-dysplasia, in which mutations in *NFKBIZ, ARID1A, ZC3H12A, PIGR,* and *ETV6* were significantly enriched. Copy number profiles were also very different between CAC/dysplasia and non-dysplasia. CNAs were obligatory and extensive in CAC, but much less common (41%) in non-dysplasia samples **(EDFig. 9a,g,h)**.

While being highly prevalent in UC non-dysplasia, *NFKBIZ* mutations were found in none of 99 CAC samples. Thus, CAC do not develop from *NFKBIZ*-mutated cells, suggesting a strong relative disadvantage of *NFKBIZ*-mutated clones in CAC that were positively selected in UC epithelium (P=~7.7×10^{-13}) **(Methods) (Fig. 4b)**. Similarly, *NFKBIZ* mutations were rarely seen in TCGA (2/356) and Japanese cases with sCRC (6/525), of which only one case had a truncating mutation **(Fig.4a, EDFig. 9b)**, also supporting the negative selection of *NFKBIZ*-mutated cells. To test this hypothesis, we transduced a lentiviral library consisting of 6 *NFKBIZ*-directed sgRNA and 4 non-specific (control) sgRNA constructs for CRISPR/Cas9-mediated gene-targeting into 5 human CRC cell lines and the enrichment of cells targeted by each sgRNA construct during 4 weeks' cell culture was evaluated **(EDFig. 10a)**. We observed a significant reduction in cell populations transduced with *NFKBIZ*-directed sgRNAs compared with control sgRNA-transduced cells in all CRC cell lines **(Fig. 4c)**. Moreover, inflammation-induced colon tumours using DSS and Azoxymethane (AOM) were significantly suppressed in *Nfkbiz*$^{fl/fl}$*VilCre* mice compared with control *Nfkbiz*$^{fl/fl}$ mice in number and size **(Fig. 4d,e and EDFig. 10b-f)**. These results of functional analyses support the negative selection of *NFKBIZ*-mutated CRC cells. A significant reduction in mutation frequency in CAC compared to UC epithelium was also observed for *ZC3H12A* mutations, which shows a sharp contrast to *TP53* and *KRAS* mutations substantially enriched in CAC **(Fig. 4b)**.

Discussion

In the steady state after intestinal development, crypts in adults still divide to balance physiological deficits of crypts. Assuming a constant crypt number in the colon, annually 50,000 crypts are estimated to be lost and replaced by new ones by fissions of adjacent crypts, which are equivalent to ~0.4% of the total intestinal crypts per year or only ~0.00413 fissions/crypt/year. In UC patients, crypt fission is dramatically accelerated (~2.16 fissions/crypt/year) to amend the large deficits of crypts, during which strong positive selection operates to allow for the dominance of clones carrying mutations in the IL-17 and other cytokine/PPR pathways. Importance of these pathways in UC pathogenesis is supported by GWAS studies reporting a significant association of UC to the *NFKBIZ* and *TRAF3IP2* loci[26,27].

Despite a significant overlap of driver genes between CAC and UC epithelia, their frequencies substantially differed between both tissues, highlighting the presence of discrete mechanisms of positive selection in cancer and non-cancer tissues. A new finding from the study on UC epithelia would be that the mechanism that allows for positive selection under chronic inflammation may not necessarily contributes to neoplastic growth. On the contrary, it may impose strong negative effects on cancer development, as evident from extremely rare mutations of *NFKBIZ* in CAC despite their pervasive nature in UC non-dysplasia.

Finally, one of the most important implications from the current study would be a possibility that chronic inflammation can evoke a tissue response by way of remodelling affected tissues by positively selecting clones that acquired mutations, to the extent that the entire rectum is almost fully reconstituted by a few clones carrying IL-17 pathway mutations. Chronic inflammation is a highly prevalent disability among the human population, potentially affecting its target organs and tissues to modify disease pictures, where a similar mechanism may operate and lead to tissue remodelling depending on the nature of inflammation. Understanding of such responses should be important for better understanding of the pathogenesis of inflammatory diseases and for the development of novel therapeutics.

REFERENCES:

1 Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883-899, doi:10.1016/j.cell.2010.01.025 (2010).

2 Hunter, P. The inflammation theory of disease. The growing realization that chronic inflammation is crucial in many diseases opens new avenues for treatment. *EMBO reports* **13**, 968-970, doi:10.1038/embor.2012.142 (2012).

3 Nathan, C. & Ding, A. Nonresolving inflammation. *Cell* **140**, 871-882, doi:10.1016/j.cell.2010.02.029 (2010).

4 Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *The New England journal of medicine* **371**, 2488-2498, doi:10.1056/NEJMoa1408617 (2014).

5 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, doi:10.1126/science.aaa6806 (2015).

6 Yokoyama, A. *et al.* Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312-317, doi:10.1038/s41586-018-0811-x (2019).

7 Burisch, J., Jess, T., Martinato, M., Lakatos, P. L. & -EpiCom, o. b. o. E. The burden of inflammatory bowel disease in Europe. *Journal of Crohn's and Colitis* **7**, 322-337, doi:10.1016/j.crohns.2013.01.010 (2013).

8 Ungaro, R., Mehandru, S., Allen, P. B., Peyrin-Biroulet, L. & Colombel, J. F. Ulcerative colitis. *Lancet (London, England)* **389**, 1756-1770, doi:10.1016/s0140-6736(16)32126-2 (2017).

9 Eaden, J. A., Abrams, K. R. & Mayberry, J. F. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut* **48**, 526-535 (2001).

10 Kang, H. & Shibata, D. Direct measurements of human colon crypt stem cell niche genetic fidelity: the role of chance in non-darwinian mutation selection. *Frontiers in oncology* **3**, 264-264, doi:10.3389/fonc.2013.00264 (2013).

11 Choi, C.-H. R., Bakir, I. A., Hart, A. L. & Graham, T. A. Clonal evolution of colorectal cancer in IBD. *Nature Reviews Gastroenterology &Amp; Hepatology* **14**, 218, doi:10.1038/nrgastro.2017.1 (2017).

12 Eto, A., Muta, T., Yamazaki, S. & Takeshige, K. Essential roles for NF-kappa B and a Toll/IL-1 receptor domain-specific signal(s) in the induction of I kappa B-zeta. *Biochemical and biophysical research communications* **301**, 495-501, doi:10.1016/s0006-291x(02)03082-6 (2003).

13 Yamamoto, M. *et al.* Regulation of Toll/IL-1-receptor-mediated gene expression by the inducible nuclear protein IkappaBzeta. *Nature* **430**, 218-222, doi:10.1038/nature02738 (2004).

14 Bruno, M. E. C., Frantz, A. L., Rogier, E. W., Johansen, F. E. & Kaetzel, C. S. Regulation of the polymeric immunoglobulin receptor by the classical and alternative NF-κB pathways in intestinal epithelial cells. *Mucosal immunology* **4**, 468, doi:10.1038/mi.2011.8 (2011).

15 Cao, A. T., Yao, S., Gong, B., Elson, C. O. & Cong, Y. Th17 cells upregulate polymeric Ig receptor and intestinal IgA and contribute to intestinal homeostasis. *Journal of immunology (Baltimore, Md. : 1950)* **189**, 4666-4673, doi:10.4049/jimmunol.1200955 (2012).

16 Bruno, M. E. *et al.* Correlation of Biomarker Expression in Colonic Mucosa with Disease Phenotype in Crohn's Disease and Ulcerative Colitis. *Digestive diseases and sciences* **60**, 2976-2984, doi:10.1007/s10620-015-3700-2 (2015).

17 Nakatsuka, Y. *et al.* Pulmonary Regnase-1 orchestrates the interplay of epithelium and adaptive immune systems to protect against pneumonia. *Mucosal immunology* **11**, 1203-1218, doi:10.1038/s41385-018-0024-5 (2018).

18 Mino, T. *et al.* Regnase-1 and Roquin Regulate a Common Element in Inflammatory mRNAs by Spatiotemporally Distinct Mechanisms. *Cell* **161**, 1058-1073, doi:10.1016/j.cell.2015.04.029 (2015).

19 Iwasaki, H. *et al.* The IkappaB kinase complex regulates the stability of cytokine-encoding mRNA induced by TLR-IL-1R by controlling degradation of regnase-1. *Nature immunology* **12**, 1167-1175, doi:10.1038/ni.2137 (2011).

20 Reznik, B., Clement, S. L. & Lykke-Andersen, J. hnRNP F complexes with tristetraprolin and stimulates ARE-mRNA decay. *PLoS One* **9**, e100992, doi:10.1371/journal.pone.0100992 (2014).

21 Ramirez-Carrozzi, V. *et al.* IL-17C regulates the innate immune function of epithelial cells in an autocrine manner. *Nature immunology* **12**, 1159-1166, doi:10.1038/ni.2156 (2011).

22 Robles, A. I. *et al.* Whole-Exome Sequencing Analyses of Inflammatory Bowel Disease-Associated Colorectal Cancers. *Gastroenterology* **150**, 931-943, doi:10.1053/j.gastro.2015.12.036 (2016).

23 Fujita, M. *et al.* Genomic landscape of colitis-associated cancer indicates the impact of chronic inflammation and its stratification by mutations in the Wnt signaling. *Oncotarget* **9**, 969-981, doi:10.18632/oncotarget.22867 (2017).

24 Din, S. *et al.* Mutational analysis identifies therapeutic biomarkers in inflammatory bowel disease-associated colorectal cancers. *Clinical Cancer Research*, clincanres.3713.2017, doi:10.1158/1078-0432.Ccr-17-3713 (2018).

25 Baker, A. M. *et al.* Evolutionary history of human colitis-associated colorectal cancer. *Gut*, doi:10.1136/gutjnl-2018-316191 (2018).

26 Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124, doi:10.1038/nature11582 (2012).

27 Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics* **47**, 979-986, doi:10.1038/ng.3359 (2015).

FIGURE LEGENDS:

Figure 1 | Somatic mutation in single crypts.

a, Correlation between the number of mutations in single crypts in WES and subjects' age in crypts from UC- and non-UC-derived crypts. Two-sided Mann-Whitney *U* test. A regression line assuming zero intercept for non-UC samples with R^2 and coefficient values (blue). R^2 (red) of the linear regression model for UC samples is also shown **(Methods)**. **b,** Average number of mutations in paired single crypts from right- and left-side colon in each non-UC and UC patient. Two-sided paired *t*-test. **c,** Clonal structure within adjacent crypts in epithelia from non-UC and UC patients. Corresponding clones in the phylogenetic trees and geographical maps are depicted by numbers and colours. Estimated time-point of UC onset, the age of the most recent common ancestral crypt in HCM09, and branches containing driver mutations are indicated **(Methods)**.

Figure 2 | Driver genes and clonal expansion in UC epithelium.

a, Distribution of *ZC3H12A* mutations in UC non-dysplasia (58 solitary single crypts and 399 bulk crypts, above) and CAC (n=99, below) samples. **b,** Landscape of driver mutations detected in clones (n=183) in UC non-dysplasia. Frequencies of each mutations are indicated on the right. Genes with asterisks are drivers in UC non-dysplasia. **c,** Mutations in IL-17 and other proinflammatory signalling pathways in non-dysplastic UC samples. Mutated components are indicated by coloured boxes. **d,** Expansion of clones within the rectum from two representative UC patients (out of 6). Histology is shown by colour in circle centre. Samples harbouring independent clones are discriminated by arch colour, while the arch size indicates the fraction of shared clone estimated by PyClone. Exceptions are samples with a grey arch; none of the mutations in each grey-arched sample are shared by other samples, even by other grey-arched samples. Driver genes harbouring each independent clone are indicated at the bottom, except in grey-arched samples, for which just the presence or absence of IL-17 pathway mutations is indicated by dark or light grey, respectively. **e,** Fraction of total sampling area occupied by clones with indicated mutations in each UC-derived rectum.

Figure 3 | Effect of driver mutations.

a, Fold changes of expression levels of *Nfkbiz* and *Pigr* (mean±SD) in colon epithelial organoids established from $Nfkbiz^{fl/fl}$ (control) and $Nfkbiz^{fl/fl}Vil^{Cre}$ (cKO) mice at indicated time points after IL-17A stimulation. Asterisks indicate $p<0.05$ (two-sided Mann-Whitney *U* test). **b,** Immunoblot analysis of HCT116 cells expressing empty, Myc-tagged wild type (WT), or ZC3H12A mutants upon stimulation with IL-17A for indicated time, using anti-Myc and β-actin antibodies (left). Data is representative of two independent experiments. Protein/mRNA ratio of

ZC3H12A (n=3 biological replicates) (mean±SD) (right). An asterisk indicates $p<0.05$ between mutant and wild type ZC3H12A (two-sided Student's t-test). For gel source data, see **Supplementary Figure 1**. **c,** Fold changes (mean±SD) of ratios of *Nfkbiz*$^{-}$ to *Nfkbiz*^{fl} alleles at indicated passages in *Nfkbiz* KO and control organoids co-cultured under the presence or absence of IL-17A. Asterisks indicate $p<0.05$ (two-sided Mann-Whitney U test). **d,** Density of LacZ-positive areas measured for indicated LacZ-positive area size in control and *Nfkbiz* cKO mice. Asterisks indicates $p<0.05$ (two-sided Mann-Whitney U test).

Figure 4 | Comparison of drivers between UC non-dysplasia and cancer.

a, Frequencies of driver mutations among clones identified in UC non-dysplasia (ND) vs. CAC samples vs. sCRC samples. Asterisks indicate $q<0.1$ (two-sided Fisher's exact test with Benjamini–Hochberg adjustment). **b,** Observed mutation frequency of indicated genes in CAC (**Fig. 4a**) and 95% CI of frequency expected from epithelial fraction carrying mutations in UC epithelium. Asterisks indicate $q<0.05$ (two-sided binomial test with Benjamini–Hochberg adjustment). **c,** Box plots of normalized fold change in frequency of indicated sgRNA sequence reads after 4 weeks' culture compared with pre-culture value in 5 human CRC cell lines (n=6 biological replicates). Significant differences between control and *NFKBIZ*-directed sgRNAs are provided (two-sided Mann-Whitney U test). **d,** Representative images of AOM-DSS-induced colon tumours in control (3 out of 25, top) and *Nfkbiz* cKO (3 out of 25, bottom) mice. **e,** Box plots of sum of tumour size per mouse. Two-sided Mann-Whitney U test. In all box plots, median, first and third quartiles, as well as outliers, are indicated with whiskers extending to the furthest value within 1.5 of the interquartile range.

METHODS:

Subjects and materials

We enrolled a total of 647 subjects who underwent surgery or endoscopy of the lower gastrointestinal tract at Kyoto University Hospital, Hyogo College of Medicine Hospital, and Cancer Institute Hospital. Among these, 76 had been diagnosed as ulcerative colitis (UC) and the remaining 571 (healthy or non-UC) were diagnosed as other disease such as colorectal polyp or cancer without associated colitis or turned out to be healthy individuals. This study was approved by the ethical committees at Kyoto University, Hyogo College of Medicine, and Cancer Institute Hospital. We have complied with all relevant ethical regulations. Informed consent was obtained from all participants. Characteristics of these subjects are summarized in **Supplementary Table 1**.

Specimens from non-UC subjects were endoscopically or surgically obtained from non-dysplastic intestinal mucosa. After methylene blue staining, they were carefully assessed under stereomicroscope to exclude samples involving dysplastic crypt lesions such as aberrant crypt foci. All specimens obtained from UC subjects were pathologically evaluated using a half of each specimen or its adjacent mucosa. Among them, 6 surgically resected rectum samples were exhaustively evaluated by 5 mm sections across the whole sampling areas according to 1cm lattices. On the basis of Riddell's classification[28], pathologic evaluation was performed by two expert pathologists (Y.T. and T.Sakurai, or T.N. and S.H.) and samples were classified into four groups: non-dysplasia, low- and high-grade dysplasia, and cancer.

To obtain single crypts and crypts in bulk, samples obtained by endoscopic biopsy were trimmed to a size of approximately 4 mm^2. From surgical specimens, samples were collected by punch biopsy of 2.5 mm in diameter (4.9 mm^2) (BP-25F, kai industries). After being separated from the submucosal layers by incubating samples in ice-cold 20mM EDTA in PBS for 20 minutes, the epithelium was mechanically dissociated using fine forceps, by which single crypts or bulk crypts samples were collected **(EDFig. 2a)**. To obtain a geographically mapped crypt cluster, the cryptic structure within the mucosa was first fixed with acrylic adhesive from the epithelial surface, then the underneath cluster of crypts was peeled off *en block* from submucosal layers after treatment with 20 mM EDTA. Otherwise, a gentle pressure was applied to the whole mucosal tissue from both sides to push up crypts after treatment with 20mM EDTA. Finally, individual crypts were isolated with their position being recorded and numbered **(EDFig. 2b)**. For the analysis of cancer, tumour cells were freshly obtained or dissected from formalin-fixed paraffin-embedded (FFPE) under stereomicroscope, from which DNA was extracted.

For whole exome sequencing (WES), genomic DNA isolated from each single crypt was split into two aliquots, each of which was independently subjected to whole genome amplification (WGA) with REPLI-g Single Cell Kit (Qiagen) and used for either WES or subsequent validation.

Other freshly obtained samples were subjected to extraction of genomic DNA and RNA simultaneously using AllPrep DNA/RNA Micro Kit (Qiagen). FFPE cancer samples were subjected to extraction of DNA using GeneRead DNA FFPE Kit (Qiagen). DNA from peripheral blood or tumour-free muscularis propria materials macro-dissected from FFPE cancer samples was extracted using Gentra Puregene Kit (Qiagen) or GeneRead DNA FFPE Kit (Qiagen), respectively, and was used as germline control. For targeted capture sequencing of sporadic colorectal cancers (sCRCs), cryo-preserved samples were subjected to extraction of genomic DNA using AllPrep DNA/RNA Mini Kit (Qiagen). Sample information is summarized in **Supplementary Tables 2 and 3**.

Whole exome sequencing

WES libraries were prepared using SureSelect Human All Exon V5 (Agilent Technologies) or xGen Exome Research Panel (IDT), followed by sequencing of enriched exon fragments on Hiseq 2500 or NovaSeq 6000 (Illumina) with 100-150 bp paired-end mode as previously described[29]. The target depth was 100×, and the actual coverage was 133x (60x-293x). The mean coverage in germline controls from 122 individuals was 140× (55×-204×). Mutation calling was performed using Genomon2 pipeline v2.6, as previously described[6]. Briefly, sequencing reads were aligned to the human genome reference (hg19) using Burrows-Wheeler Aligner, version 0.7.8 with default parameter settings. PCR duplicates were eliminated using biobambam version 0.0.191. Somatic mutations were detected by eliminating polymorphisms and sequencing errors. To achieve this, Genomon2 first discards any of low-quality, unreliable reads and variants, that are defined by the following criteria: (i) mapping quality < 20, (ii) base call quality < 15. After further excluding those variants that are not supported by a sufficient number of reads (total reads ≥ 8 and variant reads ≥ 3), variant allele frequencies (VAFs) ≥ 0.05 (for fresh and FFPE samples), ≥ 0.25 (for single-crypt WGA samples), and < 0.02 (for germline control), and strand ratio ≠ 0 or 1, the remaining variants were interrogated for the evidence that they were observed at significantly higher VAFs than expected for errors ($P \leq 10^{-4}$, $10^{-3.5}$, and 10^{-4} for fresh, FFPE, and WGA samples, respectively), where the significance is evaluated by EBCall algorithm[30], on the basis of an empirical VAF distribution as determined using WES data of non-paired peripheral blood samples (n=20). Putative germline variants were also excluded by comparing VAFs with matched control using the Fisher's test ($\leq 10^{-1}$, 10^{-1}, and 10^{-2} for fresh, FFPE, and WGA samples, respectively), which also eliminated remaining additional sequencing errors. The monoclonal origin of single-crypt cells, which are derived from a single stem cell at a given time, was evident from VAFs normally distributed around 0.5 **(EDFig. 3a,b)**.

For mapping samples of adjacent single crypts and of surgically resected rectum, each of the mutations detected in a sample was interrogated in all the remaining samples; if a variant detected in one or more samples by above mutation call was also present (VAF ≥ 0.1 and ≥ 0.05 for single-crypt mapping and rectum mapping, respectively) in another sample within the same mapping specimen, the variant was considered to be true positive, because it serves a high prior probability for true positivity.

Validation of detected mutations

Validation of mutations detected by WES in fresh and FFPE samples was performed using PCR-based deep (>500×) sequencing as previously described[29], in which 708 single nucleotide variants (SNVs) and 28 indels were randomly selected and evaluated. A mutation was considered to be validated, when (1) the sequencing depth was ≥500x in both test and germline control samples, (2) the VAF in the test sample was 5 times higher than that in the corresponding germline control sample, and (3) the VAF in the test sample should be high enough (≥0.01) to be captured by the lower threshold in WES (0.05)[6]. The overall validation rate was 99.0 % for fresh samples (390/394 mutations) and 97.4 % for FFPE samples (333/342 mutations). Validation of mutations detected by single-crypt WGA samples was performed by WES of replicated samples **(EDFig. 2c)**, in which 126 SNVs and indels detected in 3 single crypts were evaluated. A mutation was considered to be validated, when (1) the sequencing depth in the replicated sample was ≥8, (2) the VAF in the replicated sample was > 0.2. The validation rate of WGA samples was 98.4 % (124/126 mutations). The results of validation sequencing were summarized in **Supplementary Table 4**.

External dataset

WES data (bam files) of paired tumour/germline control samples from sCRC patients were downloaded from the TCGA Data Portal. Bam files from the TCGA were converted to fastq format using biobambam and processed with the same pipeline for mutation calling as applied to fresh samples obtained in this study (see above). For accuracy, sequencing data generated from tumours with WGA were excluded. Hypermutated tumours (≥ 12 mutations/10^6 bases) were also excluded from the analysis of comparison of driver genes between UC non-dysplasia and TCGA sCRC. Mutation lists of 65 colitis-associated cancer (CAC) in UC individuals were available from the four publications of UC-CAC WES studies[22-25]. Samples included in this study were summarized in **Supplementary Table 2**.

Analysis of copy number abnormalities (CNAs)

CNAs were evaluated from WES data using our inhouse pipeline “CNACS”[6]. The fractions of genomic regions with CNAs were compared between non-dysplasia, dysplasia, and CAC in UC individuals, where average fraction with CNAs was used for expanded clones observed in more than one sample.

Estimation of mutant cell fractions

Precise estimation of the cell fraction having a somatic mutation, mutant cell fraction (MCF), might be complicated when the mutation involves a gene located on the chromosomal segment showing a CNA. Most mutations detected by WES in solitary non-dysplastic colon epithelia were outside of regions with CNAs. Thus, for those mutations (7,024 / 7,072), MCF is simply estimated from their VAFs as follows:

$\mathrm{MCF} = 2 \times \mathrm{VAF}$ (for autosome and X chromosome in female), or
$\mathrm{MCF} = \mathrm{VAF}$ (for X chromosome in male)

By contrast, when the mutated locus did show abnormal copy number or allelic imbalance, estimation of MCF depends on several factors, which may not be inferred in a deterministic way, particularly in the case where a sample contains multiple clones with different mutant allele compositions. However, a mutation on a region showing copy-number loss or allelic imbalance, following calculation never overestimate MCF. Thus, we estimate MCF of such mutations (31 / 7,072) as follows:

$\mathrm{MCF} = 2 \times \mathrm{VAF} - \%\mathrm{UPD}/100$ (for UPD)
$\mathrm{MCF} = \mathrm{VAF} \times (2 - \%\mathrm{Del}/100)$ (for deletion)

where %UPD and %Del denote estimated fraction (%) of the respective abnormality. Other mutations associated with copy-number gain regions (17 / 7,072) were excluded from MCF analysis in **EDFig. 5**.

Analysis of significantly mutated (driver) genes

Significantly mutated genes, or driver genes, were investigated relying on dN/dS, which was calculated for all recurrently mutated genes using dNdScv. For surgically resected rectum specimens, where samples were obtained according to 1cm lattices, mutations from the same clone could be detected in multiple samples. Such multiple observation of mutations belonging to the same clone should be corrected before the analysis. To achieve this, we first clustered samples into a sample set in which samples shared ≥4 mutations with VAF ≥0.25 **(Supplementary Table 5)**, or otherwise, treated as single samples. Then, each mutation detected in a case was decided to be included in a sample set or a single sample which had the highest VAF of the mutation. After applying dNdScv to the mutations in non-dysplasia (36 sample sets including 152 samples, and 243 single samples) and non-hypermutated CAC (56 including our samples and publicly available data with information of synonymous mutations) in

UC individuals separately, those 14 and 6 genes that were significantly mutated (dN/dS>1, q<0.05) and found in ≥ 4% of non-dysplasia or CAC samples were considered as drivers **(Supplementary Tables 6)**. Driver genes were also analysed by using MutSigCV, where 10/14 and 4/6 genes were confirmed as significantly mutated (q<0.05) in non-dysplasia and cancer in UC individuals, respectively.

In UC non-dysplasia, we investigated loss of function mutations in genes annotated as cytokine receptor (n=108) or pattern recognition receptor (PRR) genes (n=33). Through WES of 399 non-dysplasia samples from 29 UC patients, we identified a total of 2,575 truncating mutations (including those affecting splice sites). The significance of the mutations in each gene was evaluated by calculating type I error rate, assuming a Poisson distribution with observed mutation rate for truncating mutations (2,575/33,479,141 bp) and the length of each gene. Multiple testing was corrected based on the Benjamini–Hochberg method among mutated genes. Genes with q<0.05 were considered as significantly mutated. We also calculated dN/dS values for these 141 genes. Information of cytokine receptor and PRR genes is summarized in **Supplementary Table 7**. Genes with protein-truncating mutation that belongs to IL-17 signalling pathway were considered as potential drivers. In total, 20 genes were considered to be related to positive selection in UC non-dysplasia **(Fig. 2b)**.

Targeted capture deep sequencing of driver genes

Targeted capture deep sequencing of 525 Japanese non-hypermutated sCRCs was performed according to the manufacture's protocols, using a custom bait library (xGen Predesigned Gene Capture Pools; IDT) that were designed to capture 16 genes combining drivers in UC non-dysplasia and CAC **(Supplementary Table 6)**. After hybridization capture, the enriched DNA fragments were sequenced on HiSeq 2500. The mean depth was 1,879x (364-2,545x). Relevant somatic mutations were called using Genomon2 pipeline and were filtered by excluding (i) synonymous SNVs; (ii) EBCall P-value > 10^{-4}; (iii) VAFs < 0.02, (iv) strand ratio = 0 or 1, (v) known variants listed in SNP databases (1000 Genome Project, Exome Sequencing Project 6500, Human Genetic Variation Database, and Exome Aggregation Consortium). Candidate mutations were further filtered by removing missense SNVs with a VAF of 0.4-0.6 except for SNVs that have ≥ 5 entries in the Catalogue of Somatic Mutations in Cancer (COSMIC) v70. Detected mutations were summarized in **Supplementary Table 8**.

Mutational signature

Mutational signatures in 29,409 SNVs, consisting of 1,537 non-UC-derived SNVs detected by WES of single crypts (n=43), 16,428 SNVs of UC non-dysplasia which were either directly detected from WES of single crypts or assigned to PyClone-imputed clones from bulk crypt samples (n=241), 2,022 SNVs assigned to clones from UC dysplasia (n=13), and 9,422 SNVs detected in non-hypermutated CAC (n=92) were extracted using R package pmsignature. Extracted signatures were compared with the COSMIC signatures by R package deconstructSigs. Signature A was confirmed highly similar to COSMIC signature 1 and signature B to COSMIC signature 17. The frequencies of signature B were compared between non-dysplasia from non-UC individuals, non-dysplasia, dysplasia, and CAC from UC individuals, where samples or clones with less than 10 mutations were excluded from the analysis.

Detection of crypt expansion in colon epithelia in UC individuals

Maximum MCF ≥ 0.5 were observed in 58% (77/133) of 4 mm^2 of colon epithelia (solitary bulk crypts) from UC individuals, which means crypt expansion often occurred in chronically inflamed mucosa. To investigate genetic events in such expanded crypts, clonal composition was interrogated on the basis of the Beta Binomial emission model implemented in PyClone version 0.13.0, through which a set of clones having a discrete set of mutations were imputed together with their estimated cellular prevalence. For bulk crypt samples from rectum mapping, samples sharing more than 4 mutations with VAF ≥ 0.25 were considered to contain expanded crypts from the same ancestral crypt, and these samples were treated as a set of samples. Otherwise, samples were treated as independent single samples as well as solitary samples **(Supplementary Table 5)**. PyClone algorithm was applied to these sample sets and single samples. When maximum cellular prevalence of mutational clusters was more than 0.5, the *'Pigeonhole principle'* can be applied to decide whether such dominant cell fraction included other minor mutated cell fractions or not. Thus, only samples with maximum cellular prevalence ≥ 0.5 were included for the analyses of expanded crypts, and minor mutational clusters were included only when the sum of such dominant cell fractions and minor cell fractions exceeded 1 **(EDFig. 1d)**. Driver mutations detected in individual samples in geographical mapping of 6 UC patients with their clone size are summarized in **Supplementary Table 9**.

Phylogenetic analysis

To analyse evolutional history of adjacent single crypts within microscale mapping, genotype information at every locus which was mutated in one or more samples was used to reconstruct a phylogenetic tree. The maximum parsimony method using Subtree-Pruning-Regrafting algorithm was performed by MEGA X, and the most parsimonious tree from 1,000 replicates was obtained. Timepoints of the UC onset in phylogenetic trees was estimated from the age of UC onset and mutation rate in non-UC subjects (**Fig. 1a**). The age of the most recent common ancestral crypt in UC trees was estimated from the timepoint of UC onset, height of the first branchpoint, and average number of mutations in analysed crypts within each UC patient, assuming a linear accumulation of mutations after UC onset.

In the multisampling analysis for rectum samples, phylogenetic trees were constructed for those clones observed in ≥2 samples, in which the assignment of different subclones to distinct clonal structure was accomplished according to the *'Pigeonhole principle'* on the basis of cellular prevalence of each mutational clusters as estimated from PyClone as described previously[6]. Briefly, when and only when the sum of their cellular prevalence exceeded 1, the two clusters were thought to be in the same clonal structure, where the cluster having smaller cellular prevalence was assigned to a subclone of the other cluster with larger cellular prevalence; all clusters for which clonal structures were undetermined were excluded from further analyses. Representative phylogenetic trees are shown in **EDFig. 7c-h**.

Estimation of crypt fission rate

In non-UC individuals, the rate of crypt fissions (fission/crypt/year) before 20 years of age was estimated on the basis of the number of ancestral crypts at birth and the number of crypts at 20 years of age in phylogenetic trees in three samples, according to the following formula:

Fission rate = $(\log_2(N_{\text{crypt}} \text{ at 20 years old}) - \log_2(N_{\text{crypt}} \text{ at birth})) / 20$ years

N_{crypt}: number of crypts

, where we assumed that:

1) Every crypt had <1 somatic mutation at the time of birth,
2) The intestine finishes development until 20 years of age, during which crypts will not be lost.

The fission rate after 20 years of age was estimated on the basis of the number of branching which occur after 20 years of age to the sampling time point, according to the following formula:

Fission rate = $(N_{\text{branchpoint}} / N_{\text{crypt}}) / Y_{\text{after20}}$

$N_{\text{branchpoint}}$: number of branchpoint after 20 years of age (12.2 mutations)

N_{crypt}: number of analysed crypts

Y_{after20}: years subtracting 20 from subject's age

, where we assumed that:

1) Colon size does not change after 20 years of age,
2) Crypt fissions occur only to balance the physiological deficits.

In UC patients, the fission rate was estimated based on the mutation rate after the disease onset and the time point of the most recent common ancestral (MRCA) branchpoint of analysed crypts, according to the following formula:

Fission rate = $\log_2(N_{\text{crypt}}) / Y_{\text{afterMRCA}}$

N_{crypt}: number of analysed crypts

$Y_{\text{afterMRCA}}$: years after the MRCA branchpoint to subject's age

, where we assumed that:

1) The mutation rate after the disease onset was constant,
2) Crypts were not lost after the time point of the MRCA crypt.

Comparison between mutation frequency in CACs and mutated fraction in UC epithelium

In order to test the hypothesis that each gene mutation has any selective effect on development of CAC from background UC epithelia, we compared mutation frequency of each gene in 99 CAC samples (**Fig. 4a**) with that expected from the fraction of mutated area in UC colon epithelia under the null hypothesis H_0: gene mutation has no selective effect on the development of CAC from background UC colon epithelia. First, for each driver genes, mean fraction of mutated area ($F_{UC\_epithelia}$) in actual UC colon epithelial samples (n=432) was calculated by using cellular prevalence in PyClone model. Then, 95% CI of mutation frequency in CACs expected from mutated fraction in UC colon epithelia was calculated for each gene from a binomial distribution assuming the probability ($F_{UC\_epithelia}$) and total number of CAC samples (n=99). For each driver gene, significant difference between observed and expected mutation frequency in CAC was calculated by the two-sided binomial test using total number of CAC samples (n=99), observed number of mutated CAC samples, and the probability ($F_{UC\_epithelia}$). Multiple testing was corrected based on the Benjamini–Hochberg method. Detailed results are summarized in **Supplementary Table 10**.

Cell culture

HeLa cells were maintained in DMEM (08459-64, nacalai tesque) supplemented with 10% FBS and 100 µM 2-mercaptoethanol (21418-42, nacalai tesque). HCT116, HT29, SW480, DLD1, and LoVo cells were maintained in DMEM supplemented with 10% FBS.

Wild-type and mutant ZC3H12A introduction

cDNAs of full length human wild-type *ZC3H12A* encoding an RNase: ZC3H12A and C-terminal truncation mutant (W543*) of this gene were inserted into pcDNA3.1(+) (V79020, Thermo Fisher) together with the N-terminal Myc-tag sequence. The S438L mutant ZC3H12A was made using QuikChange Lightning Site-Directed Mutagenesis Kit (210519, Agilent). Plasmid DNAs were transfected into HCT116 cells by using Lipofectamine 2000 (11668500, Thermo Fisher) according to manufacturer’s protocol.

Degradation assay of ZC3H12A upon cytokine stimulation

HCT116 cells were stimulated with 50 ng/ml recombinant human IL-17A (570502, BioLegend) or 10 ng/ml IL-1β (201-LB-005, R&D Systems). Cells were lysed in the lysis buffer (20 mM Tris-HCl (pH = 7.4), 150 mM NaCl, 0.5 % NP-40, cOmplete Mini (EDTA-free) (11836170001, Roche), 200 units/ml RNaseOUT (10777019, Thermo Fisher)). Protein samples were resolved on a 7.5 % PAGE gel (E-T7.5L, ATTO) and transferred to Immun-Blot PVDF membranes (1620177, Bio-Rad). The following antibodies were used for immunoblot analysis: monoclonal anti-c-Myc (M4439, Sigma-Aldrich), monoclonal anti-beta-Actin (sc-47778 HRP, Santa Cruz), anti-mouse IgG HRP (NA9310-1ML, GE Healthcare). Band intensities were analysed using Amersham Imager 600 Analysis Software (GE Healthcare). For the analysis of protein/mRNA ratio, total RNAs were isolated using TRIzol Reagent (15596018, Thermo Fisher) and reverse-transcribed using ReverTra Ace qPCR RT Master Mix with gDNA Remover (FSQ-301, TOYOBO) according to manufacturer's protocols. The cDNA fragments were amplified with PowerUp SYBR Green Master Mix (A25742, Thermo Fisher). Primers are described in **Supplementary Table 11**.

Degradation assay of ZC3H12A under BTRC inhibition

HeLa cells were infected with CS-TRE-Flag-BTRCΔF-PRE-Ubc-tTA-I2G lentivirus by which doxycycline (Dox) treatment induced the expression of dominant negative form of BTRC (BTRC(ΔF))[19]. After 100 ng/ml Dox administration, cells were stimulated with 50 ng/ml recombinant human IL-17A (570502, BioLegend) or 10 ng/ml IL-1β (201-LB-005, R&D Systems). Cells were lysed and protein samples were prepared as described above. The following antibodies were used for immunoblot analysis: polyclonal anti-ZC3H12A (HPA032053, Atlas Antibodies), polyclonal anti-Flag (F7425, MERCK), monoclonal anti-beta-Actin, anti-rabbit IgG HRP (NA9340V, GE healthcare).

Analysis of the effect of *NFKBIZ* on cancer cell growth

sgRNAs targeting *NFKBIZ* were designed, and knockout efficiency was confirmed by Guide-it Mutation Detection Kit (Z1448N, TAKARA). Five human colorectal cancer-derived cell lines, including, HT29, SW480, DLD1, HCT116, and LoVo cells, were infected with lentivirus library generated from pL-CRISPR.EFS.GFP (Addgene #57818) containing SpCas9 and sgRNAs targeting *NFKBIZ* (sgNFKBIZ #1-6) or non-targeting control sgRNAs (non-targeting #1-4) at a multiplicity of infection of 0.2 **(EDFig. 10a)**. GFP-positive cells were sorted using FACSAria IIIu (BD Biosciences) 3 days after infection. Half of the sorted cells were used for DNA extraction, while the remaining half were cultured in 6-well plates. After 4 weeks' culture, 50ng of genomic DNA extracted from each culture were subjected to PCR-amplification of integrated sgRNAs, together with the DNA from pre-culture cells, for amplicon sequencing using NextSeq 550. The fold change (FC^{sgRNA}) of the frequency of sgRNA-containing reads between pre- ($Freq^{sgRNA/Pre}$) and post- ($Freq^{sgRNA/Post}$) culture samples were calculated and normalized for the mean value of control sgRNAs as follows:

$$FC^{sgRNA} = Freq^{sgRNA/Post} / Freq^{sgRNA/Pre}$$

Normalized $FC^{sgRNA} = FC^{sgRNA}$ / mean FC for control sgRNAs

Significant differences between sgRNAs targeting *NFKBIZ* and non-targeting sgRNAs were calculated by two-sided Mann-Whitney *U* test. PCR primers and sgRNA sequences are provided in **Supplementary Table 11 and 12**, respectively.

Animal experiments

All *in vivo* experiments in animals were approved by Kyoto University Animal Care and Use Committee. Genetically engineered C57BL/6 mice were used. *VilCre* mice (004586) and *R26-LacZ* mice (003309) were obtained from The Jackson Laboratory. *Nfkbiz flox* mice[31] (RBRC06410) were obtained from Riken BRC. *VilCre*^{ER} mice[32] were kindly gifted by Sylvie Robine.

Cellular response to inflammatory stimulation in colon organoids

Colon organoids were generated from *Nfkbiz*$^{fl/fl}$*VilCre* (cKO) mice and *Nfkbiz*$^{fl/fl}$ (control) mice and were maintained in basement membrane matrix (Matrigel) (354234, Corning) with 50% L-WRN conditioned medium (CM) as previously described[33]. Control organoids were stimulated with 50 ng/ml IL-1β (575102, BioLegend), 50 ng/ml IL-4 (574304, BioLegend), 50 ng/ml IL-13 (575904, BioLegend), 50 ng/ml IL-17A (421-ML-025, R&D Systems), 50 ng/ml IL-19 (2915-IL-025, R&D Systems), 50 ng/ml IL-20 (1204-ML-025, R&D Systems), 50 ng/ml IL-24 (7807-ML-010, R&D Systems), 50 ng/ml IL-31 (210-31-10UG, PeproTech), 50 ng/ml Oncostatin M (762804, BioLegend), 50 ng/ml TNFα (575202, BioLegend), 50 ng/ml Lymphotoxin alpha1/beta2 (9968-LY-025/CF, R&D Systems), 10 µg/ml LPS (L6529-1MG, Sigma), 2.5 µM CpG oligodeoxynucleotides (ALX-746-001-C100, Enzo), and PBS (control) for 1 hour or 24 hours, followed by RNA extraction. Expression level of *Nfkbiz* and *Gapdh* were measured by real-time PCR. Expression level of *Nfkbiz*, *Pigr*, and *Gapdh* of organoids stimulated with IL-17A were also measured by real-time PCR at 0, 1, 4, and 24 hours after stimulation. For each gene, expression level was calculated as a ratio to *Gapdh*. Then, the fold changes of expression levels from values before IL-17A stimulation was calculated for each time point. Primer sequences are described in **Supplementary Table 11**.

Competition assay of colon organoids from *Nfkbiz*$^{-/-}$ and *Nfkbiz*$^{+/+}$ mice

An equal number of colon organoids derived from *Nfkbiz*$^{fl/fl}$*VilCre* (*Nfkbiz*$^{-/-}$) and *Nfkbiz*$^{fl/fl}$ (*Nfkbiz*$^{+/+}$) mice were co-cultured in 50% L-WRN CM with or without 50 ng/ml IL-17A stimulation. Organoids were passaged in the ratio of 1:5 every 5 days. Genomic DNA were extracted at every passage, followed by measurement of *Nfkbiz*$^{-}$ and *Nfkbiz*^{fl} alleles using real-time PCR. For each passage, a ratio of *Nfkbiz*$^{-}$ to *Nfkbiz*^{fl} alleles are calculated. Then, fold changes from ratios at start of culture were calculated for every passages. Primer sequences are described in

Supplementary Table 11.

Clonal expansion of ***Nfkbiz*** knock-out colon epithelium in chronic DSS colitis

R26$^{LacZ/wt}$*Nfkbiz*$^{fl/fl}$*VilCre*^{ER} (cKO) mice and *Rosa*$^{LacZ/wt}$*Nfkbiz*$^{wt/wt}$*VilCre*^{ER} (control) mice were co-housed after weaning. At 6 weeks of age, Cre-mediated recombination was induced by intraperitoneal injection of 2 mg of tamoxifen (T5648, SIGMA) twice a day for 2 consecutive days. At 8 to 9 weeks of age, experimental chronic colitis was induced by three cycles of 1% dextran sodium sulphate (DSS, M.W. = 36,000-50,000 Da; MP Biomedicals) (5 days DSS, followed by 16 days water). During three cycles of DSS, body weight was measured three times a week. Mice were killed before or after three cycles of DSS and colon lengths were measured. Colon was opened longitudinally and mounted, then was stained by X-gal (9031, Takara) according to manufacturer's protocol. Picture images of colon were taken under a stereomicroscope and LacZ-positive areas from middle to distal colon were measured by using ImageJ software. Histological scores of colitis in which epithelial damage and inflammatory-cell infiltration were scored from 0 to 12 as previous described[34] were counted on 5 randomly selected HPFs of HE stained slides in each mouse. The mean values of the scores were used for the analysis.

Transcriptome analysis of colon epithelial cells in DSS-induced colitis.

Colon epithelial cells were obtained from rectum mucosa from 8-week-old *Nfkbiz*$^{fl/fl}$*VilCre* and *Nfkbiz*$^{fl/fl}$ mice which underwent 4 days of 1% DSS administration. RNA was extracted by using AllPrep DNA/RNA Micro Kit (Qiagen). Libraries for RNA sequencing were prepared using the NEBNext Ultra RNA Library Prep kit for Illumina (New England Biolabs) and subjected to sequencing using HiSeq 2500 instrument (Illumina) with a standard 125-bp paired-end protocol.

The sequencing reads were aligned to the human or mouse reference genome (mm9) using STAR v2.5.3. Reads on each gene were counted with featureCounts v1.5.3 from Subread package, and edgeR package was used to identify the differentially expressed genes with false discovery rate (FDR; q-value) threshold of 0.05. The analysis was performed in genes expressed at > 1 counts per million (CPM) in two or more samples, and generalized linear models were used to compare gene expression data. GSEA (v2.2.4) was used to determine the sets of genes which were differentially expressed in the analysed cohort compared to the control cohort. For GSEA, publicly available gene sets, including downregulated genes in *IL17RA* knocked out cells[35] and downregulated genes in *Nfkbiz* knocked out cells[13,36-38], were used.

AOM-DSS induced Colitis-associated cancer model

Nfkbiz$^{fl/fl}$*VilCre* (cKO) and *Nfkbiz*$^{fl/fl}$ (control) littermates were co-housed after weaning. At 8 weeks of age, 12 mg per gram body weight of Azoxymethane (AOM) (A5486, SIGMA) were injected into peritoneum and administered 3 cycles of 2% DSS (5 days DSS, followed by 16 days water). After completion of 3 cycles of DSS, mice were sacrificed and colons were opened longitudinally. The size and number of tumours in the colon were measured. Induced tumours were evaluated histologically by HE staining, β-catenin, and Ki-67 immunohistochemistry as previously described[39]. Body weight, colon length, and histological score of colitis were also measured as described above.

Statistical analysis

Statistical analyses were performed using R, version 3.5.1. All *P* values were calculated by two-sided analysis unless otherwise specified. The Student's *t*-test, paired *t*-test, Fisher's exact test or Mann-Whitney *U* test was used for group comparisons. Multiple testing was corrected based on the Benjamini–Hochberg method. The linearity of the number of mutations in single crypts (n=43) from 22 non-UC individuals with age was evaluated on the basis of the Pearson's correlation coefficient in a linear regression model that assumed zero intercept, because somatic mutations among exome region at 0 year of age is assumed nearly 0. By the direct observation of mutations without culturing cells, the estimated mutation rate of colon epithelia was 0.61 mutations/exome/year in non-UC individuals, which is comparable to the previous report[40]. For the analysis of mutation rate of single crypts from 18 UC individuals, we included one single crypt per subject that showed the maximum number of mutations in the same UC individuals and all crypt (n=43) from 22 non-UC individuals. The mutation rate of single crypts after UC onset was estimated by linear regression model that incorporated both age of UC onset and disease duration, where non-UC individuals were treated as disease duration of 0, which showed significant improvement from a model that incorporated age of individuals alone, using an ANOVA **(Supplementary Table 13)**. For comparison between left- vs. right-side colon, the analyses were confined to those samples from the patients with sampling from both left- and right-side colons, where in each patient, mutations or MCFs were averaged for multiple samples in left- and right-side colon before paired *t*-test. Significant difference of maximum MCF of solitary bulk crypt samples between non-UC and UC individuals was tested by two-sided Mann-Whitney *U* test. For the linearity of maximum MCF in bulk crypt samples with UC duration, a linear regression model without assuming zero intercept was applied, because small number of mutations with high MCF can be observed as somatic mosaicism even in non-UC individuals. The significance of the correlation of UC duration with maximum MCF was tested using an *F* test.

References (in Methods):

28 Riddell, R. H. *et al.* Dysplasia in inflammatory bowel disease: standardized classification with

provisional clinical applications. *Human pathology* **14**, 931-968, doi:10.1016/s0046-8177(83)80175-0 (1983).

29 Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-69, doi:10.1038/nature10496 (2011).

30 Shiraishi, Y. *et al.* An empirical Bayesian framework for somatic mutation detection from cancer genome sequencing data. *Nucleic acids research* **41**, e89, doi:10.1093/nar/gkt126 (2013).

31 Okuma, A. *et al.* Enhanced apoptosis by disruption of the STAT3-IkappaB-zeta signaling pathway in epithelial cells induces Sjogren's syndrome-like autoimmune disease. *Immunity* **38**, 450-460, doi:10.1016/j.immuni.2012.11.016 (2013).

32 el Marjou, F. *et al.* Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis (New York, N.Y. : 2000)* **39**, 186-193, doi:10.1002/gene.20042 (2004).

33 Miyoshi, H. & Stappenbeck, T. S. In vitro expansion and genetic modification of gastrointestinal stem cells in spheroid culture. *Nature protocols* **8**, 2471-2482, doi:10.1038/nprot.2013.153 (2013).

34 Katakura, K. *et al.* Toll-like receptor 9-induced type I IFN protects mice from experimental colitis. *The Journal of clinical investigation* **115**, 695-702, doi:10.1172/jci22996 (2005).

35 Chen, K. *et al.* IL-17 Receptor Signaling in the Lung Epithelium Is Required for Mucosal Chemokine Gradients and Pulmonary Host Defense against K. pneumoniae. *Cell host & microbe* **20**, 596-605, doi:10.1016/j.chom.2016.10.003 (2016).

36 Johansen, C. *et al.* IkappaBzeta is a key driver in the development of psoriasis. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E5825-5833, doi:10.1073/pnas.1509971112 (2015).

37 Muller, A. *et al.* IkappaBzeta is a key transcriptional regulator of IL-36-driven psoriasis-related gene expression in keratinocytes. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 10088-10093, doi:10.1073/pnas.1801377115 (2018).

38 Hildebrand, D. G. *et al.* IkappaBzeta is a transcriptional key regulator of CCL2/MCP-1. *Journal of immunology (Baltimore, Md. : 1950)* **190**, 4812-4820, doi:10.4049/jimmunol.1300089 (2013).

39 Goto, N. *et al.* Distinct roles of Hes1 in normal stem cells and tumor stem-like cells of the intestine. *Cancer Research*, canres.3192.2016, doi:10.1158/0008-5472.Can-16-3192 (2017).

40 Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260-264, doi:10.1038/nature19768 (2016).

MAIN TEXT STATEMENTS:

Acknowledgments

This work was supported by Grants-in-Aid from the Ministry of Health, Labour and Welfare of Japan (Health and Labour Sciences Research Expenses for Commission and Applied Research for Innovative Treatment of Cancer), the Project for Development of Innovative Research on Cancer Therapeutics from the Japan Agency for Medical Research and Development, AMED (JP15cm0106056h0005 and JP16cm0106501h0001) [Kyoto; S. Ogawa], the Ministry of Education, Culture, Sports, Science and Technology of Japan, the High Performance Computing Infrastructure System Research Project (hp150232, hp160219, hp190158; This research used computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project) [Kyoto; S. Ogawa, Tokyo; S.M.], Scientific Research on Innovative Areas (15H05909) [Kyoto; S. Ogawa, Tokyo; S.M.], (15H05912) [Tokyo; S.M.], and "Stem Cell Aging and Disease" (14430052) [Nagoya; M.S.], JSPS KAKENHI (17H04157) [Kyoto; H.S.] & (18H05278) [Kyoto; O.T.], Takeda Science Foundation [Kyoto; S. Ogawa, H.M., T.Y.], Uehara Memorial Foundation [Kyoto; H.S.], and Foundation for Promotion of Cancer Research [Kyoto; H.S.]. S. Ogawa is a recipient of JSPS Core-to-Core Program, A. Advanced Research Networks. We thank Mai Nakazawa, Maki Nakamura, Takeshi Shirahari, Ai Takatsu, and Atsuko Ryu for technical assistance. We thank iLAC, Co., Ltd. for sequencing support. We thank the Centre for Anatomical, Pathological and Forensic Medical Research, Kyoto University Graduate School of Medicine, for preparing microscope slides. We also thank all staffs at the Endoscopy Unit, Kyoto University Hospital for endoscopic sampling.

Author contributions

N.K., K.Y., H.Marusawa, T.C., H.S., and S.Ogawa designed the study. M.U., K.Kawada, S.N., A.Yokoyama, S.Y., M.M., T.Horimatsu, S.Okamoto, H.O., Y.Sakai, H.I., and H.N. provided specimens. T.K., Y.T., T.Sakurai, H.H., and S.H. performed histological analysis. N.K., Y.I., Y.F., and S.K.K. performed sample preparation. M.S., E.S., and T.Sato performed sequencing. N.K., Y.O., Y.Shiozawa, T.Y., and Y.N. performed mutation calling, validated the results, and analysed copy number alterations, mutational signature, and clonal dynamics. N.K., Y.Shiraishi, K.C., H.T., and S.M. performed bioinformatics analysis. N.K., K.A., T.Hirano, Y.K., Y.W., K.Kataoka, M.M.N., A.Yoda, H.Miyoshi, M.M.T., J.O., R.I., Y.U., and O.T. performed *in vitro* functional analysis. N.K., N.G., A.K., K.Kataoka, and T.M. performed *in vivo* functional analysis. N.K., A.Yokoyama, Y.N., H.Makishima, and S.Ogawa prepared the manuscript.

Competing interests

The authors declare no competing financial interests.

DATA AND CODE AVAILABILITY STATEMENT:

Data availability

All the WES and RNAseq data have been deposited in the European Genome-Phenome Archive under accession numbers EGAS00001003801 and EGAS00001003802, respectively. Data for Figures and Extended Data Figures are available as Source Data. All other data are available from the corresponding author on reasonable request.

EXTENDED DATA FIGURE LEGENDS:

ED Figure 1 | Study design.

a,b,c, Summary of sampling types, histology of samples, subjects, and sequencing platforms are shown for single crypts (**a**), bulk crypts (**b**), and colorectal cancers (**c**). **d,** Summary of the method for determining expanded clones in epithelial samples from UC individuals.

ED Figure 2 | Sampling methods.

a, For preparation of epithelial samples, after treating intestinal mucosa in 20 mM EDTA in PBS at 4°C for 20 minutes, epithelium was dissociated from lamina propria. Then, a single crypt or epithlium (bulk crypts) was collected. **b,** For geographical clone tracking of a cluster of single crypts in colon mucosa, acrylic adhesive was put on the mucosal surface, followed by a treatment with PBS with 20 mM EDTA. After peeling off the epithelium upside down, single crypts attached on the adhesive were collected one by one under a stereomicroscope with their positions being recorded (top). To isolate crypts from UC individuals, after colorectal mucosa were treated with 20 mM EDTA in PBS, a gentle pressure was applied to the whole mucosal tissue from both sides to push up crypts, from which each single crypt was isolated in a similar manner (bottom). **c,** For single-crypt sequencing, genomic DNA was extracted from single-crypts by direct cell lysis and heat denaturation and split into two aliquots, each of which was subjected to whole-genome amplification (WGA) using phi29 DNA polymerase. One of two WGA products was subjected to WES, and the other was stocked for validation.

ED Figure 3 | Mutations detected in single crypts.

a,b, Representative VAF histograms of somatic mutations detected by WES in single crypts from non-UC (**a**) and UC (**b**) individuals. **c,** Frequencies of driver mutations among single crypts from non-UC (n=43) and UC (n=58) individuals. Genes showing significant difference in mutation frequency between non-UC and UC crypts are indicated by asterisks (two-sided Fisher's exact test with Benjamini–Hochberg adjustment; $q < 0.1$). **d,** Landscape of driver mutations is depicted for single crypts (n=58) from UC non-dysplasia (middle). Number of mutations in each single crypt is depicted on the top. The disease duration and age of subjects are also shown on the bottom.

ED Figure 4 | Phylogenetic trees in microscale colon epithelia.

a-d, Phylogenetic structure of single crypts clustered in 0.25-4 mm^2 epithelia obtained from two non-UC individuals of 84-year-old male (CR40 (**a**)) and 88-year-old female (CR43 (**b**)), and two UC patients, 56-year-old male (HCM28 (**c**)) and 42-year-old female (UC3 (**d**)). Phylogenetic trees of clones (right or bottom) are projected onto the positions of crypts, as indicated in images of specimens (left or top). Colours of branches in trees correspond to those of circles for the positions of crypts. The estimated height of UC onset is indicated on the basis of the mutation rate in non-UC crypts **(Fig.1a)**. The estimated age of the most recent common ancestral crypt is also indicated in UC individuals (**c,d**). Branches containing driver mutations are indicated. **e,** Histograms of the number of mutations (or height) of branchpoints in phylogenetic trees are plotted for non-UC (n=3, top) and UC individuals (n=3, bottom).

ED Figure 5 | Mutations detected in colon epithelia from non-UC and UC individuals.

a, Mutant cell fractions (MCFs) of mutations detected with WES of solitary bulk crypt samples from non-UC (n=136) and UC (n=133) colon epithelia. Samples belonging to each subject are separated by blank columns. MCFs of driver mutations are shown by coloured triangles as indicated. Disease state is indicated by colour as shown on the top. Colon segments of sampling, mutation status of driver genes, number of mutations, UC duration, and age of subjects are shown at the bottom. **b, c,** Average number of mutations (**b**) and average value of maximum MCFs (**c**) detected by WES of bulk crypt samples within each individual is plotted against UC duration. Non-UC individuals are plotted at zero on x-axis. Regression lines with (**b**) or without (**c**) assuming zero intercept are shown with R^2 value. *P* value is calculated for the significance of correlation between UC duration and average value of maximum MCF by using an *F* test (**c**). **d, e,** Average number of mutations (**d**) and average value of maximum MCFs (**e**) of bulk crypt samples from cecum to transvers colon (right-side) and descending colon to rectum (left-side) within the same individuals is plotted in UC individuals (n=22). Two-sided paired *t*-test.

ED Figure 6 | Functional analysis of driver genes in UC mucosa.

a, Distribution of driver gene mutations in UC non-dysplasia (58 solitary single crypts and 399 bulk crypts, top) and CAC (n=99, bottom) samples. Amino acid sequences of different species around hotspot mutations of *HNRNPF* are shown. Completely conserved amino acids across all species are indicated by asterisks. **b,** Expression levels of *Nfkbiz* in colon epithelial organoids (*n* = 3 biological replicates) treated with indicated stimulation for 1 or 24 hours. Data is normalized by the levels of PBS (control) and shown as mean ± SD. Asterisks indicate $p < 0.05$ between indicated stimulations and PBS (two-sided Student's *t*-test). PBS, phosphate buffered saline; OSM, oncostatin M; LTα1β2, lymphotoxin alpha1/beta2; ODN, CpG oligodeoxynucleotides. **c,** Immunoblot analysis of HCT116 cells expressing empty, Myc-tagged wild type (WT), or indicated ZC3H12A mutants upon stimulation with IL-1β for 0, 0.5, and 4 hours, using anti-Myc and β-actin antibodies. **d,** Immunoblot analysis of ZC3H12A, Flag-tagged dominant negative BTRC(ΔF) and β-actin in HeLa cells expressing doxycycline-inducible BTRC(ΔF) upon stimulation with IL-17A or IL-1β. Data is representative of two independent experiments. For gel source data, see **Supplementary Figure 1**.

ED Figure 7 | Clonal expansion in UC rectums.

a, Expansion of clones within the rectum from the remaining 4 out of 6 UC patients analysed, of which 2 are presented in **Fig. 2d**. All the representations follow those in **Fig. 2d**. Driver mutations including those found in the 2 samples in **Fig. 2d** are summarized in **Supplementary Table 9**. **b,** Size of area occupied by each clone and total number of mutations estimated by PyClone for each clone detected in the UC rectums. The highest pathological grade within each clone is indicated by colour. **c-h,** Phylogenetic trees of representative clones found in two or more rectum samples surgically obtained from 6 UC patients. Imputed clonal structures are depicted in violin plots at the bottom, in which the estimated cellular prevalence of imputed mutational clusters (vertical axis) are shown as the distribution of posterior probabilities (width of the violin plots) calculated according to PyClone model. Colours correspond to those in the associated branches of trees. The estimated timepoint of UC onset are indicated on the basis of the mutation rate in non-UC crypts **(Fig. 1a)**. Branches containing driver mutations are also indicated.

Extended Data 8 | Effect of *Nfkbiz* on clonal expansion in a chronic DSS-colitis model.

a, Experimental schedule of deletion of $R26^{\mathrm{LacZ/wt}}Nfkbiz^{\mathrm{wt/wt}}VilCre^{ER}$ (control) and $R26^{LacZ/wt}Nfkbiz^{fl/fl}VilCre^{ER}$ (*Nfkbiz* cKO) mice with tamoxifen (TAM) administration, followed by induction of DSS colitis. **b,e,** Images of whole-mount X-gal staining of the large intestine from control (top) and *Nfkbiz* cKO (bottom) mice after tamoxifen administration without **(b)** or with **(e)** following DSS treatment. **c,f,** Fraction of LacZ-positive areas in the middle to distal colon before **(c)** or after **(f)** induction of DSS-colitis according to genotypes. Two-sided Mann-Whitney *U* test. **d,** Density of LacZ-positive areas (/cm^{2}) measured for indicated LacZ-positive area size is plotted for control and *Nfkbiz* cKO mice before induction of DSS-colitis. **g,** Colon length (left) and histological score (right) after chronic DSS-colitis for the evaluation of severity of colitis. Two-sided Mann-Whitney *U* test. **h,** Relative weight change of *Nfkbiz* cKO and control mice during three courses of DSS-colitis. Data is shown as mean ± SD. **i,** Significant enrichment of NFKBIZ- (left) and IL-17 pathway-associated genes (right) in the down-regulated genes in rectum mucosa from *Nfkbiz* cKO compared with those in control mice 4 days after initiation of DSS administration. Data was generated from n=3 biologically independent animals for both genotypes. *P* value is described as a two-sided nominal *P* value calculated by gene set enrichment analysis (GSEA).

Extended Data 9 | Mutational profiles and copy-number alterations in non-UC and UC patients.

a, Landscape of driver mutations in CAC samples (n=99) and UC dysplasia (n=13). Genes with asterisks are drivers in CACs **(Methods)**. Copy number abnormalities of known oncogenes are depicted only for CACs from the Kyoto cohort. **b,** Frequencies of driver mutations among clones identified in UC non-dysplasia (n=183) vs. Japanese sCRCs (n=525). Asterisks indicate *q* < 0.1 (two-sided Fisher's exact test with Benjamini–Hochberg adjustment). **c,** Number (middle) and relative frequency (bottom) of mutations allocated to signatures A and B in samples of single crypts from non-UC individuals (n=43), non-dysplasia (n=241; 58 single crypts and 183 clones), dysplasia (n=13), and non-hypermutated CAC (n=92). Information about disease state, pathology, and sample type is indicated on the top. **d,** Comparison of the number of mutations detected by WES in clones identified in non-dysplasia, dysplasia samples, and CAC samples without hypermutation. Two-sided Mann-Whitney *U* test. **e,** Mutational signature of colon epithelium. **f,** Fraction of Signature B mutations in single crypts from non-UC non-dysplasia (ND), and UC ND, UC dysplasia (D), and non-hypermutated CAC containing ≥ 10 mutations. Two-sided Mann-Whitney *U* test. **g,** Box plots of fractions of genomic regions showing CNAs in clones identified in non-dysplasia (n=183), dysplasia samples (n=13), and CAC samples (n=34) from UC individuals. Two-sided Mann-Whitney *U* test. **h**, Colour-gradient maps of CNAs and uniparental disomies (UPDs) as detected by WES data are shown for clones identified in non-dysplasia (n=183), dysplasia (n=13), and colitis-associated cancer (n=34) from UC individuals. Pathological grade, fractions of genomes showing CNAs and UPDs, and TP53 mutation status are also plotted (top panels). In all box plots, median, first and third quartiles, as well as outliers, are indicated with whiskers extending to the furthest value within 1.5 of the interquartile range.

Extended Data Figure 10 | Effect of *NFKBIZ* knock-out in colon cancer cell lines and an AoM-DSS colorectal tumour model.

a, An experimental strategy for enrichment analysis using CRISPR/Cas9 **(Methods)**. **b,** Experimental schedule of induction of CAC in *Nfkbiz*$^{fl/fl}$*VilCre* (*Nfkbiz* cKO) and *Nfkbiz*$^{fl/fl}$ (control) mice with AOM, followed by three cycles of DSS. **c,** Box plots of tumour numbers according to tumour size in diameter. Two-sided Mann-Whitney *U* test. **d,** Colon length (left) and histological score (right) for the evaluation of inflammation in AOM-DSS model. No significant difference was seen between *Nfkbiz* cKO and control mice (two-sided Mann-Whitney *U* test). **e,** Relative weight change of mice during three courses of DSS administration in AOM-DSS tumour model. Data is shown as mean±SD. **f,** Histological images of HE staining (left), and immunohistochemistry for β-catenin (middle) and Ki-67 (right) of colon tumours in *Nfkbiz* cKO and control mice, which were indistinguishable between these two genotypes. Representative images are shown from n=3 biologically independent samples for both genotypes.

Figure 1

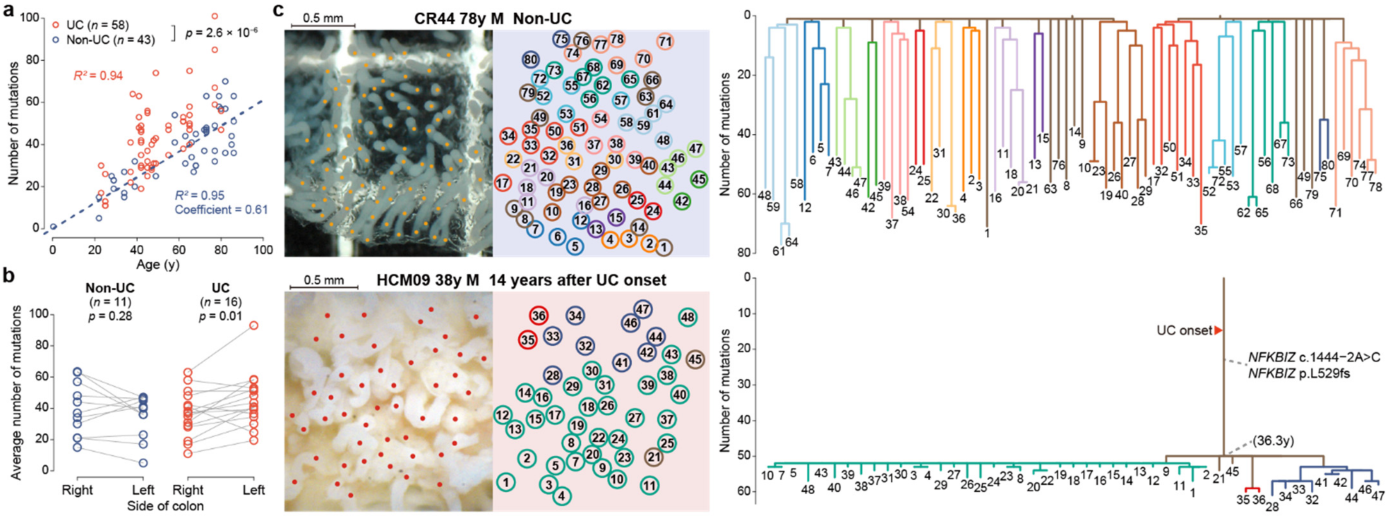


Figure 2

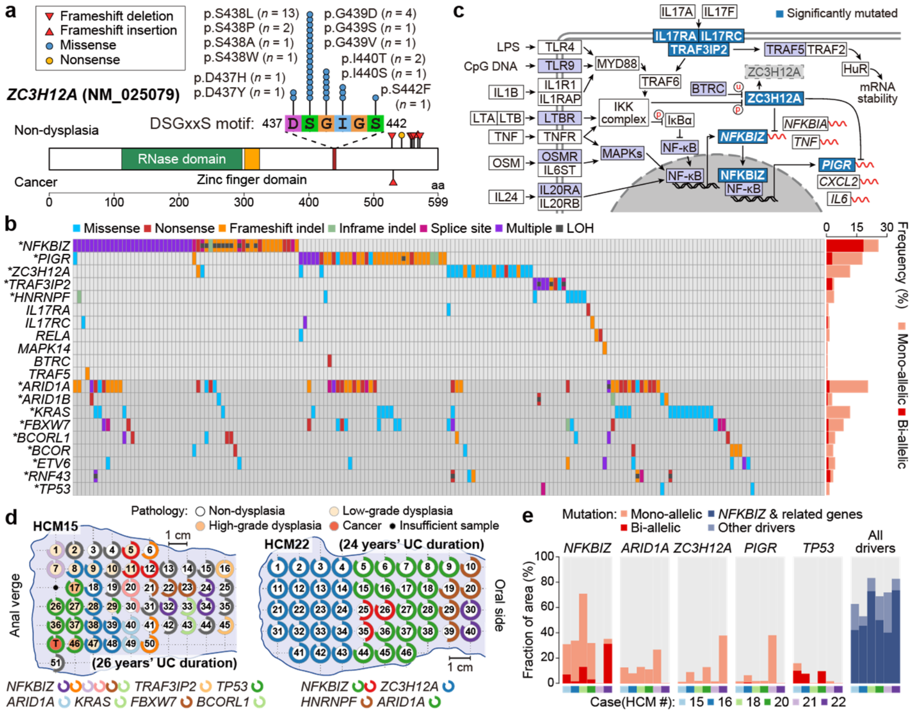


Figure 3

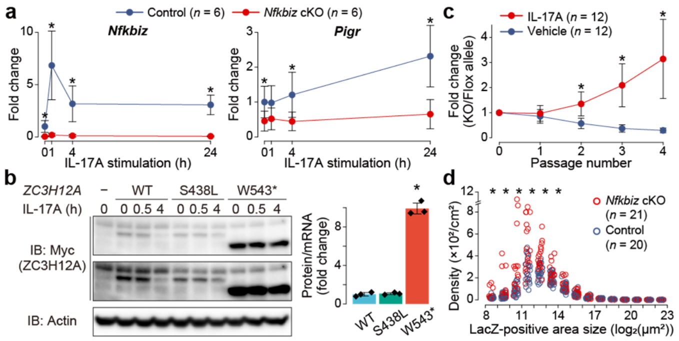
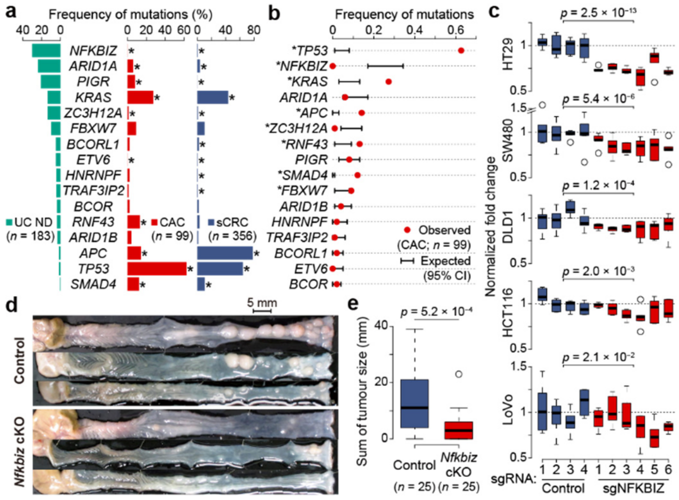
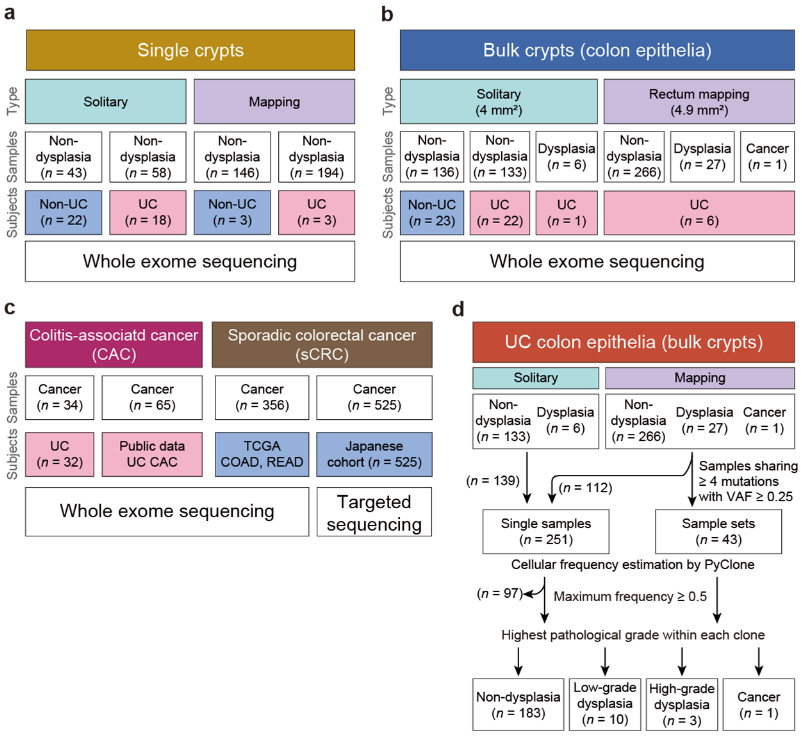


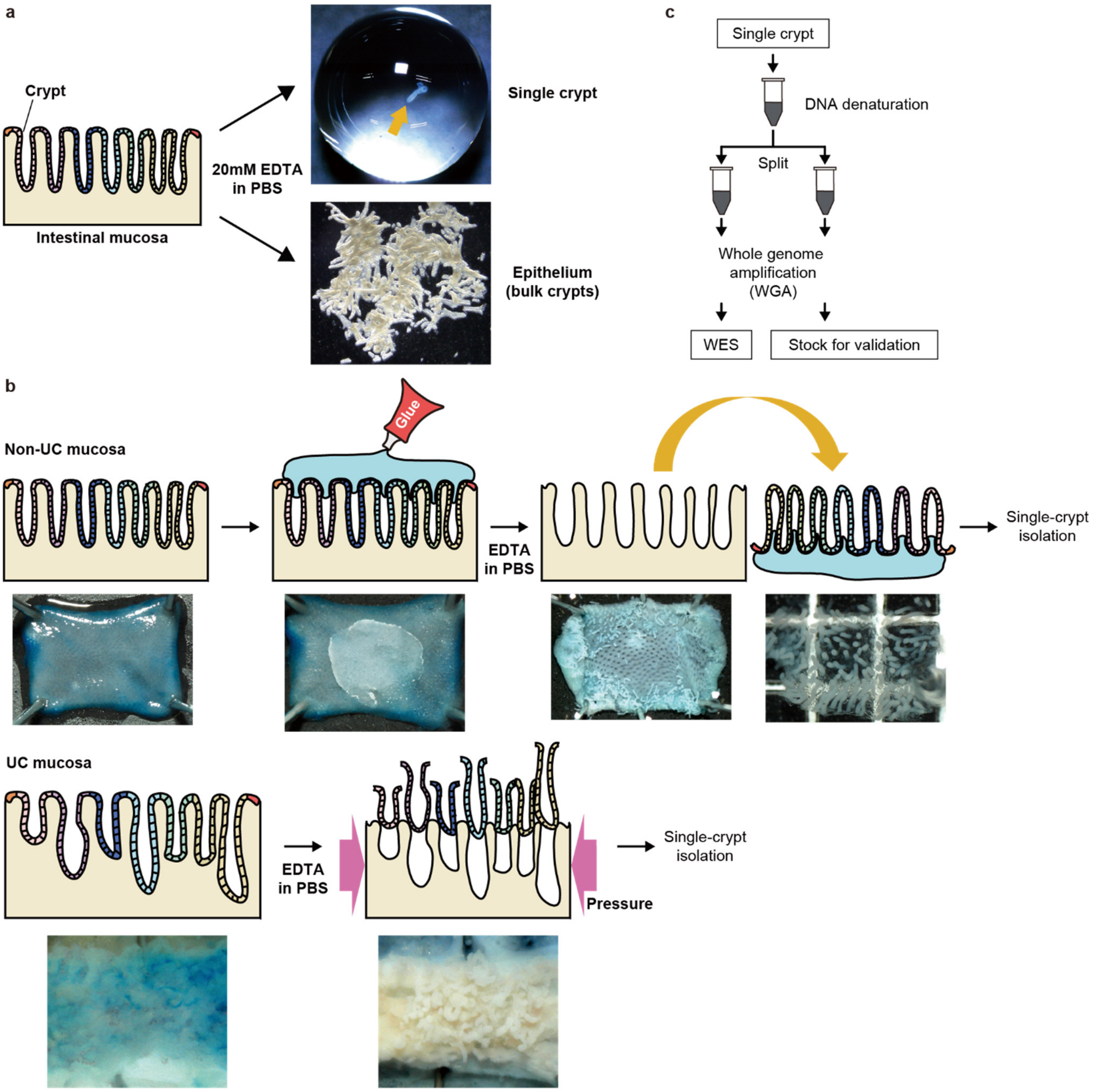
Figure 4



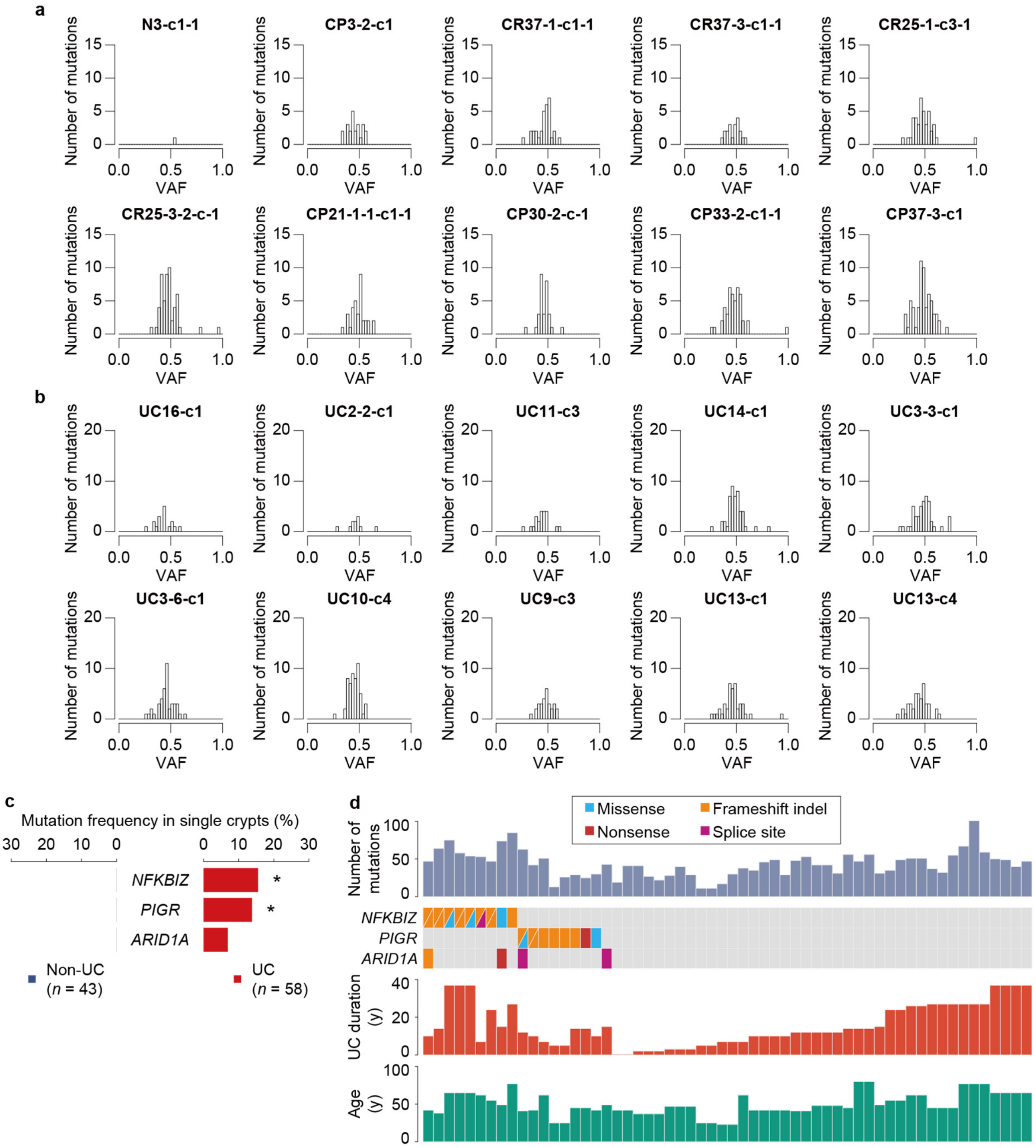
Extended Data Figure 1


a
Single crypts
Type
Solitary
Mapping
Samples
Non-dysplasia (n = 43)
Non-dysplasia (n = 58)
Non-dysplasia (n = 146)
Non-dysplasia (n = 194)
Subjects
Non-UC (n = 22)
UC (n = 18)
Non-UC (n = 3)
UC (n = 3)
Whole exome sequencing
b
Bulk crypts (colon epithelia)
Type
Solitary (4 mm^2)
Rectum mapping (4.9 mm^2)
Samples
Non-dysplasia (n = 136)
Non-dysplasia (n = 133)
Dysplasia (n = 6)
Non-dysplasia (n = 266)
Dysplasia (n = 27)
Cancer (n = 1)
Subjects
Non-UC (n = 23)
UC (n = 22)
UC (n = 1)
UC (n = 6)
Whole exome sequencing
c
Colitis-associatd cancer (CAC)
Sporadic colorectal cancer (sCRC)
Samples
Cancer (n = 34)
Cancer (n = 65)
Cancer (n = 356)
Cancer (n = 525)
Subjects
UC (n = 32)
Public data UC CAC
TCGA COAD, READ
Japanese cohort (n = 525)
Whole exome sequencing
Targeted sequencing
d
UC colon epithelia (bulk crypts)
Solitary
Mapping
Non-dysplasia (n = 133)
Dysplasia (n = 6)
Non-dysplasia (n = 266)
Dysplasia (n = 27)
Cancer (n = 1)
(n = 139)
(n = 112)
Samples sharing ≥ 4 mutations with VAF ≥ 0.25
Single samples (n = 251)
Sample sets (n = 43)
Cellular frequency estimation by PyClone
(n = 97)
Maximum frequency ≥ 0.5
Highest pathological grade within each clone
Non-dysplasia (n = 183)
Low-grade dysplasia (n = 10)
High-grade dysplasia (n = 3)
Cancer (n = 1)

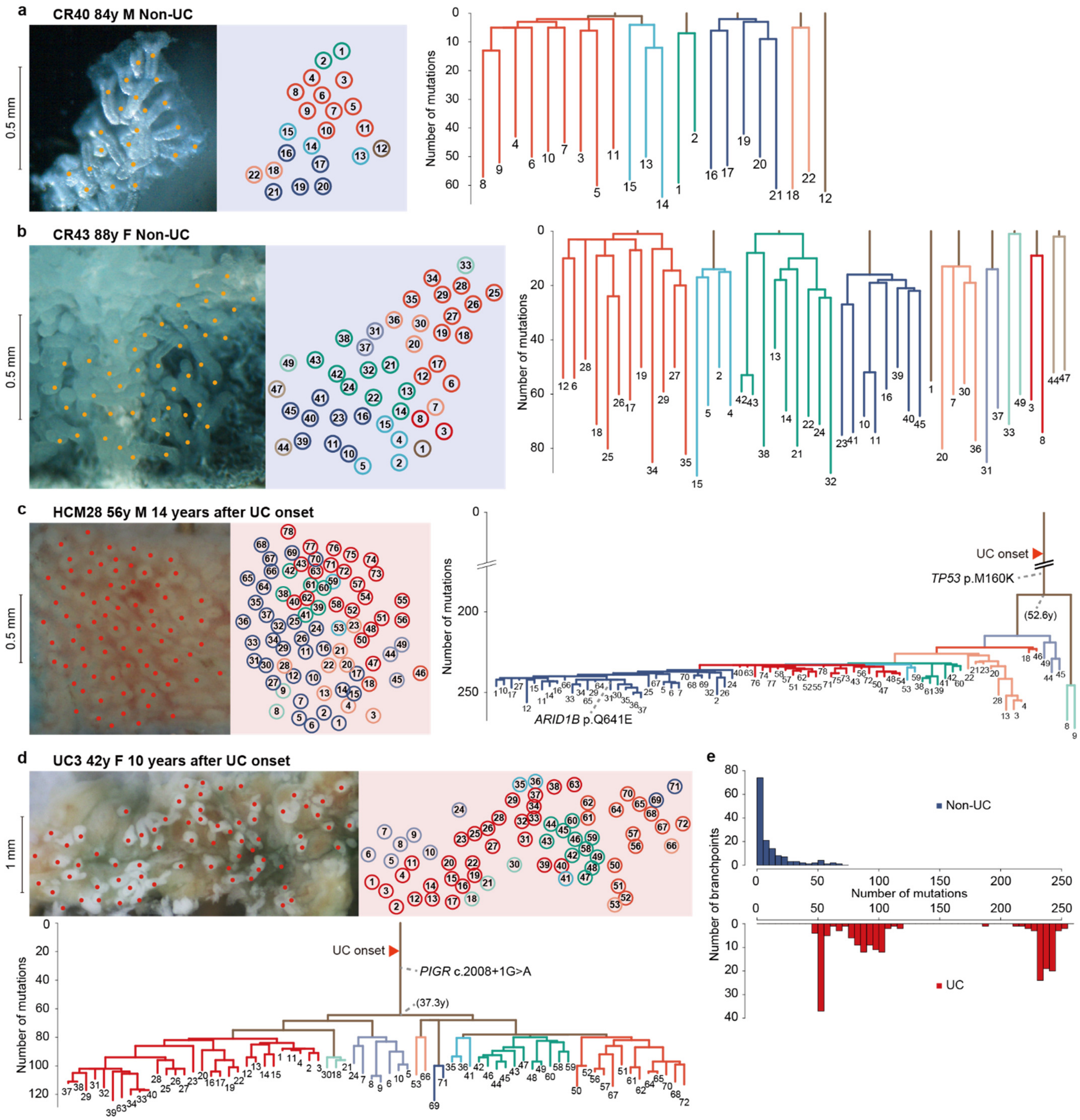
Extended Data Figure 2


a
Crypt
Intestinal mucosa
20mM EDTA in PBS
Single crypt
Epithelium (bulk crypts)
c
Single crypt
DNA denaturation
Split
Whole genome amplification (WGA)
WES
Stock for validation
b
Non-UC mucosa
Glue
EDTA in PBS
Single-crypt isolation
UC mucosa
EDTA in PBS
Pressure
Single-crypt isolation

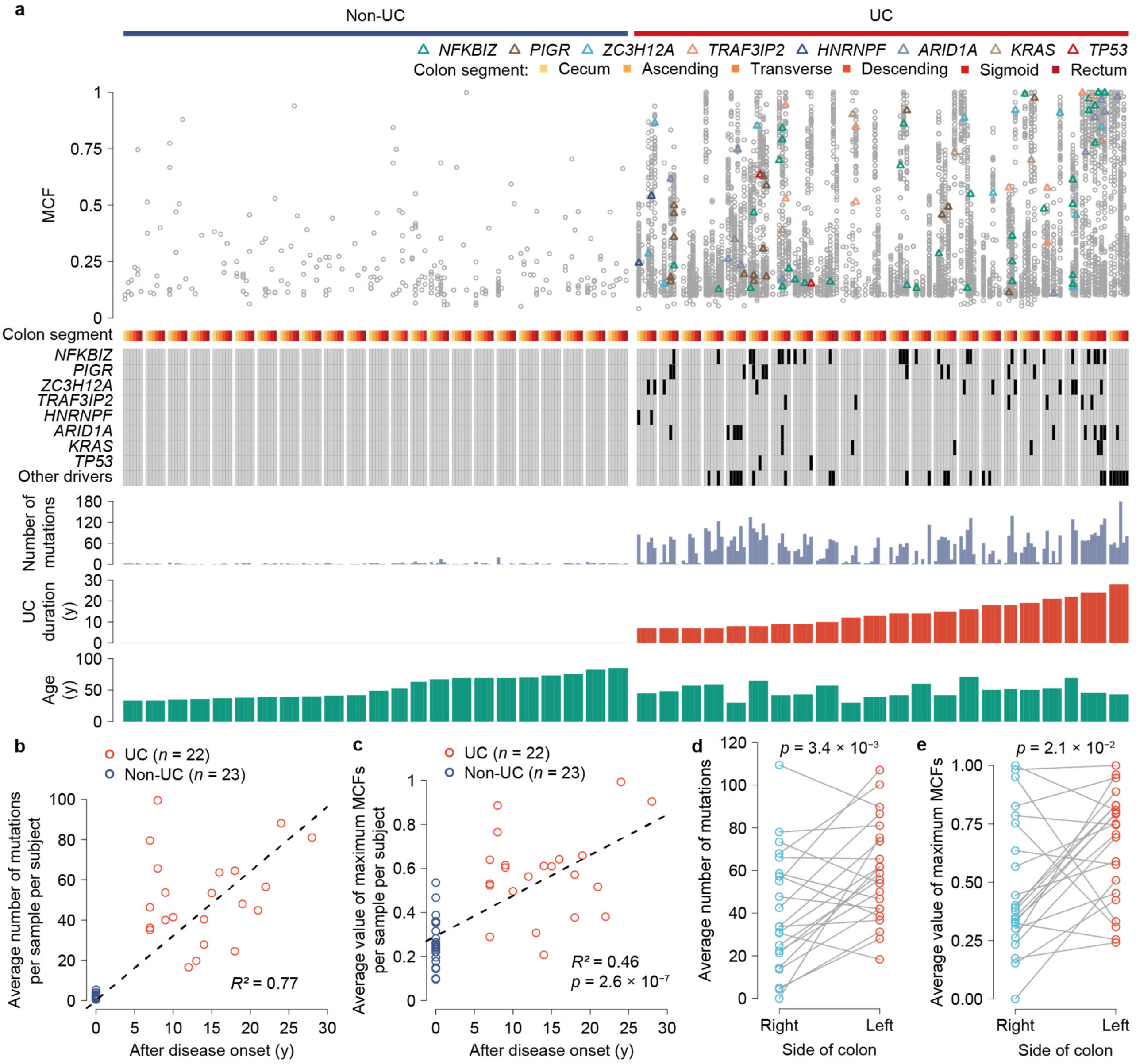
Extended Data Figure 3



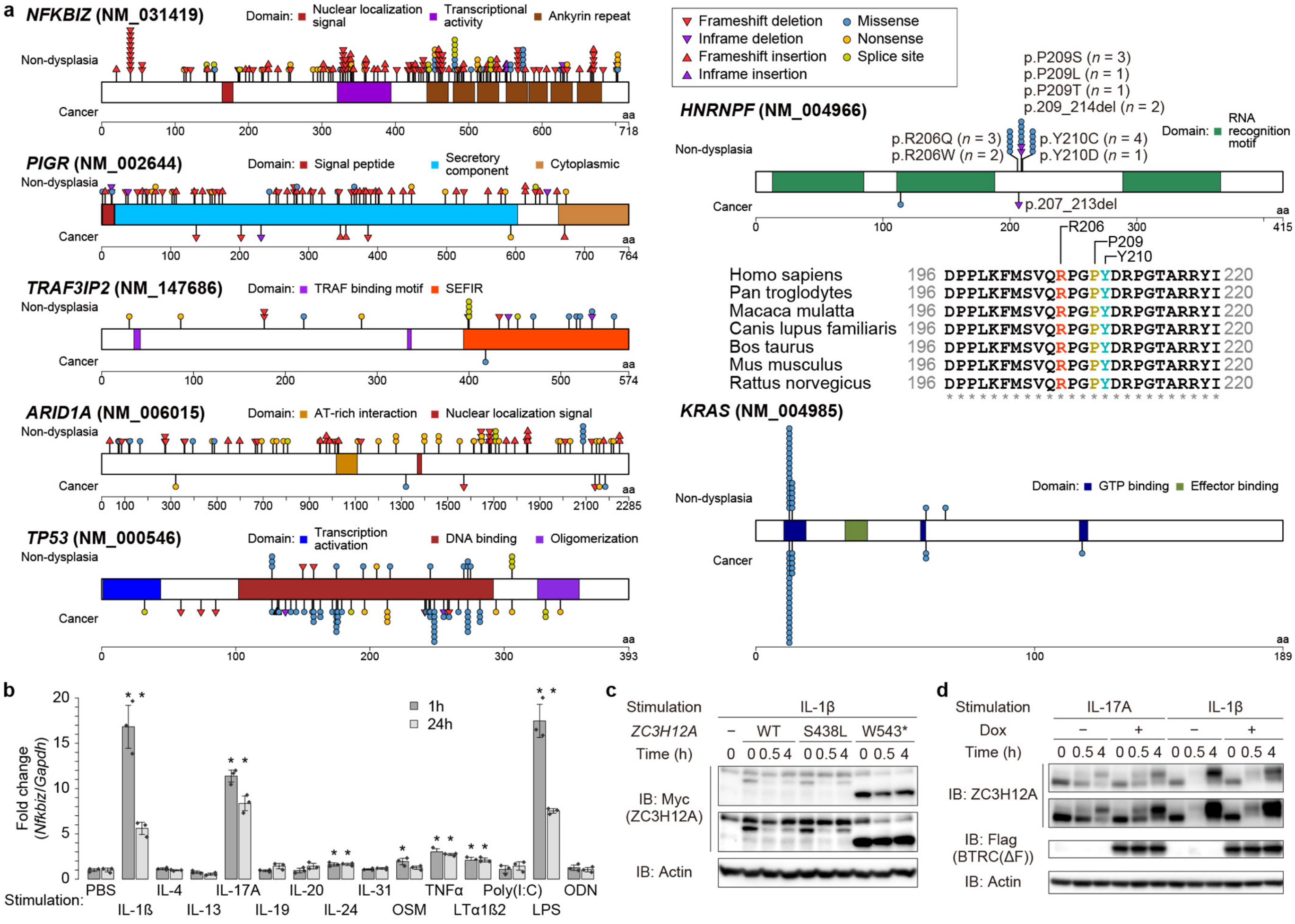
Extended Data Figure 4



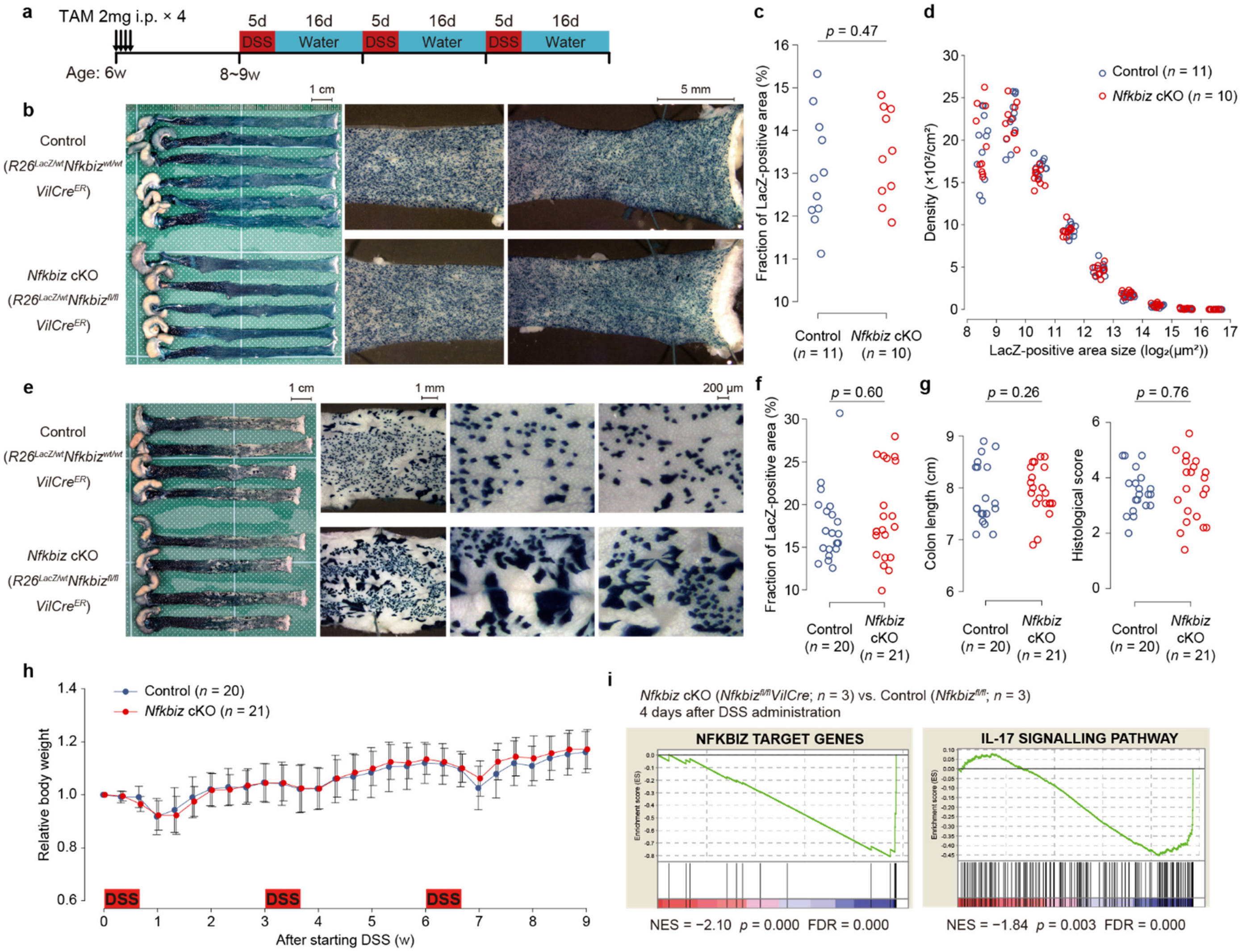
Extended Data Figure 5



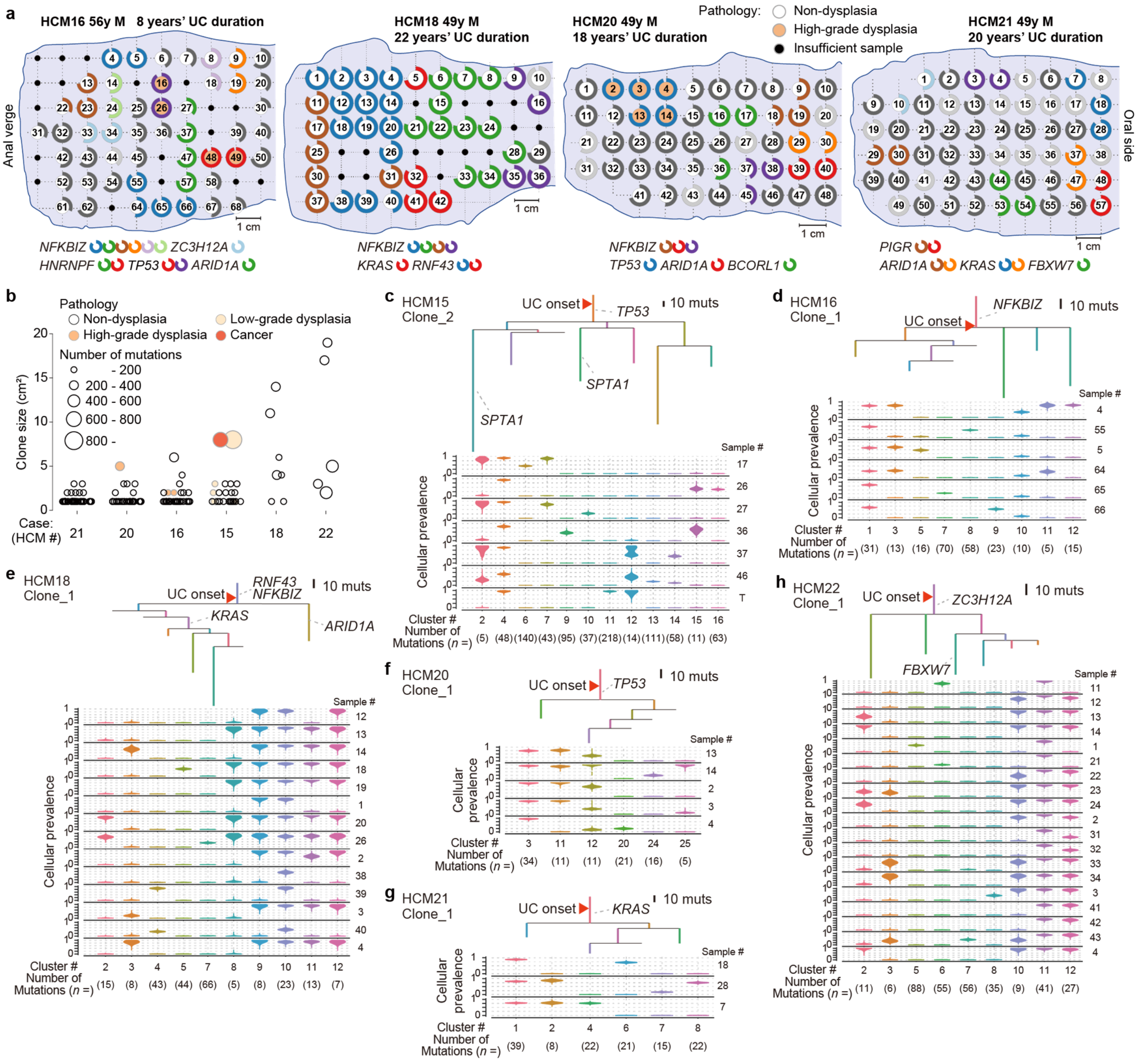
Extended Data Figure 6



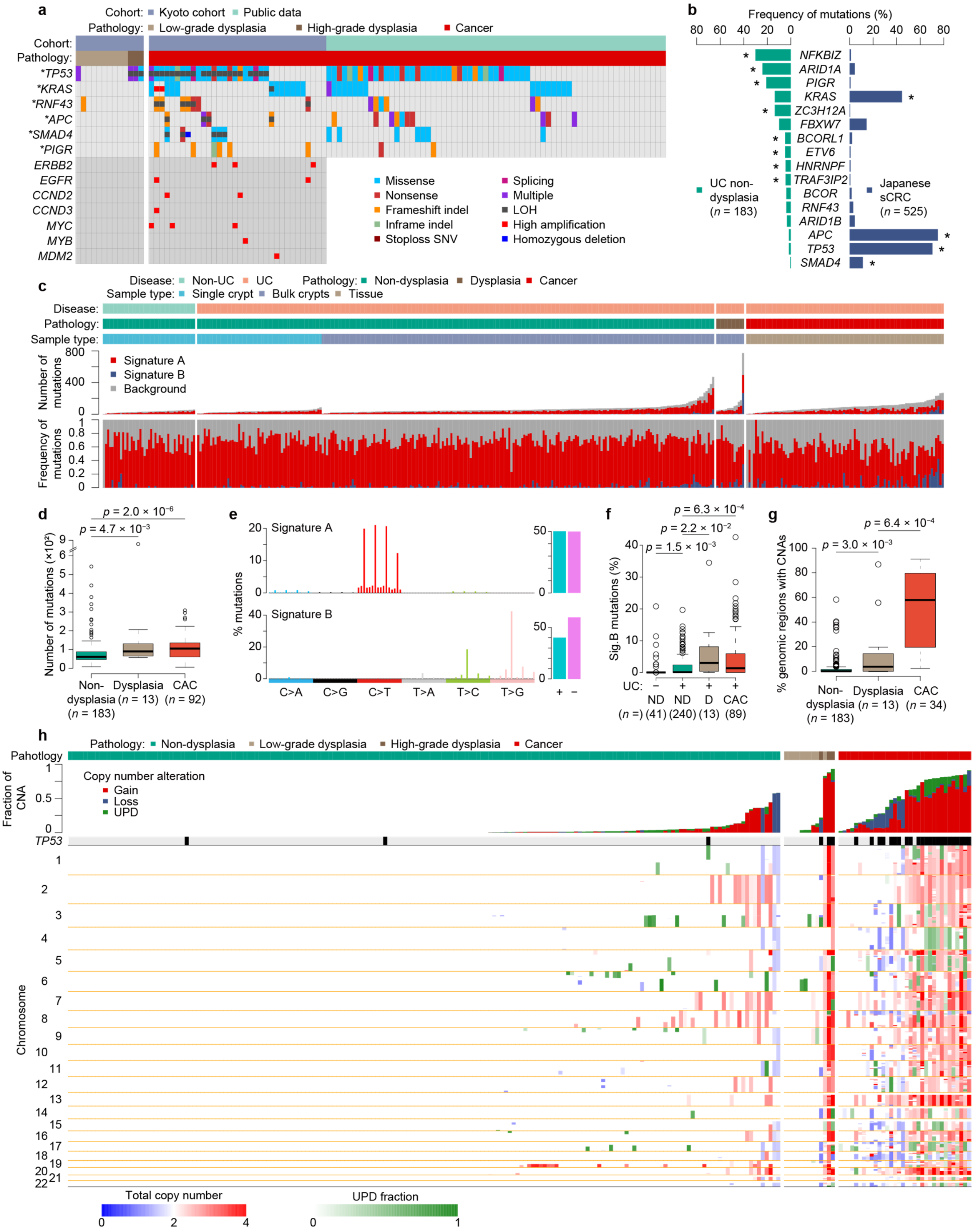
Extended Data Figure 7



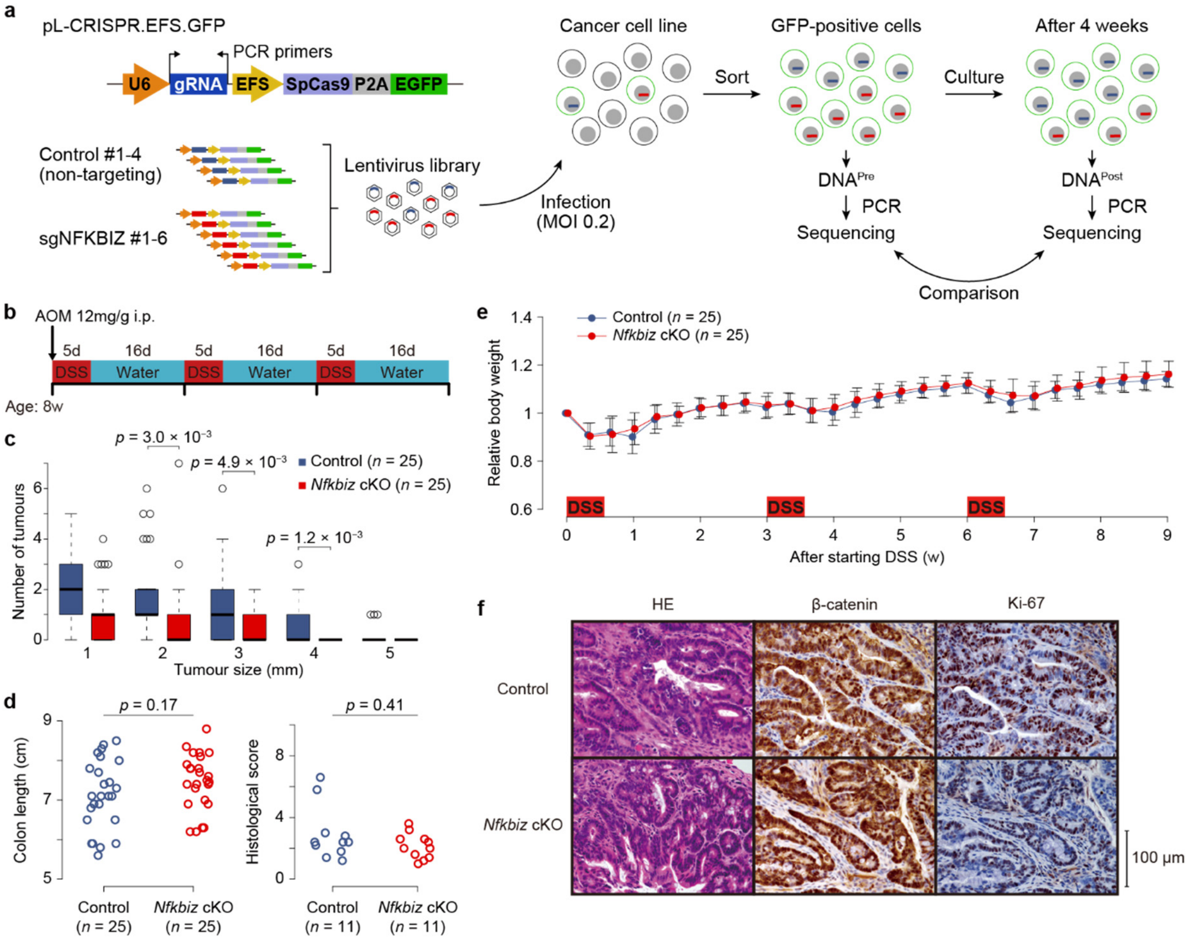
Extended Data Figure 8



Extended Data Figure 9

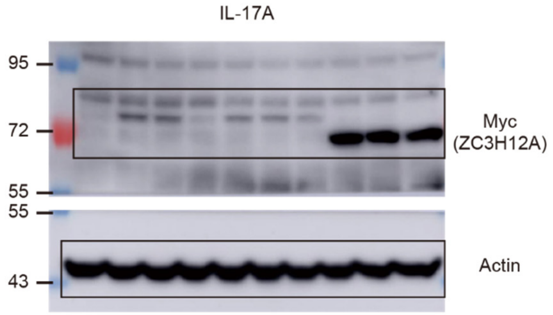


Extended Data Figure 10



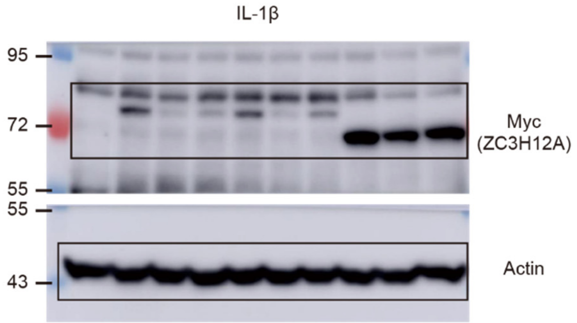
Supplementary Figure 1 | Uncropped scans with size marker indications.

Fig. 3b


IL-17A
95
72
55
55
43
Myc
(ZC3H12A)
Actin

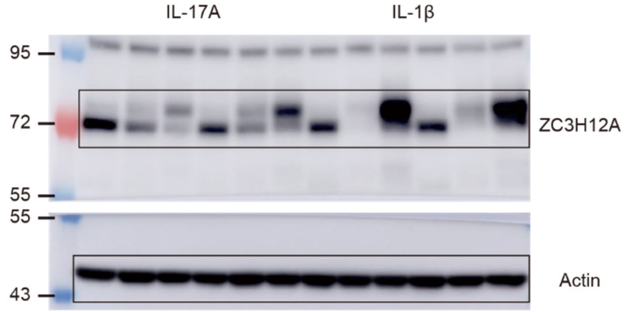
Samples were run on the same gel and blotted for Myc and Actin.

ED Fig. 6c

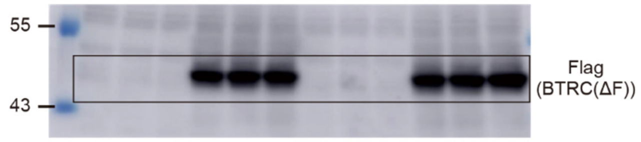

IL-1β
95
72
55
55
43
Myc
(ZC3H12A)
Actin

Samples were run on the same gel and blotted for Myc and Actin.

ED Fig. 6d


IL-17A
IL-1β
95
72
55
55
43
ZC3H12A
Actin

Samples were run on the same gel and blotted for ZC3H12A and Actin.


55
43
Flag
(BTRC(ΔF))

The same amount of samples was run on another gel and blotted for Flag.

Download https://www.dropbox.com/s/zsjealedi5uyd7a/Supplementary_Tables.xlsx

Supplementary Tables (Supplementary_Tables.xlsx)

Supplementary Table 1. Characteristics of subjects.

Supplementary Table 2. Sample information.

Supplementary Table 3. Summary of solitary single crypts according to subjects.

Supplementary Table 4. Results of validation for mutation calls.

Supplementary Table 5. Summary of sample sets for clone size estimation by PyClone.

Supplementary Table 6. Significantly mutated genes analysed by dN/dS analysis and MutSigCV.

Supplementary Table 7. Protein truncating mutations of cytokine receptors and pattern recognition receptors in UC non-dysplasia.

Supplementary Table 8. Summary of somatic mutations in targeted capture deep sequencing of Japanese sCRCs.

Supplementary Table 9. Summary of driver mutations detected in geographical mapping of UC rectums.

Supplementary Table 10. Summary of mutation frequency in CACs and 95% CI of frequency expected from mutated fraction in UC epithelium.

Supplementary Table 11. Primers for PCR.

Supplementary Table 12. sgRNA sequences used for *NFKBIZ* knockout.

Supplementary Table 13. Summary of ANOVA for estimation of mutation rate for single crypts from UC patients.