Supporting Information

Salk et al. 10.1073/pnas.0909428106

Materials and Methods

Specimens. For initial studies, biopsies were harvested from the colons of 19 UC patients: 11 with cancer or high-grade dysplasia (Progressors) and 8 without (Non-Progressors). Samples were obtained either at colonoscopy (3 UC Non-Progressors, 1 UC Progressor) or immediately following colectomy (5 UC Non-Progressors, 7 UC Progressors) in accordance with approved Human Subject's Guidelines at the University of Washington. Colectomy specimens came from surgeries performed for intractable symptoms, emergent hemorrhage, sepsis or the presence of high-grade dysplasia or cancer found during colonoscopy. Progression status was determined by histological assessment of 44–144 biopsies depending on whether samples were obtained at colonoscopy or colectomy. Biopsies were evenly spaced along the colon with diagnosis for dysplasia made according to DMSG consensus criteria (1) by a GI pathologist (author MPB). Selection of colons was based on surgical availability. All samples were frozen at −70 °C in Minimal Essential Medium with 10% DMSO until use. The mean patient age within each group was: Non-Progressors: 45.0 (range: 23–59) and Progressors: 41.2 (range: 31–61). The mean year-duration of UC within each group was: Non-Progressors: 10.8 (range: 0.25–20), Progressors: 15.5 (range: 4–29). Additional patient information is listed in Fig. S2. For the whole-colon mapping study, biopsies taken at time of colectomy were obtained from a UC Progressor of unknown age and disease duration. In all cases, colonic epithelial cells were isolated from a ≈9 mm^2 portion of each biopsy by EDTA shakeoff; this method yields 90% enrichment for epithelial cells (2). DNA from the epithelial cell isolate and residual stroma of each biopsy was extracted by silica filtration column (Qiagen) and quantified by Nanodrop UV spectroscopy (Thermo Scientific). Stromal DNA was not included in the latter two mapping studies An adjacent portion of each biopsy was fixed and sectioned for histology.

Polyguanine Tract-Length Genotyping. A BLAST search (3) of NCBI build 36 of the human genome was used to identify polyguanine tracts of 12 or more residues in length. PCR primers were designed to flank each tract (Fig. S1) and examined in silico (http://genome.ucsc.edu) to confirm predicted amplification of a unique, non-coding product between 90 and 250 base-pairs in length. Oligonucleotides were synthesized by Operon or Applied Biosystems. All forward primers integrated a 5′ fluorescent dye (6-FAM, NED or HEX), and reverse primers contained GTTTCTT on the 5′ end to minimize genotyping artifacts due to the terminal deoxynucleotidyl transferase activity of Taq DNA polymerase (4). PCR reactions (5 μl) containing 4 ng genomic DNA were prepared in robotically-loaded 384 well plates, and amplification was carried out for 42 cycles using Taq DNA polymerase (Qiagen). PCR fragments were resolved with an ABI PRISM 3730xl Genetic Analyzer. All genotypes were repeated in duplicate. Electropherograms were analyzed with ABI GeneMapper 4.0 software in a fully blinded fashion using the PeakSeeker approach (5). For each marker, the allele lengths most commonly observed in a set of samples derived from a single patient was defined as the consensus genotype. Genotypes differing from the consensus were considered mutant. If both replicates for a sample produced robust signals and yielded the same genotype, or if only one replicate generated good signal and was equal to the consensus genotype, the result was recorded. If only one replicate produced sufficient signal and suggested a mutant genotype, or if there was a discrepancy between genotypes of the two replicates, the allele lengths were resolved using three additional PCR amplifications. If neither replicate produced a robust signal, the genotype was marked as "X" and not repeated. Some amplification failures were sporadic; others were systematically related to certain PCR amplicons. Progressor and Non-Progressor DNA samples were handled identically in that they were amplified in intermingled wells on the same PCR plates and genotyped blindly.

1. Riddell RH, et al. (1983) Dysplasia in inflammatory bowel disease: standardized classification with provisional clinical applications. *Hum Pathol* 14(11):931–968.
2. Rabinovitch PS, et al. (1999) Pancolonic chromosomal instability precedes dysplasia and cancer in ulcerative colitis. *Cancer Res* 59(20):5148–5153.
3. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25(17):3389–3402.
4. Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* 20(6):1004–1006, 1008–1010.
5. Thompson JM, Salipante SJ (2009) PeakSeeker: a program for interpreting genotypes of mononucleotide repeats. *BMC Res Notes* 2:17.

Polyguanine amplicons

Amplicon #	Forward Sequence (5'-3')	Fwd 5' Dye	Reverse sequence (5'-3')	Accession #	Chromosome
22	cctatattcccagctacagctacac	6-FAM	GTTTCTTgggtatatagtgatagtggtttgtttc	NW_926839	17
41	tcttttgactctaagtcccttagcc	NED	GTTTCTTgtttatagtccgctttttgtaaagg	NW_925495	13
18	atgatggctccagagtcatagaac	6-FAM	GTTTCTTgtgcagatcaggaaggagaattg	NW_926839	17
26	gactgacactgttgtaataccaagg	6-FAM	GTTTCTTggtttcaaacattacaagatcaagg	NW_926018	16
27	ctgatgagggacaggaatctcac	6-FAM	GTTTCTTatgacccagggacaggtacag	NW_926018	16
30	ggagtattgctaggagggagttttc	HEX	GTTTCTTcgctatatgggtagtcactatctgg	NW_925884	15
47	tttggttaaggccctaaatttgaac	NED	GTTTCTTtttctgcatttttatagtgctttcc	NW_925351	12
34	gttataaaagatgcaactgctcagg	HEX	GTTTCTTgtagctactgattctggttcctctg	NW_925561	14
45	aaggtctgagataagctccagaatc	NED	GTTTCTTaccttagagttcggtgtcatgaag	NW_925351	12
54	ctaagtgttaaggacacagactgaagg	NED	GTTTCTTgagaccttacaggaacagaagaatatc	NW_924606	10
2	tcatcaggttactaggcaatattagg	HEX	GTTTCTTctctgtccctgaccaggtctac	NW_927705	X
25	cacactccttggtgacagtgc	6-FAM	GTTTCTTgtcaggtcagtccgtggttc	NW_926018	16
32	ggaggttgactaaggatctacacag	HEX	GTTTCTTctactacagggacatcacacacaac	NW_925884	15
46	ccgttattaaaaagtctcacgtttg	NED	GTTTCTTatatctaacctctcctcaggtttcc	NW_925351	12
21	tacccaggtgtaagatcttgaaaag	6-FAM	GTTTCTTaggaaacctctactcatgctgaaag	NW_926839	17
52	cagctaatttttctgtttttagtacagg	NED	GTTTCTTtgcagtcaagaacccatcac	NW_925106	11
36	gggcattcaggaccactagg	HEX	GTTTCTTgttcagagcgtctcttggtttc	NW_925561	14
89	tagtcaccttgttcagcacctaatatc	HEX	GTTTCTTgtgactatccactctttgccttg	NW_922217	4
78	caaagagtgaaacagactatcgacttc	HEX	GTTTCTTaacctttagatttacagaaaaattgagc	NT_079595	7
84	aggtgtctgagaataaagaagatgaag	HEX	GTTTCTTatggattcctggtgagatgttg	NW_922984	6
88	aattttcagtctctgagtgtgatcc	HEX	GTTTCTTcatttgcgagcaatttctctttag	NW_922217	4
58	gtaagtaaatcaatgaatgtggttgtg	6-FAM	GTTTCTTataaatttttattggattttcgtttgg	NW_924606	10
64	gtaatcaccatcaatttggcaatttac	6-FAM	GTTTCTTgactaagggaggagaatcactagaac	NW_924062	9
75	catgagttcaattgtttttatttttagc	HEX	GTTTCTTcatttctgagataagggttcaaatg	NT_079595	7
66	acatgtacattcagttcactgttaagc	6-FAM	GTTTCTTtagctttgttctagttttgtgtgtgtg	NW_923929	8
87	tacatgaaattctcaatgattacaacg	HEX	GTTTCTTaagatctattccatcccattgactc	NW_922217	4
57	ccgaatcttaaattgaaaacacaaag	6-FAM	GTTTCTTtttttagtagaaatggggtttcacc	NW_924606	10
104	agttacgacaatcaaaaatgtctctg	6-FAM	GTTTCTTgagatgcctagaccactgattctc	NW_927128	1
81	gtgaactgtgtttctgtcactacactc	HEX	GTTTCTTtacaaaaatcatggtttagttttctcc	NW_922984	6
103	gggcagtattaaaaactatagaatacccc	6-FAM	GTTTCTTtacactcttgtgcattttcctttttc	NW_921585	2
83	cagtgtctcattcatctttgtcattc	HEX	GTTTCTTcaaaaactacaaaatgtcttaatggag	NW_922984	6
105	ttaccttaacattcagtcttcctcttg	6-FAM	GTTTCTTtagatatgccacttttgtcatctacag	NW_927128	1
74	taacaagggaatgtaaaggaacttatg	6-FAM	GTTTCTTttatttagtccagatttaatgacaaagg	NT_079595	7
102	ttggtattctattatagcagcctgaac	6-FAM	GTTTCTTcattacacatacttattaccaccaggac	NW_921585	2
107	ctctcatgacctagctaaaaatgattc	6-FAM	GTTTCTTgcccagacttttatttcttattttgtc	NW_927128	1

NCI Bethesda Panel panel MSI amplicons

Amplicon #	Forward Sequence (5'-3')	Fwd 5' Dye	Reverse sequence (5'-3')	Genbank #	Chromosome
D17S250	ggaagaatcaaatagacaat	HEX	gctggccatatatatatttaaacc	177030	17
BAT26	tgactacttttgacttcagcc	6-FAM	aaccattcaacatttttaaccc	9834505	2
D2S123	aaacaggatgcctgccttta	6-FAM	ggactttccacctatgggac	187953	2
BAT25	tcgcctccaagaatgtaagt	6-FAM	tctgcattttaactatggctc	9834508	4

Fig. S1. Primer sequences and marker location for polyguanine homopolymer loci and Bethesda panel MSI markers. Capitalized bases represent a nonhomologous 5′ ''pigtail'' sequence added to limit genotyping artifacts.

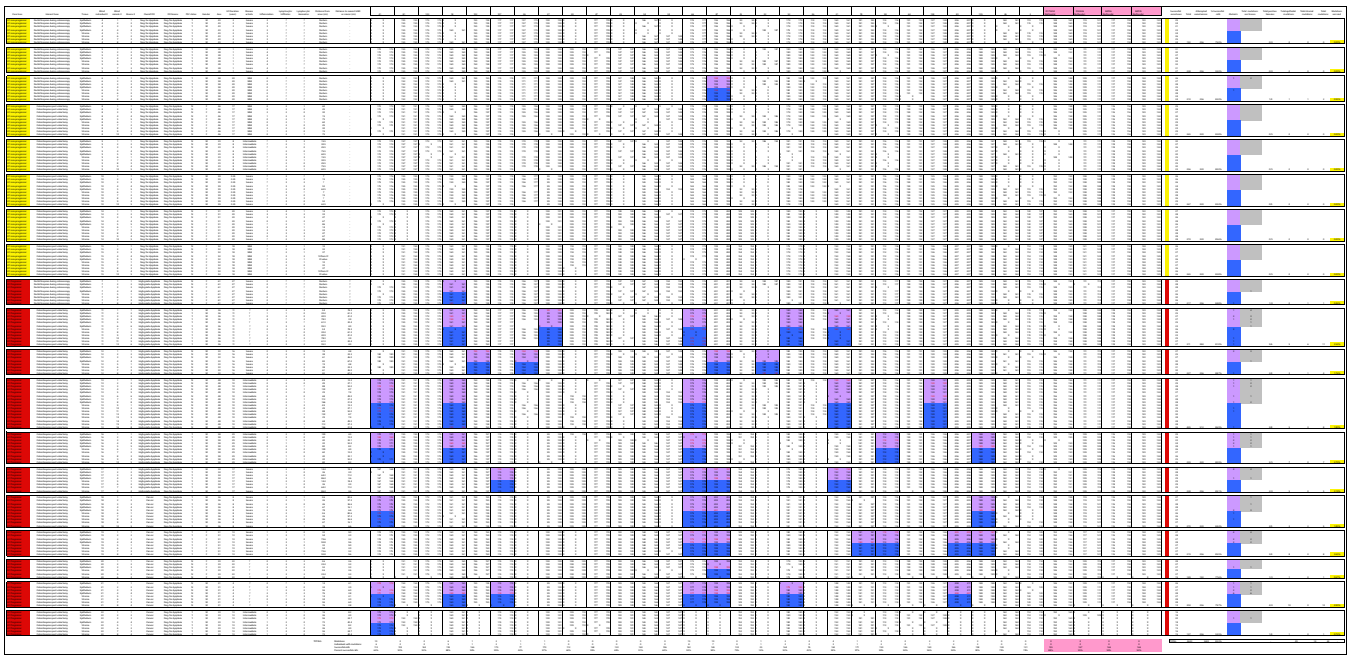
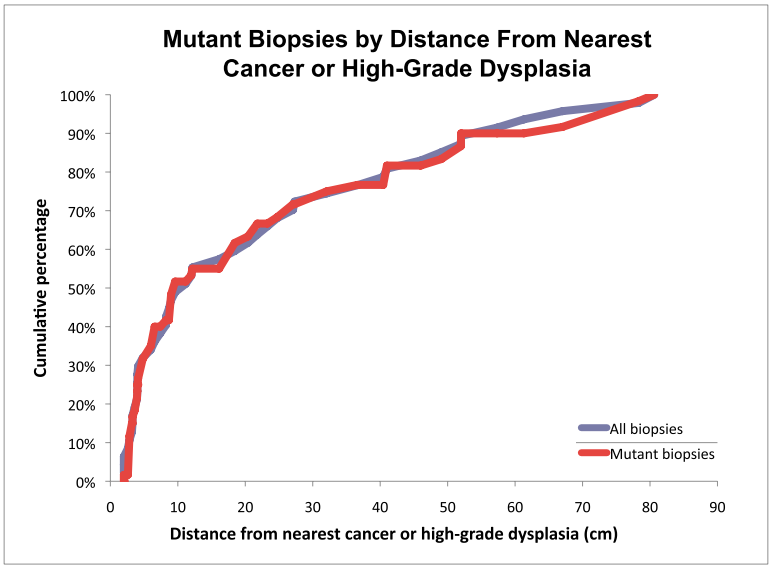
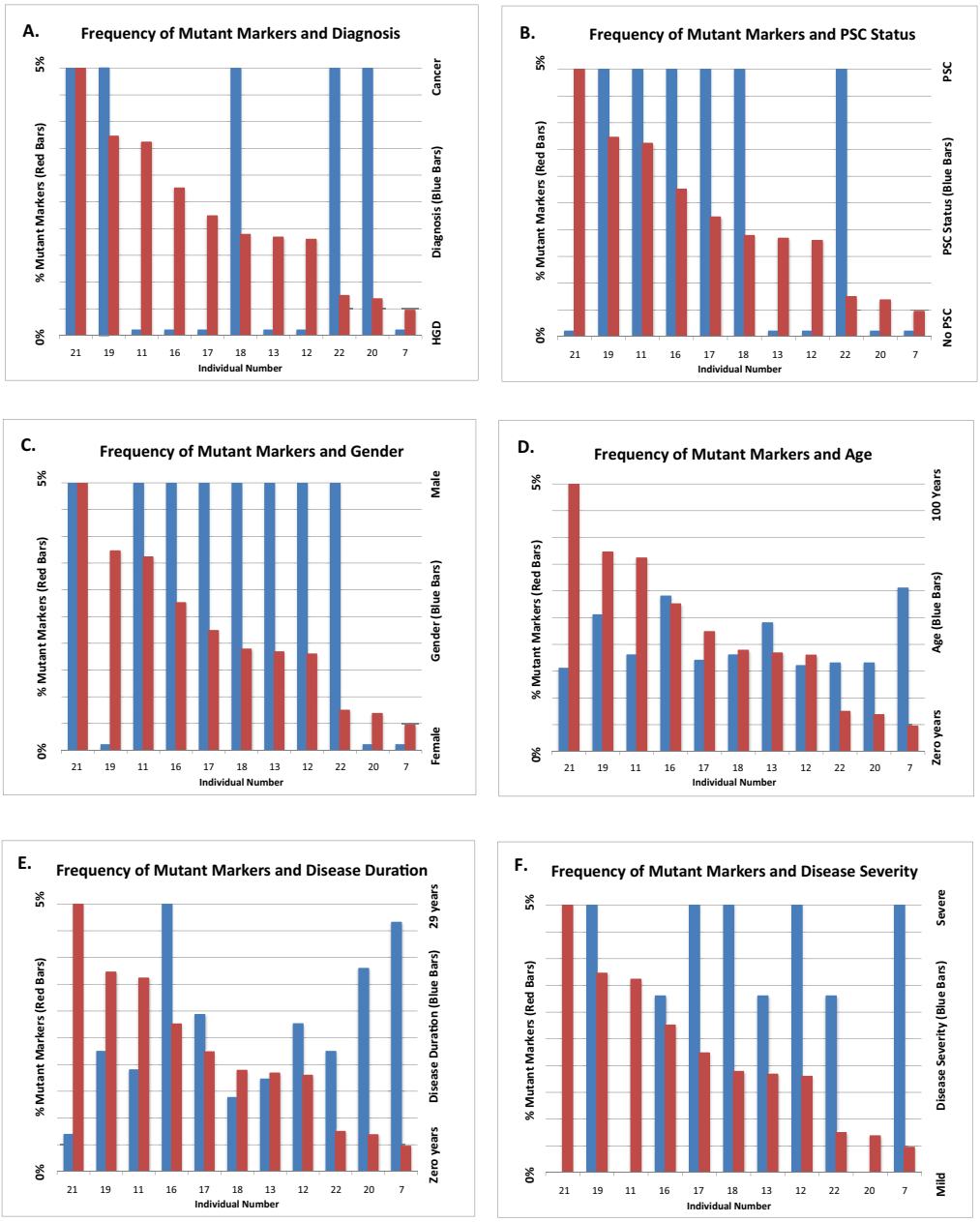


Fig. S2. Complete genotyping data set. Horizontal boxes represent individual cases grouped by disease status: UC Non-Progressor (yellow), UC Progressor (red). Biopsies were separated into epithelial and stromal fractions and genotyped separately. Rows within each case indicate specific fractions of individual biopsies. Relevant clinical data for each individual/biopsy is given. Vertical columns making up the bulk of the table indicate the length of alleles for each polyguanine marker. The four rightmost marker columns (pink headings) correspond to nonpolyguanine Bethesda Panel MSI markers. ''X'' indicates unsuccessful genotyping. Genotypes differing from the consensus genotype for that individual are shown in red. Markers where at least one mutant genotype was identified in an individual are highlighted; epithelial samples are shown in violet, stromal samples in blue. Summary data by case and by marker are listed at the far right and bottom, respectively. Genotyping was performed under fully blinded conditions.



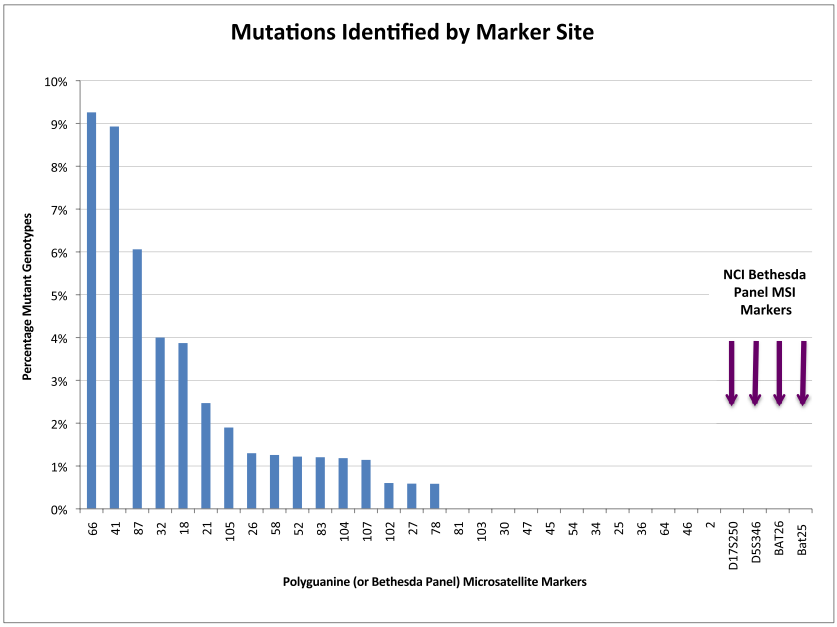
Distance to nearest HGD or cancer	Number of mutations per biopsy	Cumulative number of mutations	Cumulative percentage of biopsies	Cumulative percentage of mutations
2.0	0	0	2.1%	0.0%
2.0	0	0	4.3%	0.0%
2.0	1	1	6.4%	1.7%
2.6	0	1	8.5%	1.7%
2.8	6	7	10.6%	11.7%
3.2	2	9	12.8%	15.0%
3.3	0	9	14.9%	15.0%
3.3	1	10	17.0%	16.7%
3.6	1	11	19.1%	18.3%
3.9	2	13	21.3%	21.7%
4.0	1	14	23.4%	23.3%
4.0	1	15	25.5%	25.0%
4.0	0	15	27.7%	25.0%
4.1	1	16	29.8%	26.7%
4.8	3	19	31.9%	31.7%
6.0	2	21	34.0%	35.0%
6.6	3	24	36.2%	40.0%
7.4	0	24	38.3%	40.0%
8.2	1	25	40.4%	41.7%
8.2	0	25	42.6%	41.7%
8.7	0	25	44.7%	41.7%
9.0	4	29	46.8%	48.3%
9.6	2	31	48.9%	51.7%
11.2	0	31	51.1%	51.7%
12.0	1	32	53.2%	53.3%
12.2	1	33	55.3%	55.0%
16.1	0	33	57.4%	55.0%
18.4	4	37	59.6%	61.7%
20.4	1	38	61.7%	63.3%
21.8	2	40	63.8%	66.7%
23.3	0	40	66.0%	66.7%
24.7	1	41	68.1%	68.3%
27.1	2	43	70.2%	71.7%
27.3	0	43	72.3%	71.7%
32.1	2	45	74.5%	75.0%
36.6	1	46	76.6%	76.7%
40.4	0	46	78.7%	76.7%
41.0	3	49	80.9%	81.7%
46.0	0	49	83.0%	81.7%
49.0	1	50	85.1%	83.3%
52.0	2	52	87.2%	86.7%
52.0	2	54	89.4%	90.0%
57.3	0	54	91.5%	90.0%
61.3	0	54	93.6%	90.0%
67.0	1	55	95.7%	91.7%
78.3	4	59	97.9%	98.3%
80.6	1	60	100.0%	100.0%

Fig. S3. Relationship between location of histologically negative biopsies with clonally expanded mutations and nearest region of advanced histological disease. The cumulative percentage of biopsies (blue) and mutant genotypes (red) are plotted as a function of increasing distance from nearest biopsy with cancer or HGD. The fraction of mutant genotypes identified tracks closely with the fraction of all biopsies analyzed suggesting that the probability of identifying a clonally expanded mutant genotype is independent of distance from dysplasia.



Patient	Cancer vs HGD	PSC?	Gender	Age	Duration (yr)	Severity	% Mutants
21	Cancer	No	Male	31	4	Unknown	4.90%
19	Cancer	Yes	Female	51	13	Severe	3.65%
11	HGD	Yes	Male	36	11	Unknown	3.54%
16	HGD	Yes	Male	58	29	Moderate	2.70%
17	HGD	Yes	Male	34	17	Severe	2.19%
18	Cancer	Yes	Male	36	8	Severe	1.85%
13	HGD	No	Male	48	10	Moderate	1.80%
12	HGD	No	Male	32	16	Severe	1.76%
22	Cancer	Yes	Male	33	13	Moderate	0.73%
20	Cancer	No	Female	33	22	Unknown	0.67%
7	HGD	No	Female	61	27	Severe	0.46%

Fig. S4. UC progressors by prevalence of mutant genotypes detected and additional clinical parameters. For each panel, red bars indicate the percentage of genotypings within an individual differing from consensus, and blue bars indicate an additional clinical feature: (*A*) highest grade histological diagnosis: HGD or cancer, (*B*) presence or absence of concurrent PSC, (*C*) gender, (*D*) age at time of sampling, (*E*) duration of clinically manifest ulcerative colitis, (*F*) symptomatic severity of ulcerative colitis.



Marker	# Mutant Genotypes	# Successful Genotypes	% Mutant Genotypes
66	15	162	9.3%
41	10	112	8.9%
87	10	165	6.1%
32	1	25	4.0%
18	6	155	3.9%
21	4	162	2.5%
105	3	158	1.9%
26	1	77	1.3%
58	2	159	1.3%
52	2	164	1.2%
83	2	166	1.2%
104	2	169	1.2%
107	2	175	1.1%
102	1	166	0.6%
27	1	170	0.6%
78	1	171	0.6%
81	0	159	0.0%
103	0	163	0.0%
30	0	112	0.0%
47	0	168	0.0%
45	0	122	0.0%
54	0	160	0.0%
34	0	106	0.0%
25	0	132	0.0%
36	0	76	0.0%
64	0	166	0.0%
46	0	128	0.0%
2	0	131	0.0%
D17S250	0	155	0.0%
D5S346	0	167	0.0%
BAT26	0	166	0.0%
BAT25	0	166	0.0%

Fig. S5. Prevalence of clonally expanded mutant genotypes as a function of microsatellite marker. Among marker sites with detectable mutations, a few were positive in 5–10% of biopsy genotypings although this was <4% for the majority. In approximately half of the polyguanine markers, and all four Bethesda panel MSI markers tested, no mutant genotypes were identified.

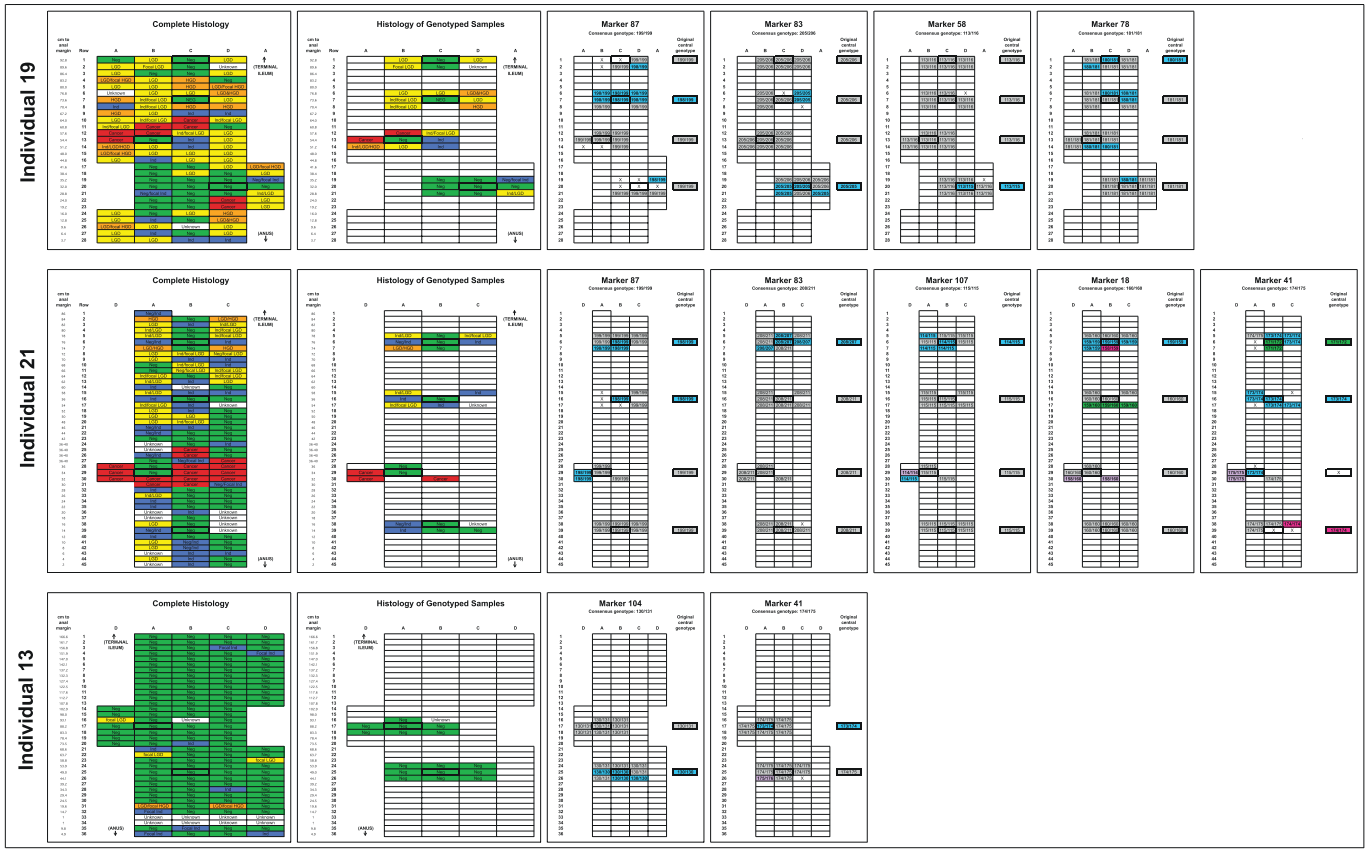


Fig. S6. Clonal patches identified by polyguanine mapping (complete data). Epithelium from biopsies surrounding each of 10 samples used in the initial study were genotyped at markers sites previously found to be mutant in a subset of the original 10 samples (outlined boxes). Longitudinally opened colectomy specimens from three individuals are diagrammed separately. Colectomy diagrams are divided into boxes representing individual biopsies measuring ≈9 mm^2, taken at evenly spaced intervals within an alphanumeric grid. The histological diagnosis from each biopsy is indicated at left: NEG, negative for dysplasia; IND, indefinite for dysplasia; LGD, low grade dysplasia; HGD, high grade dysplasia; CAN, cancer; ?, no data. The genotypes of biopsies for specific polyguanine markers are indicated at right. Empty boxes represent biopsies not genotyped, ''X'' indicates unsuccessful genotyping. Gray fields indicate biopsies with the consensus genotype for the marker, and different colors represent distinct mutant genotypes within an individual. Clustering of identical mutant genotypes in adjacent biopsies suggests large, clonally derived patches. Within each panel, the previously determined genotypes of the initial biopsies are listed to the right. Genotypes of surrounding biopsies and repeat genotypes of central biopsies are shown. Repeat testing of central biopsies uniformly obtained the same genotype as previously identified.

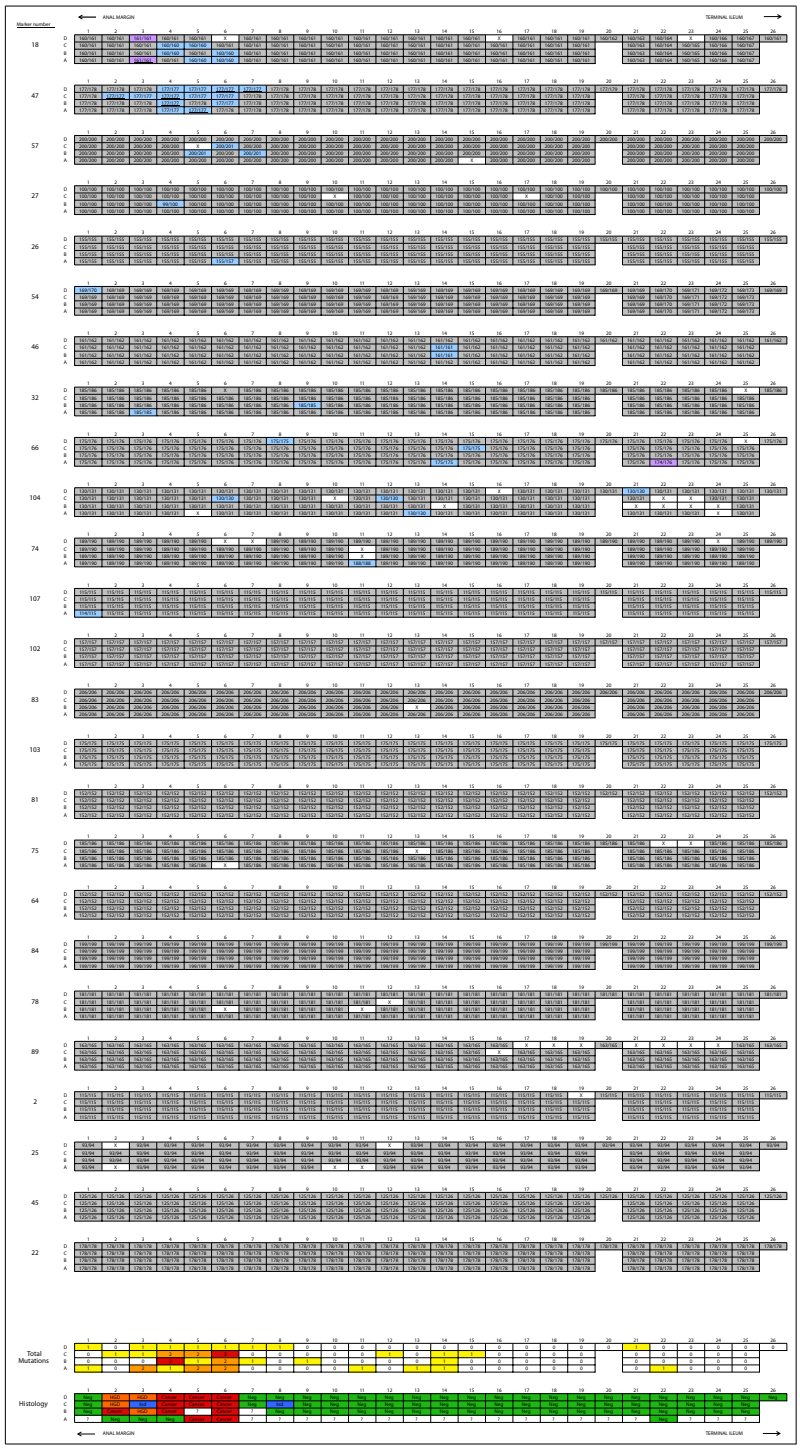


Fig. S7. Full data set from polyguanine mapping of a complete UC Progressor colon. A longitudinally opened colectomy specimen is diagramed with small boxes representing evenly spaced biopsies measuring ≈9 mm^2 within an alphanumeric grid. The histological diagnosis of each biopsy is indicated at far right: NEG, negative for dysplasia; IND, indefinite dysplasia; HGD, high grade dysplasia; CAN, cancer; ?, no data. Biopsy genotypes for various polyguanine marker are indicated in separate grids. ''X'' indicates unsuccessful genotyping. Gray fields indicate biopsies with the consensus genotype for the marker, and different colors represent distinct mutant genotypes. Underlined genotypes indicate biopsies where a mixture of consensus and mutant genotypes were observed, suggesting a mixed population of cells with different genotypes. The total number of mutant genotypes identified across all markers is reported for each biopsy (heat map). Large, clonally derived patches identified by three markers were observed near the cancer site. Numerous smaller patches were detected throughout the nondysplastic portions of the colon.

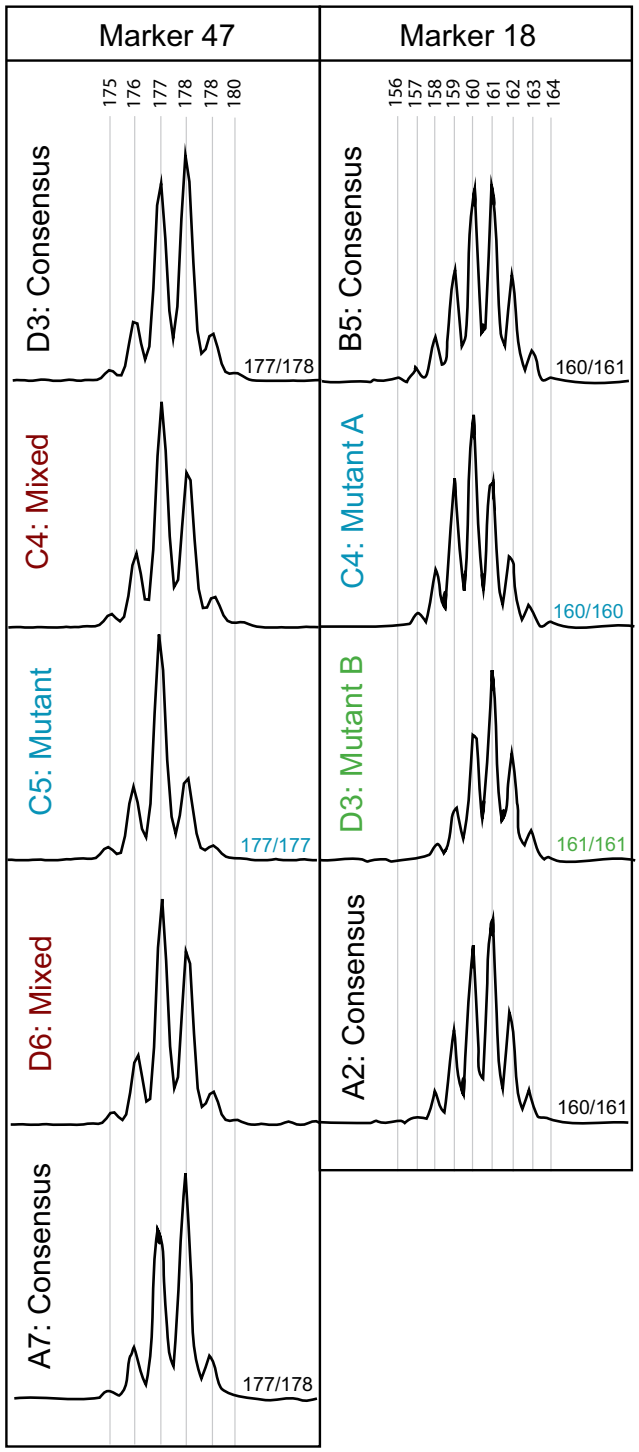


Fig. S8. Example electropherograms from a complete colon map showing consensus, mutant and mixed polyguanine tract genotypes. X-axis indicates product length (bp), y-axis represents signal intensity. Allele lengths are indicated, with mutant alleles in red. Non-indicated peaks are an artifact of PCR amplification (''stutter''). For each polyguanine marker, the ''consensus'' genotype is that most commonly observed among biopsies from a single patient. Mutant genotypes are those that differ from the consensus with respect to the length of at least one allele. ''Mixed'' genotypes reflect a combination of consensus and mutant genotypes, and likely occur when a biopsy contains a mixture of cells with both consensus and a mutant genotype. Product length scale is listed at the top of each marker column. The alphanumeric code to the left of each electropherogram corresponds to the biopsy grid position in Fig. 4 and Fig. S7. Genotype calls are indicated to the right. Genotype assignments are based on the major peak(s) of each electropherogram. Adjacent ''stutter'' peaks are an artifactual consequence of PCR amplification.