Monoallelic silencing and haploinsufficiency in early murine intestinal neoplasms

James M. Amos-Landgraf[a], Amy A. Irving[a,b], Cory Hartman[a], Anthony Hunter[a], Brianna Laube[a], Xiaodi Chen[a], Linda Clipson[a], Michael A. Newton[c,d], and William F. Dove[a,e,1]

[a]McArdle Laboratory for Cancer Research, Department of Oncology, [b]Molecular and Environmental Toxicology Center, [c]Statistics, [d]Biostatistics and Medical Informatics, and [e]Laboratory of Genetics, University of Wisconsin – Madison, Madison, WI 53706

Contributed by William F. Dove, December 16, 2011 (sent for review December 3, 2011)

Studies of tumors from human familial adenomatous polyposis, sporadic colon cancer, and mouse and rat models of intestinal cancer indicate that the majority of early adenomas develop through loss of normal function of the *Adenomatous polyposis coli* (*APC*) gene. In murine models of familial adenomatous polyposis, specifically the multiple intestinal neoplasia mouse (Min) and the polyposis in the rat colon (Pirc) rat, most adenomas have lost their WT copy of the *Apc* gene through loss of heterozygosity by homologous somatic recombination. We report that large colonic adenomas in the *Pirc* rat have no detectable copy number losses or gains in genomic material and that most tumors lose heterozygosity only on the short arm of chromosome 18. Examination of early mouse and rat tumors indicates that a substantial subset of tumors shows maintenance of heterozygosity of *Apc* in genomic DNA, apparently violating Knudson's two-hit hypothesis. Sequencing of the *Apc* gene in a sampling of rat tumors failed to find secondary mutations in the majority of tumors that maintained heterozygosity of *Apc* in genomic DNA. Using quantitative allele-specific assays of *Apc* cDNA, we discovered two neoplastic pathways. One class of tumors maintains heterozygosity of $Apc^{Min/+}$ or $Apc^{Pirc/+}$ RNA expression and may involve haploinsufficiency for Apc function. Another class of tumors exhibits highly biased monoallelic expression of the mutant *Apc* allele, providing evidence for a stochastic or random process of monoallelic epigenetic silencing of the tumor suppressor gene *Apc*.

epigenetics | genomic stability | loss of imprinting | X-inactivation

Understanding the earliest genetic, epigenetic, and environmental events in adenoma formation may illuminate the range of mechanisms of tumor initiation and subsequent pathways of progression. In most sporadic cases of colon cancer, the first genetic lesion involves the gatekeeper tumor suppressor gene *Adenomatous polyposis coli* (*APC*), leading to growth of the adenoma (1). Studies of tumors and normal epithelium from patients with familial adenomatous polyposis (FAP), heterozygous for an *APC* mutation, indicate that a single genetic lesion may itself have a biological effect—the one-hit hypothesis (2). Frank tumor formation seems to require a balance of mutant *APC* alleles—the just-right hypothesis (3). Early colonic adenomas with *APC* mutations can persist for up to a decade without further advancement (4) and may only progress to invasive metastatic disease by acquiring additional mutations or genomic instability.

Loss of function of the *APC* gene can occur through multiple alternative genetic mechanisms representing many different pathways of pathogenesis (5–7). The acquisition of some form of genetic instability may be required to accumulate sufficient mutations or imbalances for cancer (8). Reports of such instability include amplifications, deletions, gene fusions, and microsatellite instability caused by germ-line or somatic mutations in mismatch repair genes leading to errors in replication. Advanced colonic cancers in humans often exhibit chromosomal instability with large losses and gains of genetic material (9). Except for microsatellite instability (10), genomic instability is remarkably missing in most mouse models of intestinal cancer (11, 12). Homologous somatic recombination is a more subtle form of genomic instability observed in human FAP and its murine models (13, 14). Conservative somatic recombination between the centromere and the *APC* locus causes loss of heterozygosity (LOH); the resultant adenoma carries two mutant copies of *APC*/*Apc*. Is such conservative somatic recombination elevated over the whole genome in the cancer lineage?

Beyond genetic and genomic changes that trigger the initiation and progression of cancer, it is important to identify epigenetic differences associated with the cancer lineage. Epigenetic changes involve transmissible changes in phenotype that are not directly caused by changes in DNA sequence. On one hand, an epigenetic change can be caused by a developmentally programmed alteration of an upstream effector (i.e., a transcription factor) acting on downstream targets to control the cancer phenotype (14). On the other hand, epimutation—either silencing or activation—can involve a somatically heritable effect on a single allele of a gene that impacts cancer.

The work by Grady et al. (15) reported an example of silencing of the E-cadherin gene in gastric cancer reflected by 100% levels of methylation of key CpG sites within the promoter region. The work also reported that a subset of tumors exhibits 50% methylation in these sites (15). This observation may reflect monoallelic silencing in the neoplastic lineage or an admixture of normal stroma in the tumor. Programmed epigenetic change is expected to affect both alleles at a locus, whereas rare stochastic epigenetic change is expected to affect only a single allele. The heterozygous diploid condition can distinguish between these alternatives.

The involvement of epigenetic instability in cancer is shown by monoallelic changes in expression in the absence of changes in DNA sequence. The silent allele at an imprinted locus is activated rarely in colon cancer (16). This change in state implies the loss of a maintenance function. By contrast, this report involves cases in which the genomically stable multiple intestinal neoplasia (Min) mouse and polyposis in the rat colon (Pirc) rat models of familial intestinal neoplasia generate adenomas through random epigenetic silencing of the WT *Apc* allele. Our detailed analysis of these exceptional cases shows that the WT *Apc* allele is silenced and maintained in a stable, monoallelically active state without detectable changes in DNA sequence.

Results

Genomic and Epigenetic Changes at the *Apc* Locus in the Rb9-Min Mouse. To assess LOH in genomic DNA (gLOH) at the heterozygous *Apc* locus, we examined the percent contribution of each allele in gDNA from tumors arising in the Rb9-Min mouse (*Materials and Methods*). A total of 52 tumors and 10 normal adjacent tissue samples were examined for the ratio of their *Apc*

Author contributions: J.M.A.-L. and W.F.D. designed research; J.M.A.-L., A.A.I., C.H., A.H., B.L., and X.C. performed research; M.A.N. contributed new reagents/analytic tools; J.M.A.-L., L.C., and M.A.N. analyzed data; and J.M.A.-L., L.C., and W.F.D. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: dove@oncology.wisc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1120753109/-/DCSupplemental.

alleles. The ratios of *Min* to WT center near 1:1 in gDNA from the control normal tissues, signaling maintenance of heterozygosity in gDNA (gMOH). In tumors, by contrast, the contribution of the WT allele varied from 8% to 50% without a clear delineation between the gLOH and gMOH classes of tumors. In an earlier report (17), the distribution of allelic ratios had been bimodal.

Tumors were then examined for allelic biases in expression by assaying cDNAs generated from tumor RNA. Two tumors scored as gMOH yielded no detectable cDNA from the *Apc* gene, whereas cDNA from the control housekeeping gene ribonuclear protein *Rpl10a* was present (Fig. S1). Thus, the level of APC RNA in these two tumors lay below the detection limit of the RT-PCR pyrosequencing assay. For the majority of the gMOH Rb9-Min tumors, the percent contribution of WT allele in both gDNA and cDNA varied from 8% to 45%. Interestingly, in 1 of these 52 tumors, the contribution of the WT allele in cDNA was significantly reduced compared with the contribution in gDNA (Fig. 1*A*). We have designated this exceptional class of tumor as gMOH/cLOH.

Genomic and Epigenetic Changes in Colonic Tumors of the F1-Pirc Rat.

Stability of copy number and heterozygosity. The predominant early somatic genetic event in the formation of early adenomas in the Pirc ($Apc^{Pirc/+}$) rat is also somatic recombination with LOH at the *Apc* locus (13, 18). We have investigated whether these conservative events at the *Apc* locus are accompanied by genome-wide changes that affect diploid copy number. Array comparative genomic hybridization analysis was performed on DNA from two large (0.8 and 1.0 cm) colonic adenomas from two F344-Pirc animals (Fig. S2). No significant alterations in copy number were detected.

Relative copy number can be conserved in the face of widespread LOH involving homologous somatic recombination. To determine whether LOH at the *Apc* locus is accompanied by genome-wide LOH, we examined the incidence of LOH across the genome in F1-Pirc rat tumors using pyrosequencing of allelic ratios at informative SNP sites (*Materials and Methods*). We chose SNP sites located at the distal ends of chromosomes to capture all proximal somatic recombination events. We previously showed that LOH in tumors in the Pirc rat involves the short (p) arm of chromosome (Chr) 18, resulting in two copies of the mutant allele in the tumor lineage. We tested a mode of 23 tumors from each of three or more independent F1-Pirc animals, examining 13 SNP sites on seven chromosomes (1, 4, 7, 10, 15, 16, and 18q). This survey revealed that nearly all of the SNP sites maintained heterozygosity, showing allelic ratios close to 1:1, including SNPs on the long arm of chromosome 18 (Fig. S3 and Table S1). One exceptional tumor showed loss of the ACI alleles for most of the distal arm of Chr 1 (Table S1). Thus, even very large adenomas in the Pirc rat can develop without detectable widespread copy number instability or LOH at loci other than *Apc*.

Allelic ratios of gDNA and cDNA in colonic tumors of the F1-Pirc rat. We analyzed the percent WT contribution in gDNA and cDNA from normal colonic tissue and tumors from F1-Pirc rats. Material came from either spontaneously arising tumors in untreated rats or tumors in animals that were treated with the ulcerogenic inflammatory agent dextran sulfate sodium (DSS) to increase the number of colonic tumors. After harvesting a portion of each tumor for RNA and DNA extraction, the remaining portion of each tumor was fixed and sectioned for histological analysis. One of the gMOH tumors from an untreated rat was determined by histological examination to contain nearly 90% stromal cells, with very little epithelial contribution, and therefore, it was removed from further analysis. DNA from 67 tumors (34 DSS-treated and 33 spontaneous tumors) revealed that the majority (47) lost the WT *Apc* allele (gLOH), whereas 20 maintained heterozygosity of the *Apc* locus (gMOH) (Fig. 1*B*). The distribution between gLOH and gMOH classes was similar between untreated vs. DSS-treated F1-Pirc rats.

Quantitative pyrosequencing of the allelic ratio at the *Pirc* site was performed on cDNAs paired with each tumor gDNA sample. These assays showed that, regardless of treatment, each tumor that lost heterozygosity in gDNA also showed significant allelic loss in cDNA. Among the gMOH class of F1-Pirc rat tumors, one-half (10/20) were also cMOH, with nearly equal contributions of mutant and WT *Apc* alleles. The remaining one-half (10/20) were gMOH/cLOH tumors with expression at the *Apc* locus that was dramatically skewed to the mutant *Pirc* allele.

RT-PCR quantitation of mRNA levels. The relative level of transcripts from the *Apc* gene was measured by quantitative RT-PCR (qRT-PCR) (Table 1). gMOH/cLOH tumors showed a relative reduction in Apc transcripts compared with gMOH/cMOH tumors (Wilcoxon rank sum test, $P = 0.003$) and gLOH/cLOH tumors ($P = 0.05$). Interestingly, gMOH/cMOH tumors did not show a significant difference in the relative level of Apc transcripts compared with gLOH/cLOH tumors ($P = 0.84$). This observation reflects the conservative maintenance of somatic *Apc* copy number by the homologous somatic recombination process.

Search for secondary* Apc *mutations in gMOH F1-Pirc tumors. To determine whether a second genetic hit had occurred in the gMOH tumors, we used traditional fluorescent Sanger sequencing to analyze gDNA from 6 of 10 gMOH/cMOH and 4 of 10 gMOH/cLOH F1-Pirc rat tumors. Tumor DNA was amplified using PCR with primers for all 15 exons of the rat *Apc* gene plus exon–intron boundaries and the minimal promoter (~1 kb) region. Beyond the *Pirc* mutation, two of gMOH/cMOH tumors carried mutations in *Apc* exon 15 that created premature stop codon mutations—one from a nontreated animal and one from a DSS-treated F1-Pirc animal (Table S2). No sequence changes in the *Apc* gene were found in gDNA from the other four gMOH/cMOH tumors. Of the four gMOH/cLOH tumors, a single mutation in *Apc* intron 13 was found in one tumor (Table S2), creating a splice acceptor that led to an alternatively spliced transcript that was shown by RT-PCR (Fig. 2). This transcript carries a premature stop codon in exon 14 that is highly likely to be subject to nonsense-mediated decay, unlike the two exon 15

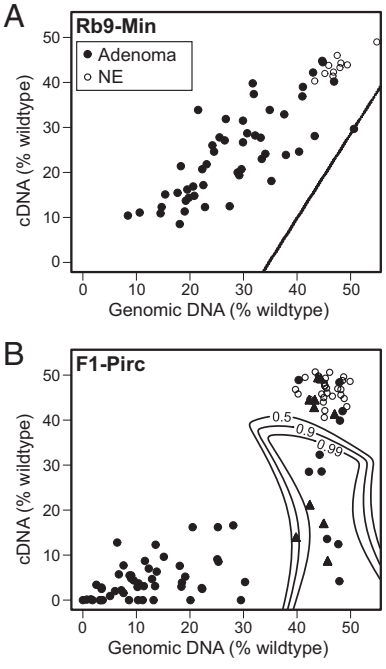

Fig. 1. Percentage of the WT *Apc* allele compared with the *Pirc* mutant allele in gDNA (*x* axis) and cDNA (*y* axis) from either Rb9-Min (*A*) or F1-Pirc (*B*) adenomas (filled symbols) or normal epithelium (NE; open circles). Tumors within the designated significance boundaries maintain heterozygosity of the gDNA but have lost heterozygosity in their cDNA. Rb9-Min significance lines (0.5–0.99) are superimposed on each other. F1-Pirc tumors that were sequenced for *Apc* are represented as filled triangles.

Table 1. qRT-PCR results for rat tumor classes

Type	*N*	ΔC_t (mean ± SD)
gLOH/cLOH	3	6.02 ± 0.52
gMOH/cMOH	8	5.97 ± 0.44
gMOH/cLOH	7	7.24 ± 0.65

Results are expressed as ΔC_t using *Gapdh* as the control gene for F1-Pirc tumors for each allele-specific class. $P = 0.007$ for the null hypothesis that all of the samples are taken from the same population (Kruskal–Wallis). Pairwise *P* values are 0.05 for gLOH/cLOH vs. gMOH/cLOH, 0.003 for gMOH/cLOH vs. gMOH/cMOH, and 0.84 for gLOH/cLOH vs. gMOH/cMOH (Wilcoxon rank sum test).

nonsense mutations described above and the *Min* and *Pirc* nonsense alleles. No sequence changes in the *Apc* gene were found in any of the three remaining gMOH/cLOH tumors, providing strong evidence for the existence of an epigenetic monoallelic silencing process.

Search for extragenic second mutations in gMOH/cMOH F1-Pirc rat tumors. The molecular basis for the interesting gMOH/cMOH F1-Pirc rat tumors was investigated. In the absence of *Apc* mutations, constitutive mutations in the β-catenin gene (*Ctnnb1*) have been shown to activate the WNT pathway. However, no sequence changes were found in the region of *Ctnnb1* associated with such activating mutations. Similarly, no mutations in *Kras* were found in any of the sequenced tumors. None of the gMOH/cMOH F1-Pirc rat tumors gave evidence for biallelic silencing by qRT-PCR assays of cDNA. Possible molecular bases for these gMOH/cMOH tumors will be discussed.

Histopathology of F1-Pirc and Rb9-Min tumors. Histological analysis of adenomas has failed to find major differences between these gLOH and gMOH classes of F1-Pirc rat adenomas. Immunohistochemical staining for β-catenin showed similar enhanced accumulation of the antigen in the cytoplasm and nucleus, with coordinate loss of the lateral membrane-bound localization. Additionally, the pattern of staining by H&E showed no pronounced difference between the classes, with similar low levels of normal epithelial cell contribution except in the single case cited above where there was substantial stromal and minimal tumor epithelial contributions. In Rb9-Min tumors, staining for WT APC protein showed that the pervasive loss of normal APC antigen and increase in cytoplasmic and nuclear β-catenin are similar between gMOH/cLOH and gLOH/cLOH cases (Fig. 3). Significantly, APC staining was undetectable in the nonepithelial stromal contribution within the tumor or surrounding normal tissue.

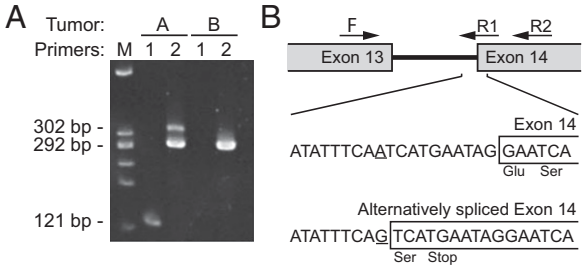

Fig. 2. RT-PCR confirmation of alternative splice forms of the Apc transcript generated by a splice acceptor mutation in intron 13. (*A*) RT-PCR of tumors A and B using primers spanning intron 13. The PCR product from the normal splice form (primer set 2) is shown in tumor B (292 bp); it is also present in tumor A. Primer set 1 is specific to the alternative splice form created by the A–G transition mutation in intron 13. The 121-bp band is present only in tumor A. The differences in band intensity for primer set 2 in tumor A indicate that the alternative splice form is less abundant than the normal splice form. (*B*) Map and PCR primers spanning intron 13. Sequencing of the alternative splice form identified the sequence and position of the mutation that creates a splice acceptor.

Discussion

Conservative Genome of Early Rb9-Min and F1-Pirc Neoplasms. The early stages of colonic neoplasia in the F1-Pirc rat model of familial colon cancer show no evidence of extensive changes in copy number over the genome (Fig. S2). These observations mirror prior studies in the Min mouse, including invasive stages of tumors in the small intestine of long-lived (SWR × B6)F1-Min mice (11). The present study extends these observations of the stable genome to investigate conservative LOH over the genome in F1-Pirc rat tumors.

Previous studies showed that LOH at the *APC* locus by somatic recombination was often found in early human FAP adenomas (19). Similarly, we have reported in Rb9-Min mouse and F1-Pirc rat tumors a similarly high proportion of LOH events at the *Apc* locus involving homologous somatic recombination (13, 18). In this report, we tested whether early adenomas that lose heterozygosity of *Apc* involve genome-wide LOH as in meiosis—a class of somatic instability. Examination of polymorphic SNP loci on multiple chromosomes by quantitative allele-specific pyrosequencing in F1 hybrids, including Chr 18, showed that LOH was largely restricted to the short arm of Chr 18 harboring the *Apc* locus. This finding indicates that somatic recombination over the whole genome is not enhanced in the tumor lineage but that the LOH events that occur on the short arm of Chr 18 are strongly selected in early adenomagenesis.

Interestingly, in one F1-Pirc tumor, we detected LOH on the long arm of Chr 1 (Table S1). This event may be a passenger event functionally independent of the LOH event at the *Apc* locus on Chr 18. Alternatively, this region of the genome may contain one or more modifier loci. We note that rat Chr 1 contains many genes with effects on cancer (*Igf2*, *Igf2R*, *Dkk1*, *Rab38*, and *Hras*). We note also that many regions on rat Chr 1 are

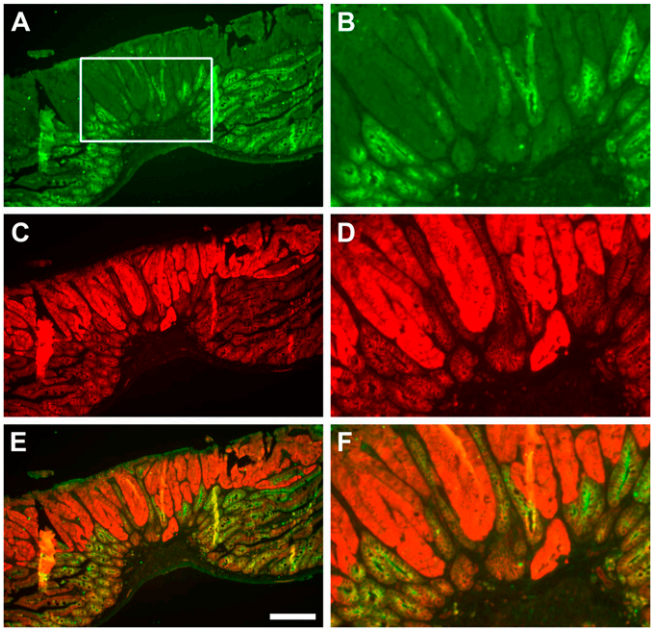

Fig. 3. Immunohistochemistry was performed with antibodies to (*A* and *B*) Apc protein (green) and (*C* and *D*) β-catenin (red; merged in *E* and *F*). *B*, *D*, and *F* are enlargements of the box in *A* and equivalent areas in *C* and *E*, respectively. Loss of Apc protein and an increase in cytoplasmic and nuclear β-catenin in tumor epithelial cells were observed for the gMOH/cLOH tumor as previously reported for MOH and LOH tumors (13). Apc protein cannot be detected in normal stroma; it is detectable primarily in normal intestinal epithelium. (Scale bar: *A*, *C*, and *E*, 200 μm.)

syntenic to imprinted regions in the human genome, including human Chrs 6 (*IGF2R*), 11 (*IGF2*/*H19*), 15 (*SNRPN* and *UBE3A*), and 19 (*PEG3*). Many of these genes regulate the growth of cancers. In this single case of LOH across the q arm of rat Chr 1, the maternal (ACI) allele was lost, plausibly duplicating the paternal expression of many of the imprinted growth-enhancing genes (*Igf2* and *Igf2r*). Indeed, cases of loss of imprinting may instead involve LOH through somatic recombination (20).

Tumors That Maintain Heterozygosity at the *Apc* Locus (gMOH). The most interesting groups of tumors in the Rb9-Min and F1-Pirc rat are those groups that maintain heterozygosity of *Apc* in their gDNA (gMOH). Under Knudson's two-hit hypothesis, one genetic hit alone would be insufficient to cause tumor formation. Interestingly, the gLOH vs. gMOH ratios are more evenly distributed in the set of Rb9-Min tumors (Fig. 1*A*) than in the F1-Pirc rat tumors (Fig. 1*B*). This finding may reflect the increased level of normal epithelial and stromal admixture in the nonpedunculate mouse tumors that are dissected from the small intestine. Alternatively, this broad distribution of allelic ratios may reflect the known polyclonality of Min tumors (21) or differences between tumors of the small intestine and the colon. By contrast, the high concordance between the majority of gLOH vs. cLOH allelic ratios in F1-Pirc tumors and between the gMOH vs. cMOH allelic ratios in normal epithelium indicates that these pedunculate colonic rat tumors are dissected with less admixture of normal epithelial and stromal tissue, resulting in more highly defined LOH and MOH classes.

Analysis of Admixture in Rb9-Min and F1-Pirc Tumors. To make rigorous distinctions between MOH and LOH classes, we investigated statistically the role of admixture in the analysis of these tumor classes for both the Rb9-Min small intestinal tumors and the F1-Pirc colonic tumors. The probability that a tumor is composed of any particular mixture of cells from the three classes was estimated by likelihood methods (Fig. 4). This analysis indicates that the Rb9-Min tumors contain a significant admixture between gMOH/cMOH and gLOH/cLOH components (Fig. 4*A*). It is estimated that 33% of the Rb9-Min tumors have a majority of cells from the gMOH/cMOH class, and 45% have a majority contribution from the gLOH/cLOH class. (The corresponding probabilities for plurality contributions are 45% and 51%, respectively.) These estimates may also reflect a high level of normal epithelial cell contribution to dissected Rb9-Min tumors. The estimated contributions to F1-Pirc tumors are more distinct (Fig. 4*B* and Table S3). These tumors show a greater gMOH/cLOH contribution admixed with both gMOH/cMOH and gLOH/cLOH classes. Thus, admixture between LOH and MOH components is observed in adenomas of both the Rb9-Min mouse and F1-Pirc rat and may reflect the polyclonal nature of early intestinal cancer (21, 22). A full statistical analysis of admixture is presented in *SI Materials and Methods* and Fig. S4; custom R code is contained in Dataset S1.

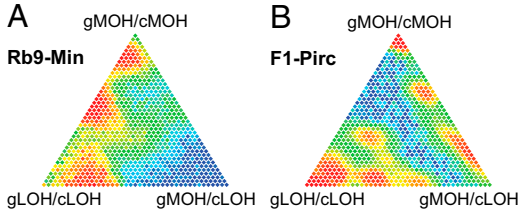

Fig. 4. Admixture models for Rb9-Min (*A*) and F1-Pirc (*B*). Likelihood methods were used to estimate the distribution of tumor composition from cells of three distinct types (corners of the triangles). Plotted here in barycentric coordinates are estimated probabilities that a tumor has a particular fraction of its cells from each cell type, with low probabilities in blue to highest probabilities in red. The estimated distribution also indicates the chance that a tumor is comprised of a majority or plurality of each cell type.

Absence of Detectable Mutations in F1-Pirc Tumors. The analysis of genomic *Apc* sequence in gMOH/cMOH F1-Pirc rat tumors revealed only two tumors with secondary mutations in *Apc* (Table S2). No new mutations in the gene for β-catenin were found, suggesting that there are alternative pathways to early adenoma formation not involving the WNT, *Apc*, β-catenin axis but requiring either the Apc^{Min} (mouse) or Apc^{Pirc} (rat) predisposing mutations. Thus, we consider the interesting possibility that the gMOH/cMOH class of adenomas commonly develops through haploinsufficiency of *Apc*. The population of *Apc* RNAs quantified from these tumors (Table 1) is composed of a 1:1 ratio of mutant to WT molecules. The Min phenotype is enhanced by heterozygosity for a KO allele of *p27* (23). For example, tumor phenotypes are observed in *p53*/+ heterozygotes without loss of the WT *p53* allele (24). The genomically stable processes of neoplasia in both the Min mouse and Pirc rat models for familial intestinal cancer in the human have opened an opportunity to discover noncomplementing genetic and epigenetic changes that would enhance any haploinsufficiency created by heterozygosity for a mutation in *Apc* (2).

Silencing in the Absence of Detectable Mutation in *Apc*. We investigated the possibility that adenoma formation in the gMOH tumors was owing to loss of function of *Apc* through epigenetic silencing. The two Rb9-Min tumors that lacked sufficient cDNA from *Apc* to amplify in the RT-PCR assay may have silenced both alleles through a programmed epigenetic mechanism. However, the limited amount of material obtained from Min tumors limited additional resequencing. F1-Pirc rat gMOH/cMOH tumors did not show a significant reduction of *Apc* transcript levels as judged by qRT-PCR (Table 1). This finding excludes the possibility that these F1-Pirc tumors arose by biallelic silencing of *Apc*.

The remaining 10 MOH/cLOH F1-Pirc rat tumors provide the strongest evidence for monoallelic epigenetic silencing. This class did show a twofold reduction in the expression of *Apc* by qRT-PCR (Table 1). Allele-specific analysis of cDNAs by pyrosequencing indicated that, as expected, the mutant *Pirc* allele was preferentially expressed. Sequence analysis of 4 of 10 tumors found that only 1 of these gMOH/cLOH tumors carried a new mutation in the *Apc* locus: an intronic mutation that leads to an alternative splice form. This mutant transcript likely involves the WT allele and would be subjected to nonsense-mediated decay, resulting in two prematurely truncated alleles. *APC* mutations resulting in alternatively spliced transcripts have been observed in FAP patients, with some of these mutations subject to nonsense-mediated decay (25). Although we sequenced the exons and exon/intron boundaries of *Apc* in only four gMOH/cLOH tumors, it seems unlikely that the remaining six F1-Pirc gMOH/cLOH tumors all have intronic mutations that lead to nonsense-mediated decay. The major result is finding three independent cases of monoallelic epigenetic silencing.

Possible Mechanisms of Monoallelic Epigenetic Silencing. Do the cases of *Apc* silencing represent programmed events or the result of a stochastically initiated silencing event? Global changes in gene expression through a developmental programming may explain the two Rb9-Min gMOH tumors that failed to express Apc mRNA. By contrast, a stochastic epigenetic event would occur at random, be fixed early in tumorigenesis, and act in *cis* at the allele level. We interpret the adenomas that exhibit monoallelic expression of *Apc* to be a product of such a stochastic silencing event that was then fixed and propagated with sufficient stability to support selection for the neoplastic process.

Monoallelic expression of a locus can reflect one of several conditions: constitutional polymorphism for a regulatory element controlling the level of the transcript from the locus, an evolved regulatory mechanism that generates allelic exclusion, or an altered structure of one allele that can be somatically propagated. The fact that the *Apc* locus is expressed from both alleles in normal epithelial tissue from the intestine of the Rb9-Min

mouse and F1-Pirc rat rules out a functional constitutional polymorphism. An evolved mechanism of allelic exclusion (in the tumor lineage) is unlikely, because many of the tumors maintain heterozygous expression of APC transcripts. Finally, the failure to detect a new mutation in the WT *Apc* allele in three independent colonic tumors of the F1-Pirc rat forces us to consider a *cis*-acting epigenetic mechanism that can be propagated in tumors.

X inactivation in eutherian mammals is the paradigm for random, monoallelic epigenetic silencing that can be stably propagated in the soma (26). The molecular events associated with X inactivation involve expression from the inactive X chromosome, Xi, of the long, noncoding RNA X-inactive specific transcript (XIST) that coats the inactive X (27). Interesting recent studies of reprogramming in mouse epiblast stem cells through nuclear transfer show that, when XIST is removed from the inactive X, genes on the inactive X are not reactivated. Instead, the Xi state may be maintained through the histone variant macroH2A that also coats the inactive X but is not removed during reprogramming (28). Finally, cytosine residues in CpG dinucleotide sequences are often methylated at key regulatory sequences and act as a *cis*-limited mark of a silent allele. Cytosine methylation of the *APC* gene has been reported in microsatellite-unstable tumors in humans (29). However, examination of the mouse *Apc* promoter provides no evidence of methylation changes in tumors or during aging (30, 31). It remains to be determined if a differential methylation mark is present at the rat *Apc* locus. Importantly, de novo CpG methylation is not required for the initiation or propagation of X inactivation in mammals (32). Alternatively, long-range protein-based silencing mechanisms (for example, with macroH2A) invoke cooperative binding. A paradigm for this behavior is the clustered binding of phage T4's gp32 protein to its substrate (33). Altogether, a range of molecular mechanisms can be investigated to explain the *cis*-acting monoallelic silencing observed in the gMOH/cLOH tumors. This report illustrates that the range of possible monoallelic and biallelic silencing mechanisms can be uncovered first through the analysis of cDNA from F1 genotypes.

We have developed informative models for familial colon cancer in both the mouse and the rat and have used quantitative assays for LOH in gDNA and cDNA of these tumors. Beyond the previously documented pathway of tumorigenesis involving LOH by homologous somatic recombination, we provide evidence for tumorigenesis by monoallelic and biallelic silencing of the *Apc* tumor suppressor locus in the absence of genomic instability. The F1-Pirc rat provides a large proportion of colon tumors that are highly enriched for epithelial tumor material. The residual level of admixture between LOH and MOH classes may be owing to the nonmedullary nature of all intestinal tumors, or it may reflect polyclonality as documented in the mouse and human (21, 22, 34).

The discovery of monoallelic silencing at the *Apc* locus indicates that this platform can be extended to investigate the occurrence and phenotypic consequences of allele-specific genetic and epigenetic events over the entire genome. Informative F1 genotypes enhance the dynamic range by which to quantify allele-specific genetic or epigenetic events. It will be interesting to expand the narrow window of the *Apc* locus in this study to a genome-wide epigenetic analysis. Identifying the full range of neoplastic pathways and stages that involve monoallelic silencing can power studies of the genetic and environmental influences on this apparently random epigenetic process. Continued investigation by whole-genome and transcriptome sequencing of the gMOH tumor classes discovered in this study is also called for to determine the spectrum of mutations that would combine synthetically with the proposed haploinsufficient state of *Apc*.

Materials and Methods

Mouse and Rat Maintenance. We bred mice carrying the *Min* allele in *cis* to the Rb9 (7:18) translocation on the C57BL/J background (Rb9-Min) to increase the proportion of adenomas that maintain heterozygosity at the *Apc* locus (13).

Rats carrying the *Pirc* allele were bred on an F1 hybrid background that permits examination of allele-specific genetic and epigenetic events at all informative loci in the genome. Inbred F344N/Tac (Taconic) male rats coisogenic for the Apc^{Pirc} mutation (F344-Pirc) were bred to ACI/Hsd (ACI; Harlan) females to generate (ACI × F344)F1 $Apc^{Pirc/+}$ (F1-Pirc) hybrids. The *Pirc* allele was genotyped as previously described (18).

Rats and mice were fed 5020 chow (Purina) with access to an automatic supply of acidified water. One-half of the rats were treated with DSS in the drinking water. These animals were given two 7-d cycles of 4% DSS (500,000 kDa molecular mass) starting at 40 d of age, with a 7-d break between treatments. Animals were maintained in standard cages under a university-approved animal protocol in a facility approved by the Association for Assessment and Accreditation of Laboratory Animal Care.

Tumor Harvesting. We designed the protocol for tumor dissection to optimize both the molecular and histological analysis of each tumor. Speed of dissection, quantity of material, and minimization of nontumor tissue are each important. Tumors were harvested at necropsy immediately after CO_2 asphyxiation. Colons in the rat and both colons and small intestines in the mouse were washed with Dulbecco's PBS. One-half of each small-intestinal mouse tumor (not including the muscularis) and a one-quarter wedge of tumor from the upper one-half of each rat colonic tumor were put into RLTplus buffer (Qiagen) and homogenized. This method reduced the extent of nontumor tissue in the resulting RNA and DNA extractions. Intestines with the remaining tumors were washed with 70% ethanol, fixed in 10% buffered formalin for ~16 h (mouse) or 24–40 h (rat), and then stored in 70% ethanol. When a portion of a tumor had been taken into RLTplus buffer, the residual tissue was excised and fixed for histological analysis. Purity of the dissected material was assessed from sections of the fixed tissue either stained with H&E or analyzed by immunochemistry for enhanced β-catenin expression.

Extraction of DNA and RNA from Tumors. DNA and RNA were extracted from each RLTplus sample using the standard Qiagen RNA/DNA extraction kit with the manufacturer's protocol. DNA contamination of RNA samples was reduced to background levels by DNase treatment of the RNA during the extraction using the recommended on the column DNase treatment (Qiagen). cDNA was then generated from the resulting RNA using the Invitrogen SuperScript III reverse transcriptase kit (Invitrogen).

Identification and Confirmation of Polymorphisms and Mutations in the Rat Genome. Standard fluorescent Sanger sequencing was used to confirm SNPs in the rat genome and identify new mutations in particular genes of interest in colonic tumors of the F1-Pirc rat. We identified SNPs using the Rat Genome Database SNPlotyper and chose those SNPs near the telomeric ends of chromosomes to capture any proximal somatic recombination events. To identify new mutations in tumors, all 15 exons, exon–intron boundaries, and 1 kb of the upstream promoter and alternative promoter of the rat *Apc* locus were sequenced as well as the hotspot regions of *K-Ras* (the sequence around the codon for amino acid G12) and β-catenin *(Ctnnb1*; exons 2, 3, and 4). All primers were designed using Primer 3 v0.4.0 (Table S4) (35). PCR protocols were optimized for Mg^{2+} concentration and annealing temperature. Products were confirmed by agarose gel electrophoresis and treated with exonuclease I (0.5 U) and shrimp alkaline phosphatase (0.25 U) at 37 °C for 30 min followed by heat inactivation at 80 °C for 15 min. A volume of 1–2 μL was used to sequence in both forward and reverse directions with Applied Biosystems Big Dye v3.1, and the products were purified on Sephadex G50 columns. All products were run by the University of Wisconsin Biotechnology Core Sequencing Facility.

Determination of Allelic Ratios. To follow genetic and epigenetic changes at the *Apc* locus, we used pyrosequencing to quantify the ratios of the *Min* and *Pirc* single-base *Apc* mutations vs. their WT alleles. Confirmed rat SNPs covering the genome were used to detect any changes of allelic ratios in F1-Pirc tumor samples. For pyrosequencing, gDNA and cDNA samples were used in a PCR mix with the following final concentrations: 1× GoTaq clear buffer, 1.2 mM $MgCl_2$, 0.2 mM dNTPs, 264 pM each primer, 0.6 U GoTaq Flexi (Promega), 8 μL DNA, and ddH_2O to 50 μL (18, 36). The PCR cycling profile was 94 °C for 3 min followed by 50 cycles of 94 °C for 15 s, 57 °C for 1.5 min, and 72 °C for 2 min, with a final elongation step at 72 °C for 10 min. Pyrosequencing was performed according to the manufacturer's protocols using Pyro Gold Reagents with a PSQ96 or Qiagen Pyromark MD machine and PSQ 96 v2.1 software (Biotage) (Table S4); 40 μL PCR product were used per well. For samples run in the PSQ96, only sequence reads with single-base peak heights of over 120 units were included. All pyrosequencing assays

were run in duplicate. If results differed by more than 5%, samples were rerun.

Statistical Analysis of Allelic Ratios. Two statistical methods—cluster and admixture analyses—were developed and applied to pyrosequencing allelic ratio data from gDNA and cDNA. In the cluster analysis, tumors were separated into three classes—gMOH/cMOH, gLOH/cLOH, and gMOH/cLOH—using computational and mathematical tools from the theory of Gaussian mixture models (37, 38). Computations were performed in the R system (39). Specifically, the gDNA, cDNA data pair from each tumor was considered to be a random draw from a mixture of three bivariate normal components, one for each of the named classes (*SI Materials and Methods*). From the fitted probability model, various predictions were computed, including contours indicating the posterior probability that a tumor arose from the gMOH/cLOH class given the gDNA/cDNA data. This mixture of bivariate normal distributions extends a simpler univariate normal mixture used previously for MOH/LOH analysis to available data (17).

In the admixture data analysis, we recognized that each tumor sample could be an unequal mixture of three pure cell types—the same three types (classes) as in the cluster analysis, except that now these classes were viewed as properties of constituent cells rather than whole tumors. Furthermore, we reasoned that measurement error in the gDNA and cDNA channels should to be uncorrelated and that apparent within-class correlation in the first analysis would be attributable to admixture of different cell types within each tumor. Unknown parameters in the admixture model include class-specific mean values for the three bivariate pure cell classes constrained as in the cluster analysis, channel-specific measurement error variance, and tumor-specific mixing proportions over the three classes. To facilitate inference, we treated each tumor's vector of three mixing proportions as a draw from an unknown distribution over proportion vectors (a distribution that characterizes the tumor population in question [Rb9-Min or F1-Pirc]). Markov chain Monte Carlo computations were then developed to estimate the unknown parameters of the admixture model (*SI Materials and Methods*).

qRT-PCR. We performed qRT-PCR to determine whether overall APC RNA expression was diminished in MOH tumors. We used predesigned assays for APC RNA (TaqMan:Rn00560714 FAM-dye-labeled MGB probe; Applied Biosystems) and control GAPDH RNA (4352338E VIC dye-labeled MGB probe). The two assays were combined and run together for each sample. We used the ABI Taqman Gene Expression Master Mix at 1× concentration in 20-μL reactions with cDNA generated as described above. The samples were run in duplicate and analyzed on an ABI 7900HT machine using the recommended cycling conditions of 40 cycles with a 15-s 95 °C denaturing step and a 1-min 60 °C annealing/extension. Data were analyzed using the ABI program DataAssist.

ACKNOWLEDGMENTS. The University of Wisconsin Comprehensive Cancer Center's Histology Shared Service has provided talented support, both technical (Jane Weeks and Harlene Edwards) and conceptual (Ruth Sullivan). We thank Norman Drinkwater and James Shull for critical comments and Alexandra Shedlovsky and Jennifer Pleiman for clarification of this report. We thank two reviewers for critiques that have significantly enhanced this report. Finally, we pay tribute to the long-term impact of Mary Lyon on mammalian genetics and epigenetics. This research was supported by American Cancer Society Postdoctoral Fellowship PF-08-077-01 (to J.M.A.-L.), a training grant from the National Institute on Environmental Health Sciences (to A.A.I.), National Cancer Institute Grants R01 CA63677 and R01 CA125591 (to W.F.D.), and National Institutes of Health/National Cancer Institute Grant P30 CA014520 (University of Wisconsin Comprehensive Cancer Center Support). This is publication no. 3647 from the Laboratory of Genetics.

1. Fearnhead NS, Britton MP, Bodmer WF (2001) The ABC of APC. *Hum Mol Genet* 10: 721–733.
2. Yeung AT, et al. (2008) One-hit effects in cancer: Altered proteome of morphologically normal colon crypts in familial adenomatous polyposis. *Cancer Res* 68: 7579–7586.
3. Lewis A, et al. (2010) Severe polyposis in Apc(1322T) mice is associated with submaximal Wnt signalling and increased expression of the stem cell marker Lgr5. *Gut* 59:1680–1686.
4. Jones S, et al. (2008) Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci USA* 105:4283–4288.
5. Cheng YW, et al. (2008) CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. *Clin Cancer Res* 14:6005–6013.
6. Georgiades IB, Curtis LJ, Morris RM, Bird CC, Wyllie AH (1999) Heterogeneity studies identify a subset of sporadic colorectal cancers without evidence for chromosomal or microsatellite instability. *Oncogene* 18:7933–7940.
7. Toyota M, et al. (1999) CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA* 96:8681–8686.
8. Loeb LA, Loeb KR, Anderson JP (2003) Multiple mutations and cancer. *Proc Natl Acad Sci USA* 100:776–781.
9. Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* 396:643–649.
10. Edelmann W, et al. (1999) Tumorigenesis in Mlh1 and Mlh1/Apc1638N mutant mice. *Cancer Res* 59:1301–1307.
11. Halberg RB, et al. (2009) Long-lived Min mice develop advanced intestinal cancers through a genetically conservative pathway. *Cancer Res* 69:5768–5775.
12. Fodde R, et al. (2001) Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat Cell Biol* 3:433–438.
13. Haigis KM, Dove WF (2003) A Robertsonian translocation suppresses a somatic recombination pathway to loss of heterozygosity. *Nat Genet* 33:33–39.
14. Sieber OM, et al. (2002) Homologous somatic recombination is a more subtle form of genomic instability observed in human FAP and in its murine models. *Proc Natl Acad Sci USA* 99:16910–16915.
15. Grady WM, et al. (2000) Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nat Genet* 26:16–17.
16. Cui H, et al. (2003) Loss of IGF2 imprinting: A potential marker of colorectal cancer risk. *Science* 299:1753–1755.
17. Shoemaker AR, et al. (1998) A resistant genetic background leading to incomplete penetrance of intestinal neoplasia and reduced loss of heterozygosity in $Apc^{Min/+}$ mice. *Proc Natl Acad Sci USA* 95:10826–10831.
18. Amos-Landgraf JM, et al. (2007) A target-selected Apc-mutant rat kindred enhances the modeling of familial human colon cancer. *Proc Natl Acad Sci USA* 104:4036–4041.
19. Crabtree M, et al. (2003) Refining the relation between 'first hits' and 'second hits' at the APC locus: The 'loose fit' model and evidence for differences in somatic mutation spectra among patients. *Oncogene* 22:4257–4265.
20. Baba Y, et al. (2010) Hypomethylation of the IGF2 DMR in colorectal tumors, detected by bisulfite pyrosequencing, is associated with poor prognosis. *Gastroenterology* 139: 1855–1864.
21. Thliveris AT, et al. (2005) Polyclonality of familial murine adenomas: Analyses of mouse chimeras with low tumor multiplicity suggest short-range interactions. *Proc Natl Acad Sci USA* 102:6960–6965.
22. Thirlwell C, et al. (2010) Clonality assessment and clonal ordering of individual neoplastic crypts shows polyclonality of colorectal adenomas. *Gastroenterology* 138: 1441–1454.
23. Payne SR, Kemp CJ (2005) Tumor suppressor genetics. *Carcinogenesis* 26:2031–2045.
24. Hagstrom SA, Dryja TP (1999) Mitotic recombination map of 13cen-13q14 derived from an investigation of loss of heterozygosity in retinoblastomas. *Proc Natl Acad Sci USA* 96:2952–2957.
25. De Rosa M, et al. (2007) Alternative splicing and nonsense-mediated mRNA decay in the regulation of a new adenomatous polyposis coli transcript. *Gene* 395:8–14.
26. Lyon MF (1961) Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190:372–373.
27. Brown CJ, et al. (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349:38–44.
28. Pasque V, Gillich A, Garrett N, Gurdon JB (2011) Histone variant macroH2A confers resistance to nuclear reprogramming. *EMBO J* 30:2373–2387.
29. Derks S, et al. (2008) Integrated analysis of chromosomal, microsatellite and epigenetic instability in colorectal cancer identifies specific associations between promoter methylation of pivotal tumour suppressor and DNA repair genes and specific chromosomal alterations. *Carcinogenesis* 29:434–439.
30. Linhart HG, et al. (2007) Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing. *Genes Dev* 21:3110–3122.
31. Maegawa S, et al. (2010) Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res* 20:332–340.
32. Sado T, Okano M, Li E, Sasaki H (2004) De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. *Development* 131: 975–982.
33. Delius H, Mantell NJ, Alberts B (1972) Characterization by electron microscopy of the complex formed between T4 bacteriophage gene 32-protein and DNA. *J Mol Biol* 67: 341–350.
34. Novelli MR, et al. (1996) Polyclonal origin of colonic adenomas in an XO/XY patient with FAP. *Science* 272:1187–1190.
35. Rozen S, Skaletsky HJ (2000) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, eds Krawetz S, Misener S (Humana Press, Totowa, NJ), pp 365–386.
36. Kwong LN, et al. (2007) Identification of *Mom7*, a novel modifier of $Apc^{(Min/+)}$ on mouse chromosome 18. *Genetics* 176:1237–1244.
37. McLachlan G, Peel D (2000) *Finite Mixture Models* (Wiley, New York).
38. Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* 97:611–631.
39. R Development Core Team (2011) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).

Supporting Information

Amos-Landgraf et al. 10.1073/pnas.1120753109

SI Materials and Methods

1. Model. Two probability models were used to analyze bivariate allelic ratios of mutant genomic DNA (gDNA) and mutant cDNA. The first model was a discrete Gaussian mixture model that provides for model-based clustering of the samples into three tumor classes. The class-specific mean parameters were constrained, and therefore, the expected gDNA value in the maintenance of heterozygosity of *adenomatous polyposis coli* (*Apc*) in gDNA and cDNA (gMOH/cMOH) class equaled the expected gDNA value in the gMOH/complementary loss of heterozygosity (cLOH) class; similarly, the expected cDNA values in classes gLOH/cLOH and gMOH/cLOH were constrained to be equal, which was justified by the measurement process. Otherwise, each class was allowed to have an unconstrained 2×2 covariance matrix. Owing to mean constraints, standard software could not be applied to fit the probability model; custom R code was developed to implement the expectation maximization algorithm and thus, estimate the means, covariances, and mixing proportions by the method of maximum likelihood. Data from normal tissues were included in the estimation but forced to arise from the gMOH/cMOH class.

A second model was developed to analyze the allelic ratios. This admixture model aimed to go farther than the discrete mixture model by representing each tumor as a mixture of cells of three different pure cell types. Computations in this admixture model were more complex than the computations required for the first model, and they are developed fully here. We analyzed the polyposis in the rat colon (Pirc) and Rb9 data separately. The n tissue samples from one experiment provided bivariate data, $\mathrm{DAT} = \{(X_i, Y_i): i = 1, 2, \ldots, n\}$, measuring mutant allele ratios in gDNA and cDNA, respectively. In the admixture model, the samples $\{i\}$ produced mutually independent data points. Furthermore, sample i was considered to be comprised of fractions of three pure cell types, with the fractions $U_{i,j} \geq 0$ and $\Sigma_{j=1}^{3} U_{i,j} = 1$. None of these fractions were observed, except for cases i that were normal tissue controls, for which we assumed $U_i = (U_{i,1}, U_{i,2}, U_{i,3}) = (1, 0, 0)$. For each tumor, the admixture vector U_i itself was viewed as a random draw from an unknown distribution $\pi(u)$ over the simplex $\mathcal{S}_3 = \{u = (u_1, u_2, u_3): u_j \geq 0, \sum_j u_j = 1\}$. Computations were enabled by taking a finite-grid approximation to $\mathcal{S}_3$ and thus, a vector approximation to π.

	Cell type	Fraction
1	gMOH/cMOH	$U_{i,1}$
2	gLOH/cLOH	$U_{i,2}$
3	gMOH/cLOH	$U_{i,3}$

We supposed that measurements of gDNA and cDNA on a pure tissue comprised only of cells of one type would be bivariate normal, with means and variances that depended on the cell type but without any correlation in the measurement error. The lack of correlation between gDNA and cDNA measurements on pure tissue was a key assumption that allowed us to infer admixture rates in the actual tissue samples. It was justified considering that pyrosequencing was performed separately for gDNA and cDNA material and that variation attributable to the initial isolation of all nucleic acids from the tissue was probably negligible. We also assumed that the distribution of data in one channel (gDNA or cDNA) was not affected by state of the other channel. In other words, for mean parameters a_1, a_2, b_1, and b_2, the mean gDNA measurement in pure gMOH cells was a_1, and the mean gDNA measurement in gLOH cells was a_2. Similarly the mean cDNA measurement in cMOH cells was b_1, and the mean cDNA measurement in cLOH cells was b_2. Fig. S4 shows these parameters as estimated from the Pirc dataset.

The contribution to gDNA/cDNA measurements (X_i, Y_i) from the pure cells of one type was assumed to be normally distributed, where the bivariate mean depended on the type and there was one variance parameter for gDNA and one variance parameter for cDNA, regardless of type. A key element was having independent measurement errors in these pure cells (the apparent marginal correlation in data arose from admixture). Thus, X_i and Y_i were treated as independent given admixture rates U_i, with gDNA data following (Eq. **S1**)

$$X_i|(U_i = u) \sim \mathrm{Normal}\{\mathrm{mean} = (u_1 + u_3)a_1 + u_2 a_2, \quad \mathrm{variance} = \sigma_a^2(u_1^2 + u_2^2 + u_3^2)\} \qquad \textbf{[S1]}$$

and cDNA data following (Eq. **S2**)

$$Y_i|(U_i = u) \sim \mathrm{Normal}\{\mathrm{mean} = u_1 b_1 + (u_2 + u_3)b_2, \quad \mathrm{variance} = \sigma_b^2(u_1^2 + u_2^2 + u_3^2)\}. \qquad \textbf{[S2]}$$

Conditional on the admixture rates, the unknown parameters were $\theta = (a_1, b_1, a_2, b_2, \sigma_a^2, \sigma_b^2)$.

2. Inference. The unknown objects were mean and variance parameters in the vector θ, admixture vectors $\{U_i = (U_{i,1}, U_{i,2}, U_{i,3})\}$, and the probability distribution $\pi(u)$ over the simplex $\mathcal{S}_3$. We placed a prior distribution on θ and $\pi(u)$ and developed Bayesian computations by Markov chain Monte Carlo to simulate the distribution of these unknowns conditional on data *DAT*. We approximated the simplex $\mathcal{S}_3$ by a finite grid of proportion vectors, denoted ugrid in the R code. This matrix was size 861×3 holding row vectors that represent possible realizations of each admixture vector U_i. $K = 861$ came by taking a regular 40×40 grid over the unit square, keeping coordinates for which the values sum to less than one and considering these two values to be the first two entries of a possible admixture vector (the third being one minus the sum). Thus, π was a length $K = 861$ vector holding the probability distribution governing the U_i values. In the Bayesian analysis, we placed a conjugate exchangeable Dirichlet prior over π, using a small prior mass $\alpha = 1$ (**S3**):

$$\pi \sim \mathrm{Dirichlet}_K\left(\frac{\alpha}{K}, \frac{\alpha}{K}, \ldots, \frac{\alpha}{K}\right). \qquad \textbf{[S3]}$$

Next, we placed a noninformative flat prior on the mean parameters a_1, a_2, b_1, b_2 in θ, except that we insisted that all mean values exceed zero. We placed a weakly informative conjugate inverse-Gamma prior for the two variance parameters σ_a^2 and σ_b^2. We used a prior guess $\sigma_0^2 = 25$, with $n_0 = 1$. That is, the prior for each inverse variance was Gamma with shape $n_0/2$ and rate $n_0\sigma_0^2/2$. Evidently, the posterior information was relatively

high, and we would have computed very similar estimates for a wide range of prior hyperparameters.

We developed a systematic scan Gibbs sampler with separate updates for the means and variances in θ, the admixture vectors, and the mixing proportions π. (The following three subsections describe details of the Gibbs sampler.) After preliminary testing, we ran the sampler for 500,000 scans, saving 1 in each 100 states and basing posterior estimates on 5,000 putative draws from the joint posterior distribution. Output analysis indicated very good mixing of the Markov chain. Mean values of the output were used for parameter estimation.

2.1. Updating θ. The gDNA and cDNA measurements X_i and Y_i were conditionally independent given the admixture vector U_i. The parameters a_1, a_2, σ_a^2 refer to the distribution of X_i, and b_1, b_2, σ_b^2 refer to the distribution of Y_i. Considering the similarity of gDNA and cDNA models, we show here only the Gibbs sample update rules for the parameters governing X_i. Re-expressing the observation model from the previous section, we have (**S4**)

$$X_i|\text{else} \sim \text{Normal}\left(c_i^T\mu, d_i\sigma_a^2\right), \qquad \textbf{[S4]}$$

where $d_i = U_{i,1}^2 + U_{i,2}^2 + U_{i,3}^2$ multiplied by variance $\mu = (a_1, a_2)^T$ holds the pure cell means, and (Eq. **S5**)

$$c_i = \begin{pmatrix} c_{i,1} \\ c_{i,1} \end{pmatrix} = \begin{pmatrix} U_{i,1} + U_{i,3} \\ U_{i,2} \end{pmatrix}. \qquad \textbf{[S5]}$$

(In both c_i and d_i, the roles of $U_{i,1}$ and $U_{i,2}$ are reversed when considering the cDNA data Y_i.) Using standard Bayesian arguments and a flat prior, we find the full-conditional distribution for the mean parameters a_1 and a_2 is (**S6**)

$$\begin{pmatrix} a_1 \\ a_1 \end{pmatrix}\Bigg|\text{else} \sim \text{Normal}\left\{\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{bmatrix} s_{1,1} & s_{1,2} \\ s_{1,2} & s_{2,2} \end{bmatrix}\right\}, \qquad \textbf{[S6]}$$

where posterior means are (Eq. **S7**)

$$m_1 = \frac{1}{1 - \dfrac{C^2}{B_1B_2}}\left(\frac{A_1}{B_1} - \frac{A_2C}{B_1B_2}\right) \qquad \textbf{[S7]}$$

and (Eq. **S8**)

$$m_2 = \frac{1}{1 - \dfrac{C^2}{B_1B_2}}\left(\frac{A_2}{B_2} - \frac{A_1C}{B_1B_2}\right). \qquad \textbf{[S8]}$$

The posterior covariance matrix is (Eq. **S9**)

$$\begin{bmatrix} s_{1,1} & s_{1,2} \\ s_{1,2} & s_{2,2} \end{bmatrix} = \frac{1}{B_1B_2 - C^2}\begin{bmatrix} B_2 & -C \\ -C & B_1 \end{bmatrix}. \qquad \textbf{[S9]}$$

The contributing quantities are (Eq. **S10**)

$$A_1 = \sum_{i=1}^{n} \frac{1}{d_i\sigma_a^2} x_i c_{i,1}, \qquad \textbf{[S10]}$$

(Eq. **S11**)

$$A_2 = \sum_{i=1}^{n} \frac{1}{d_i\sigma_a^2} x_i c_{i,2}, \qquad \textbf{[S11]}$$

(Eq. **S12**)

$$B_1 = \sum_{i=1}^{n} \frac{1}{d_i\sigma_a^2} c_{i,1}^2, \qquad \textbf{[S12]}$$

(Eq. **S13**)

$$B_2 = \sum_{i=1}^{n} \frac{1}{d_i\sigma_a^2} c_{i,2}^2, \qquad \textbf{[S13]}$$

and (Eq. **S14**)

$$C = \sum_{i=1}^{n} \frac{1}{d_i\sigma_a^2} c_{i,1}c_{i,2}. \qquad \textbf{[S14]}$$

Thus, the Gibbs update of a_1 and a_2 arose from the bivariate normal posterior given in expression **S6**. We sampled it by sampling the marginal of a_1 and then, the induced conditional of a_2 given a_1. We imposed the constraint of $a_1, a_2 > 0$ in both updates.

Under the conjugate inverse-Gamma prior indicated above, the variance parameter σ_a^2 has an inverse-Gamma full-conditional distribution. More specifically (Eq. **S15**),

$$\frac{1}{\sigma_a^2}\Bigg|\text{else} \sim \text{Gamma}\left\{\text{shape} = \frac{n_0 + n}{2}, \text{rate} = \frac{1}{2}\left(n_0\sigma_0^2 + \sum_{i=1}^{n}\frac{\left(x_i - c_i^T\mu\right)^2}{d_i}\right)\right\}. \qquad \textbf{[S15]}$$

Analogous updates for the Y_i parameters b_1, b_2, and σ_b^2 are as above but with the roles of $U_{i,1}$ and $U_{i,2}$ reversed.

2.2. Updating admixture vectors. Tumor i has admixture vector U_i, which has a complicated but discrete conditional distribution over the K possible rows in ugrid given the data and the parameters θ and $\pi = (\pi_u)$. For a row u of ugrid, we have (Eq. **S16**)

$$P(U_i = u|\text{DAT}, \theta, \pi) \propto p(x_i|u, \theta)p(y_i|u, \theta)\pi_u. \qquad \textbf{[S16]}$$

In our implementation, we find the logarithm of the right-hand side by invoking the normal model from *SI Materials and Methods, 2.1.* and the estimated sampling model for U_i. We renormalize for each i to get the full conditional distribution for each U_i and then run the Gibbs update by sampling these discrete distributions one time each in parallel.

2.3. Updating mixing proportions. Given everything else, the vector π depends only on how many admixture vectors take each of the possible values in ugrid. By conjugacy of the Dirichlet relative to these multinomial counts (**S17**),

$$\pi|\text{else} \sim \text{Dirichlet}_K\left(\frac{\alpha}{K} + s_1, \frac{\alpha}{K} + s_2, \ldots, \frac{\alpha}{K} + s_K\right), \qquad \textbf{[S17]}$$

where s_j counts how many U_i values take value ugrid[j,]. This sample is by renormalizing independent and properly Gamma-distributed variables.

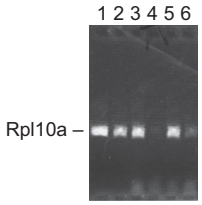

Fig. S1. RT-PCR for the ribonuclear protein subunit *Rpl10a* in mouse Rb9-*Multiple intestinal neoplasia* (*Min*) tumors to assess cDNA qualitatively. Lanes 1 and 2 are tumors that maintained heterozygosity of *Apc* in the DNA fraction but were below the level of detection by pyrosequencing of cDNA for *Apc*. Lanes 3–6 are other mouse tumors, both positive and negative for cDNA.

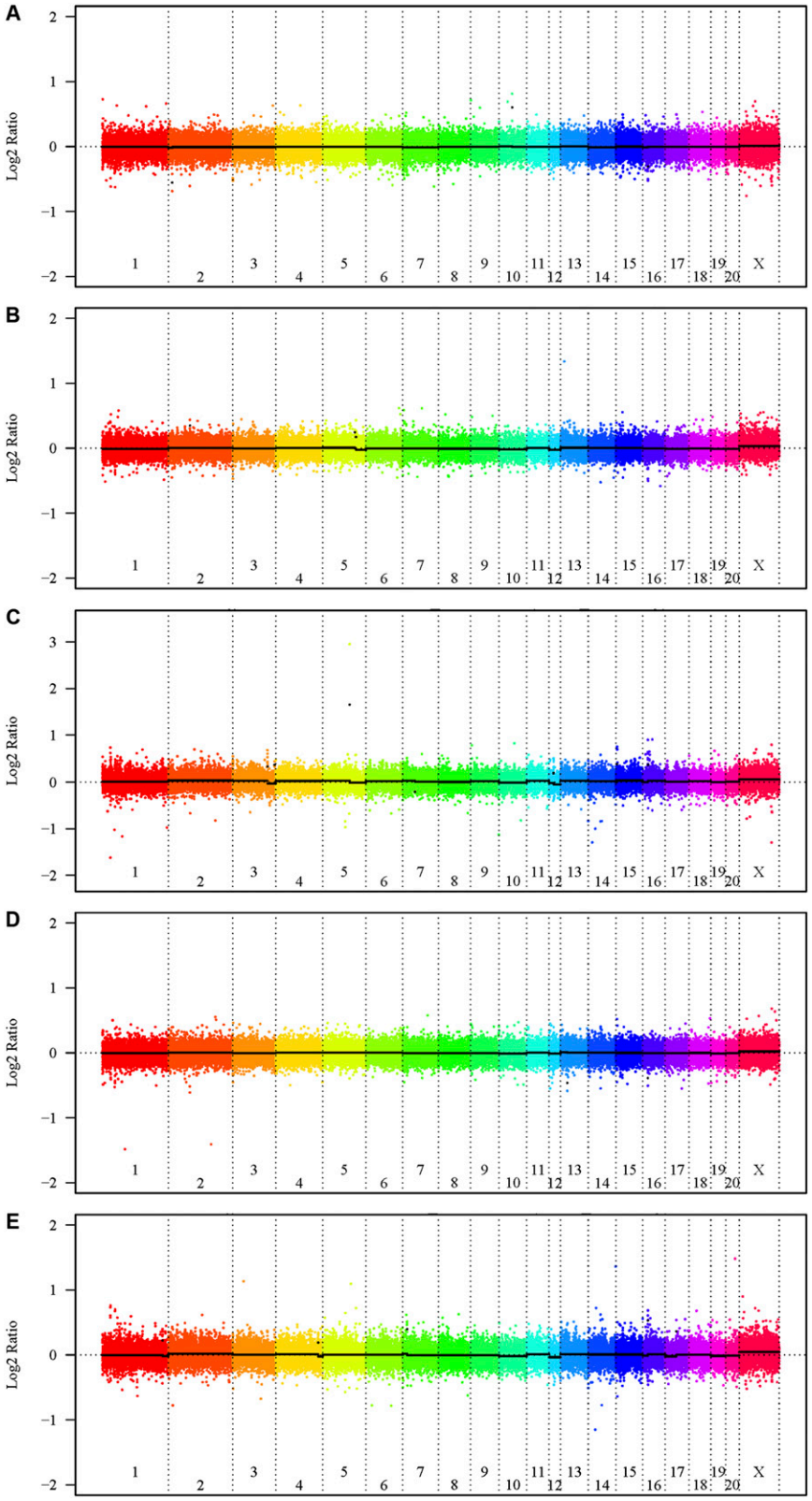

Fig. S2. Array comparative genomic hybridization (aCGH) was performed using Nimblegen rat whole-genome arrays and two-color labeling (Cy5 and Cy3) with tumor and normal spleen DNA. Two male F344-Pirc animals and one age-matched male F344/Tac WT rat were killed at 9 mo of age. From each F344-Pirc animal (P1 and P2), DNA was isolated immediately after necropsy, starting with 50 mg from the largest single colonic tumor and spleen. DNA was extracted using the Qiagen DNeasy tissue kit. Comparative genomic hybridizations were performed by Nimblegen Systems in their manufacturing facilities in Iceland using the Nimblegen RGSC 3.4 isothermal rat aCGH chip with 385,000 unique sequence features and a median probe density of 5,303 bp. One aliquot of tumor (T) and spleen (S) DNA from each animal was labeled with Cy5, and a second aliquot was labeled with Cy3. Spleen DNA sample P1S was hybridized against WT spleen DNA (*A*). Reciprocal hybridizations of Cy5 and Cy3 were performed: P1T vs. P1S (*B* and *C*). Similarly, Reciprocal hybridizations of P2T and P2S were performed (*D* and *E*). aCGH plots were generated using NimbleScan software, and the data were analyzed using SignalMap v1.8.

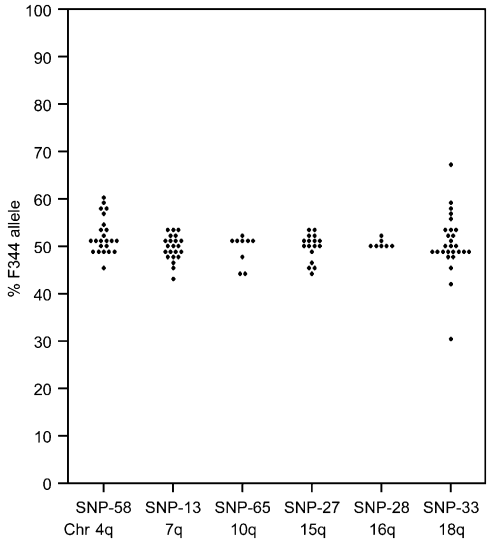

Fig. S3. LOH analysis of F1-Pirc tumors for markers on chromosomes 4, 7, 10, 15, 16, and 18q, not including markers for chromosome 1 (Table S2) using SNPs listed in Table S1. Each point represents a tumor. Not all tumors were tested for all markers. No significant loss of the F344 or ACI alleles was seen in any tumors.

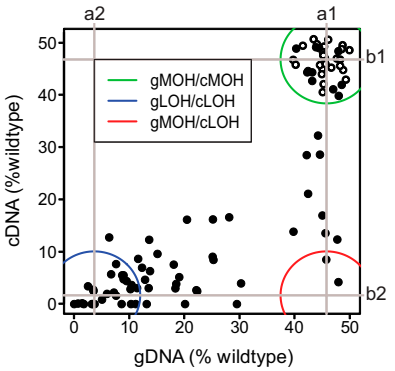

Fig. S4. Bivariate pyrosequencing data from normal tissue (open circles) and Pirc tumors (filled circles). Components of the fitted admixture model are shown. Pure cells of each of three types would yield bivariate Gaussian measurements from the three bivariate normal distributions (contours show regions holding 95% probability in the fitted component). The means are constrained (gray lines). Each tumor sample is viewed as an unknown mixture of the three cell types.

Table S1. LOH analysis of gDNA from F1-Pirc tumors using quantitative allele-specific pyrosequencing

			Contribution of F344 allele (%)	
Marker	Chr	Position (Mb)	Tumor X	Other tumors (mean ± SD)
SNP-76	1p	7	48.1	49.3 ± 3.0
SNP-102	1q	46	47.8	ND
SNP-82	1q	131	**84.7**	53.4 ± 0.6
SNP-75	1q	159	**78.7**	49.5 ± 1.4
SNP-83	1q	232	**78.3**	43.5 ± 1.8
SNP-2	1q	233	**85.8**	54.8 ± 3.1
Pirc	18p	27	**85.0**	83.5 ± 15.1*
SNP-33	18q	83	49.4	51.2 ± 3.3

SNPs across the genome were used to detect significant deviations from the expected 50% contribution from each of the parental strains. Tumor X showed LOH on Chr 1q, with a cross-over between 46 and 131 Mb. Allele ratios in bold show gLOH on one arm of the salient chromosome. Seven other tumors from the same animal serve as controls. Chr, chromosome; ND, not done.
*Six tumors were evaluated for Pirc.

Table S2. Mutations to *Apc* in gMOH/cMOH F1-Pirc tumors

Mutation type	Position on Chr. 18	Position in *Apc*	Treatment
T to C	26,782,134	Intron 13	DSS
C to T	26,783,847	Exon 15	None
C to T	26,784,615	Exon 15	DSS

The mutation type and base pair position (rat genome build Baylor 3.4, November 2004) of the three identified *Apc* mutations of the 12 gMOH tumors sequenced. Chr, chromosome; DSS, dextran sulfate sodium.

Table S3. Admixture probabilities for Rb9-Min and F1-Pirc tumors

Model	Tumor category	Probability (majority)	Probability (plurality)
Rb9-Min	gMOH/cMOH	0.33	0.45
Rb9-Min	gLOH/cLOH	0.43	0.51
Rb9-Min	gMOH/cLOH	0.02	0.04
F1-Pirc	gMOH/cMOH	0.19	0.19
F1-Pirc	gLOH/cLOH	0.63	0.68
F1-Pirc	gMOH/cLOH	0.12	0.13

The admixture probabilities for the majority or plurality of cells of each of the three tumor cell classes present in a given sample for both Rb9-Min and F1-Pirc as described in *SI Materials and Methods*.

Table S4. Primers

Primer name	Location	Forward	Reverse	Sequencing	SNP with surrounding sequence
Apc Alt-F	chr18: 26,777,147 + 26,777,166	GCAAGTGTTTTGCGGAATTT	NA	NA	NA
Apc Alt-R1	chr18: 26,782,136 + 26,782,155	NA	AGAGTTGATTCCTATTCATG	NA	NA
Apc Alt-R2	chr18: 26,782,307 + 26,782,326	NA	ACATTCCGTAATATCCCACC	NA	NA
APC Alt Promoter	chr18: 26,720,834 + 26,721,818	CCAGTCACTCGGTTCTGTCA	TACCACTGCCTGTGTCCTTG	NA	NA
Apc Promoter	chr18: 26,731,661 + 267,32,322	GAATGGCTGGAGTTCTTTGG	TGCCTCAGTTTCCAGTTCTGT	NA	NA
Apc Exon 1	chr18: 26,732,010 + 26,732,410	CCCATGTTCCGTCCTCTTAG	TGTCTTAGCTATCCCAAGTTTACAA	NA	NA
Apc Exon 2	chr18: 26,739,223 + 26,739,623	TGGCATGTTGTCAGTGTGAA	GATCTGAGGGGAAACGTCAA	NA	NA
Apc Exon 3	chr18: 26,739,951 + 26,740,394	CTGACTTTGCAGGGCAAGTT	CAGGCTGCCACCTATCATCT	NA	NA
Apc Exon4	chr18: 26,742,943 + 26,743,399	TAGCAGGGAAGGAGCACACT	AATGACGACGGTGACACTCA	NA	NA
Apc Exon 5	chr18: 26,747,107 + 26,747,506	TCATAGTTTAGGATTCCTTGGAAGA	TTTGATGGTGCTGCAGGTTA	NA	NA
Apc Exon 6	chr18: 26,751,039 + 26,751,389	TTGGAAAGCAGAAGCCAAAG	CGCATGTAACAAAAGGGGACT	NA	NA
Apc Exon 7	chr18: 26,762,128 + 26,762,608	TGAGAGGAGCACATCTGGTG	TCCGGAGCATAAAACATTCC	NA	NA
Apc Exon 8	chr18: 26,766,510 + 26,766,983	CTGCAGTCTGCTGGTTGTGT	GTCTGGAGCTTAGGCTGCTC	NA	NA
Apc Exon 9	chr18: 26,768,835 + 26,769,655	TGACCTTTAACTGATGCTGACAA	TTTTTCAGGGAAATGCAGATT	NA	NA
Apc Exon 10	chr18: 26,770,811 + 26,771,210	ACTGCTGGCTGTTCTCCTGT	ATATGGATGGAAGCAGCTGAA	NA	NA
Apc Exon 11	chr18: 26,775,431 + 26,775,923	CTGGCAGGGAGTGAAATTGT	CAGTGTGAGAAAGGGGCTTC	NA	NA
Apc Exon 12	chr18: 26,776,233 + 26,776,784	TTTCATTTGGGCAGTTTGTTC	TTGGGTTAGGAATGCTCCAC	NA	NA
Apc Exon 13	chr18: 26,776,833 + 26,777,287	GGGAAAACCTATGTTCTGAAGC	CTATTCTCGTACCCCACATTTT	NA	NA
Apc Exon 14	chr18: 26,782,006 + 26,782,505	TGCATCATGGAAAGATACGTG	TGGAGTTGATAGTGGTCTTCACA	NA	NA
Apc Exon 15-1	chr18: 26,783,492 + 26,784,011	ACGATTGCGATGTGAACGTA	CCCCATAGCGATCATTTTGT	NA	NA
Apc Exon 15-2	chr18: 26,783,179 + 26,784,469	TGCTCACAAGTGCTTTGGAC	GAAGAGGTTCCCGGATTTTC	NA	NA
Apc Exon 15-3	chr18: 26,784,077 + 26,785,123	GTCTCCTGGCTCAAGTTTGC	TGCTGTTGGAACTTGAGGTG	NA	NA
Apc Exon 15-4	chr18: 26,785,044 + 26,786,394	AGGCAATCAAGAAGCCAGAA	AGGGCTCATCCAGACTCAGA	NA	NA
Apc Exon 15-5	chr18: 26,786,008 + 26,787,066	CTTCTGTCAGCTCCCTGGAC	GCTTGTGGCTTCTTCCTTTG	NA	NA
Apc Exon 15-6	chr18: 26,787,047 + 26,788,273	CAAAGGAAGAAGCCACAAGC	GACTTCAGGGACAGGATGGA	NA	NA
Apc Exon 15-7	chr18: 26,788,000 + 26,789,006	AGAGGAGGCCTTCCAGACTC	CAAGCCTGTTTGTTTGCTGA	NA	NA
Apc Exon 15-8	chr18: 26,788,977 + 26,790,026	CAGCAAAACCTCAGCAAACA	AAAGGAGTTCAGGCGGTTTT	NA	NA
Apc Exon 15-9	chr18: 26,789,791 + 26,790,663	AGATGGCACCTGCTGTTTCT	AAAGGACAATGGAACGGATG	NA	NA
SNP-76	chr1: 7,422,419 − 7,422,495	Biotin-TCCTCCAAAGTGTTCAACCA	GAGGAGGTCTGGTCGTGTTG	CTTCACTATCTGGATCTG	GAAACC[G/T]CGTGGTTGAA CACTT
SNP-102	chr1: 45,963,406 − 45,963,533	Biotin-ATGAGGGCTCAAAAA AATTACCTA	GTGTGCTTATTATTTGGGGCTATG	CTGTTATTTCCTTCTACCAT	G[C/T]AGATTCAGGGGATGAATT CATGTCA
SNP-82	chr1: 131,354,285 − 131,354,366	Biotin-GGGAGCAGGAAC AACTGAAG	CTCTTGGGTACCTGGCAAAG	TCCATCTGCCTCTGGCTG	AGC[T/C]GCTCTGCATCCCTCT
SNP-75	chr1: 158,844,845 + 158,844,927	Biotin-AAGAGAGGCGG TTCTCAGGT	ACAGAGAGCCGGGAGGAG	GAGGGCAGCGATGACGAC	ACCTC[G/A]CTCACCTGAGAA CCGCCT
SNP-83	chr1: 231,882,065 − 231,882,156	Biotin-CCACTGCGAAG AGATTCAC	AGTGCGGAACTCACTGTCTG	TGCCTTCAGCTTCCTCGA	GCCTCTT[T/C]GGTGAATCTCTTCG
SNP-2	chr1: 232,855,852 − 232,855,932	Biotin-ACCCCCATTGTTGT GTGACT	CCTGAGACGGAAAAACTGGA	AGGTAACTCTGGAAGGGA	GTGACTGCCCCCCAT[T/C]CTCC CTTCC
SNP-58	chr4: 75,039,070 + 75,039,137	CCAAGTCTGCTACTTCCCTCA	Biotin-GAACATCCATGCGACATCC	AGGGGCGACTCAGTGCTT	ACTCAGTGCTTT[C/G]TTTCCA AGCGGAT

Table S4. Cont.

Primer name	Location	Forward	Reverse	Sequencing	SNP with surrounding sequence
SNP-61	chr7: 71,279,026 + 71,279,107	GCATGCCCTGTATCAAGACC	Biotin-GCTTCCTTTCTTA TGTCTTGTTGG	GACCATGTATGCAGTGCT	ATGTATGCAGTGCTA[C/G]TGTC CTGGAGT
SNP-13	chr7: 128,110,496 − 128,110,607	Biotin-TGGCTAAGAAGTGAA TCACGAA	GCAGAGCCTCAGGCAATTTA	TTCATCTTCTTTGGGGTA	CAATGACTTAGA[A/G]GTACCCC AAAGAA
SNP-65	chr10: 25,359,997 + 25,360,102	Biotin-CCTTCAAGAGCTC CAAGCTG	GGGAATGAGAAGGGAGTTGA	TTGTGAACTAAAAGACAT	CTTCACTACCCC[A/G]TATGTCTTTT
SNP-27	chr15: 13,867,393 − 13,867,502	Biotin-ATTTGCTGGAATGG CTTCTC	GGAGGTTTCCTAATGGAGCTG	GAATTAAAGGACAGGTAA	CTTTCTTCTCTGAATCAC[A/G]C TTACCTG
SNP-28	chr16: 9,709,784 + 9,709,903	Biotin-GCCAAATACAACGC AGACCT	GGCATTGCCAGACTCTTGAC	GGACAGACTTGGCTGAAA	GTGCTCCATACTCAG[A/G]GATGT ACTTG
SNP-33	chr18: 838,18,221 − 83,818,328	Biotin-AGGCAAGGGA TGATTCTCCT	TGTCACTTGCCAATGAGGAC	AAAAGTTCCAGCTGTCAG	CCAGCCTGTCAGA[T/A]GGCCCCT AGGAG
Apc Min (Mouse)	chr18: 34,472,100 + 34,472,293	TTTTGACGCCAATCGACATG	Biotin-GATGGTAAGCACTGAG GCCAATA	CGTTCTGAGAAAGACA GAAG	
Apc Pirc (Rat)	chr18: 26,785,219 + 26,785,307	ATGTGAACCAGTCTTTGTG TCAG	Biotin- ATGCTGTTCTTCCTCAG AATAACG	GGAAGACGACTATGAAGAT	

The list of PCR primers used for either the mutation detection in tumors or confirmation and allele-specific pyrosequencing of SNPs. Alt primers were used to confirm the alternative splicing between exons 13 and 14 (Fig. 2). Chr, chromosome.

Other Supporting Information Files

Dataset S1 (PDF)

R source with main MCMC function

```
##################### mcmc.R ################################################

mcmc <- function(gDNA, cDNA, tissue, ugrid, nsave, nskip, hyper )
 {
  # inital state
        n <- length(gDNA)  ##number of tumors
        K <- nrow(ugrid)
        pi <- rep( 1/K, K )            ## simplex distribution for U's
        theta <- c(50,50,5,5,25,25)  ## 4 means and 2 variances (one cDNA and one gDNA variance
        mu.mat <- cbind( theta[1:2], theta[3:4], theta[c(1,4)] )
        sig2.mat <- cbind( theta[5:6], theta[5:6], theta[(5:6)] )
        id.norm <- nrow(ugrid)  ### (1,0,0) = normal tissue case


  # run parameters
        nscan <- nskip*nsave ## number of scans
        skipcount <- 0
        isave <- 1

  # storage info
        pisave <- matrix(NA, nsave,K)     ## storage for pi's
        thetasave <- matrix(NA, nsave,6)  ## storage for theta's

  # hyper-parameters
        alpha <- hyper$alpha    ## total mass of Dirichlet prior on pi
        sig0 <- hyper$sig0      ## prior guess at measurement standard deviation
        n0 <- hyper$n0       ## like prior sample size on variance

  ## run
  for( iscan in 1:nscan )
   {
    skipcount <- skipcount+1
    ########################################################################
    #
    # update U's [these are the mixing proportions over the three states
    # Do this by Gibbs

    mu <- mu.mat %*% t(ugrid)   ## 2 x ngrid  , rows for gDNA, cDNA means given U
    sdev <- sqrt( sig2.mat %*% t(ugrid)^2 ) ## 2 x ngrid,rows for gDNA, cDNA variance given U
                         ## note gDNA independent of cDNA given U, and both normal
    xx2 <- -0.5*( outer( mu[1,], gDNA, "-" )/sdev[1,] )^2
    yy2 <- -0.5*( outer( mu[2,], cDNA, "-" )/sdev[2,] )^2

    # ngrid x n
    logp.gDNA  <- xx2 - log( sdev[1,] )
```

```
logp.cDNA  <- yy2 - log( sdev[2,] )

logp <- logp.gDNA + logp.cDNA +  log(pi)
mm <- apply(logp,2,max)
foo <- exp( t(logp) - mm  )
foo.s <- apply(foo,1,sum)
pp <- foo/foo.s
cpp <- t( apply( pp, 1, cumsum ) )  ## cumulative dist of each U over grid
uu <- runif( n )
bar <- cpp > uu
U.id<- apply( bar, 1, which.max )  ## the first '1'
U.id[tissue=="Normal"] <- id.norm   ## use the known label for normal tissue
U  <- t( ugrid[U.id,] )         ## tumor sample mixing rates


#########################################################################
# Update theta:
# first the two horizontal means [...theta[1] and theta[3]...]

# compute the statistics for the full conditionals
sig2 <- sig2.mat %*% U^2  ## 2 x n  , rows for gDNA, cDNA variance given U
A1 <- sum( (1/sig2[1,])*(U[1,]+U[3,])*gDNA )
A2 <- sum( (1/sig2[1,])*(U[2,])*gDNA )
B1 <- sum( (1/sig2[1,])*(U[1,]+U[3,])^2 )
B2 <- sum( (1/sig2[1,])*(U[2,])^2 )
C <- sum( (1/sig2[1,])*(U[1,]+U[3,])*(U[2,]) )
mean.1 <- (A1*B2-A2*C)/(B1*B2-C^2)
mean.2 <- (A2*B1-A1*C)/(B1*B2-C^2)
v.1 <- B2/(B1*B2-C^2)
v.2 <- B1/(B1*B2-C^2)
rho <- -C/(B1*B2-C^2)
#Gibbs (from the bivariate normal posterior)
theta[1] <- rnorm(1, mean=mean.1, sd=sqrt(v.1) )
csd <- sqrt( B1*(1-C^2/(B1*B2))/(B1*B2-C^2) )
theta[3] <- cnorm(mu=(mean.2-C*(theta[1]-mean.1)/B2), sigma=csd )
# same thing (could be in parallel) for cDNA means theta[2], theta[4]
A1 <- sum( (1/sig2[2,])*(U[1,])*cDNA )
A2 <- sum( (1/sig2[2,])*(U[2,]+U[3,])*cDNA )
B1 <- sum( (1/sig2[2,])*(U[1,])^2 )
B2 <- sum( (1/sig2[2,])*(U[2,]+U[3,])^2 )
C <- sum( (1/sig2[2,])*(U[2,]+U[3,])*(U[1,]) )
mean.1 <- (A1*B2-A2*C)/(B1*B2-C^2)
mean.2 <- (A2*B1-A1*C)/(B1*B2-C^2)
v.1 <- B2/(B1*B2-C^2)
v.2 <- B1/(B1*B2-C^2)
rho <- -C/(B1*B2-C^2)
#Gibbs (from the bivariate normal posterior)
theta[2] <- rnorm(1, mean=mean.1, sd=sqrt(v.1) )
```

```
csd <- sqrt( B1*(1-C^2/(B1*B2))/(B1*B2-C^2) )
theta[4] <- cnorm(mu=(mean.2-C*(theta[2]-mean.1)/B2), sigma=csd )

# update matrix version
mu.mat <- cbind( theta[1:2], theta[3:4], theta[c(1,4)] )


## now the variances (one for gDNA theta[5], and one for cDNA theta[6], common to componen

## A Gibbs sample, using an inverse chi-square prior
shape.g <- (n0+n)/2
tmp.mean <- (U[1,] + U[3,])*theta[1] + U[2,]*theta[3]
bar <- U[1,]^2 + U[2,]^2 + U[3,]^2
tmp.stat <- sum( (gDNA -tmp.mean)^2/bar  )
rate.g <- (1/2)*( n0*sig0^2 + tmp.stat  )
theta[5] <- 1/rgamma(1,shape=shape.g,rate=rate.g)

shape.c <- (n0+n)/2
tmp.mean <- U[1,]*theta[2] + (U[2,]+U[3,])*theta[4]
tmp.stat <- sum( (cDNA -tmp.mean)^2/bar )
rate.c <- (1/2)*( n0*sig0^2 + tmp.stat )
theta[6] <- 1/rgamma(1,shape=shape.c,rate=rate.c)

# update matrix version
sig2.mat <- cbind( theta[5:6], theta[5:6], theta[5:6] )


#########################################################################
# Update pi  by Gibbs (ignore the normals...)

    tmp <- table(U.id[!(tissue=="Normal")])
    cnts <- rep(0,K)    ## the empirical distribution of U's on their grid
    names(cnts) <- 1:K
    cnts[ match(names(tmp),1:K ) ] <- tmp
    gg <- rgamma(K, shape=(cnts+alpha/K) )
    pi <- gg/sum(gg)

#########################################################################
#
# Store summary statistics periodically..
if( skipcount == nskip )
 {
  skipcount <- 0
  pisave[isave,] <- pi
  thetasave[isave,] <- theta
    ## maybe some posterior info for each U?
  print( isave )
  isave <- isave+1
```

```
     }

    }
   out <- list( data=cbind(gDNA,cDNA,tissue), mcmc=c(nsave,nskip), hyper=hyper,
                theta=thetasave, pi=pisave, ugrid=ugrid )
   out
  }


# a function to simulate a normal given it is positive
cnorm <- function(mu,sigma,nsim=1)
 {
  u <- runif(nsim)
  tmp <- pnorm( -mu/sigma )
  bar <- u*(1-tmp) + tmp
  foo <- qnorm(bar)
  # a bailout if bar=1 (i.e. if numerical error makes it hard to get the conditioned normal)
  x <- ifelse( foo < Inf,  mu + sigma*foo, sigma*log(1/runif(nsim)) )
  x
 }

###############################################################################
```

R source running posterior computations for Pirc.

```
################## pirc-2.R ##############################################
# Data:
        dat <- read.delim("data1.txt",header=TRUE)
        tissue <- dat[-37,3]  ## 37 is an extreme outlier
        gDNA <- dat[-37,7]    ## replicate-averaged numbers
        cDNA <- dat[-37,10]    ## replicate-averaged numbers

## grid of values supporting U
        load("grid.RData")  ## thresholded 40x40 --> 861 grid points
         ## p1keep, p2keep
        ugrid <- cbind( p1keep, p2keep, 1-p1keep-p2keep )
        ugrid[ugrid<0] <- 0
        ugrid[ugrid>1] <- 1  ## trim some round-off error

source("mcmc.R")  ## the main function

fit <- mcmc( gDNA=gDNA, cDNA=cDNA, tissue=tissue, ugrid=ugrid, nsave=5000,
        nskip=100, hyper=list( alpha=1, sig0=5, n0=1 ) )

save(fit, file="results/fit-pirc-2-long.RData" )
#########################################################################
```

R source to plot the estimated admixture distribution (Pirc shown)

```
############################ plot-pi.R ########################################

load("results/fit-pirc-2-long.RData") ## 500,000 run
pihat <- apply(fit$pi,2,mean)

load("grid.RData")

ugrid <- fit$ugrid
bary <-t(  T %*% t( ugrid[,2:3] ) + c(1/2, sqrt(3)/2)  )  ## barycentric coordinates

broman <- rev( rainbow(256, start=0, end=2/3 ) )

n <- length(pihat)
cls <-  rev( rainbow(n, start=0, end=2/3) )

ord <- order( pihat )

pdf( file="plots/pihat-pirc.pdf" )
par( mar=c(0,0,2,0) )
plot( bary[ord,1], bary[ord,2], pch=18, col=cls, axes=FALSE, xlab="", ylab="",
      xlim=c(-1/6, 7/6), ylim=c(-1/6, 1+ 1/6), cex=1.6 , main="Pirc: estimated admixture")
eps <- 0
text( 0, -1/16, labels="gLOH/cLOH", cex=1.5 )
text( 1, -1/16, labels="gMOH/cLOH", cex=1.5 )
text( 1/2, sqrt(3)/2+1/16, labels="gMOH/cMOH", cex=1.5 )
dev.off()

## get some summaries
# majority prob

ok1 <- ugrid[,1] > 1/2
ok2 <- ugrid[,2] > 1/2
ok3 <- ugrid[,3] > 1/2
p.maj <- c( sum( pihat[ok1] ), sum(pihat[ok2]), sum( pihat[ok3] ) )
names(p.maj) <- c("gMOH/cMOH", "gLOH,cLOH","gMOH/cLOH" )
foo <- apply(ugrid,1,which.max)
p.plur <- c( sum( pihat[foo==1] ), sum( pihat[foo==2] ), sum( pihat[foo==3] ) )
names(p.plur) <- names(p.maj)
probs.pirc <- cbind( p.maj, p.plur )
dimnames(probs.pirc)[[2]] <- c("Pr( majority )","Pr( plurality )" )

###############################################################################
```