NIH Public Access
Author Manuscript
Genes Chromosomes Cancer. Author manuscript; available in PMC 2013 August 13.

Published in final edited form as:
Genes Chromosomes Cancer. 2009 January ; 48(1): 1–9. doi:10.1002/gcc.20614.

Oncogenetic Tree Model of Somatic Mutations and DNA Methylation in Colon Tumors

Carol Sweeney[1,*], **Kenneth M. Boucher**[1], **Wade S. Samowitz**[1], **Roger K. Wolff**[1], **Hans Albertsen**[1], **Karen Curtin**[1], **Bette J. Caan**[2], and **Martha L. Slattery**[1]

[1]Health Sciences Center, University of Utah, Salt Lake City, UT

[2]Division of Research, Kaiser Permanente Medical Care Program, Oakland, CA

Abstract

Our understanding of somatic alterations in colon cancer has evolved from a concept of a series of events taking place in a single sequence to a recognition of multiple pathways. An oncogenetic tree is a model intended to describe the pathways and sequence of somatic alterations in carcinogenesis without assuming that tumors will fall in mutually exclusive categories. We applied this model to data on colon tumor somatic alterations.

Methods—An oncogenetic tree model was built using data on mutations of *TP53, KRAS2, APC*, and *BRAF* genes, methylation at CpG sites of *MLH1* and *p16* genes, methylation in tumor (MINT) markers, and microsatellite instability (MSI) for 971 colon tumors from a population-based series.

Results—Oncogenetic tree analysis resulted in a reproducible tree with three branches. The model represents methylation of MINT markers as initiating a branch and predisposing to MSI, methylation of *MHL1* and *p16*, and *BRAF* mutation. *APC* mutation is the first alteration in an independent branch and is followed by *TP53* mutation. *KRAS2* mutation was placed a third independent branch, implying that it neither depends on, nor predisposes to, the other alterations. Individual tumors were observed to have alteration patterns representing every combination of one, two, or all three branches.

Conclusions—The oncogenetic tree model assumptions are appropriate for the observed heterogeneity of colon tumors, and the model produces a useful visual schematic of the sequence of events in pathways of colon carcinogenesis.

Over two decades, the model of molecular events in colon cancer has evolved from a schematic of a linear path (Vogelstein et al., 1988; Fearon and Vogelstein, 1990; Kinzler and Vogelstein, 1996) to a concept of a more complex set of contributing pathways resulting in heterogeneous tumors (Smith et al., 2002; Wong et al., 2007). The Vogelstein model estimated the sequence of events over time and provided a useful visual schematic. More recent data on somatic alterations in colon cancer continue to support the conceptual model that one alteration can predispose to a subsequent specific alteration (Samowitz et al., 2001b, 2005, 2007; Smith et al., 2002; Iacopetta et al., 2006). For example, several independent data sets are in agreement that mutations of the *BRAF* gene are much more frequent among tumors exhibiting CpG island methylation phenotype (CIMP) than in tumors without CIMP (Kambara et al., 2004; Nagasaka et al., 2004; Samowitz et al., 2005; Li et al., 2006).

[*]Correspondence to: Carol Sweeney, Division of Epidemiology, 30 N 1900 E, AC221, Salt Lake City, UT 84132. Phone: 801-581-5865; Fax: 801-581-3623; e-mail: carol.sweeney@hsc.utah.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Fearon and Vogelstein described their model as a "preferred sequence", recognizing that the earlier events were not always essential for the later ones to occur (Fearon and Vogelstein, 1990). As the number of recognized somatic alterations has increased, a linear sequence no longer seems an adequate description of the pathways. A classification tree modeling approach identified three colon tumor categories that explained much of the clustering of somatic mutations and DNA methylation (Shen et al., 2007). Other modeling strategies can be applied as tools to further explore the timing and relations of somatic events in colon carcinogenesis.

Oncogenetic tree modeling applies an algorithm to data on somatic alterations to infer a model of the sequence of genetic events in tumor progression (Desper et al., 1999). A tree structure illustrating relationships among alterations is produced by an algorithm applied to the data, without a priori specification of pathways or groups. The oncogenetic tree approach assumes that there may be multiple pathways, or branches, of alterations in evolution of the tumor. The modeling objective is to define pathways, not to place tumors in mutually exclusive categories. The model allows that more than one carcinogenic pathway may have taken place in an individual tumor. This model has been applied to analysis of comparative genomic hybridization data describing gain or loss of chromosomal regions for several tumor types (Desper et al., 1999; Jiang et al., 2000; Radmacher et al., 2001; Huang et al., 2002; Rennstam et al., 2003; Huang et al., 2004; Cremer et al., 2005; Gunawan et al., 2007; Jiang et al., 2007), but to our knowledge has not been applied to colon cancer, nor to data on mutation and methylation of specific genes. We examine application of this model to existing data describing somatic mutations, DNA methylation, and microsatellite instability (MSI) in colon cancer.

MATERIALS AND METHODS

The oncogenetic tree model was applied to existing data on somatic alterations in a population-based series of colon cancer cases. Characteristics of the study subjects, associations between the somatic alterations and epidemiologic variables in this population, and pairwise relationships between some alterations have been reported elsewhere (Samowitz et al., 2000, 2001b, 2005, 2007; Slattery et al., 2000a, 2000b, 2001a, 2001b, 2002, 2007). The methods for determining subject eligibility, for obtaining tumor tissue, and for laboratory assays have been previously described and are summarized briefly below.

Study Subjects and Characterization of Somatic Alterations

Eligible subjects were individuals ages 30 to 79 years with a first diagnosis of colon cancer (ICD02 18.0 and 18.2–18.9) between October, 1991 and September, 1994; subjects were identified as cases for a population-based case-control study through the Utah Cancer Registry, the Northern California Cancer Registry, and the Sacramento Cancer Registry (Slattery et al., 1997, 2003). In northern California, the study population was restricted to members of the Kaiser Permanente Medical Care Program (KPMCP). This research was conducted in accordance with human subjects research protocols approved at each responsible institution. Tumors thought to be HNPCC-related, i.e. with a deleterious mutation detected in the germline DNA sequence of the *MLH1* or *MSH2* genes (Samowitz et al., 2001a), were excluded from the data set.

Tumor Tissue and Laboratory Assays

Archived, paraffin-embedded tumor material for eligible cases was obtained from hospitals in Utah and from KPMCP (Slattery et al., 2000c). Tumor tissue was microdissected from sections for DNA extraction. Of 1917 eligible cases, tumor tissue was obtained and useable DNA was extracted (i.e. at least one assay was successful) for 1539 or 80.3%.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Microsatellite instability (MSI), mutations of the *TP53, KRAS2, BRAF*, and *APC* genes, and methylation of selected CpG sites were assayed for these samples as detailed in previous reports for this study population. Briefly, microsatellite instability (MSI) was evaluated by assaying for a panel of 10 tetranucleotide repeats (Samowitz and Slattery, 1997a), and for mutations in two poly-A mononucleotide repeats known to indicate MSI, the A10 region of the transforming growth factor beta type 2 gene (*TGFβRII*) (Samowitz and Slattery, 1997b), and BAT-26 (Samowitz et al., 1999); instability was recognized by the length of polymerase chain reaction (PCR) products, compared to normal DNA, when separated on polyacrylamide gels. Mutations in codons 12 and 13 of the *KRAS2* gene were detected by PCR amplification and sequencing (Samowitz et al., 2000). Mutations of exons and intron–exon boundaries of exons 5–8 of *TP53* were detected by PCR amplification and single-strand conformational polymorphism (SSCP) analysis followed by sequencing of DNA corresponding to abnormal bands from SSCP (Samowitz et al., 2002). Methylation-specific PCR of sodium bisulfite-modified DNA was used to assay for methylation at CpG sites in promoter regions of the *MLH1* and *p16 (CDKN2A)* genes, and at three non-gene MINT (methylated in tumor) sites, MINT1, MINT2, and MINT31, that are considered to be markers of CpG island methylation phenotype (CIMP) (Samowitz et al., 2005). Methylation-specific PCR procedures included primers for both methylated and unmethylated products. Mutations of the adenomatous polyposis coli (*APC*) gene were detected, for a subset of 90 tumors, by PCR amplification and sequencing of the entire coding region of the gene (Samowitz et al., 2007).

Data Analysis

Each tumor was categorized as positive or negative for each somatic alteration. For MSI, the majority of tumors were classified as MSI-positive or MSI-negative based on Bat-26 status; tumors which did not have a Bat-26 result were classified using *TGFβRII*; tumors which had neither Bat-26 nor *TGFβRII* results were classified using the panel of 10 tetranucleotide repeats (Slattery et al., 2000a). A tumor was coded as having a *TP53* mutation if any non-synonymous mutation in exons and intron–exon boundaries of exons 5–8 of *TP53* was detected. *KRAS2* was scored as mutated if any mutation was present in codons 12 or 13. *APC* mutation was scored as positive if a nonsense or out-of-frame insertion or deletion mutation was detected. Methylation at CpG sites in the promoter of the *MLH1* gene or the *p16* gene were categorized as present or absent and represented by a dichotomous variable for each gene. A single variable was used to indicate methylation at one or more of the MINT CpG sites because these non-gene sites had been selected as collective markers of a phenotype.

The oncogenetic tree model requires that every variable be defined for each observation. Data from one or more DNA methylation assays were missing for n = 490 tumors. Insufficient quantity of DNA extracted from small tumors was often the limitation for methylation assays, and as a result small tumors are under-represented in the analysis. In addition, 37 cases were missing data for TP53 mutation, 65 missing KRAS2 mutation, 34 missing MSI, and 74 missing BRAF mutation. The remaining 971 cases were evaluated in the oncogenetic tree model.

Pairwise correspondence between the seven alterations with complete data was evaluated by calculating the tetrachoric correlation coefficient, a measure appropriate for representing correlation between dichotomous variables (Kraemer, 2006). Correlations with *APC* could not be calculated because the sample with *APC* data was non-representative. A primary oncogenetic tree model, n = 971, was developed based on data seven dichotomous variables: *TP53* mutation, MINT methylation, *KRAS2* mutation, *p16* methylation, *MLH1* methylation, MSI, and *BRAF* mutation. A second tree was developed for these seven alterations and, in

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

addition, *APC* mutation. The second model was based on the n = 90 tumors with *APC* data and used weighting of observations to account for non-representativeness.

Oncogenetic Tree Model

An oncogenetic tree is a rooted directed tree with edges pointing away from the root (Desper et al., 1999; Desper et al., 2000). The root presents the state of tissue with none of the measured somatic alterations. Each of the other nodes is associated with a particular somatic alteration. According to the oncogenetic tree model, for an alteration to occur in a particular tumor, all of the alterations corresponding to nodes that lie on the directed path from the root must be present in the tumor. Each edge is labeled by the "transition probability", i.e. the conditional probability that a tumor will acquire the alteration at the next node away from the root, given that the alteration at the node proximal to the root is already present. When two edges emanate from the same node, neither of the alterations on the next node is required for the other to occur. Each tumor corresponds to a "subtree" of the oncogenetic tree containing all the alterations in the tumor. The method used to fit the oncogenetic tree is described in greater detail elsewhere (Szabo and Boucher, 2002; Szabo and Boucher, 2008).

The oncogenetic trees were constructed using a simple algorithm. The parent of each node was found by maximizing the weight function $w(v_i, v_j)=\log\frac{p_{ij}}{p_i(p_i+p_j)}$. If the true oncogenetic tree is not skewed, meaning that no excessively high positive correlations are observed between alterations on different branches, the algorithm is guaranteed to reconstruct the correct tree. The Desper model proceeds to a second stage, in which transition probabilities and false positive and false negative error probabilities are estimated by minimizing the error function $Max_i|O_i - E_i|$, where O_i represents the observed frequency of the i^{th} alteration and E_i represents the expected frequency of the i^{th} alteration under the model. This stage of the model assumes that pairs of alterations may either be positively correlated or independent. Another approach is to directly estimate, from observed data, the probability for each alteration conditional on the proximal alteration.

A nonparametric bootstrap resampling method was used to estimate the amount of sampling variability in the fitted tree. The bootstrap method samples from the rows of the data with replacement N times, where N is the size of the original sample. An oncogenetic tree is fit to the resampled data. We repeated the procedure M = 1000 times.

RESULTS

Among the seven somatic alterations assayed for 971 tumors, the *TP53* mutation was the most frequent and was detected in 47% of colon cancers. *BRAF* mutation was the least common, found in 9% (Table 1). Pairwise tetrachoric correlations revealed both positive and negative relations among the seven somatic alterations (Table 1). Methylation of *MLH1* was highly correlated with MSI, $\rho = 0.93$. *BRAF* mutation exhibited strong positive correlations, $\rho = 0.67$ to 0.74, with each of four other somatic alterations: methylation of MINT markers, methylation of the *p16* gene, methylation of the *MLH1* gene, and MSI. The latter four alterations were mutually positively correlated. *KRAS2* mutation was strongly negatively correlated with *BRAF* mutation, $\rho = -0.72$, and *KRAS2* was negatively correlated with every other alteration except MINT methylation. *TP53* mutation was negatively correlated with every other alteration examined, although the magnitude of the negative correlations with p16 methylation, MINT methylation, and *KRAS2* were small. The negative correlations between some pairs of alterations violate an assumption used in calculating edge weights in the second stage of the Desper (Desper et al., 1999) model. Therefore transition probabilities shown on the oncogenetic tree were based on observed data rather than the second-stage model.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The oncogenetic tree resulting from these data showed three main branches departing from the root (Figure 1). Figure 1 primarily presents information from the n = 971 model, shown in solid lines. The placement of APC in the model (dashed lines) is based on the n = 90 secondary model. *KRAS2* mutation was located on one branch emanating from the root. The *KRAS2* branch did not continue to any other alteration. A second branch led to *APC* mutation and continued to *TP53* mutation. In the n = 971 model, without *APC* data, this branch led from the root directly to *TP53* mutation. The *APC-TP53* branch did not continue beyond the *TP53* node. The third main branch emanating from the root led to methylation of MINT markers as a first node, then branched again to *p16* methylation and MSI. The MSI sub-branch continued with high probability to *MLH1* methylation and then to *BRAF* mutation. Because this main branch includes MINT and other DNA methylation variables, we will refer to it as the CIMP branch.

In the 1000 replicate bootstrap resampling procedure for the primary, n = 971 tree, the original tree was replicated 462 times. Among trees represented at least 30 times from the 1000 resampled oncogenetic trees, all showed the same positions for *KRAS2, TP53*, MINT, and *p16* as seen in Figure 1 (solid lines). That is to say, *KRAS2, TP53*, and MINT were always on three independent branches emanating from the root, and *p16* was always a direct descendent of MINT. On all bootstrap trees, *MSI, MLH1*, and *BRAF* were direct or indirect descendents of MINT. In the second most frequent tree in the bootstrap resampling, n = 206, BRAF descended from p16 rather than from MSI. Other configurations of *BRAF, p16*, and MSI were observed in less frequent trees.

We examined the combinations of alterations present in individual tumors to consider what subtrees of the main tree were represented. The subtrees were based on 7 alterations (omitting *APC*). No tumor had all seven of the alterations. The nine most frequent subtrees account for the alteration patterns observed in 75% of tumors (Figure 2). Subtrees containing alterations from only one of each of the three main branches were observed, as were subtrees with all possible combinations of two main branches, and subtrees with alterations from all three main branches. A subtree containing *TP53* mutation, with no alterations from the other two branches, was the most common subtree, accounting for 19.6% of tumors. The next most common subtrees, in order, represent the 13.2% of tumors that had none of the 7 alterations examined, 7.6% with *KRAS2* mutation only, 7.4% with *KRAS2* and MINT methylation, 6.5% with *TP53* and *KRAS2*, and 5.9% with MINT methylation only. If an alteration that the model places farther from the root was present in a tumor but one or more alterations proximal to the root on the same branch was not detected, the pattern of alterations does not form a subtree consistent with the main tree. Combinations of alterations inconsistent with the main tree were observed in 108 of 971 tumors (11%). Many of these, n = 49, had one or more alterations in the CIMP branch but without the first alteration in that pathway, MINT methylation. Another group of tumors, n = 41, had a BRAF mutation but did not have MSI.

DISCUSSION

We applied an oncogenetic tree model to data representing somatic alterations in colon cancer. The resulting model showed three main branches. The branch leading to *KRAS2* and terminating at that node indicates that, according to this model, *KRAS2* mutations do not depend on or predispose to any of the other somatic alterations for which data were available. The interpretation of the *APC-TP53* branch, based on the oncogenetic tree assumptions, is that *APC* alteration occurs before, and as a precondition for, *TP53*. The structure of the third branch is consistent with a mechanism in which CIMP, indicated here by MINT markers, is initiated before, and is a precondition for, MSI, which is usually accompanied by *MLH1* methylation. *BRAF* mutation appears farther from the root on the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

same branch, characterizing it as a later and less frequent event in the CIMP pathway. *p16* methylation may be present in tumors with the CIMP pathway but its presence does not predispose to or depend on MSI. The tree model of the sequence and interdependence of these alterations can be regarded as robust, because bootstrap resampling of the large data set reliably reproduced the three major pathways.

The oncogenetic tree structure was drawn based on an algorithm applied to colon cancer somatic alteration data, and does not represent any *a priori* groupings of alterations. The resulting model is consistent with existing concepts of pathways of colon tumorigenesis, e.g. that *APC* mutation is an early event and predisposes to other alterations (Fearon and Vogelstein, 1990), and that epigenetic changes predispose to certain genetic changes (Baylin and Ohm, 2006). A departure from early colon cancer models is that *APC* is not represented as predisposing to *KRAS2* mutation. *APC* mutation and *KRAS2* mutation were positively related in this data set when considered in a pairwise comparison, as previously reported (Samowitz et al., 2007), but both *APC* and *KRAS2* are negatively correlated with *BRAF* mutation; apparently when *BRAF* is taken into account in a separate pathway, occurrence of *KRAS2* mutation does not appear to be conditional on *APC* mutation. The independence of the *APC* branch from both the *KRAS2* and the CIMP/MSI/*BRAF* branches is in keeping with the more recent concept that some tumors containing MSI and *BRAF* mutations, or *KRAS2* mutations, develop from hyperplastic polyps and are distinct from those that develop from *APC*-mutated adenomatous polyps (O'Brien et al., 2006; Jass, 2007). Obtaining complete data on *APC* mutational status is problematic because of the large size of the gene, and therefore a data gap exists in this and other data sets; additional data on *APC* mutation in relation to *KRAS2* and *BRAF* mutations in sporadic tumors would be of interest to better understand these relationships.

In the CIMP branch, MSI appears nearer to the root than *MLH1* methylation. This is inconsistent with the accepted sequence of events in which suppression of *MLH1* expression disrupts mismatch repair and leads to MSI. MSI is thought to be dependent on *MLH1* methylation in the majority of MSI-positive tumors, but MSI can also occur in the absence of *MLH1* methylation. Of the 117 MSI positive tumors in this data set, 27 (23%) were not methylated at *MLH1*. (Another study reported that 34% of MSI positive tumors were unmethylated at *MLH1* (de Vogel et al., 2008).) MSI tumors without *MLH1* methylation represent alternative pathways to MSI, including HNPCC cases and an uncommon pathway in which MSI occurs independently of *MLH1* methylation and CIMP (Samowitz et al., 2005). Tumors identified as HNPCC-related (Samowitz et al., 2001a) were omitted from this data set, but HNPCC patients whose inherited defect was not detected by sequencing of *MLH1* and *MSH2* exons would be unrecognized. The empirically derived oncogenetic tree model interprets the 27 MSI positive, *MLH1* methylation negative tumors as indicating that MSI can occur independently of *MLH1* methylation and places MSI before *MLH1* methylation in the pathway. The complexity of multiple pathways to MSI can not be taken into account by the model.

The novel aspect of the oncogenetic tree model applied to colon cancer is that the model's goal is to define pathways rather than categories of tumors. This distinguishes it from classification tree analysis of tumor data. A recent classification tree analysis applied to colon tumor alterations resulted in a useful model, and overall similarities to our results can be noted, e.g. clustering of CIMP characteristics, with or without MSI and *BRAF* mutations, and relative independence of *TP53* and *KRAS2* mutations from CIMP (Shen et al., 2007). The oncogenetic tree model assumes that pathways shown as diverging are independent and may or may not co-occur in the same tumor. Subtrees of the oncogenetic model represented by individual tumors were diverse, with every possible combination of one, two, or all three of the main branches represented. Thus the observed heterogeneity of the individual tumors

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

is suited to the oncogenetic tree model assumption that pathways may occur together or separately in a tumor.

The presence of distinct pathways that are present in different combinations fits well with the conceptual model that in order to become a tumor a cell must develop a set of "acquired capabilities" (Hanahan and Weinberg, 2000). Hanahan and Weinberg proposed that there would be several possible gene defects that could result in a cell acquiring the phenotype for each capability. It is interesting to note that among the subtrees in our data set that exhibited alterations in only one of the three main branches, that branch was most commonly the pathway including a mutation of *TP53*, an alteration which is thought to contribute to several of the acquired capabilities. A subtree with the CIMP branch alone was less frequent than subtrees with a combination of CIMP pathway and *TP53* or *KRAS2* alteration. An interpretation of this subtree pattern is that epigenetic alterations contribute to the carcinogenetic process but require additional genetic alterations to confer complementary capabilities.

Desper's model assumed that alterations in oncogenesis would only be positively related or independent. In the data from these colon tumors, negative correlations were observed. This departure from the model assumptions has no effect on development of the tree structure, which is non-parametric, but has implications for the estimation of transition probabilities. Because the model assumption was not met, we presented transition probabilities based on observed data rather than on the second stage of the oncogenetic tree model. Negative correlations are unexpected if taken at face value, i.e. as indicating that presence of one alteration prevented another. However, tumor tissue does not represent all alterations that occurred in colon cells, but only those combinations of alterations that resulted in tumors. Negative correlations can be explained by a selection process. For example, *KRAS2* and *BRAF* mutations do probably occur independently in many cells, most of which never become tumors. These mutational events would be rare in the underlying cell populations, and the probability of a cell having both would be small. The normal functions of the *KRAS2* and *BRAF* gene products fall in the same biological pathway, and therefore it seems plausible that one or the other of these mutations, but not both, are needed to confer a certain acquired capability, i.e. to compromise the normal signaling mechanism. Selection of cells with one of these mutations, in the presence of additional alterations, including ones not assayed in our data set, to develop into tumors would result in the observed high frequency, and apparent negative correlation, of the two mutations among tumors.

Our analysis is based on a relatively small number of somatic alterations characterized in a large set of tumors. This is in contrast to current research using array methods, which result in data sets with large numbers of genes represented (e.g. data reported by Wood et al. (Wood et al., 2007)), but typically for small numbers of tumors. For our population-based study, paraffin-embedded, preserved tissue was the only available source of tumor DNA. The limited quantity and fragmented nature of DNA extracted from paraffin placed limitations on the number and types of alterations that could be investigated. *APC* mutational status has not been characterized in the entire data set due to limitations of resources and of quantity of DNA for each tumor. We previously reported that 71.5% of CIMP-low, microsatellite stable tumors contained an *APC* mutation (Samowitz et al., 2007). Therefore it is probable that most tumors in subtrees 1 and 2 in Figure 2, showing *TP53* mutation only or showing none of the seven alterations that were assessed for the complete data set, do have *APC* mutations.

Another potential limitation of the data is that alterations may have been missed. As has been discussed in previous reports on this study population (Samowitz et al., 2000, 2002, 2005, 2007), the estimated prevalence of these alterations has differed among studies in the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

literature. Differences may be related to selection of the study population as well as to differences in methods for detecting alterations. *TP53* mutation in our study population was somewhat more frequent than in another large study (Soong et al., 2000) but very similar in frequency to other reports (Borresen-Dale et al., 1998; Kahlenberg et al., 2000). Methylation markers used to define CIMP have differed across studies; the fraction that was determined to be CIMP high in our population is similar to that in other large studies (Hawkins et al., 2002; van Rijnsoever et al., 2002). The 33% prevalence of mutations in *KRAS2* this study population is lower than in some studies, particularly compared to studies that sequenced additional regions of the gene beyond codons 12 and 13 (Breivik et al., 1994; Wadler et al., 1997; Ahnen et al., 1998; Andreyev et al., 1998; Hardingham et al., 1998). The prevalence of *BRAF* mutations is somewhat lower than what was observed in sporadic tumors in one study (Kambara et al., 2004) but higher than reported by another (Deng et al., 2004). The proportion of colon tumors carrying *APC* mutations in the present study is essentially identical to the 60% originally reported by Powell et al. (Powell et al., 1992). The methods used to detect mutations or methylation in DNA extracted from paraffin-embedded material are unlikely to indicate an alteration if none in fact is present, but it is possible that an assay could find no alteration when one actually was present. A false negative could be present if the mutation was a large deletion, if a normal allele was amplified but PCR failed for the mutated allele, or if a mutation was outside the region of the gene assayed. The reliability of the model on resampling indicates that the overall model structure is correct, even in the presence of false negative data points.

The 11% of individual tumors with alteration patterns inconsistent with the main tree may be evidence that the model assumption that an alteration farther from the root is always dependent on the proximal alteration in the same branch above is overly strict. However, most tumors did form subtrees consistent with the main tree, so that even with this strict assumption, the model is useful. Some of the 49 tumors with no methylation in MINT markers but positive for other alterations in the CIMP branch may represent false negative results from the methylation assay, but they also may be cases of departure from the preferred sequence, a situation acknowledged by Fearon and Vogelstein (Fearon and Vogelstein, 1990). The 41 tumors with *BRAF* mutation but without evidence of MSI are notable; given that there were only 109 tumors with *BRAF* mutation in overall dataset, it seems likely that these 41 do not represent laboratory errors but instead an indicate alternative pathway not accounted for by the strict model assumption that each alteration appears only once in the tree. The assumption that each alteration will appear once also does not allow the alternative pathways of *MLH1* followed by MSI vs. MSI independent of *MLH1* to appear in the schematic. However, given the type of data used in the present example, presence and absence of alterations measured at one undetermined time, this simplifying assumption is necessary, and does result in a model that is consistent with the patterns of alterations in the majority of tumors.

Future useful extensions of oncogenetic tree modeling approach would be to incorporate stage at diagnosis data when reconstructing the sequence of events, and to allow missing data for individual alterations. Oncogenetic tree models could be applied to other somatic alteration data sets, and for comparisons of trees and subtrees for different tumor sites or subsites. The oncogenetic tree approach provides an explanatory model that describes a biologically plausible sequence of events for the major pathways of colon carcinogenesis, fits the heterogeneity of colon cancer well, and presents a useful visual representation of the pathways.

Acknowledgments

The authors thank Michael Hoffman and Erica Wolff for technical assistance with this study.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Supported by: U.S. National Institutes of Health grants CA48998 and CA61757. Additional support from the Biostatistics Core, NCI Cancer Center Support Grant P30CA042014, and Survey Methods and Data Collection Core, Huntsman Cancer Institute/Huntsman Cancer Foundation.

REFERENCES

Ahnen DJ, Feigl P, Quan G, Fenoglio-Preiser C, Lovato LC, Bunn PA Jr. Stemmerman G, Wells JD, Macdonald JS, Meyskens FL Jr. Ki-ras mutation and p53 overexpression predict the clinical behavior of colorectal cancer: a Southwest Oncology Group study. Cancer Res. 1998; 58:1149–1158. [PubMed: 9515799]

Andreyev HJ, Norman AR, Cunningham D, Oates JR, Clarke PA. Kirsten ras mutations in patients with colorectal cancer: the multicenter "RASCAL" study. J Natl Cancer Inst. 1998; 90:675–684. [PubMed: 9586664]

Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat Rev Cancer. 2006; 6:107–116. [PubMed: 16491070]

Borresen-Dale AL, Lothe RA, Meling GI, Hainaut P, Rognum TO, Skovlund E. TP53 and long-term prognosis in colorectal cancer: mutations in the L3 zinc-binding domain predict poor survival. Clin Cancer Res. 1998; 4:203–210. [PubMed: 9516972]

Breivik J, Meling GI, Spurkland A, Rognum TO, Gaudernack G. K-ras mutation in colorectal cancer: relations to patient age, sex and tumour location. Br J Cancer. 1994; 69:367–371. [PubMed: 8297737]

Cremer FW, Bila J, Buck I, Kartal M, Hose D, Ittrich C, Benner A, Raab MS, Theil AC, Moos M, Goldschmidt H, Bartram CR, Jauch A. Delineation of distinct subgroups of multiple myeloma and a model for clonal evolution based on interphase cytogenetics. Genes Chromosomes Cancer. 2005; 44:194–203. [PubMed: 16001433]

de Vogel S, Bongaerts BW, Wouters KA, Kester AD, Schouten LJ, de Goeij AF, de Bruine AP, Goldbohm RA, van den Brandt PA, van Engeland M, Weijenberg MP. Associations of dietary methyl donor intake with MLH1 promoter hypermethylation and related molecular phenotypes in sporadic colorectal cancer. Carcinogenesis. 2008

Deng G, Bell I, Crawley S, Gum J, Terdiman JP, Allen BA, Truta B, Sleisenger MH, Kim YS. BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. Clin Cancer Res. 2004; 10:191–195. [PubMed: 14734469]

Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, Schaffer AA. Inferring tree models for oncogenesis from comparative genome hybridization data. J Comput Biol. 1999; 6:37–51. [PubMed: 10223663]

Desper R, Jiang F, Kallioniemi OP, Moch H, Papadimitriou CH, Schaffer AA. Distance-based reconstruction of tree models for oncogenesis. J Comput Biol. 2000; 7:789–803. [PubMed: 11382362]

Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990; 61:759–767. [PubMed: 2188735]

Gunawan B, von Heydebreck A, Sander B, Schulten HJ, Haller F, Langer C, Armbrust T, Bollmann M, Gasparov S, Kovac D, Fuzesi L. An oncogenetic tree model in gastrointestinal stromal tumours (GISTs) identifies different pathways of cytogenetic evolution with prognostic implications. J Pathol. 2007; 211:463–470. [PubMed: 17226762]

Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000; 100:57–70. [PubMed: 10647931]

Hardingham JE, Butler WJ, Roder D, Dobrovic A, Dymock RB, Sage RE, Roberts-Thomson IC. Somatic mutations, acetylator status, and prognosis in colorectal cancer. Gut. 1998; 42:669–672. [PubMed: 9659162]

Hawkins N, Norrie M, Cheong K, Mokany E, Ku SL, Meagher A, O'Connor T, Ward R. CpG island methylation in sporadic colorectal cancers and its relationship to microsatellite instability. Gastroenterology. 2002; 122:1376–1387. [PubMed: 11984524]

Huang Q, Yu GP, McCormick SA, Mo J, Datta B, Mahimkar M, Lazarus P, Schaffer AA, Desper R, Schantz SP. Genetic differences detected by comparative genomic hybridization in head and neck

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

squamous cell carcinomas from different tumor sites: construction of oncogenetic trees for tumor progression. Genes Chromosomes Cancer. 2002; 34:224–233. [PubMed: 11979556]

Huang Z, Desper R, Schaffer AA, Yin Z, Li X, Yao K. Construction of tree models for pathogenesis of nasopharyngeal carcinoma. Genes Chromosomes Cancer. 2004; 40:307–315. [PubMed: 15188453]

Iacopetta B, Grieu F, Li W, Ruszkiewicz A, Caruso M, Moore J, Watanabe G, Kawakami K. APC gene methylation is inversely correlated with features of the CpG island methylator phenotype in colorectal cancer. Int J Cancer. 2006; 119:2272–2278. [PubMed: 16981189]

Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology. 2007; 50:113–130. [PubMed: 17204026]

Jiang F, Desper R, Papadimitriou CH, Schaffer AA, Kallioniemi OP, Richter J, Schraml P, Sauter G, Mihatsch MJ, Moch H. Construction of evolutionary tree models for renal cell carcinoma from comparative genomic hybridization data. Cancer Res. 2000; 60:6503–6509. [PubMed: 11103820]

Jiang HY, Huang ZX, Zhang XF, Desper R, Zhao T. Construction and analysis of tree models for chromosomal classification of diffuse large B-cell lymphomas. World J Gastroenterol. 2007; 13:1737–1742. [PubMed: 17461480]

Kahlenberg MS, Stoler DL, Rodriguez-Bigas MA, Weber TK, Driscoll DL, Anderson GR, Petrelli NJ. p53 tumor suppressor gene mutations predict decreased survival of patients with sporadic colorectal carcinoma. Cancer. 2000; 88:1814–1819. [PubMed: 10760757]

Kambara T, Simms LA, Whitehall VL, Spring KJ, Wynter CV, Walsh MD, Barker MA, Arnold S, McGivern A, Matsubara N, Tanaka N, Higuchi T, Young J, Jass JR, Leggett BA. BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. Gut. 2004; 53:1137–1144. [PubMed: 15247181]

Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell. 1996; 87:159–170. [PubMed: 8861899]

Kraemer HC. Correlation coefficients in medical research: from product moment correlation to the odds ratio. Stat Methods Med Res. 2006; 15:525–545. [PubMed: 17260922]

Li WQ, Kawakami K, Ruszkiewicz A, Bennett G, Moore J, Iacopetta B. BRAF mutations are associated with distinctive clinical, pathological and molecular features of colorectal cancer independently of microsatellite instability status. Mol Cancer. 2006; 5:2. [PubMed: 16403224]

Nagasaka T, Sasamoto H, Notohara K, Cullings HM, Takeda M, Kimura K, Kambara T, MacPhee DG, Young J, Leggett BA, Jass JR, Tanaka N, Matsubara N. Colorectal cancer with mutation in BRAF, KRAS, and wild-type with respect to both oncogenes showing different patterns of DNA methylation. J Clin Oncol. 2004; 22:4584–4594. [PubMed: 15542810]

O'Brien MJ, Yang S, Mack C, Xu H, Huang CS, Mulcahy E, Amorosino M, Farraye FA. Comparison of microsatellite instability, CpG island methylation phenotype, BRAF and KRAS status in serrated polyps and traditional adenomas indicates separate pathways to distinct colorectal carcinoma end points. Am J Surg Pathol. 2006; 30:1491–1501. [PubMed: 17122504]

Powell SM, Zilz N, Beazer-Barclay Y, Bryan TM, Hamilton SR, Thibodeau SN, Vogelstein B, Kinzler KW. APC mutations occur early during colorectal tumorigenesis. Nature. 1992; 359:235–237. [PubMed: 1528264]

Radmacher MD, Simon R, Desper R, Taetle R, Schaffer AA, Nelson MA. Graph models of oncogenesis with an application to melanoma. J Theor Biol. 2001; 212:535–548. [PubMed: 11597184]

Rennstam K, Ahlstedt-Soini M, Baldetorp B, Bendahl PO, Borg A, Karhu R, Tanner M, Tirkkonen M, Isola J. Patterns of chromosomal imbalances defines subgroups of breast cancer with distinct clinical features and prognosis. A study of 305 tumors by comparative genomic hybridization. Cancer Res. 2003; 63:8861–8868. [PubMed: 14695203]

Samowitz WS, Albertsen H, Herrick J, Levin TR, Sweeney C, Murtaugh MA, Wolff RK, Slattery ML. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. Gastroenterology. 2005; 129:837–845. [PubMed: 16143123]

Samowitz WS, Curtin K, Lin HH, Robertson MA, Schaffer D, Nichols M, Gruenthal K, Leppert MF, Slattery ML. The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. Gastroenterology. 2001a; 121:830–838. [PubMed: 11606497]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Samowitz WS, Curtin K, Ma KN, Edwards S, Schaffer D, Leppert MF, Slattery ML. Prognostic significance of p53 mutations in colon cancer at the population level. Int J Cancer. 2002; 99:597–602. [PubMed: 11992552]

Samowitz WS, Curtin K, Schaffer D, Robertson M, Leppert M, Slattery ML. Relationship of Ki-ras mutations in colon cancers to tumor location, stage, and survival: a population-based study. Cancer Epidemiol Biomarkers Prev. 2000; 9:1193–1197. [PubMed: 11097226]

Samowitz WS, Holden JA, Curtin K, Edwards SL, Walker AR, Lin HA, Robertson MA, Nichols MF, Gruenthal KM, Lynch BJ, Leppert MF, Slattery ML. Inverse relationship between microsatellite instability and K-ras and p53 gene alterations in colon cancer. Am J Pathol. 2001b; 158:1517–1524. [PubMed: 11290569]

Samowitz WS, Slattery ML. Microsatellite instability in colorectal adenomas. Gastroenterology. 1997a; 112:1515–1519. [PubMed: 9136829]

Samowitz WS, Slattery ML. Transforming growth factor-beta receptor type 2 mutations and microsatellite instability in sporadic colorectal adenomas and carcinomas. Am J Pathol. 1997b; 151:33–35. [PubMed: 9212728]

Samowitz WS, Slattery ML, Potter JD, Leppert MF. BAT-26 and BAT-40 instability in colorectal adenomas and carcinomas and germline polymorphisms. Am J Pathol. 1999; 154:1637–1641. [PubMed: 10362787]

Samowitz WS, Slattery ML, Sweeney C, Herrick J, Wolff RK, Albertsen H. APC mutations and other genetic and epigenetic changes in colon cancer. Mol Cancer Res. 2007; 5:165–170. [PubMed: 17293392]

Shen L, Toyota M, Kondo Y, Lin E, Zhang L, Guo Y, Hernandez NS, Chen X, Ahmed S, Konishi K, Hamilton SR, Issa JP. Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. Proc Natl Acad Sci U S A. 2007; 104:18654–18659. [PubMed: 18003927]

Slattery ML, Anderson K, Curtin K, Ma K, Schaffer D, Edwards S, Samowitz W. Lifestyle factors and Ki-ras mutations in colon cancer tumors. Mutat Res. 2001a; 483:73–81. [PubMed: 11600135]

Slattery ML, Anderson K, Curtin K, Ma KN, Schaffer D, Samowitz W. Dietary intake and microsatellite instability in colon tumors. Int J Cancer. 2001b; 93:601–607. [PubMed: 11477566]

Slattery ML, Caan BJ, Benson J, Murtaugh M. Energy balance and rectal cancer: an evaluation of energy intake, energy expenditure, and body mass index. Nutr Cancer. 2003; 46:166–171. [PubMed: 14690792]

Slattery ML, Curtin K, Anderson K, Ma KN, Ballard L, Edwards S, Schaffer D, Potter J, Leppert M, Samowitz WS. Associations between cigarette smoking, lifestyle factors, and microsatellite instability in colon tumors. J Natl Cancer Inst. 2000a; 92:1831–1836. [PubMed: 11078760]

Slattery ML, Curtin K, Anderson K, Ma KN, Edwards S, Leppert M, Potter J, Schaffer D, Samowitz WS. Associations between dietary intake and Ki-ras mutations in colon tumors: a population-based study. Cancer Res. 2000b; 60:6935–6941. [PubMed: 11156393]

Slattery ML, Curtin K, Ma K, Edwards S, Schaffer D, Anderson K, Samowitz W. Diet activity, and lifestyle associations with p53 mutations in colon tumors. Cancer Epidemiol Biomarkers Prev. 2002; 11:541–548. [PubMed: 12050095]

Slattery ML, Curtin K, Sweeney C, Levin TR, Potter J, Wolff RK, Albertsen H, Samowitz WS. Diet and lifestyle factor associations with CpG island methylator phenotype and BRAF mutations in colon cancer. Int J Cancer. 2007; 120:656–663. [PubMed: 17096326]

Slattery ML, Edwards SL, Palmer L, Curtin K, Morse J, Anderson K, Samowitz W. Use of archival tissue in epidemiologic studies: collection procedures and assessment of potential sources of bias. Mutat Res. 2000c; 432:7–14. [PubMed: 10729707]

Slattery ML, Potter J, Caan B, Edwards S, Coates A, Ma KN, Berry TD. Energy balance and colon cancer--beyond physical activity. Cancer Res. 1997; 57:75–80. [PubMed: 8988044]

Smith G, Carey FA, Beattie J, Wilkie MJ, Lightfoot TJ, Coxhead J, Garner RC, Steele RJ, Wolf CR. Mutations in APC, Kirsten-ras, and p53--alternative genetic pathways to colorectal cancer. Proc Natl Acad Sci U S A. 2002; 99:9433–9438. [PubMed: 12093899]

Soong R, Powell B, Elsaleh H, Gnanasampanthan G, Smith DR, Goh HS, Joseph D, Iacopetta B. Prognostic significance of TP53 gene mutation in 995 cases of colorectal carcinoma. Influence of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

tumour site, stage, adjuvant chemotherapy and type of mutation. Eur J Cancer. 2000; 36:2053–2060. [PubMed: 11044641]

Szabo A, Boucher K. Estimating an oncogenetic tree when false negatives and positives are present. Math Biosci. 2002; 176:219–236. [PubMed: 11916510]

Szabo, A.; Boucher, K. Oncogenetic Trees. In: Hanin, LG.; Tan, W-T., editors. Handbook of Cancer Models with Applications to Cancer Screening, Cancer Treatment and Risk Assessment. Singapore and River Edge, NJ: World Scientific; 2008. in press

van Rijnsoever M, Grieu F, Elsaleh H, Joseph D, Iacopetta B. Characterisation of colorectal cancers showing hypermethylation at multiple CpG islands. Gut. 2002; 51:797–802. [PubMed: 12427779]

Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. Genetic alterations during colorectal-tumor development. N Engl J Med. 1988; 319:525–532. [PubMed: 2841597]

Wadler S, Bajaj R, Neuberg D, Agarwal V, Haynes H, Benson AB 3rd. Prognostic implications of c-Ki-ras2 mutations in patients with advanced colorectal cancer treated with 5-fluorouracil and interferon: a study of the eastern cooperative oncology group (EST 2292). Cancer J Sci Am. 1997; 3:284–288. [PubMed: 9327152]

Wong JJ, Hawkins NJ, Ward RL. Colorectal cancer: a model for epigenetic tumorigenesis. Gut. 2007; 56:140–148. [PubMed: 16840508]

Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B. The genomic landscapes of human breast and colorectal cancers. Science. 2007; 318:1108–1113. [PubMed: 17932254]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

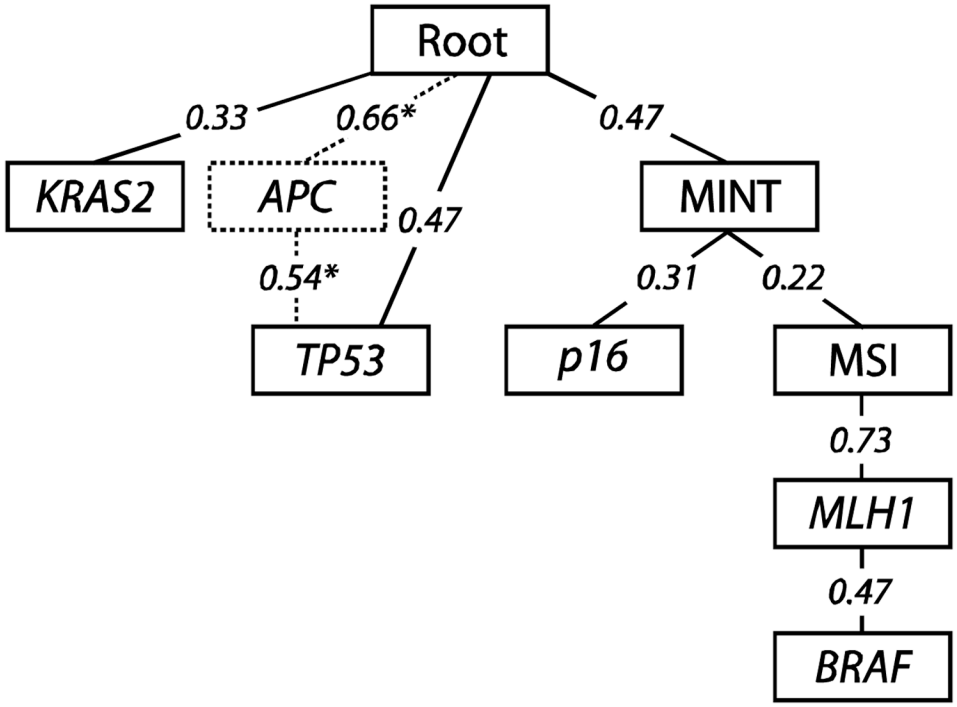

Figure 1.

Oncogenetic tree model for somatic alterations in 971 colon tumors from a population-based series. "Root" represents tissue with no alterations. Alterations are *APC* mutation, *TP53* mutation, *KRAS2* mutation, methylation at methylation in tumor (MINT) markers, methylation of *p16*, methylation of *MLH1*, microsatellite instability (MSI), and *BRAF* mutation. The numbers represent the probability of the next alteration on a branch, farther from the root, conditional on the presence of the alteration nearer the root. Two versions of the TP53 branch are shown: solid line, based on n = 971 samples without data on *APC* mutational status, and dashed lines, from an n = 90 subset with *APC* data.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

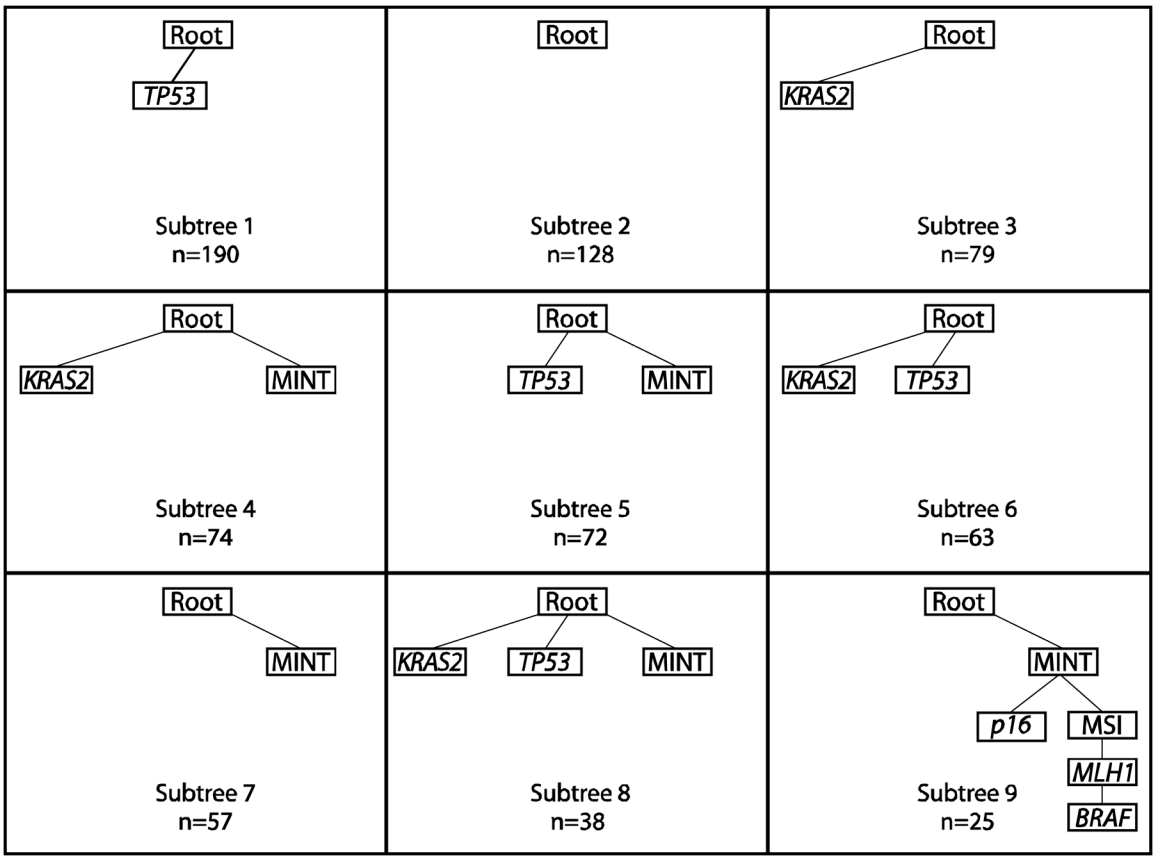

Figure 2.

Subtrees of the oncogenetic model represented by individual tumors: for *TP53* mutation, *KRAS2* mutation, methylation at methylation in tumor (MINT) markers, methylation of *p16*, methylation of *MLH1*, microsatellite instability (MSI), and *BRAF* mutation, the nine most common subtrees (of 56 total) are shown, based on n = 971 for the seven alterations. *APC* mutation is not included in the subtrees because *APC* data were unavailable for most tumors.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Genes Chromosomes Cancer. Author manuscript; available in PMC 2013 August 13.

Table 1

Somatic alterations in tumor tissue from a population-based sample: frequencies and correlations

			Correlation Coefficient*						
Alteration	**Type of Change**	**Frequency**	***TP53***	**MINT**	***KRAS2***	***p16***	***MLH1***	**MSI**	***BRAF***
TP53	mutation	0.47	1.00						
MINT	methylation	0.45	−0.16	1.00					
KRAS2	mutation	0.33	−0.12	0.09	1.00				
p16	methylation	0.16	−0.01	0.58	−0.01	1.00			
MLH1	methylation	0.11	−0.37	0.70	−0.54	0.47	1.00		
MSI	instability	0.11	−0.50	0.44	−0.49	0.35	0.93	1.00	
BRAF	mutation	0.09	−0.24	0.71	−0.72	0.69	0.74	0.67	1.00

*Pairwise ρ, estimated as the tetrachoric correlation coefficient.