

Published in final edited form as:

Science. 2009 March 27; 323(5922): 1747–1750. doi:10.1126/science.1163040.

## A Transposon-Based Genetic Screen in Mice Identifies Genes Altered in Colorectal Cancer\*

Timothy K. Starr<sup>1,\*</sup>, Raha Allaei<sup>1</sup>, Kevin A. T. Silverstein<sup>2</sup>, Rodney A. Staggs<sup>2</sup>, Aaron L. Sarver<sup>2</sup>, Tracy L. Bergemann<sup>3</sup>, Mihir Gupta<sup>4</sup>, M. Gerard O'Sullivan<sup>5</sup>, Ilze Matise<sup>5</sup>, Adam J. Dupuy<sup>6</sup>, Lara S. Collier<sup>7</sup>, Scott Powers<sup>8</sup>, Ann L. Oberg<sup>9</sup>, Yan W. Asmann<sup>9</sup>, Stephen N. Thibodeau<sup>9</sup>, Lino Tessarollo<sup>10</sup>, Neal G. Copeland<sup>11</sup>, Nancy A. Jenkins<sup>11</sup>, Robert T. Cormier<sup>12</sup>, and David A. Largaespada<sup>1,\*</sup>

<sup>1</sup>Department of Genetics, Cell Biology and Development, Arnold & Mabel Beckman Center for Transposon Research, the Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA.

<sup>2</sup>Biostatistics and Informatics, Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA.

<sup>3</sup>Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA.

<sup>4</sup>Harvard University, Cambridge, MA.

<sup>5</sup>College of Veterinary Medicine, University of Minnesota, St. Paul, MN 55108

<sup>6</sup>Anatomy and Cell Biology, University of Iowa, Iowa City, IA, 52242, USA.

<sup>7</sup>School of Pharmacy, University of Wisconsin, Madison, WI 53705, USA.

<sup>8</sup>Cold Spring Harbors Laboratory, Cold Spring Harbor, NY 11724, USA.

<sup>9</sup>Mayo Clinic College of Medicine, Rochester, MN 55905, USA.

<sup>10</sup>Neural Development Group, Mouse Cancer Genetics program, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702

<sup>11</sup>Institute of Molecular and Cell Biology, Singapore, 138673.

<sup>12</sup>University of Minnesota Medical School, Duluth, MN 55812, USA.

### Abstract

Human colorectal cancers (CRCs) display a large number of genetic and epigenetic alterations, some of which are causally involved in tumorigenesis (drivers) and others that have little functional impact (passengers). To help distinguish between these two classes of alterations, we used a transposon-based genetic screen in mice to identify candidate genes for CRC. Mice harboring mutagenic Sleeping Beauty (SB) transposons were crossed to mice expressing SB transposase in gastrointestinal tract epithelium. Most of the offspring developed intestinal lesions including intraepithelial neoplasia, adenomas, and adenocarcinomas. Analysis of over 16,000 transposon insertions identified 77 candidate CRC genes, 60 of which are mutated and/or dysregulated in human CRC and thus are most likely to drive tumorigenesis. These genes include *APC*, *PTEN* and *SMAD4*. The screen also

\***Publisher's Disclaimer:** This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. Their manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS.

\*To whom correspondence should be addressed: E-mail: star0044@umn.edu (T.K.S.); larga002@umn.edu (D.A.L.).

identified 17 new candidate genes that had not previously been implicated in CRC, including *POLI*, *PTPRK*, and *RSPO2*.

Recent genomic studies have revealed that human colorectal cancers (CRCs) undergo numerous genetic and epigenetic alterations (1–4). These alterations likely derive from a mixture of “drivers” that play a causal role in tumor formation and progression, and “passengers” that have little or no effect on tumor growth. The design of targeted therapeutics for CRCs is dependent on the ability to distinguish drivers from passengers.

To help identify potential driver genes in CRC we developed a forward genetic screen in mice by using a Sleeping Beauty (SB) system for generating insertional mutations. To confine transposition to the gastrointestinal tract, SB11 transposase cDNA preceded by a LoxP-flanked stop cassette, was knocked into the *Rosa26* locus (fig. S1) (5). These mice were then crossed with *Villin-Cre* transgenic mice to activate SB transposase in epithelial cells of the gastrointestinal tract (6). Once expressed, SB transposase catalyzed the transposition of *T2/Onc*, a mutagenic SB transposon (7) (Fig. 1A). *T2/Onc* contains a murine stem cell virus long terminal repeat and splice donor site (MSCV-LTR-SD), which can deregulate the expression of a nearby proto-oncogene. *T2/Onc* also carries splice acceptor sites in both DNA strands and a bidirectional poly(A) signal, which can inactivate the expression of a tumor suppressor gene. Since SB transposition is biased toward reintegration of the transposon into the same chromosome as the donor transposon – a phenomenon referred to as “local hopping” – we used two *T2/Onc* transgenic lines that each carried approximately 25 copies of the *T2/Onc* transposon in a concatamer on different donor chromosomes (Chr 1 and 15) (7).

A histochemical analysis of the triple transgenic mice (*Rosa26-LsL-SB11*, *T2/Onc*, *Villin-Cre*) showed that SB transposase was strongly expressed in epithelial cells of the gut and pancreas, but undetectable in other tissues (fig. S2). We created a cohort of 28 triple transgenic mice and 72 double transgenic control mice carrying all possible dual combinations of the three transgenes. Mice in this first cohort were monitored daily for 18 months. We generated a second cohort of 50 triple transgenic mice that were maintained in a separate facility for 12 months and also monitored daily.

Triple transgenic mice became moribund at a faster rate than double transgenic controls, beginning around one year of age (Fig. 1B). Examination of the gastrointestinal tract of moribund animals revealed discrete raised lesions from 2 mms to as large as 5 mms in diameter in the small and large intestine. In the first cohort, 100% (12/12) of the experimental mice that became moribund before 18 months harbored intestinal lesions (table S1), while none of the control mice sacrificed before 18 months had lesions. In the second cohort, 72% (36/50) of triple transgenic mice had intestinal lesions, with an average of 1.9 intestinal lesions in the small intestines and 0.2 lesions in the large intestine.

We performed histopathologic analyses of tumor tissue sections from 11 animals. These analyses identified 39 and 16 intraepithelial neoplasias, 50 and 15 adenomas and 3 and 0 adenocarcinomas in the small and large intestines, respectively (Fig. 1C–E). An additional adenocarcinoma was identified for which the site of origin was undetermined. We also selected six large tumors ( $\geq 5$  mms) from six additional mice and found three were adenocarcinomas and three were adenomas.

For use in DNA isolation and sequencing experiments, we harvested 135 tumors: 42 tumors from 11 of the triple transgenic mice from the first cohort (dataset 1) and 93 tumors from 36 mice in the second cohort (dataset 2). The majority of tumors were small and the entire tumor was used for DNA isolation. This precluded our ability to perform histological analysis and to link the molecular data to the histopathology of specific tumors. However, given the

distribution of frank intestinal lesions from the histopathological analysis, the majority of tumors were likely to be adenomas. We then performed linker-mediated PCR on DNA from these 135 harvested tumors to generate PCR products containing transposon-genomic junction fragments. We sequenced over 195,000 of these PCR products, of which 99,624 could be uniquely mapped to TA dinucleotides in the mouse genome, consistent with SB insertion-site requirements. After combining duplicate insertions within a given tumor, we found that 45% of the insertions mapped to the same chromosome as the donor concatamer (Chr 1 or 15), consistent with local hopping seen in previous SB screens. We removed these insertions from further analysis to eliminate statistical bias due to local hopping. We also eliminated insertions mapping to the precise TA dinucleotide in tumors from two or more different mice because these insertions could represent a PCR artifact. The consummate total of 16,690 mapped, non-redundant genomic loci (table S2) equates to an average of 124 mapped insertions per tumor.

To define common insertion sites (CISs), we performed Monte Carlo simulations using randomly assigned insertions. Genomic window sizes were chosen based on simulations that used the same number of insertions as the datasets, such that one would not expect to find a single CIS after randomly distributing transposon insertions throughout the genome (Expected value,  $E < 1$ ). For example, in a random assignment of 16,690 insertions we would not expect to find a single cluster of five insertions within 25 Kb, six insertions within 50 Kb, seven insertions within 80 Kb, etc. (8). Any cluster of insertions meeting or exceeding these parameters was defined as a CIS. We removed one CIS from this list because it was composed entirely of insertions from a single mouse, indicating that those tumors may be related.

As a final control, we amplified and sequenced 15,556 SB insertions present in tail DNA derived from 89 double transgenic weanling mice. These mice contained a ubiquitously expressed *SB11 transposase* transgene and the *T2/Onc* transposon concatamers (7,9). Because there was no selection pressure for tumor outgrowth in these mice and because SB integration does not have a strong bias for any individual TA dinucleotide (10), we expected the insertions to be distributed randomly throughout the genome, except for local hopping. From this control dataset we identified six CISs. This was more than expected, but considerably less than that observed in tumor DNA (table S3). These CISs could be previously unknown hotspots for transposon integration. Alternatively, they could reflect incipient clonal neoplastic growth, as these genetically manipulated mice eventually develop lymphoma (9). Two of these six CISs were also identified in the tumor datasets, thus they were eliminated from the list of tumor CISs, leaving us with 77 CISs (table S4).

Candidate genes were assigned to the 77 CISs according to the percentage of insertions in or near a gene within the CIS boundaries. Insertions were mainly located within introns (51%), with only 2% in an exon and the remaining 47% either upstream or downstream of a coding region. The top 10 CISs, based on number of insertions found, are listed in Table 1.

The goal of this study was to identify genes that are drivers of tumorigenesis in order to identify new candidate genes whose mutational status in human CRC can then be tested. We compared our list of mouse CIS genes to the human genes listed in the Catalog of Somatic Mutations in Cancer (COSMIC) database (11). Among our list of 77 CIS genes, 38 have human homologs present in the COSMIC database, and 18 (47%) of these 38 homologs have documented non-silent mutations in human cancers (table S5), which would not be expected by chance ( $P < 0.05$ ). Furthermore, if we limit our analysis of COSMIC to genes mutated in human CRC, the overlap has a lower probability of being due to chance ( $p < 0.001$ ) (8).

Similarly, our CIS list overlaps with a recent large-scale exon-resequencing project that cataloged mutations in 18,191 human genes in 11 colorectal tumor samples (1). That project identified 848 human gene mutations, 140 of which were considered likely to be driver

mutations for CRC. Of the 77 CIS mouse genes identified in our study, 74 have human homologs that were included in the exon-resequencing study. Among these 74 homologs, 10 had a mutation and four were identified as candidate driver mutations in human CRC (table S6) making these findings highly significant ( $p < 0.005$ ) (8).

We then investigated whether the human homologs of the mouse CIS genes were amplified or deleted in human CRC. We examined a dataset that identified 482 deletions and 224 amplifications in human CRCs (8). The human homologs of 10 CIS genes were located in deleted regions and 23 were located in amplified regions (table S7 & table S8). This represents a significant overlap ( $p < 0.05$ ) (8) and suggests that the candidates found in this screen are relevant to human CRC.

Finally, we analyzed cDNA microarray data to determine whether CIS genes were differentially expressed in human CRC versus normal colonic tissue. The Oncomine database (12) contains five microarray datasets that compare gene expression levels in 138 CRCs and 88 normal samples. Of the 77 CIS human homologs, 50 were identified as being differentially regulated ( $p < 0.05$ ) in one or more of these studies (table S9).

By comparing our list of mouse CIS genes with human genes that are (i) mutated in CRC, (ii) listed in COSMIC, (iii) amplified or deleted in CRC, (iv) aberrantly expressed in CRC or (v) known cancer genes identified by the Cancer Genome Project (CGP) (13), we identified 15 CIS genes that are the most likely to be driver mutations in human CRC (Table 2) by virtue of being present in at least three of the five categories listed above. Among these 15 genes is *adenomatous polyposis coli*, *Apc*, a member of the Wnt signaling pathway and the most commonly mutated gene in human CRC (70–80%) (14). Also included in this list are *Bmpr1a*, *Smad4*, and *Pten*, which are responsible for juvenile polyposis syndrome, juvenile intestinal polyposis, and Cowden disease, respectively. Another gene on the list, *Fbxw7*, is a component of the SCF ubiquitin ligase complex, which is mutated in 11.5% of human CRCs (15). Thus, among the 15 prioritized genes in our study, five are validated human CRC genes and together represent some of the most commonly mutated genes identified in human CRC.

Three other genes on the complete CIS list (table S4) are also implicated in human CRC: *CDK8*, *MCC* and *SND1*. *CDK8*, which encodes cell division protein kinase 8, is commonly amplified in human CRC and plays a direct role in  $\beta$ -catenin-driven cell transformation (16). *MCC* encodes the colorectal mutant cancer protein and, in addition to finding somatic mutations in *MCC* (17), a recent study found that 50% of primary CRCs exhibited *MCC* promoter methylation (18). Furthermore, *MCC* interacts with  $\beta$ -catenin and its re-expression in CRC cells inhibits Wnt signaling and proliferation, suggesting that *MCC* is a tumor suppressor (19). *SND1*, a component of the RNA-induced silencing complex, is highly expressed in CRC and its overexpression in rat epithelial cells leads to a loss of contact inhibition and increased cell growth (20). Interestingly, *SND1* overexpression leads to a downregulation of APC protein even though mRNA levels are unchanged.

In addition to identifying genes whose human homologs are known to be altered in cancer, our screen identified a number of novel candidate CRC genes that could, on the basis of their function, be drivers of CRC. These candidate CRC genes include *POLI*, *PPP1R13B*, and *RSPO2*, which affect DNA stability, p53-induced apoptosis, and Wnt signaling, respectively. *POLI*, the product of *POLI*, is an error-prone DNA polymerase responsible for the high frequency of UV-induced mutations in xeroderma pigmentosum variant cells (21). *PPP1R13B*, the product of *PPP1R13B*, enhances the ability of p53 to stimulate the expression of pro-apoptotic genes (22). *RSPO2* is a member of a novel family of Wnt-signaling regulators, the R-Spondins (23). Finally, two microRNA genes not previously associated with CRC, *Mir181b-2* and *Mir181a-2*, reside within an intron of *Nr6a1*, one of the CIS genes identified

in our screen. It is possible that these microRNAs, and not the gene *Nr6a1*, are the genes affected at this CIS. Both of these microRNAs are aberrantly expressed in CRC (24,25) and function as tumor suppressors in glioma (26).

Our transposon-based forward genetic screen encountered some limitations. We believe the screen was unable to recapitulate the effect of certain activating point mutations, such as the *Kras*<sup>G12V</sup> mutation that is found in a large percentage of CRC. In addition, random transposon insertions could potentially miss small genetic loci such as microRNAs. By design, the statistical method we used to determine CISs in order to identify likely candidate driver mutations ignores the majority of mapped transposon insertions that occurred in only one or two tumors. These non-CIS insertions may also have contributed to carcinogenesis by creating CRC driver or cooperating mutations or by causing some other level of genomic instability.

In summary, our transposon-mediated forward genetic screen in mice identified genetic mutations that lead to the development of an epithelial cancer. The discovery of a significant overlap of mouse candidate genes and human genes that are altered in cancer indicates that this mouse model will be useful for distinguishing between driver and passenger mutations. In addition, the large number of CISs uncovered in this screen affirms the hypothesis that the growth of human CRC is driven by a few commonly mutated genes and a much larger number of genes that are rarely mutated (1).

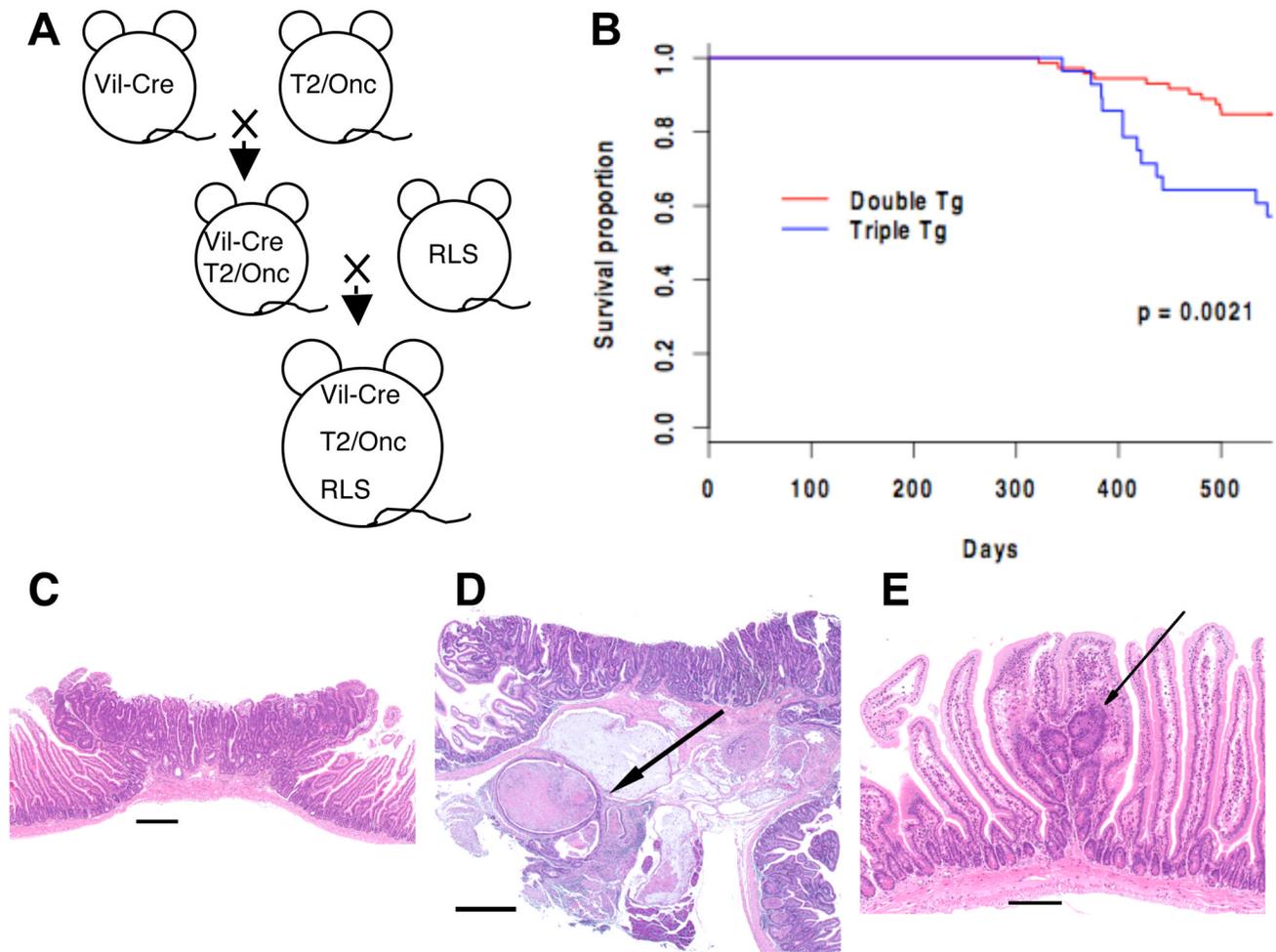
## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References and notes

1. Wood LD, et al. *Science* 2007 Nov 16;318:1108. [PubMed: 17932254]
2. Stoler DL, et al. *Proc Natl Acad Sci U S A* 1999 Dec 21;96:15121. [PubMed: 10611348]
3. Staub E, et al. *Mol Cancer* 2007;6:79. [PubMed: 18081933]
4. Hayashi H, et al. *Hum Genet* 2007 Jan;120:701. [PubMed: 17024368]
5. Zambrowicz BP, et al. *Proc Natl Acad Sci U S A* 1997 Apr 15;94:3789. [PubMed: 9108056]
6. el Marjou F, et al. *Genesis* 2004 Jul;39:186. [PubMed: 15282745]
7. Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA. *Nature* 2005 Jul 14;436:272. [PubMed: 16015333]
8. Materials and methods are available as supporting material on Science online.
9. Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA. *Nature* 2005 Jul 14;436:221. [PubMed: 16015321]
10. Yant SR, et al. *Mol Cell Biol* 2005 Mar;25:2085. [PubMed: 15743807]
11. Forbes S, et al. *Br J Cancer* 2006 Jan 30;94:318. [PubMed: 16421597]
12. Rhodes DR, et al. *Neoplasia* 2004 Jan-Feb;6:1. [PubMed: 15068665]
13. Futreal PA, et al. *Nat Rev Cancer* 2004 Mar;4:177. [PubMed: 14993899]
14. Segditsas S, Tomlinson I. *Oncogene* 2006 Dec 4;25:7531. [PubMed: 17143297]
15. Rajagopalan H, et al. *Nature* 2004 Mar 4;428:77. [PubMed: 14999283]
16. Firestein R, et al. *Nature* 2008 Sep 25;455:547. [PubMed: 18794900]
17. Kinzler KW, et al. *Science* 1991 Mar 15;251:1366. [PubMed: 1848370]
18. Kohonen-Corish MR, et al. *Oncogene* 2007 Jun 28;26:4435. [PubMed: 17260021]
19. Fukuyama R, et al. *Oncogene*. 2008 Jun 30;
20. Tsuchiya N, et al. *Cancer Res* 2007 Oct 1;67:9568. [PubMed: 17909068]
21. Wang Y, et al. *Cancer Res* 2007 Apr 1;67:3018. [PubMed: 17409408]
22. Sullivan A, Lu X. *Br J Cancer* 2007 Jan 29;96:196. [PubMed: 17211478]
23. Kim KA, et al. *Cell Cycle* 2006 Jan;5:23. [PubMed: 16357527]

24. Lu J, et al. *Nature* 2005 Jun 9;435:834. [PubMed: 15944708]
25. Bandres E, et al. *Mol Cancer* 2006;5:29. [PubMed: 16854228]
26. Shi L, et al. *Brain Res.* 2008 Jul 30;
27. Insertion data are deposited in the Retrovirus Tagged Cancer Gene Database (RTCGD). We thank S. Singh for help with GSFlex 454 sequencing; K. Akagi for maintaining the RTCGD; A. Weber-Main and K. Stone for editing assistance, and the Minnesota Supercomputing Institute for providing computational resources. Research was funded by the American Cancer Society fellowship PF-06-282-01-MGO (to T.K.S.), NIH grant R01CA113636-01A1 (to D.A.L.), NCI grant K01CA122183-03 (to L.S.C.), the Biomedical Research Council of A\*STAR (Agency for Science, Technology and Research) Singapore (N.A.J. and N.G.C.), and the NCI Intramural Research Program of the NIH. (L.T.). The University of Minnesota AHC Faculty Development Grant provided additional funding to D.A.L. and R.T.C. D.A.L. is a co-founder of, and has an equity interest in, Discovery Genomics, Inc., a biotechnology company that is pursuing Sleeping Beauty (SB) technology for human gene therapy. The University of Minnesota has filed a patent related to the work described in this paper.



**Fig. 1.** Triple transgenic mice develop intestinal tumors and become moribund faster than double transgenic controls. (A) Breeding scheme for generating triple transgenics. (B) Kaplan-Meier Survival curve comparing triple transgenics to double transgenic controls. Photomicrographs of hematoxylin and eosin stained representative tissue showing an adenoma (C), an adenocarcinoma (D) with arrow pointing to a cluster of glands that has invaded through the serosa, and gastrointestinal intraepithelial neoplasia (E) with arrow indicating a cluster of dysplastic glands accompanied by fusion of villi. Vil = *Villin* promoter, RLS = *Rosa26-Lox-stop-Lox-Sleeping Beauty* 11, Tg = Transgenic, bars = 250  $\mu$ m (C), 500  $\mu$ m (D), and 100  $\mu$ m (E).

**Table 1**

Top 10 CIS candidate genes, ranked according to the number of unique insertions defining the CIS

Candidate Gene	Chromosome	# of unique insertions	Function
<i>Apc</i>	18	61	Wnt signaling
<i>Rspo2</i>	15	17	Wnt signaling
<i>Kcnq1</i>	7	14	Potassium channels
<i>Tcf12</i>	9	13	Transcription factor
<i>Wac</i>	18	13	Novel adaptor
<i>Fbxw7</i>	3	11	Ubiquitination
<i>Ptprk</i>	10	11	Cell adhesion
<i>Cugbp1</i>	2	10	RNA binding
<i>Nr6a1</i>	2	10	Orphan nuclear receptor
<i>Zcchc7</i>	4	10	Novel zinc finger

**Table 2**  
Candidate CIS genes likely to be drivers of human CRC

Gene Symbol	Mutated in human cancer <sup>*</sup>	Amplified or deleted in human CRC <sup>†</sup>	Aberrantly expressed in human CRC <sup>‡</sup>	Known human cancer gene <sup>§</sup>
<i>ANKRD11</i>	x	x	x	
<i>APC</i>	x		x	x
<i>BMPRIA</i>	x	x	x	x
<i>DSTN</i>	x	x	x	
<i>EVII</i>	x		x	x
<i>FBXW7</i>	x			x
<i>GPBP1</i>	x	x	x	
<i>NOTCH1</i>	x	x	x	x
<i>NSD1</i>	x		x	x
<i>PPP1R12A</i>	x		x	
<i>PTEN</i>	x	x	x	x
<i>RREB1</i>	x	x	x	
<i>SMAD4</i>	x	x	x	x
<i>TCF12</i>	x	x	x	x
<i>TNPO3</i>	x	x	x	

\* A somatic mutation was identified in COSMIC or by exon-resequencing of human CRCs by Wood *et. al.*, (2).

† Gene is located in a region known to be deleted or amplified in human CRC.

‡ mRNA level was significantly different ( $p < 0.05$ ) in human CRC compared to normal colon tissue.

§ Gene is identified as a human cancer gene by the Cancer Genome Project (13).