Genome-wide association and fine mapping of genetic loci predisposing to colon carcinogenesis in mice

Pengyuan Liu1,3, Yan Lu1,3, Hongbo Liu3, Weidong Wen1, Dongmei Jia1, Yian Wang1 and Ming You2,3

1Department of Physiology and the Cancer Center, Medical College of Wisconsin, Milwaukee, WI 53226, USA
2Department of Pharmacology and Toxicology and the Cancer Center, Medical College of Wisconsin, Milwaukee, WI 53226, USA
3Department of Surgery and the Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO 63110, USA

Running title: Fine mapping of genetic loci affecting colon cancer

Key words: Association, colon cancer, fine mapping, linkage, mice

Corresponding author:
Ming You
Department of Pharmacology and Toxicology and the Cancer Center
Medical College of Wisconsin
Milwaukee, WI 53226, USA.
Tel: 414-955-2565; Fax: 414-955-6058
E-mail: myou@mcw.edu
Abstract

To identify the genetic determinants of colon tumorigenesis, 268 male mice from 33 inbred strains derived from different genealogies were treated with Azoxymethane (AOM) (10 mg/kg) once a week for six weeks to induce colon tumors. Tumors were localized exclusively within the distal colon in each of the strains examined. Inbred mouse strains exhibit a large variability in genetic susceptibility to AOM-induced colon tumorigenesis. The mean colon tumor multiplicity ranged from 0 to 38.6 (mean = 6.5 \pm 8.6) and tumor volume ranged from 0 to 706.5 mm³ (mean = 87.4 \pm 181.9) at 24 weeks after the first dose of AOM. AOM-induced colon tumor phenotypes are highly heritable in inbred mice, and 68.8% and 71.3% of total phenotypic variation in colon tumor multiplicity and tumor volume, respectively, are attributable to strain-dependent genetic background. Using 97,854 SNPs, we performed a genome-wide association study (GWAS) of AOM-induced colon tumorigenesis and identified a novel susceptibility locus on chromosome 15 (rs32359607, P = 6.31 \times 10^{-6}). Subsequent fine mapping confirmed five (Scc3, Scc2, Scc12, Scc8 and Ccs1) of 16 linkage regions previously found to be associated with colon tumor susceptibility. These five loci were refined to less than 1-Mb genomic regions of interest. Major candidates in these loci are Sema5a, Fmn2, Grem2, Fap, Gsg1l, Xpo6, Rabep2, Eif3c, Unc5d and Gpr65. In particular, the refined Scc3 locus shows high concordance with the human GWAS locus that underlies Hereditary Mixed Polyposis Syndrome. These findings increase our understanding of the complex genetics of colon tumorigenesis, and provide important insights into the pathways of colorectal cancer development and might ultimately lead to more effective individually targeted cancer-prevention strategies.
Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related mortality in the United States. In 2010, there will be an estimated 50,000 deaths associated with this disease (1). Genetic factors play important roles in CRC development and account for ~ 35% of CRC risk (2). Highly penetrant germline mutations in \textit{APC}, \textit{SMAD4}, \textit{BMPRIA}, \textit{MUTYH}, \textit{STK11} and DNA mismatch repair genes are estimated to account for ~ 6% of all CRC cases (3). In addition to these rare variants, much of CRC inherited susceptibility is likely attributable to multiple low-penetrance common variants. Direct evidence for common variants for CRC is highlighted by recent genome-wide association studies (GWAS). These GWAS on CRC have identified 10 independent loci that confer risk of CRC, including those on chromosomes 8q24.21, 11q23, 18q21.1, 8q23.1 15q, 19q13.1, 20q12.3, 14q22.2, 16q22.1 and 10p14 (4-10). However, risk associated with these common variants is modest and only a small proportion of CRC risk can be explained by currently identified loci. Other genetic factors underlying CRC remain unidentified and this strongly supports the continued search for novel CRC susceptibility genes.

Inbred strains of laboratory mice have been valuable in the identification of tumor susceptibility genes because they display a wide range of spontaneous and chemically induced tumor incidence. A number of linkage mapping studies using cross-breeding experiments were performed for chemically induced colon tumors in mice (11-16). As a result, 16 quantitative trait loci (QTLs) responsible for chemically induced colon tumors have been mapped on the mouse genome, implying a very complex picture of inherited susceptibility of colon tumorigenesis exists. However, a major obstacle of identifying QTL genes is the difficulty of resolving these chromosomal regions (10~20 cM) into sufficiently small intervals to make positional cloning possible. These colon tumor QTLs span a total of 178 cM that corresponds to 351 Mb, covering
13% of the mouse genome. Recent advances in genomic sequence analysis and SNP discovery have provided researchers with the necessary resources to explore a wide range of genetic variation in laboratory inbred mice (17). The use of dense SNP maps in laboratory inbred mice has proven successful in the refinement of previous QTL regions and the identification of new genetic determinants of complex traits (18, 19).

Here, we carried out a genome-wide association study to map novel susceptibility loci and refine previous QTL regions for colon tumorigenesis in inbred mice. A total of 268 male mice from 33 inbred strains were treated with Azoxymethane (AOM), an organotropic colon carcinogen, to induce colon tumors. We identified a novel genetic susceptibility locus on mouse chromosome 15 and narrowed five of 16 previous linkage regions into less than 1-Mb genomic regions of interest in which candidate genes were identified.

Materials and Methods

Inbred mouse strains and SNP Data

Thirty-three inbred mouse strains were used in our colon cancer study. The chosen inbred strains were derived from different genealogies and include 16 Castle's mice (129S1/SvImJ, 129S4/SvJae, 129X1/SvJ, A/J, AKR/J, BALB/cByJ, C3H/HeJ, CBA/J, DBA/1J, DBA/2J, I/LnJ, LP/J, NZB/BINJ, NZW/LacJ, SEA/GnJ, SM/J), 4 C57-related strains (C57BL/6J, C57BLKS/J, C57L/J, C58/J), 4 Swiss mice (FVB/NJ, NON/ShiLtJ, SJL/J, and SWR/J), 2 wild-derived strains (CAST/EiJ, PERA/EiJ), 1 strain derived from colonies from China and Japan (KK/HJ) and 6 other inbred strains (BTBR T+ tf/J, BUB/BnJ, CE/J, LG/J, PL/J and RIII/JS/J). This brought a wide range of variation in colon tumorigenesis between inbred mouse strains. The SNP data were obtained from the Mouse Phenome Database (MPD) (http://phenome.jax.org/), which
contains 190,903 SNPs on commonly used mouse inbred strains. These SNP data were further filtered by removing SNPs with less than 26 strains typed or without genetic mapping information. To be included, each SNP allele also had to be present in five or more inbred strains. The resulting data consisted of 97,854 SNPs, spanning the mouse genome at an average density of approximately 28 kb per SNP.

Colonic tumorigenesis assays

The inbred mouse strains were purchased from The Jackson Laboratory (Bar Harbor, ME). Animals were housed in plastic cages with hardwood bedding and dust covers, in a HEPA filtered, environmentally controlled room (24 ± 1°C, 12/12-h light/dark cycle). Animals were given Rodent Lab Chow, #5001 (Purina, St. Louis, MO) and water ad libitum. Animals received administration of Azoxymethane (AOM) (Sigma, St. Louis, MO) with a dose of 10 mg/kg in 0.1 ml of PBS at 5 weeks of age. All mice were injected intraperitoneally with AOM once a week for 6 weeks. Animals were killed 24 weeks after the first dose of AOM by CO₂ asphyxiation. Immediately after sacrifice, the colons (proximal and distal) were flushed with ice-cold phosphate buffered saline to remove fecal material, opened longitudinally and placed flat on a filter paper. In general, 5-10 mice per strain were treated and phenotyped. The flushed colons were fixed in Tellyesniczky's solution (20) overnight followed by 70% ethanol. The fixed colons were evaluated by at least two investigators under a dissecting microscope to obtain the fixed surface tumor counts (i.e. tumor multiplicity), and individual tumor volume was measured based on the following formula: \( V = \frac{4}{3}\pi r^3 \) where \( r \) is radius of tumor. The total tumor volume was obtained by adding the individual tumor volume per mouse. Tumor multiplicity and volume are widely used phenotypes in cancer studies, which are more informative than the binary phenotype (case versus disease-free control).
Genome-wide SNP Association Analysis

To correct for population structure and genetic relatedness among inbred strains, we used a recently developed method, efficient mixed model association (EMMA) approach, to assess association of colon tumor susceptibility with SNPs (21). Specifically, the mixed model in the EMMA method can be represented by:

\[ y = X\beta + Zu + e \]

where \( y \) is an \( n \times 1 \) vector of observed phenotypes (i.e., colon tumor multiplicity and volume), and \( X \) is an \( n \times q \) matrix of fixed effects including mean, SNPs, and other covariate variables. \( \beta \) is a \( q \times 1 \) vector representing coefficients of the fixed effects. \( Z \) is an \( n \times t \) incidence matrix mapping each observed phenotype to one of \( t \) inbred strains. \( u \) is the random effect (i.e., strain effects) of the mixed model with \( \text{Var}(u) = 2K\sigma_g^2 \) where \( K \) is the \( t \times t \) kinship matrix inferred from genotypes, and \( e \) is an \( n \times n \) matrix of residual effect such that \( \text{Var}(e) = I\sigma_e^2 \). The overall phenotypic variance–covariance matrix can be represented as \( V = 2KZ^T\sigma_g^2 + I\sigma_e^2 \). The kinship matrix \( K \) based on genetic similarity was inferred from SNP genotype data.

An R package implementation of the EMMA method is publicly available (http://mouse.cs.ucla.edu/emma/). A two-sided \( p \)-value from the EMMA for each SNP was obtained for testing hypothesis of no association between the SNP and colon tumor phenotypes. Prior to the statistical analysis, colon tumor multiplicity and volume were converted into normal data by Box-cox transformation.

Two hundred permutations were used to establish a genome-wide threshold (a global \( p \) value of 0.05) for declaring significant associations in the association analysis, which take into account linkage disequilibrium among SNPs on the genome (22). Specifically, colon tumor
phenotypes were randomly reshuffled among subjects while fixing the genotypes. For each permutation, the EMMA approach described above was implemented, and the most significant –log_{10}(P) was recorded. Sorting the maximum –log_{10}(P) from large to small, the 5% quantile of the empirical distribution was taken as the genome-wide threshold (a global p value of 0.05 and 0.10). We also used the above simulation procedure to determine a region-wide threshold for each colon susceptibility locus from previous linkage studies. The only difference here is to perform association tests in that linkage region instead of the whole genome. Briefly, we first identified one LOD (logarithm of the odds) supporting interval for each of the previous linkage regions. Then we performed permutation analysis as described above in the identified one LOD supporting interval.

**Detection of differential expression in colon tumor models and human colon cancer studies**

Mouse cDNA array data were downloaded from the Gene Expression Omnibus database (GSE5261). We analyzed 14 tumor samples from the AOM-induced mouse colon model (23), 9 tumor samples from the \( Apc^{Min/+} \) mouse model (24) and 3 adult normal mouse colon. The expression data were normalized by Lowess intensity-dependent normalization as implemented in GeneSpring 7.2. Then, the samples were referenced to expression levels of normal control colon samples (25). Two-sample \( t \)-tests were used to detect differential expression between tumor and control samples.

Human colon cancer microarray data were downloaded from the Gene Expression Omnibus database (GSE10950). The microarray hybridization of 24 colon normal and tumor pairs was performed using the Illumina Gene Expression Sentrix®BeadChip HumanRef-8_V2 and the expression data was normalized by cubic spline normalization in GeneSpring 7.2. Pair wise \( t \)-tests were used to detect differential expression between tumor and normal tissues. Fold
changes of gene expression in tumor versus normal tissues were also recorded. Flow charts of the study design and data analysis and fine mapping of previous linkage regions were described in Supplementary fig. S1.

Results

Colon tumorigenesis in inbred strains

Two hundred sixty-eight male mice from 33 inbred mouse strains were measured for colon tumor multiplicity and volume 24 weeks after injection with AOM. In each of the strains examined, the tumors were localized exclusively within the distal colon and not detected grossly or histologically within the proximal colon. Inbred mouse strains had a large variability in AOM-induced colon tumor multiplicity and volume (Fig. 1). Mean colon tumor multiplicity ranged from 0 to 38.6 (mean ± SD, 6.5 ± 8.6) and tumor volume ranged from 0 to 706.5 mm$^3$ (mean ± SD, 87.4 ± 181.9) at 24 weeks after the first dose of AOM. Several strains show marked sensitivity to AOM-induced colon tumorigenesis and developed more than 10 tumors/mouse with diameter $\geq$ 2mm including C57L/J, FVB/NJ, BTBR T$^{+}$ tf/J, A/J, NON/ShiLtJ, SM/J, KK/HJ and I/LnJ. In contrast, strains AKR/J, PERA/EiJ, RIIIS/J, DBA/1J, C57BL/6J and DBA/2J had less than 1 tumor/mouse after treatment with AOM. Tumor multiplicity is highly correlated with tumor volume in colon tumorigenesis in inbred mice ($r^2 = 0.74$), suggesting some degree of common genetic components between these two tumor phenotypes. The between-strain variance accounts for 68.8% and 71.3% of total phenotypic variation in colon tumor multiplicity and tumor volume, respectively, implying that most of the variations we observed in colon tumorigenesis are heritable.
Genome-wide association analysis

To identify the genetic basis of colon tumorigenesis in the AOM model, we performed a genome-wide association analysis on 33 strains of inbred mice using colon tumor multiplicity and volume as the phenotypes (Fig. 2). To correct for population structure and genetic relatedness in inbred mouse strains (18), we used the EMMA approach to assess association of colon tumor susceptibility with SNPs. The distribution of observed P values was similar to the expected distribution, indicating no inflation of test statistics from population structure or any other form of bias (Supplementary fig. S2). Therefore, any bias due to population structure and genetic relatedness among inbred strains has been largely eliminated by the EMMA approach. Using the permutation analysis, we established genome-wide significance levels of 0.05 and 0.10, which corresponds to a point-wise P value of $3.39 \times 10^{-6}$ and $7.29 \times 10^{-6}$ for the analysis of colon tumor multiplicity and corresponds to P values of $2.53 \times 10^{-6}$ and $6.04 \times 10^{-6}$ for tumor volume (Fig. 2).

The GWAS identified 20 SNPs ($P < 10^{-4}$) potentially associated with colon tumor susceptibility (Table 1). In particular, two SNPs on chromosome 15, rs32359607 and rs32137981, are strongly associated with colon tumor multiplicity ($P = 6.31 \times 10^{-6}$ and $P = 7.80 \times 10^{-6}$, respectively). These two SNPs achieved genome-wide significance level of 0.10. Candidate genes nearby these two SNPs include Tas2r119, Snord123 and Sema5a (Fig. 3). This is a novel susceptibility locus on the proximal mouse chromosome 15 and has not been reported in previous linkage studies (26).

Fine mapping of previous linkage regions

A total of 16 QTLs responsible for chemically induced colon tumors have been previously mapped on the genome by linkage analysis of intercross and/or backcross of inbred
mouse strains (Supplementary fig. S3) (26). Here, we systematically investigated genetic association signals in these colon tumor susceptibility loci. We first identified microsatellite markers flanking one LOD supporting interval in each susceptibility locus based on previous linkage mapping studies, and thus determined genomic locations of these loci. Using permutation analysis, we then established region-wide thresholds for declaring significant associations in each locus (Table S1). The region-wide association analysis identified a number of SNPs showing significant associations with colon tumor susceptibility in five of 16 susceptibility loci, including susceptibility to colon cancer locus 3 (Scc3), Scc2, Scc12, Scc8 and colon-cancer susceptibility locus 1 (Ccs1) (Supplementary fig. S4). These linkage regions were generally refined into less than 1-Mb genomic regions by association mapping. They are located on chromosomes 1, 2, 7, 8 and 12, respectively.

Due to high linkage disequilibrium within the mouse genome, we also checked candidate genes in 500 kb flanking regions on either side of these susceptibility loci (Fig. 4). Candidate genes in these refined linkage regions were prioritized based on location of the most significant SNPs, mRNA expression in tumors, and their functional relevance in the literature. The most significant association in the Scc3 was located at 176.7 Mb on chromosome 1 (rs32121685, $P = 3.47 \times 10^{-4}$) (Table 2). The refined Scc3 locus includes ten annotated genes, of which Fmn2 and Grem2 are major candidates. The Scc2 locus on chromosome 2 (rs28039498, $P = 2.09 \times 10^{-4}$) covers seven genes, of which Fap, Ifih1, and Gca are strong candidates. The Scc12 locus was narrowed to approximately 1 Mb on chromosome 7 (rs31119421, $P = 4.22 \times 10^{-5}$), which encompassing 18 genes of which Xpo6 is a strong candidate. Only two candidates, Gm3920 and Unc5d, were identified in the Scc8 locus on chromosome 8 (rs33399853, $P = 3.33 \times 10^{-4}$). The Ccs1 locus on chromosome 12 also had significant region-wide associations with AOM-induced...
Colon tumorigenesis (rs29164033, $P = 2.62 \times 10^{-4}$). Eight candidate genes were identified in this locus: *Gm2417*, *Galc*, *Gpr65*, *Kcnk10*, *Spata7*, *Ptpn21*, *Zc3h14*, *Zc3h14* and *Eml5*.

**Gene expression of candidates in the identified loci**

To identify if candidate genes were significantly activated or repressed in colon tumorigenesis, we analyzed gene expression profiles of colon tumors from *Apc* $^{Min/+}$ (24) and AOM-induced (23) mouse colon tumor models compared to normal colon. Ten candidate genes had significant differential expression in both models of colon tumorigenesis compared to normal colon controls ($P < 0.05$) (*Table S2*). *Sema5a*, *Grem2* and *Rabep2* were the most significantly decreased transcripts in both *Apc* $^{Min/+}$ and AOM mouse models; while *Eif3c* and *Tufm* were the most increased transcripts (*Fig. 5*). When comparing human colon tumors with matched normal tissues, GREM2, SEMA5A and TUFM were altered in the consistent trends with the above mouse models (*Table S2*).

**Discussion**

In the present study, we performed large-scale phenotyping to systematically evaluate the sensitivity of 33 inbred mouse strains to AOM-induced colon carcinogenesis (*Fig. 1*). The strains used in the study were derived from genetically diverse genealogies and most of them are on the priority list of the publicly accessible Mouse Phenome Database (MPD) ([http://phenome.jax.org/pub-cgi/phenome/mpdcgi](http://phenome.jax.org/pub-cgi/phenome/mpdcgi)). Many of these strains have not been characterized previously for AOM colon carcinogenesis. The relative AOM sensitivities for a large set of inbred mouse strains established in the study will facilitate future studies of colon tumorigenesis. AOM is an organotropic colon carcinogen that is commonly used to induce colon
tumors in rodents (27). In inbred mice, the high frequency of AOM-induced tumors was observed exclusively within the distal colon; the majority of mouse colon tumors arise as aberrant crypt foci, progress to adenomas, and ultimately result in adenocarcinomas, paralleling the progression of CRC in humans (27). Notably, a recent study showed that APC protein is aberrant in AOM-induced mouse colon adenomas and carcinomas (28). These support the use of mouse models for studying the genetics and pathogenesis of colon cancer. In general, AOM-induced colon cancer in rodents can recapitulate in a highly reliable way the phases of initiation and progression of tumor that occur in humans. However, it should be noted that p53 mutations are rarely observed and the tendency to metastasize is low in AOM-induced colon cancer mouse model (29). Other mouse models such as genetically modified animals are also very useful for studying diverse human CRC.

Using these AOM-induced colon tumor data, we performed a genome-wide association analysis in 268 mice from 33 inbred strains, and subsequently fine mapping of 16 previous linkage regions using the association data. The GWAS also identified a novel susceptibility locus on chromosome 15 in which Sema5a is a major candidate (Fig. 3). Sema5a belongs to a large family of proteins involved in the patterning of both the vascular and the nervous systems. Interestingly, Sema5a expression was significantly repressed in both Apc Min/+ and AOM-induced colon tumors (Fig. 5). Fine mapping association analysis narrowed 5 of 16 QTLs for colon tumorigenesis into less than 1-Mb genomic regions of interest. This makes positional identification of candidate genes in these susceptibility loci more feasible (Fig. 4 and Table 2).

The Scc3 locus on mouse chromosome 1 was initially mapped in the linkage analysis of 192 (BALB/c × CcS-19) F2 mice (16) and was further narrowed to less than 500-Kb by our association mapping. The most significant SNP in the refined Scc3, rs32121685, is located near
two genes, *Fmn2* and *Grem2*. *Grem2* mRNA expression was significantly repressed in both *Apc<sup>Min<sup>−/−</sup></sup>* and AOM-induced colon tumors, while we did not observed significant changes in *Fmn2* expression. A recent GWAS identified common genetic variants at CRAC1 (HMPS) locus on human chromosome 15q13.3 that confers CRC risk in the Ashkenazi population (9). The CRAC1 locus was initially characterized in the classical linkage analysis of families with hereditary mixed polyposis syndrome (HMPS). The HMPS is a Mendelian condition characterized by multiple colorectal polyps and CRC (30). Interestingly, two candidate genes near the strongest SNP association were FMN1 and GREM1. Syntenic regions containing FMN1/GREM1 and FMN2/GREM2 are highly conserved between mice and humans. This has been suggested to arise from an ancient gene duplication event (Supplementary fig. S5) (31).

Genes around the Scc3 locus in mice are present in the same order and orientation in humans. These data suggest FMN1/GREM1 and FMN2/GREM2 may be lineage-specific susceptibility genes for colon cancer.

Three major candidates were identified in the refined Scc2 locus: *Fap*, *Ifih1*, and *Gca*. *Fap* encodes a homodimeric integral membrane gelatinase belonging to the serine protease family. It is selectively expressed in reactive stromal fibroblasts of epithelial cancers, granulation tissue of healing wounds, and malignant cells of bone and soft tissue sarcomas (32, 33). This gene is involved in the control of fibroblast growth or epithelial-mesenchymal interactions during development, tissue repair, and epithelial carcinogenesis. Abrogation of *Fap* enzymatic activity attenuates tumor growth in HEK293 cells (34). Increased expression of *Fap* is associated with lymph node metastasis in colorectal, esophageal, ovarian and pancreatic cancers (35-38).

The refined Scc12 locus is a gene-rich region in which *Gsg1l*, *Spns1*, *Rabep2*, and *Sh2b1* are interesting candidates. Of them, *Spns1* play roles in programmed cell death in Drosophila
melanogaster and has orthologs in nematode, mouse, and human (39). *Rabep2* is a RAB GTPase binding effector protein which encodes a member of AP-1 family of transcription factors involved in cell proliferation, differentiation, apoptosis, and other biological processes. *Rebep2* was up-regulated in hyperplastic and neoplastic breast disorders (40). *Sh2b1* mediates activation of various kinases and may function in cytokine and growth factor receptor signaling and cellular. A recent study found that SH2-B beta specifically activates JAK2 and functions as an adapter protein that cross-links actin filaments, leading to modulation of cellular responses in response to JAK2 activation (41, 42). Another study demonstrates an essential role of SH2-B beta in the activation of the Src kinase and the resulting mitogenic response, causing phenotypic cell transformation involving the Src substrate STAT3 (43). The refined Scc8 locus only contained one predicted gene *Gm3920* and one known gene *Unc5d* (i.e., Unc5h4). *UNC5H4* is a netrin-1 receptor UNC5H family member and is a direct transcriptional target of p53 that is induced during DNA damage-mediated apoptosis (44).

*Gpr65*, *Ptpn21* and *Eml5* are major candidates in the refined Ccs1 locus. *Gpr65* encodes a pro-apoptotic G protein-coupled receptor that promotes glucocorticoid-induced apoptosis. Activation of *Gpr65* by its agonist psychosine markedly enhanced dexamethasone-induced apoptosis in a Gpr65-dependent manner (45). Up-regulation of *Gpr65* in human tumors is involved in driving or maintaining tumor formation (46). *Ptpn21* is a member of the protein tyrosine phosphatase (PTP) family that are known to regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. Frameshift mutations in coding repeats of PTP genes were frequently observed in colorectal tumors with microsatellite instability (47).
In summary, we identified a novel susceptibility locus on mouse chromosome 15 for AOM-induced colon tumorigenesis through a genome-wide association study in inbred mice. Subsequent fine mapping analysis further identified five of 16 previous linkage regions to be associated with colon tumor susceptibility. These susceptibility loci were narrowed to less than 1-Mb regions of interest in which candidate genes were identified. These findings will provide important insights into the pathways of CRC development and may ultimately lead to more effective individually targeted cancer-prevention strategies.
Acknowledgments

We thank the Broad Institute of Harvard and MIT, the Wellcome Trust Center for Human Genetics, Perlegen Sciences and The Jackson Lab for releasing/collecting inbred laboratory mouse SNP data. We thank Nancy Saccone and Weimin Duan for their assistance in the data analysis.
References


Table 1 Top SNPs associated with colon tumorigenesis in inbred mice.

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Chr</th>
<th>Pos a</th>
<th>Alleles b</th>
<th>Maf c</th>
<th>P value d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tumor multiplicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs32359607</td>
<td>15</td>
<td>32327123</td>
<td>A/G</td>
<td>0.424</td>
<td>6.31 × 10^{-6}</td>
</tr>
<tr>
<td>rs32137981</td>
<td>15</td>
<td>32369170</td>
<td>A/G</td>
<td>0.364</td>
<td>7.80 × 10^{-6}</td>
</tr>
<tr>
<td>rs31836718</td>
<td>1</td>
<td>77495091</td>
<td>G/A</td>
<td>0.469</td>
<td>2.89 × 10^{-5}</td>
</tr>
<tr>
<td>rs31119421</td>
<td>7</td>
<td>133221746</td>
<td>A/G</td>
<td>0.438</td>
<td>4.22 × 10^{-5}</td>
</tr>
<tr>
<td>rs33125466</td>
<td>17</td>
<td>57041453</td>
<td>C/T</td>
<td>0.400</td>
<td>4.83 × 10^{-5}</td>
</tr>
<tr>
<td>rs31940065</td>
<td>1</td>
<td>82034606</td>
<td>G/A</td>
<td>0.424</td>
<td>5.17 × 10^{-5}</td>
</tr>
<tr>
<td>rs13482505</td>
<td>15</td>
<td>30566973</td>
<td>A/C</td>
<td>0.375</td>
<td>5.29 × 10^{-5}</td>
</tr>
<tr>
<td>rs3677347</td>
<td>7</td>
<td>133844466</td>
<td>T/A</td>
<td>0.414</td>
<td>6.48 × 10^{-5}</td>
</tr>
<tr>
<td>rs30919908</td>
<td>19</td>
<td>21656425</td>
<td>G/C</td>
<td>0.485</td>
<td>6.63 × 10^{-5}</td>
</tr>
<tr>
<td>Tumor volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6182695</td>
<td>4</td>
<td>107061698</td>
<td>G/A</td>
<td>0.267</td>
<td>2.33 × 10^{-5}</td>
</tr>
<tr>
<td>rs31836718</td>
<td>1</td>
<td>77495091</td>
<td>G/A</td>
<td>0.469</td>
<td>6.14 × 10^{-5}</td>
</tr>
<tr>
<td>rs33204842</td>
<td>17</td>
<td>38672775</td>
<td>C/A</td>
<td>0.406</td>
<td>6.81 × 10^{-5}</td>
</tr>
<tr>
<td>rs32137981</td>
<td>15</td>
<td>32369170</td>
<td>A/G</td>
<td>0.364</td>
<td>7.51 × 10^{-5}</td>
</tr>
<tr>
<td>rs28127134</td>
<td>4</td>
<td>107084372</td>
<td>G/A</td>
<td>0.25</td>
<td>8.28 × 10^{-5}</td>
</tr>
<tr>
<td>rs28127052</td>
<td>4</td>
<td>107097861</td>
<td>A/G</td>
<td>0.25</td>
<td>8.28 × 10^{-5}</td>
</tr>
<tr>
<td>rs32014282</td>
<td>4</td>
<td>107076284</td>
<td>G/T</td>
<td>0.296</td>
<td>8.30 × 10^{-5}</td>
</tr>
<tr>
<td>rs28127136</td>
<td>4</td>
<td>107084304</td>
<td>A/G</td>
<td>0.273</td>
<td>8.34 × 10^{-5}</td>
</tr>
<tr>
<td>rs32359607</td>
<td>15</td>
<td>32327123</td>
<td>A/G</td>
<td>0.424</td>
<td>8.81 × 10^{-5}</td>
</tr>
<tr>
<td>rs28127240</td>
<td>4</td>
<td>107055865</td>
<td>C/T</td>
<td>0.242</td>
<td>9.64 × 10^{-5}</td>
</tr>
<tr>
<td>rs32695065</td>
<td>4</td>
<td>107066262</td>
<td>C/T</td>
<td>0.242</td>
<td>9.64 × 10^{-5}</td>
</tr>
</tbody>
</table>

a The SNP positions (bp) were based on the NCBI mouse genome build 37.1.
b Allele in bold is minor allele of SNPs
c Minor allele frequency in the sample of inbred mice
d P = 3.39 × 10^{-6} and P = 7.29 × 10^{-6} correspond to 5% and 10% genome-wide thresholds in tumor multiplicity, respectively; and P = 2.53 × 10^{-6} and P = 6.04 × 10^{-6} correspond to 5% and 10% genome-wide thresholds in tumor volume.
Table 2 Candidate genes and SNPs in the refined previous linkage regions

<table>
<thead>
<tr>
<th>QTL</th>
<th>Chr</th>
<th>Region (Mb)</th>
<th>dbSNP^a</th>
<th>P-value</th>
<th>Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scc3</td>
<td>1</td>
<td>174.9-191.1</td>
<td>rs32121685</td>
<td>$3.47 \times 10^{-4}$</td>
<td>Fmn2, Grem2</td>
</tr>
<tr>
<td>Scc2</td>
<td>2</td>
<td>34.8-65.3</td>
<td>rs28039498</td>
<td>$2.09 \times 10^{-4}$</td>
<td>Fap, Ifih1, Gca</td>
</tr>
<tr>
<td>Sc212</td>
<td>7</td>
<td>126.8-149.6</td>
<td>rs31119421</td>
<td>$4.22 \times 10^{-5}$</td>
<td>Gsg1l, Spns1, Rabep2, Sh2b1, Eif3c</td>
</tr>
<tr>
<td>Scc8</td>
<td>8</td>
<td>7.2-35.7</td>
<td>rs33399853</td>
<td>$3.33 \times 10^{-4}$</td>
<td>Unc5d</td>
</tr>
<tr>
<td>Ccs1</td>
<td>12</td>
<td>80.1-101.3</td>
<td>rs29164033</td>
<td>$2.62 \times 10^{-4}$</td>
<td>Gpr65, Ptpn2l</td>
</tr>
</tbody>
</table>

^a The most significant SNP in the linkage region was presented.
Figure legend

**Figure 1** AOM-induced colon tumorigenesis in inbred strains. (a) Tumor multiplicity. (b) Tumor volume. 278 mice from 33 inbred mouse strains were measured for tumor multiplicity and volume at 24 weeks after the first dose of AOM.

**Figure 2** Results from genome-wide association analysis of AOM-induced colon tumor multiplicity (a) and tumor volume (b) in inbred mice. Scatter plot of P values in –log scale for 97,854 SNPs. The two red dash lines are genome-wide thresholds of P = 0.05 and 0.10 for colon tumor phenotypes.

**Figure 3** A genetic locus on chromosome 15 affecting colon tumorigenesis in inbred mice. Blue and red dots represent association results from AOM-induced colon tumor multiplicity and volume, respectively. Genes shown on the upper side of the chromosome (turquoise lines) are transcribed in the – orientation (from right to left), and those on the lower side (pink lines) in the + orientation (from left to right).

**Figure 4** Fine mapping of previous linkage regions. Associations achieved region-wide significance in five previous linkage regions, including Scc3, Scc2, Scc12, Scc8 and Ccs1. Blue and red dots represent association results from the analysis of colon tumor multiplicity and volume, respectively. The x axis is physical distance of mouse genome (Mb) (NCBI mouse build 37.1).

**Figure 5** Expression of Sema5a, Grem2, Rabep2 and Eif3c in colon tumors. All the samples were referenced to expression levels of normal control colon samples.
Figure 1

(a)

(b)
Figure 2

(a)

(b)
Figure 3
Figure 4
Figure 5

**Sema5a**

- **Control**
- **AOM**
- **APC (Min/+)**
  - $P = 0.002$
  - $P = 1.09 \times 10^{-6}$

**Grem2**

- **Control**
- **AOM**
- **APC (Min/+)**
  - $P = 9.30 \times 10^{-17}$
  - $P = 2.11 \times 10^{-10}$

**Rabep2**

- **Control**
- **AOM**
- **APC (Min/+)**
  - $P = 1.05 \times 10^{-5}$
  - $P = 5.36 \times 10^{-5}$

**Eif3c**

- **Control**
- **AOM**
- **APC (Min/+)**
  - $P = 1.21 \times 10^{-6}$
  - $P = 0.005$
Molecular Cancer Research

Genome-wide association and fine mapping of genetic loci predisposing to colon carcinogenesis in mice

Pengyuan Liu, Hong-Bo Liu, Yan Lu, et al.

Mol Cancer Res  Published OnlineFirst November 29, 2011.

Updated version  Access the most recent version of this article at:
doi:10.1158/1541-7786.MCR-10-0540

Supplementary Material  Access the most recent supplemental material at:
http://mcr.aacrjournals.org/content/suppl/2012/02/24/1541-7786.MCR-10-0540.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.