Genome-Wide Association and Fine Mapping of Genetic Loci Predisposing to Colon Carcinogenesis in Mice

Pengyuan Liu1,3, Yan Lu1,3, Hongbo Liu3, Weidong Wen3, Dongmei Jia3, Yian Wang3, and Ming You2,3

Abstract

To identify the genetic determinants of colon tumorigenesis, 268 male mice from 33 inbred strains derived from different genealogies were treated with aminopyrene (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 ± 8.6) and tumor volume ranged from 0 to 706.5 mm3 (mean = 87.4 ± 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 single-nucleotide polymorphisms, we carried out a genome-wide association study (GWAS) of AOM-induced colon tumorigenesis and identified a novel susceptibility locus on chromosome 15 (rs32359607, P = 6.31 × 10−9). Subsequent fine mapping confirmed five (Scs3, Scs2, Scs12, Scs8, and Csf1) 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, Fmm2, Grem2, Gpr15, Xp06, Rabep2, Eif3c, Unc5d, and Gpr65. In particular, the refined Scs3 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. Mol Cancer Res; 10(1); 66–74. ©2011 AACR.

Introduction

Colorectal cancer 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 colorectal cancer development and account for approximately 35% of colorectal cancer risk (2). Highly penetrant germline mutations in mismatch repair genes are estimated to account for approximately 6% of all colorectal cancer cases (3). In addition to these rare variants, much of colorectal cancer inherited susceptibility is likely attributable to multiple low penetrance common variants. Direct evidence for common variants for colorectal cancer is highlighted by recent genome-wide association studies (GWAS). These GWAS on colorectal cancer have identified 10 independent loci that confer risk of colorectal cancer, including those on chromosomes 8q24.21, 11q23, 18q21.21, 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 colorectal cancer risk can be explained by currently identified loci. Other genetic factors underlying colorectal cancer remain unidentified and this strongly supports the continued search for novel colorectal cancer 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 with cross-breeding experiments were carried out for chemically induced colon tumors in mice (11–16). As a result, 16 quantitative trait loci (QTL) 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.

Note: Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).

Corresponding Author: Ming You, Department of Pharmacology and Toxicology and the Cancer Center, Medical College of Wisconsin, Milwaukee, Wisconsin; Phone: 414-955-2565; Fax: 414-955-6058; E-mail: myou@mcw.edu

doi: 10.1158/1541-7786.MCR-10-0540
©2011 American Association for Cancer Research.

66 Mol Cancer Res; 10(1) January 2012
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 single-nucleotide polymorphism (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 GWAS 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 organotropism colon carcinogen, to induce colon tumors. We identified a novel genetic susceptibility locus on mouse chromosome 15 and narrowed 5 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, 129 x 1/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, and SM/J), 4 Castle’s mice (C57BL/6J, C57BLKS/J, C57L/J, and C58/J), 4 Swiss mice (FVB/NJ, NON/ShiLtJ, SJL/J, and SWR/J), 2 wild-derived strains (CAST/EiJ and PERA/EiJ), 1 strain derived from colonies from China and Japan (KK/HlJ), and 6 other inbred strains (BTBR T1 + d/J, BUB/BnJ, CE/J, LG/J, PL/J and RIJSS/J). This brought a wide range of variation in genome-wide SNP data for testing hypothesis of no association between the SNP and 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), \( X \) is an \( n \times q \) matrix of fixed effects. For each permutation, the EMMA approach de-

Colon tumorigenesis assays

The inbred mouse strains were purchased from The Jackson Laboratory. Animals were housed in plastic cages with hardwood bedding and dust covers, in a high-efficiency particulate air (HEPA) filtered, environmentally controlled room (24°C ± 1°C, 12/12-hour light/dark cycle). Animals were given Rodent Lab Chow, #5001 (Purina) and water ad libitum. Animals received administration of AOM (Sigma) 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 PBS to remove fecal material, opened longitudinally, and placed flat on a filter paper. In general, 5 to 10 mice per strain were treated and phenotyped. The flushed colons were fixed in Tellyesnickyz’s solution (20) overnight followed by 70% ethanol. The fixed colons were evaluated by at least 2 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: \( V = 4/3\pi r^3 \), 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 of animal models, which are more informative than the binary phenotype (case vs. 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. An R package implementation of the EMMA method is publicly available (http://mouse.cs.ucla.edu/emma/). A 2-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 \( P = 0.05 \)) 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
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 carry out association tests in that linkage region instead of the whole genome. Briefly, we first identified one logarithm of the odds (LOD) supporting interval for each of the previous linkage regions. Then, we conducted 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 \( Ap \) 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 carried out by the Illumina Gene Expression SentrixBeadChip HumanRef-8_V2, and the expression data were 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 \( \pm \) SD, 6.5 \( \pm \) 8.6) and tumor volume ranged from 0 to 706.5 mm\(^3\) (mean \( \pm \) SD, 87.4 \( \pm \) 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 per mouse with diameter of 2 mm or more including C57L/J, FVB/NJ, BTBR T\(^{+}\)tf/J, A/J, NON/ShiLtJ, SM/J, KK/HIJ, and I/LnJ. In contrast, strains AKR/J, PERA/Eij, RIHS/J, DBA/1J, C57BL/6J, and DBA/2J had less than 1 tumor per 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 conducted a genome-wide association analysis on 33 strains of inbred mice with 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 correspond to a point-wise \( P = 3.39 \times 10^{-6} \) and \( 7.29 \times 10^{-6} \) for the analysis of colon tumor multiplicity and corresponds to \( P = 2.53 \times 10^{-6} \) 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, 2 SNPs on chromosome 15, rs32359607, and rs32137981, are strongly associated with colon tumor multiplicity (\( P = 6.31 \times 10^{-6} \) and \( 7.80 \times 10^{-6} \), respectively). These 2 SNPs achieved genome-wide significance level of 0.10. Candidate genes nearby these 2 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; ref. 26). Here, we systematically investigated genetic association signals in these colon tumor susceptibility loci. We first identified microsatellite markers flanking 1 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 (Supplementary Table S1). The region-wide association analysis identified a number of SNPs showing significant associations with colon tumor susceptibility in 5 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, \( P = 3.47 \times 10^{-6} \); Table 2). The refined Scc3 locus includes 10 annotated
genes, of which *Fmn2* and *Grem2* are major candidates. The *Scc2* locus on chromosome 2 (rs28039498, \( P = 2.09 \times 10^{-4} \)) covers 7 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 *Xp06* is a strong candidate. Only 2 candidates, *Gm3920* and *Umc5d*, were identified in the *Scc8* locus on chromosome 8 (rs33399853, \( P = 3.33 \times 10^{-5} \)). The *Cas1* 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*, *Spata7*, *Pip21*, *Z3h14*, *Z3h14*, and *Emil5*.

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

To identify whether 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 with normal colon. Ten candidate genes had significant differential expression in both models of colon tumorigenesis compared with normal colon controls (\( P < 0.05 \); Supplementary Table S2). *Sema5a*, *Grem2*, and *Rabep2* were the most significantly decreased transcripts in both Apc^{Min/+} and AOM mouse models; whereas *Eif5a* and *Tufm* were the most increased transcripts (Fig. 5). When comparing human colon tumors with matched normal tissues, *GREM2*, *SEMA5A* and *TUFM* were

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

<table>
<thead>
<tr>
<th>dbSNP</th>
<th>Chr</th>
<th>Pos*</th>
<th>Allelesb</th>
<th>Mafc</th>
<th>Pd</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs32359607</td>
<td>15</td>
<td>32327123</td>
<td>A/G</td>
<td>0.424</td>
<td>( 6.31 \times 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 \times 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 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs31119421</td>
<td>7</td>
<td>133221746</td>
<td>A/G</td>
<td>0.438</td>
<td>( 4.22 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs33125466</td>
<td>17</td>
<td>57041453</td>
<td>C/T</td>
<td>0.400</td>
<td>( 4.83 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs31940065</td>
<td>1</td>
<td>82034606</td>
<td>G/A</td>
<td>0.424</td>
<td>( 5.17 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs13482505</td>
<td>15</td>
<td>30566973</td>
<td>A/C</td>
<td>0.375</td>
<td>( 5.29 \times 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 \times 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 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs6182695</td>
<td>4</td>
<td>107061698</td>
<td>G/A</td>
<td>0.267</td>
<td>( 2.33 \times 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 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs32204842</td>
<td>17</td>
<td>38672775</td>
<td>C/A</td>
<td>0.406</td>
<td>( 6.81 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs32137981</td>
<td>15</td>
<td>32369170</td>
<td>A/G</td>
<td>0.364</td>
<td>( 7.51 \times 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 \times 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 \times 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 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs28127136</td>
<td>4</td>
<td>107084304</td>
<td>A/G</td>
<td>0.273</td>
<td>( 8.34 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs32359607</td>
<td>15</td>
<td>32327123</td>
<td>A/G</td>
<td>0.424</td>
<td>( 8.81 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs28127240</td>
<td>4</td>
<td>107055865</td>
<td>C/T</td>
<td>0.242</td>
<td>( 9.64 \times 10^{-6} )</td>
</tr>
<tr>
<td>rs32695065</td>
<td>4</td>
<td>107066262</td>
<td>C/T</td>
<td>0.242</td>
<td>( 9.64 \times 10^{-6} )</td>
</tr>
</tbody>
</table>

**Abbreviations:** Maf, minor allele frequency; Pos, SNP positions.

*The SNP positions in base pairs were based on the National Center for Biotechnology Information mouse genome build 37.1.

bAllele in bold is minor allele of SNPs.

cMinor allele frequency in the sample of inbred mice.

dMinor allele frequency in the sample of inbred mice. \( P = 3.33 \times 10^{-6} \) and \( 7.29 \times 10^{-6} \) correspond to 5% and 10% genome-wide thresholds in tumor multiplicity, respectively, and \( P = 2.53 \times 10^{-6} \) and \( 6.04 \times 10^{-6} \) 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</th>
<th>P</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 × 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 × 10^{-4}</td>
<td>Fap, Ifh1, Gca</td>
</tr>
<tr>
<td>Scc12</td>
<td>7</td>
<td>126.8–149.6</td>
<td>rs31119421</td>
<td>4.22 × 10^{-6}</td>
<td>Gsg1l, Spns1, Rabep2, Sh2b1, Eif3c</td>
</tr>
<tr>
<td>Scc8</td>
<td>8</td>
<td>7.2–38.7</td>
<td>rs33399853</td>
<td>3.33 × 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 × 10^{-4}</td>
<td>Gpr65, Ptpn21</td>
</tr>
</tbody>
</table>

*The most significant SNP in the linkage region was presented.*
altered in the consistent trends with the above mouse models (Supplementary Table S2).

**Discussion**

In this study, we carried out 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 MPD (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 colorectal cancer 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 metasize 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 colorectal cancer.

Using these AOM-induced colon tumor data, we conducted a genome-wide association analysis in 268 mice from 33 inbred strains, and subsequently fine mapping of 16 previous linkage regions with the association data. The GWAS 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*<sup>Min/</sup> 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, Table 2).

The *Scc3* locus on mouse chromosome 1 was initially mapped in the linkage analysis of 192 (BALB/c × C57BL/6J) 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 2 genes, *Fmn2* and *Grem2*. *Grem2* mRNA expression was significantly repressed in both *Apc*<sup>Min/</sup> and AOM-induced colon tumors, although we did not observe significant changes in *Fmn2* expression. A recent GWAS identified common genetic variants at CRAC1 (HMPS) locus on human chromosome 15q13.3 that confers colorectal cancer risk in the Ashkenazi population (9). The *CRAC1* locus was initially

![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.](image-url)
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 colorectal cancer (30). Interestingly, 2 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; ref. 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, Sh2b1, and Gia. 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 Gg1l, Spin1, Rabep2, and Sh2b1 are interesting candidates. Of them, Spin1 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. Rabep2 was upregulated 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-B 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 shows an essential role of SH2-B 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., Unc5b4). UNCSH4 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 Em15 are major candidates in the refined Ces1 locus. Gpr65 encodes a proapoptotic 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). Upregulation 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 is known to regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. Framed 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 GWAS in inbred mice. Subsequent fine mapping analysis further identified 5 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 colorectal cancer development and may ultimately lead to more effective individually targeted cancer prevention strategies.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank the Broad Institute of Harvard and MIT, the Wellcome Trust Center for Human Genetics, Perlegen Sciences, The Jackson Laboratory for releasing/collecting inbred laboratory mouse SNP data, and Nancy Saccone and Weimin Duan for their assistance in the data analysis. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 30, 2010; revised October 4, 2011; accepted November 21, 2011; published OnlineFirst November 29, 2011.

References

Genome-Wide Association and Fine Mapping of Genetic Loci Predisposing to Colon Carcinogenesis in Mice

Pengyuan Liu, Yan Lu, Hongbo Liu, et al.


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

Cited articles
This article cites 47 articles, 11 of which you can access for free at: http://mcr.aacrjournals.org/content/10/1/66.full.html#ref-list-1

Citing articles
This article has been cited by 5 HighWire-hosted articles. Access the articles at: /content/10/1/66.full.html#related-urls

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.