Infrequent Detection of Germline Allele-Specific Expression of TGFBR1 in Lymphoblasts and Tissues of Colon Cancer Patients

Kishore Guda1, Leanna Natale1, James Lutterbaugh1,2, Georgia L. Wiesner1,3, Susan Lewis3, Stephan M. Tanner4, Jerneja Tomsic4, Laura Valle4, Albert de la Chapelle4, Robert C. Elston5, Joseph Willis6, and Sanford D. Markowitz1,2

1 Department of Medicine, Ireland Cancer Center, University and Case Medical Center, Cleveland, Ohio
2 Howard Hughes Medical Institute, University and Case Medical Center, Cleveland, Ohio
3 Department of Genetics and Center for Human Genetics, University and Case Medical Center, Cleveland, Ohio
4 Human Cancer Genetics Program, Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio
5 Department of Epidemiology and Biostatistics, University and Case Medical Center, Cleveland, Ohio
6 Department of Pathology, Case Western Reserve, University and Case Medical Center, Cleveland, Ohio

Abstract

Recently, germline allele-specific expression (ASE) of the gene encoding for transforming growth factor-β type I receptor (TGFBR1) has been proposed to be a major risk factor for cancer predisposition in the colon. Germline ASE results in a lowered expression of one of the TGFBR1 alleles (>1.5-fold), and was shown to occur in ~20% of informative familial and sporadic colorectal cancer (CRC) cases. In the present study, using the highly quantitative pyrosequencing technique, we estimated the frequency of ASE in TGFBR1 in a cohort of affected individuals from familial clusters of advanced colon neoplasias (cancers and adenomas with high grade dysplasia), and also from a cohort of individuals with sporadic CRCs. Cases were considered positive for the presence of ASE if demonstrating an allelic expression ratio <0.67 or >1.5. Using RNA derived from lymphoblastoid cell lines, we find that of 46 informative Caucasian advanced colon neoplasia cases with a family history, only two individuals display a modest ASE, with allelic ratios of 1.65 and 1.73, respectively. Given that ASE of TGFBR1, if present, would likely be more pronounced in the colon compared to other tissues, we additionally determined the allele ratios of TGFBR1 in the RNA derived from normal-appearing colonic mucosa of sporadic CRC cases. We however found no evidence of ASE in any of 44 informative sporadic cases analyzed. Taken together, we find that germline ASE of TGFBR1, as assayed in lymphoblastoid and colon epithelial cells of colon cancer patients, is a relatively rare event.

Keywords

TGFbeta; pyrosequence; colon cancer; allele
Introduction
Colon cancer is one of the major leading causes of cancer-related deaths in developed nations. While most of the colorectal cancer (CRC) cases are sporadic in nature, up to 30% of the cases show familial clustering of the disease (1). Two well characterized, highly penetrant inherited syndromes, familial adenomatous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC), together account for ~5% of all CRC cases (1). However, in 20–25% of additional colon cancer cases with a family history, the underlying inherited genetic risk factors for cancer predisposition remain largely unknown (1).

The transforming growth factor \( \beta \) (TGF-\( \beta \)) signaling pathway is a key tumor suppressor pathway that is commonly targeted for inactivation in colon cancers, with one-third of colon cancers acquiring somatic mutations in TGF-\( \beta \) type II receptor (2), and many of the remaining colon cancers acquiring somatic mutations and deletions in the downstream signaling proteins, SMAD elements (3). While genetic aberrations in the TGF-\( \beta \) type I receptor (TGFBR1) have not been identified thus far in colon cancer, a recent study reported that peripheral blood leukocytes from ~20% of informative sporadic and non-syndromic familial CRC cases demonstrate reduced expression of the TGFBR1 gene owing to a lowered expression one of the TGFBR1 alleles (4). This differential allele expression, referred to as germline allele-specific expression (ASE), was suggested to be dominantly inherited and proposed to confer a substantially increased risk for CRC (4). In order to verify the extent of ASE of TGFBR1 in colon cancer, we used a panel of 5 single nucleotide polymorphisms (SNPs) in the 3′ UTR region of the TGFBR1 gene to determine the allele expression ratio in a group of familial and sporadic CRC cases. Using DNA and RNA from lymphoblastoid cells and from colonic mucosa cells, this study finds that germline ASE of TGFBR1 is less common in colon neoplasia than originally proposed (4).

Materials and Methods

Patients

**Familial colon neoplasia cases**—For this study, colon neoplasia kindreds were obtained from the Colon Neoplasia Sibling Study (CNSS) cohort that were each selected to have at least one pair of siblings in which both individuals had by age 65 been diagnosed with either a colon cancer or an adenoma with high grade dysplasia (5). A total of 102 of these affected siblings (98% Caucasian, 2% African American) and 44 disease-free familial member controls (98% Caucasian, 2% African American) were chosen based on availability of EBV transformed lymphocyte cell lines. The average age at diagnosis for the CRC cases was 51.2 years, and the average age of having documented a clean colonoscopy for the unaffected family members was 55.4 years. All advanced colon neoplasms arising in the CNSS cohort were microsatellite stable, with no evidence for FAP or HNPCC, as described earlier (5). The familial cases and controls included 53% and 61% females, respectively.

**Sporadic CRC cases**—A random sample of 96 CRC cases (74% Caucasian, 21% African American and 5% unknown ethnicity) with no family history of colon cancer were accrued under an IRB approved protocol at the Case Medical Center. The average age at diagnosis for the sporadic CRC cases was 69 years, and included 51% females.

**DNA/RNA extraction and cDNA synthesis**

DNA and RNA from the familial CRC cases and controls were extracted from lymphoblastoid cell lines as described earlier (5). DNA and RNA from the sporadic CRC cases were extracted from the normal-appearing mucosal layer of the colon as described earlier (6). Total RNA was
treated with DNAse (TURBO DNA-free kit, Ambion, Austin, TX) prior to cDNA synthesis (Invitrogen, Carlsbad, CA). Reverse transcriptase (RT)-negative control reactions were performed in parallel on the DNAse-treated RNAs for each sample.

**PCR primers and conditions**

Five SNPs, rs868, rs334348, rs334349, rs420549 and rs1590, located in the TGFBR1 3′UTR were used for quantifying individual allele expression. Three of these SNPs (rs334348, rs334349, rs1590) were selected from the group characterized by Valle et al.(4), with two additional SNPs also added to increase the number of informative cases that could be analyzed for ASE. The range of minor allele frequencies for these 5 SNPs was 9% to 38% (http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp). The PCR and sequencing primers for the analysis were designed using the PSQ Assay Design software (Biotage, Charlottesville, VA). One of the PCR primer pair for each SNP was biotinylated at the 5′-end, and purified using HPLC. The sequences are as follows.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Forward (5′→3′)</th>
<th>Reverse (5′→3′)</th>
<th>Size</th>
<th>Seq (5′→3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs868</td>
<td>ctaggctttacgacatgctggtt</td>
<td>biotin-ctgctgcaaauggagcaata</td>
<td>193</td>
<td>ggttttaatttgaggtg</td>
</tr>
<tr>
<td>rs334348</td>
<td>biotin-tctgaaaatgcctttctcacc</td>
<td>tgcacacaagttgagaa</td>
<td>117</td>
<td>caaacatgaccatgct</td>
</tr>
<tr>
<td>rs334349</td>
<td>biotin-ccaatcctttgctccgacattg</td>
<td>ggttcacacaactcctcaactg</td>
<td>148</td>
<td>eggtgtgtaaggaggag</td>
</tr>
<tr>
<td>rs420549</td>
<td>biotin-lggtaacttggcttgatttgg</td>
<td>aagggtagtgattttgtgtgc</td>
<td>124</td>
<td>tgttgacactaaccg</td>
</tr>
<tr>
<td>rs1590</td>
<td>ttatactggttggggagt</td>
<td>biotin-tttcaccccccagagacta</td>
<td>172</td>
<td>aagggtagttggtgcttt</td>
</tr>
</tbody>
</table>

The PCR conditions for DNA included 50 cycles of 95°C for 45s, 60°C for 45s and 72°C for 45s. For RNA, similar PCR conditions as above were used except that the reaction was terminated at 35 cycles. For each SNP, the PCR reactions for DNA and RNA were performed in triplicates in a 50μl reaction volume using AmpliTaq Gold polymerase (Applied Biosystems, Foster City, CA). In addition, PCR reactions were performed on the RT-negative DNAse-treated RNA samples as an internal control.

**Pyrosequencing and allele quantitation**

Following PCR, the DNA and RNA amplification products were sequenced using the sequencing primers listed above on a PyroMark MD pyrosequencing instrument (Biotage) as per the manufacturer’s instructions, and the frequencies of individual alleles for each SNP were obtained using the PyroMark MD software package (Biotage). The ratio of the common versus rare allele frequency in the RNA was then calculated, and normalized to their respective ratio in the DNA using the formula: Allele Expression Ratio= RNA level of (common allele expression/rare allele expression) divided by genomic DNA level of (common allele expression/rare allele expression). Samples were considered positive for ASE if the median value of this ratio as determined among all informative SNPs was <0.67 or >1.5, as suggested by the earlier study (4). As a quality control for this method, we also assessed ASE for these 5 SNPs in mixtures of varying proportions of DNA and RNA from two individuals who were homozygous for opposite SNP alleles in TGFBR1 3′ UTR. Pyrosequencing yielded a ratio of allele specific expression of these mixtures that was ≥97% of the predicted ratio (data not shown).

**Results and Discussion**

**ASE of TGFBR1 is rare in familial colon neoplasia**

The recent study by Valle et al., reports that ASE of TGFBR1 occurs in ~ 20% of the informative familial colon cancers (4). While most cases of ASE showed an imbalance in allele expression
ratio of <2-fold, individuals with ASE show a marked deregulation in the downstream TGF-
β signaling in four out of the four cases tested (4). To determine the prevalence of ASE in an
independent group of familial colon neoplasia cases, we assayed for the presence of ASE of
TGFBR1 in lymphoblastoid cells from a well characterized cohort of familial colon neoplasia
cases (5). Of the 102 familial cases and 44 familial controls selected, 46 and 17 individuals,
respectively, were heterozygous for informative TGFBR1 SNPs. These 46 familial cases
included 20 individuals with colon carcinoma, and 26 individuals with advanced colon
adenomas that showed high grade dysplasia (carcinoma in situ) (5). As defined by the earlier
report, samples were considered positive for ASE if the allele ratio is >1.5 fold (4). Based on
the above criteria, we found that of the 46 informative familial cases, only two individuals
showed ASE with allele ratios slightly above the cut-off value (allele ratios, 1.65 and 1.73,
respectively), one of whom was affected by colon cancer, and the other by an advanced
adenoma with high grade dysplasia (Figure 1). These two cases belonged to two different
families. Thus, ASE as assayed in this cohort was infrequent (4.3% of cases, CI (95%) = 0.5%,
14.8%), and did not account for a substantial number of the familial colon neoplasia cases that
we investigated.

One intriguing observation regarding ASE is that the TGFBR1 gene on chromosome 9q22.33
lies in the middle of genomic region in which we and others have previously found linkage to
risk of familial colon neoplasia (5,7). Furthermore, Valle et al., observed that within four
families that they studied, ASE appeared to segregate as a dominant genetic trait inherited in
consort with markers of high risk TGFBR1 alleles (4). Of the 46 familial cases that we studied,
31 cases (derived from 22 families) were from sibships showing excess sharing of genomic
markers across the 9q22.2-31.2 interval, consistent with disease in these families being linked
to a 9q22.2-31.2 pathogenic allele (5). Both of the individuals in whom we detected ASE were
from families belonging to this 9q22.2-31.2 linked cohort (5). However, even among these 9q
linked families, ASE was an uncommon finding, accounting for only 2 of 31 cases (6.5%, CI
(95%) = 0.8%, 21.4%) and 2 of 22 families (9.1%, CI (95%) = 1.1%, 29.2%). Thus we conclude
that ASE of TGFBR1 is unlikely to be the major driver of linkage of some colon neoplasia
families to the 9q22.2-31.2 region.

**Absence of TGFBR1 ASE in sporadic colon cancers**

In addition to familial colon cancers, ASE was also found to occur in up to 20% of informative
sporadic CRCs included in the initial study that defined this phenomenon (4). We did not have
available a collection of lymphoblastoid cells from individuals with sporadic colorectal
cancers. However, we reasoned that if ASE of TGFBR1 were a driver of sporadic colon cancers,
the lower expression of one TGFBR1 allele should likely also be evidenced in colon epithelial
cells from affected individuals. Accordingly, we determined the frequency of ASE in normal
colon mucosa resections from a group of 96 sporadic CRC cases. As per the analysis of ASE
in lymphoblastoid cells, we utilized RNA and DNA extracted from the normal-appearing
colic mucosa of these CRC cases for subsequent pyrosequencing analysis. Of the 96 sporadic
CRCs, 44 were heterozygous for informative TGFBR1 SNPs. Pyrosequencing analysis
revealed that none of these 44 informative sporadic cases show allele ratios >1.5-fold in their
colic mucosa (Figure 1). These results further suggest that ASE of TGFBR1 is infrequently
found in individuals with sporadic colon cancers.

In summary, our current findings in a Caucasian-dominated study population did not find
TGFBR1 ASE as being a common inherited risk factor for colon cancer. Of the total 90 colon
neoplasia cases we studied, only 2 showed evidence for ASE. This was true despite enriching
our study with 31 cases of familial colon neoplasia linked to the 9q22.2-31.2 locus. One
possible concern was that the original detection of ASE was based on the use of SNaPshot
assays for quantitation of expression of individual TGFBR1 alleles (4); whereas, the current
study employed pyrosequencing for this purpose. To directly compare these assays, a set of twelve samples from the current study, including two with ASE values above 1.5 and ten with ASE values below 1.5, were re-assayed in a blinded fashion using SNaPshot assays under the same conditions as employed in the initial study of ASE (4). For these samples, SNaPshot yielded ASE values essentially identical to those determined by pyrosequencing. Our findings also differ from the study of Valle et al., in that the current study includes some familial cases who had developed advanced adenomas by age 65 or younger; whereas in Valle et al., (4) familial cases were individuals who had at any age developed colon cancers. It is possible that this difference in ascertainment criteria might alter the proportion of familial cases in whom ASE could be detected. Furthermore, whereas the initial study by Valle et al. to determine ASE was based on assays of archived pellets of peripheral blood leukocytes (4), the current study utilized either EBV immortalized lymphocytes or normal colon mucosal samples to determine ASE. Accordingly, a biologically intriguing hypothesis would be that ASE of TGFBR1 might be a phenotype that is tissue specific, as reported for some other genes (8,9). An alternative hypothesis would be that ASE might be more common among individuals who carry minor alleles for certain specific TGFBR1 SNPs, and so might be more or less frequent depending upon the precise SNP panel used for identifying informative persons. A third possibility would of course be that the differences between the current and the previous study may simply reflect some still unappreciated differences between the two specific populations tested. Last, it is possible that the true frequency of ASE lies somewhere in between the estimates provided by these first two cohorts that have been studied. We currently cannot be certain which one of the above hypotheses, or some alternative hypothesis, underlies the differences between the current and the previous findings. Future studies should therefore be specifically designed to address the basis for this current discordance in findings.

Acknowledgments

Supported by PHS awards CA130901-01 to SDM, CA043703 to the Case Western Reserve University Comprehensive Cancer Center, and CA16058 to the Ohio State University Comprehensive Cancer Center. SDM is an investigator of the Howard Hughes Medical Institute.

References

Figure 1. Analysis of TGFBR1 allele ratios in familial and sporadic colon neoplasia
Shown in the figure is the allele expression ratio of TGFBR1 for each sample as determined in lymphoblastoid cells from affected individuals from colon neoplasia kindreds (Familial Cases), lymphoblastoid cells from unaffected members of colon neoplasia kindreds (Familial Controls), and normal colon mucosa from individuals with sporadic colorectal cancers (Sporadic CRCs). Samples with ratios <0.67 or >1.5 cut-off values (indicated by horizontal dashed lines) were considered positive for ASE. The solid lines within each scatter plot indicate the mean and standard error values. The triangles and spheres within the Familial Cases data set indicate individuals with colon cancers and colon adenomas with high-grade dysplasia, respectively.