Common and Distinct Genomic Events in Sporadic Colorectal Cancer and Diverse Cancer Types

Eric S. Martin,1,3 Giovanni Tonon,1,3 Raktim Sinha,1 Yonghong Xiao,2 Bin Feng,2 Alec C. Kimmelman,1,4 Alexei Protopopov,1 Elena Ivanova,3 Cameron Brennan,1 Kate Montgomery,8 Raju Kucherlapati,1 Gerald Bailey,1 Mark Redston,6 Lynda Chin,2,3 and Ronald A. DePinho1,2,3

1Department of Medical Oncology and 2Center for Applied Cancer Science, Belfer Institute for Innovative Cancer Science, Dana-Farber Cancer Institute, Department of Genetics and Medicine, 3Harvard Radiation Oncology Program, and 4Department of Dermatology, Harvard Medical School, 5Department of Pathology, Brigham and Women’s Hospital, and 6Harvard Partners Center for Genetics and Genomics, Boston, Massachusetts; and 7Neurosurgery Service, Memorial Sloan-Kettering Cancer Center, Department of Neurosurgery, Weill Cornell Medical College, New York, New York

Abstract
Colorectal cancer (CRC) is a major cause of cancer morbidity and mortality, and elucidation of its underlying genetics has advanced diagnostic screening, early detection, and treatment. Because CRC genomes are characterized by numerous non-random chromosomal structural alterations, we sought to delineate regions of recurrent amplifications and deletions in a collection of 42 primary specimens and 37 tumor cell lines derived from chromosomal instability neoplasia and microsatellite instability neoplasia CRC subtypes and to compare the pattern of genomic aberrations in CRC with those in other cancers. Application of oligomer-based array-comparative genome hybridization and custom analytic tools identified 50 minimal common regions (MCRs) of copy number alterations, 28 amplifications, and 22 deletions. Fifteen were highly recurrent and focal (<12 genes) MCRs, five of them harboring known CRC genes including EGFR and MYC with the remaining 10 containing a total of 65 resident genes with established links to cancer. Furthermore, comparisons of these delimited genomic profiles revealed that 22 of the 50 CRC MCRs are also present in lung cancer, glioblastoma, and/or multiple myeloma. Among 22 shared MCRs, nine do not contain genes previously shown genetically altered in cancer, whereas the remaining 13 harbor 35 known cancer genes, of which only 14 have been linked to CRC pathogenesis. Together, these observations point to the existence of many yet-to-be discovered cancer genes driving CRC development, as well as other human cancers, and show the utility of high-resolution copy number analysis in the identification of genetic events common and specific to the development of various tumor types. [Cancer Res 2007;67(22):10736–43]

Introduction
Colorectal cancer (CRC) is the third most commonly diagnosed cancer and ranks second in cancer mortality, with ~106,680 new cases and an estimated 55,170 deaths in the United States in 2006 alone.9 Extensive genetic and genomic analysis of human CRC has uncovered germ line and somatic mutations relevant to CRC biology and malignant transformation. These mutations have been linked to well-defined disease stages from aberrant crypt proliferation or hyperplastic lesions to benign adenomas, to carcinoma in situ, and finally to invasive and metastatic disease, thereby establishing a genetic paradigm for cancer initiation and progression (1).

Genetic and genomic instability are catalysts for colon carcinogenesis (2). CRC can present with two distinct genomic profiles termed (a) chromosomal instability neoplasia (CIN), characterized by rampant structural and numerical chromosomal aberrations driven in part by telomere dysfunction (3) and mitotic aberrations (2) and (b) microsatellite instability neoplasia (MIN), characterized by near-diploid karyotypes with alterations at the nucleotide level due to mutations in mismatch repair (MMR) genes (4). Germ line MMR mutations are highly penetrant lesions that drive the MIN phenotype in hereditary nonpolyposis CRCs, accounting for 1% to 5% of CRC cases (4). Although CIN and MIN are mechanistically distinct, their genomic and genetic consequences emphasize the requirement of dominant mutator mechanisms to drive intestinal epithelial cells toward a threshold of oncogenic changes needed for malignant transformation.

A growing number of genetic mutations have been identified and functionally validated in CRC pathogenesis. Activation of the WNT signaling pathway is an early requisite event for adenoma formation. Somatic alterations are present in APC in >70% of non-familial sporadic cases and seem to contribute to genomic instability and induce the expression of MYC and CCND1 (5), whereas activating CTNNB1 mutations represent an alternative means of WNT pathway deregulation in CRC (6). KRAS mutations occur early in neoplastic progression and are present in ~50% of large adenomas (7). The BRAF serine/threonine kinase and PIK3CA lipid kinase are mutated in 5% to 18% and 26% of sporadic CRCs, respectively (8, 9). BRAF and KRAS mutations are mutually exclusive in CRC, suggesting overlapping oncogenic activities (10).

Mutations associated with CRC progression, specifically the adenoma-to-carcinoma transition, target the TP53 and the transforming growth factor-β (TGF-β) pathways. Greater than 50% of CRCs harbor TP53 inactivating mutations (7), and 30% of

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

Requests for reprints: Ronald A. DePinho, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney Street, M413, Boston, MA 02115. Phone: 617-632-6085; Fax: 617-632-6069; E-mail: ron_depinho@dfci.harvard.edu and Lynda Chin, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney Street, M446, Boston, MA 02115. Phone: 617-632-6472; Fax: 617-582-8169; E-mail: lynda_chin@dfci.harvard.edu.

©2007 American Association for Cancer Research. doi:10.1158/0008-5472.CAN-07-2742

http://www.cancer.org/
cases possess TGFβ-RII mutations (11). MIN cancers consistantly inactivate TGFβ-RII by frameshift mutations, whereas CIN cancers target the pathway via inactivating somatic mutations in the TGFβ-RII kinase domain (15%) or in the downstream signaling components of the pathway, including SMAD4 (15%) or SMAD2 (5%) transcription factors (12).

Numerous molecular, cytogenetic, copy number analyses, and resequencing efforts have pointed to a large number of genetic and genomic events that may underlie CRC pathogenesis. Recent resequencing of >13,000 coding sequences in breast cancer and CRC identified 189 genes with somatically acquired, nonsynonymous mutations (so-called cancer genes), the majority of which were not previously implicated in the neoplastic process (13). Similarly, high-resolution copy number alteration (CNA) analyses employing bacterial artificial chromosome (BAC)–based array-comparative genome hybridization (aCGH) have defined focal events, frequent gains of 8q, 13q, and 20q, and losses of 5q, 8p, 17p, and 18q (14–19). The pathogenetic relevance of these amplifications and deletions is inferred by their recurrence, presence of known cancer genes at these loci, and alternative mechanisms targeting resident genes by mutation or epigenetic means (13, 14, 16, 18).

Together, these observations point to the possible existence of many genetic aberrations driving CRC development, the majority of which have yet to be defined. Establishment of robust oligomer-based aCGH and computational methods has enabled high-resolution genome-wide analysis of CNAs using full complexity genomic DNA (20–22). This approach has identified a large number of highly focal recurrent amplifications and deletions in diverse human cancers, including non–small cell lung cancer, glioblastoma, and multiple myeloma. Here, we have generated comparable human CRC genomic profiles and integrated these profiles with those from other cancers in an effort to identify common genes and loci driving the pathogenesis of CRC and other human cancer types.

**Materials and Methods**

**Cell lines and primary tumors.** All of the primary tumors were acquired from the Brigham and Women’s Hospital tissue bank (Boston) under an approved institutional protocol. The tumor histology was confirmed by pathologists (M.R., G.B.) before inclusion in this study. All of the cell lines were obtained from the American Type Culture Collection. The characteristics of the primary tumors and cell lines are detailed in Supplementary Table S1.

**aCGH profiling on oligonucleotide microarrays.** Genomic DNAs from cell lines and primary tumors were extracted according to the manufacturer’s instructions (Gentra Systems). Genomic DNA was fragmented and random-prime labeled as described (20, 22, 23) and hybridized to human oligonucleotide microarrays. The oligonucleotide array contains 22,500 elements designed for expression profiling (Human 1A V2, Agilent Technologies), for which 16,097 unique map positions were defined (National Center for Biotechnology Information (NCBI) Build 35). The median interval between mapped elements is 54.8 kb, 96.7% of intervals are <1 Mb, and 99.5% are <3 Mb. Fluorescence ratios of scanned images of the arrays were calculated as the average of two paired arrays, and the raw aCGH profiles were processed to identify statistically significant transitions in copy number by using a segmentation algorithm (20, 22, 23). In this study, significant copy-number changes are determined on the basis of segmented profiles only.

**Automated minimal common region definition.** Loci of amplification and deletion are evaluated across samples with an effort to define minimal common regions (MCRs) targeted by overlapping events in two or more
samples. An algorithmic approach is applied to the segmented data, as described in Supplementary Methods.

Sequencing, mutation, and immunohistochemical analysis. For mutation analysis of the KRAS, BRAF, and PIK3CA genes, coding exons were PCR amplified, purified, and sequenced using standard protocols at the Harvard Partners Center for Genetics and Genomics (primers available upon request). Chromatograms were assembled using the program SEQUENCHER (Gene Codes) and manually compared with the NCBI table.
<table>
<thead>
<tr>
<th>Chromosome MCR position (Mb)</th>
<th>HRF locus</th>
<th>Known cancer genes*</th>
<th>Can-genes †</th>
<th>Candidate cancer genes</th>
<th>miRNA (hsa-mir-X)</th>
<th>Lung MCR ‡</th>
<th>GBM MCR ‡</th>
<th>MM MCR ‡</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amplification</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>143.18–158.37</td>
<td>BCL9, ARNT, AF1Q, TPM3, MUC1, PRCC, NTRK1, IRTA1, FCGR2B</td>
<td>THBS3, IRTA2</td>
<td>9-1*</td>
<td>150.95–151.92</td>
<td>142.60–152.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>101.56–101.95</td>
<td>+</td>
<td>TFG</td>
<td></td>
<td></td>
<td></td>
<td>95.26–103.06</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>39.26–41.10</td>
<td>+</td>
<td>CCND3</td>
<td></td>
<td></td>
<td></td>
<td>49.91–52.16, 45.99–53.24</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>50.91–54.32</td>
<td>+</td>
<td>TFEB</td>
<td></td>
<td></td>
<td></td>
<td>53.47–55.73</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>133.13–135.58</td>
<td>MYB</td>
<td>EYA4</td>
<td></td>
<td></td>
<td></td>
<td>0.17–0.81, 3.81–5.48</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>44.99–48.46</td>
<td>+</td>
<td>EGFR</td>
<td></td>
<td></td>
<td></td>
<td>120.50–126.52</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>105.26–107.15</td>
<td>+</td>
<td>PIK3CG</td>
<td></td>
<td></td>
<td></td>
<td>128.5–129.02, 128.50–146.20</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>100.03–103.64</td>
<td>COX6C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>128.5–129.02</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>120.92–121.53</td>
<td>+</td>
<td>MTBP</td>
<td></td>
<td></td>
<td></td>
<td>128.5–129.02</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>128.31–129.14</td>
<td>+</td>
<td>MYC</td>
<td></td>
<td></td>
<td></td>
<td>128.5–129.02</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>130.92–136.68</td>
<td>30b, 30dα</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>128.5–146.20</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.17–6.67</td>
<td>CCND2, ELKS, ZNF584</td>
<td>KRAS</td>
<td>KRAS</td>
<td>24.61–26.95</td>
<td>15.02–37.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>24.61–27.46</td>
<td>+</td>
<td>ZNF198, CDX2, FLT3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>19.14–33.30</td>
<td>+</td>
<td>KLF5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>97.47–114.11</td>
<td>+</td>
<td>FGF14, IRS2</td>
<td></td>
<td>111.77–112.97, 113.25–114.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>100.06–100.27</td>
<td>+</td>
<td>DLK1</td>
<td></td>
<td></td>
<td></td>
<td>0.24–11.19</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>45.52–47.13</td>
<td>+</td>
<td>PHKB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>63.53–65.35</td>
<td>+</td>
<td>CDH11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>21.83–23.84</td>
<td>+</td>
<td>CDH5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.94–0.98</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2.26–3.84</td>
<td>+</td>
<td>PTPRA, CDC25B</td>
<td>103-2φ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>25.22–29.52</td>
<td>+</td>
<td>PLAGL2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>30.08–30.49</td>
<td>+</td>
<td>GNAS, SS18L1</td>
<td>GNAS</td>
<td>296, 1-1, 133a-2, 124a-3</td>
<td>57.04–60.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>21.54–21.72</td>
<td>+</td>
<td>MSH2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>42.46–47.52</td>
<td>+</td>
<td>PRDM5, SPYR1</td>
<td></td>
<td></td>
<td>52.587–68.443</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>57.51–57.73</td>
<td>+</td>
<td>REST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>121.97–129.40</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>92.95–94.05</td>
<td>+</td>
<td>LOC157697</td>
<td>DLC1</td>
<td>124a-1</td>
<td>0.295–0.356</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.29–0.35</td>
<td>+</td>
<td>PPP2CR, DUSP4, DUSP26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>12.00–13.99</td>
<td>+</td>
<td>WRN</td>
<td>LOC157697, DLC1</td>
<td>200</td>
<td>28.74–33.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>21.60–38.38</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>39.08–41.23</td>
<td>+</td>
<td>SFRP1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>42.99–43.17</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>104.51–108.33</td>
<td>+</td>
<td>TCF7L2, CASP7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>114.82–115.79</td>
<td>+</td>
<td>TCF7L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on the following page)
Results

Recurrent CNAs in sporadic CRC. Forty-two primary CRC tumors and 37 CRC cell lines were subjected to copy number analysis employing a well-established oligomer-based aCGH platform and a modified circular binary segmentation methodology (see Supplementary Methods). The clinical and histopathologic characteristics of these samples, including microsatellite instability (MSI) status, are summarized in Supplementary Table S1, and the aCGH data are available online.10 The overall pattern of genomic aberrations defined in our sample set agrees well with the frequencies of previously reported CRC amplifications and deletion (14, 15, 17–19, 25–29), many of which are consistent with patterns of allelic changes often observed in CRC (ref. 30; Fig. 1). In addition, the analysis of \( \text{KRAS}, \text{BRAF}, \text{TP53} \), and \( \text{PIK3CA} \) mutations indicated that mutation frequencies in our primary tumor and cell line sample set are in line with those of published databases (Supplementary Table S2), although the \( \text{PIK3CA} \) mutation frequency of 5% in our primary tumors is significantly less than that of 28% in our cell lines (\( P = 0.0095, \) Fisher's exact test) and 32% in a published report (31). Also, consistent with previous reports (8, 10), we observed a mutually exclusive pattern of \( \text{BRAF} \) and \( \text{KRAS} \) mutations with \( \text{BRAF} \) mutations predominant in MIN tumors [five of six BRAF mutant primary tumors show MSI, and all are \( \text{MLH1} \) deficient (data not shown)].

A prominent feature of the CRC genomic profiles was the large number of complex CNAs \( (n = 251) \). The presence of CNAs targeting the same locus across different tumor samples and cell lines enabled the definition of a minimal common region (MCR) of gain/amplification or loss/deletion. To identify those MCRs with potentially greater pathogenic relevance, MCRs were further selected based on the occurrence of at least one CNA in a primary tumor and at least one high-amplitude event at that locus (see Supplementary Methods; ref. 22). This approach identified and delimitated 50 MCRs consisting of 28 recurrent amplifications with a median size of 1.86 Mb (range, 0.04–16.64 Mb) containing a total of 1,225 known genes (median of 19 per MCR) and 22 recurrent deletions with a median size of 1.31 Mb (range, 0.06–19.07 Mb) containing a total of 802 known genes (median of 11 per MCR; Table 1). A total of 8 of 42 (20%) primary tumors were determined to be MMR deficient (see Materials and Methods; Supplementary Table S1; data not shown), and their aCGH profiles were classically MIN, containing fewer gross genomic changes and/or focal amplifications and deletions relative to CIN samples (refs. 16, 18, 32; Supplementary Fig. S1).

The veracity of the aCGH data was confirmed by quantitative PCR, fluorescence in situ hybridization (FISH), and/or spectral karyotyping analysis of randomly selected MCRs (\( n = 25 \); Supplementary Fig. S2, data not shown) and by the presence of all previously identified human CRC loci (refs. 14, 15, 17–19; Table 1), including those targeting signature CRC genes like \( \text{KRAS}, \text{MYC} \), and \( \text{EGFR} \). Moreover, other known cancer genes or their homologues not previously implicated in CRC resided within our CRC MCRs.

Table 2. CRC MCR candidate gene list and cross-tumor MCR overlap of CRC, lung adenocarcinoma, glioblastoma, and multiple myeloma MCRs (Cont’d)

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>MCR position (Mb)</th>
<th>HRF locus</th>
<th>Known cancer genes*</th>
<th>Can-genes †</th>
<th>Candidate cancer genes</th>
<th>miRNA (hsa-mir-X)</th>
<th>Lung MCR ‡</th>
<th>GBM MCR †</th>
<th>MM MCR †</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>132.78–133.61</td>
<td>BNIP3</td>
<td></td>
<td>TCERG1L,</td>
<td></td>
<td>128.9–135.241</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PPP2R2D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>73.73–75.24</td>
<td></td>
<td></td>
<td>PHLDA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>44.04–44.65</td>
<td>+</td>
<td>C1orf155</td>
<td>BTBD5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>98.93–99.26</td>
<td>+</td>
<td></td>
<td>CCNK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>53.90–72.97</td>
<td>TCF12</td>
<td>SMAD3,</td>
<td>NEDD4</td>
<td>90.27–101.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>30.70–30.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.66–2.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>5.22–14.67</td>
<td></td>
<td></td>
<td>PTPRM,</td>
<td>6.95–10.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>75.32–76.01</td>
<td></td>
<td></td>
<td>PTPN2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>47.62–48.55</td>
<td></td>
<td>CEACAM1</td>
<td>47.78–48.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Known cancer genes were derived from the Cancer Gene Census and/or have been previously implicated in CRC. Can-genes are from Sjoblom et al. (13), whereas non-CRC MCRs have been previously reported (20, 22, 23).

*http://www.sanger.ac.uk/genetics/CGP/Census/.
†Sjoblom et al. (13).
‡Tonon et al. (22).
§Maher et al. (21).
∥Carrasco et al. (20).
¶Volinia et al. (34).

10 http://www.ncbi.nlm.nih.gov/geo/; GEO Accession # GSE7604
(Table 2). Additionally, 11 of 50 (22%) MCRs contain common proviral integration sites (data not shown), and 8 MCRs encompass 13 microRNAs (Table 2), several of which, hsa-mir-9-1, hsa-mir-30d, and hsa-mir-103-2, are known to be aberrantly expressed in multiple cancer types (33, 34), suggesting that these genomic events may promote deregulated expression of these dominant and recessive microRNAs in CRC.

We identified 15 highly recurrent (>10%) and focal (≤12 MCR resident genes) events designated high-recurrence-focal (HRF) MCRs: 8 amplicons and 7 deletions. The HRF MCRs show a recurrent 35% among primary tumors and cell lines and contain 101 known genes (Table 2). Several HRF MCRs contain known CRC genes, including SFRP1 (35), MYC (36), and EGFR (37), as well as REST, a tumor suppressor gene identified in a genetic screen and found to be deleted and/or mutated in CRC (38). In addition, the 10q25.2-10q26.11 HRF MCR spans nine genes, including the WNT signaling mediator TCF7L2/TCF4 (39) and the effector CASP7 (40). Among the remaining 10 HRF MCRs with a total of 65 genes (Table 1 and 2), there are genes encoding cancer-relevant functions. For example, the focal 13q21.33-13q22.3 amplification that contains the intestinal-enriched kruppel-like factor (KLF5), a positive regulator of cellular proliferation and transformation (41), and the focal 6p21.2-6q12 amplification, which contains DAAM2, a homologue of DAAM1, a mediator of the WNT induced Dishevelled and Rho complex, a key regulator of cytoskeletal architecture and cell polarity (42).

**Relationship of CRC MCRs to somatic can-gene mutations.**

In addition to known classic CRC mutations, we also examined the extent to which the CRC MCR gene list (Table 2) concurs with the list of 69 CRC can-gene somatic mutations identified in the recent re-sequencing study of 13,023 CCDS genes (13). Of the 2,027 CRC MCR genes, 1,082 (53%) were part of the CCDS gene set. Notably, only 7 of 69 CRC can-genes are present on the list of 1,082 CRC MCR genes (Table 2) with no enrichment for can-genes in the CRC MCRs relative to the CCDS gene set (P = 0.249, Fisher’s exact test). This observation suggests that the majority of cancer genes identified in these two data sets preferentially use distinct mutational mechanisms, or alternatively, that the majority of can-genes or MCR resident genes are not driving the tumorigenic process. Arguing against the latter, at least for the aCGH analysis, is the presence of known cancer genes within the MCRs identified in our data set previously linked to colon and other cancers, cancer-relevant microRNAs, and common proviral integration sites in our data set as noted above. Next, we asked whether there was any enrichment of CRC can-genes versus breast cancer can-genes in our CRC MCR resident gene list. The Sjoblom et al. re-sequencing study identified 122 can-genes in breast cancer, of which only 2 were also CRC can-genes (13). Subsequently, 4 of 122 (3.3%) breast can-genes mapped to the CRC MCRs versus 7 of 69 (10.1%) CRC can-genes (P = 0.0588, Fisher’s exact test). The modest enrichment for CRC versus breast-specific can-genes within genomic regions of alterations defined in CRC suggests that these mutation patterns, and, by inference, our MCRs, may reflect distinct biologically relevant processes in these specific cancer types.

**Comparison of CNAs in CRC and other human cancer types.**

Finally, we asked whether any of our CRC MCRs overlap with, or are distinct from, similarly derived MCR lists of other cancer types. Comparison with non–small cell lung cancer, glioblastoma, and multiple myeloma (20–22) MCRs revealed that 22 (13 amplifications and 9 deletions) of the 50 CRC MCRs (44%) matched with one MCR in at least one of the other tumor types and include the EGRF, MYC, and KRAS loci (Table 2). Eight of these CRC MCRs (6 of 28 amplifications; 2 of 22 deletions) overlapped with an MCR present in two of the other tumor types (Table 2). The overlap of our novel CRC MCRs with an MCR in other cancer types supports their cancer relevance.

Given the shared cross-tumor type MCRs, we asked whether this overlap might delimit further the CRC MCRs, using information from other cancers, and select a more limited list of genes with potential cancer relevance to be enlisted in functional validation. Examination of Table 2 reveals that this approach allows for the reduction in size of several MCRs. For example, the chromosome 6 CRC amplicon (50.91–54.32 Mb) is targeted in non–small cell lung cancer (49.91–52.16 Mb) and multiple myeloma (45.99–53.24 Mb), generating a common MCR (50.91–52.16 Mb) that contains the CRC can-gene, PKHD1 (13), and the somatically altered ICK (43). Such data integration not only strengthens the case for PKHD1 and/or ICK in CRC pathogenesis, but also points to its potential role in non–small cell lung cancer and multiple myeloma. A second case is a focal deletion of chromosome 10 in CRC (132.78–133.61 Mb), which enabled refinement of a large MCR (128.9–135.24 Mb) in multiple myeloma that was delimited to a region with only three genes: PPP2R2D, BNIP3, and TCEB1. Thus, cross-tumor comparisons can be useful in delimiting regions of potential interest for additional in-depth analysis.

**Discussion**

In this study, we defined regions of recurrent CNAs in a collection of sporadic primary CRC specimens and tumor cell lines. The 50 CRC MCRs identified were notable for the presence of well-known CRC genes (KRAS, EGFR, and MYC) and several CRC can-genes (13) with a modest enrichment for CRC can-genes over breast can-genes, presence of known cancer-relevant microRNA genes (hsa-mir-9-1, hsa-mir-30d, and hsa-mir-103-2), and overlap with MCRs in other cancer types (22 of 50 CRC MCRs). As a whole, these data indicate that many additional loci drive CRC pathogenesis and suggest that both common and distinct loci drive the biological processes needed for various cancer types to achieve their malignant end point.

Chromosomal gains and losses previously identified using conventional and lower resolution aCGH were further resolved in our study. Frequent gains of chromosomes 7p, 7q, 8q, 13q, 20p, and 20q have each been previously documented at resolutions nearing ~1 Mb using a BAC aCGH platform (15) and appear with frequencies ranging between 43% and 67% in our sample set (Table 1). The most frequent chromosomal gain, 13q (>62%), contained an MCR (13q21-q22, 72.22–73.15 Mb) with only seven resident genes, including the transcription factor KLF5. KLF5 is highly expressed in epithelia in regions of active proliferation and has been shown to promote cellular proliferation (44). KLF5 binds directly to the 5’ regulatory region of EGFR, which leads to the transcriptional up-regulation of EGFR and the subsequent activation of MEK/ERK signaling (44). Furthermore, the regulation of proliferation by KLF5 is dependent on EGFR and MEK/ERK signaling because the proliferative response to KLF5 is blocked by pharmacologic inhibition of EGFR or MEK. Inhibition of EGFR or MEK also decreases KLF5 expression. Thus, KLF5 regulates MEK/ERK signaling via EGFR and is also downstream of MAPK signaling, providing a novel mechanism for signal amplification or suppression and control of proliferation in epithelial cells (44).
An additional event of potential strong relevance to CRC pathogenesis is a highly focal (0.41 Mb) and highly recurrent (65%) MCR on chromosome 20 (20q12-20q13.33, 30.08–30.39) that contains only nine resident genes including the putative oncogene PLAGL2. PLAGL2 has been shown to cooperate with the CBFB-MYH11 fusion gene in vivo in a mouse model of acute myelogenous leukemia (AML) and to promote S-phase entry and expansion of hematopoietic progenitors and increased cell renewal in vitro (45). PLAGL2, and its family member PLAG1, are overexpressed in 20% of human AML samples (45).

Amplifications targeting chromosome 8q24 was also a very recurrent event (>48%) in our data set and could be additionally resolved into two smaller regions (<0.85 Mb) of high-amplitude gain at 120.92–121.53 and 128.31–129.14 Mb. One region of amplification contains the MYC proto-oncogene whose deregulated expression and activity in CRC has been linked to the aberrant activation of WNT signaling (46) and/or genomic amplification. The other, more centromeric, 8q24 MCR contains five genes including MTBP, which encodes a protein capable of binding to and stabilizing MDM2 and promoting MDM2-mediated degradation of TP53 (47). Amplification and/or overexpression of MTBP locus may provide a cooperative or alternative mechanism to the inactivation of TP53, a hallmark of CRC pathogenesis.

Complex cytogenetic and genomic rearrangements of distal chromosome 8p that include allelic loss via mitotic recombination/loss of heterozygosity, translocation, and/or copy number loss are one of the most frequent events across a wide spectrum of epithelial tumors (48). Recently, high-resolution aCGH analysis of several tumor types including colon cancers revealed several highly resolved, complex CNAs along chromosome 8p (17, 48). Collectively, FISH, breakpoint mapping, and aCGH studies (48) have implicated many possible candidate tumor suppressor genes along distal 8p, including the WRN helicase on 8p12 (31 Mb). In our analysis, copy number loss of distal chromosome 8p ranged from 23% to 49%, with two MCRs surviving the HRF criteria (Table 1). Although the informative deletion of 8p12-p11 (39.08–41.23) in our data lies centromeric to the previously defined breakpoint cluster of 8p12 (48), only 11 genes reside in this 2-Mb MCR. Among these genes, SFRP1, is of particular interest due to its preferential hypermethylation in CRC (49) and the capacity of enforced SFRP1 expression in CRC cells to attenuate WNT signaling even in the presence of downstream mutations (50).

Because most bona fide cancer genes are subject to alterations by multiple mechanisms, a priori selection of genes residing within regions of CNAs for focused re-sequencing represents a plausible strategy. Subsequently, we determined the extent of convergence of somatically mutated candidate cancer genes (can-genes) identified by Sjoblom et al. (13) and recurrent CNAs defined by our analysis. We found that the majority of their can-genes (13) did not reside within our MCRs. However, the failure to establish a statistically significant enrichment, or lack thereof, of somatically mutated genes in regions of genomic alterations may be due the use of small, exploratory data sets and/or a limited number of genes sequenced.

Interestingly, among the MCR-resident somatically mutated genes identified by Sjoblom et al. (13), they were more likely to be CRC can-genes when compared with breast-specific can-genes.

Finally, our comparison of CRC MCRs with those similarly defined in lung adenocarcinoma, glioblastoma, and multiple myeloma identified 23 CRC MCRs common to at least one of these tumor types (20–22). The use of cross-tumor comparisons in defining highly informative MCRs can provide a useful means in narrowing a list of candidate genes and in identifying genes that, when targeted by either CNA and/or somatic mutation, may impact a broad spectrum of tumors. Furthermore, the prioritization of these MCRs may provide a high-yield entry point for the discovery of novel genes important in the development of a wide spectrum of cancers.

Acknowledgments

Received 7/19/2007; revised 8/23/2007; accepted 9/25/2007.

Grant support: NIH grants to R.A. DePinho (U01 CA84313 and R01CA84628) and to L. Chin (ROI CA099041 and ROI CA93947) and NCI grant P50 CA127003-01 (R.A. DePinho and L. Chin). Genome-wide aCGH profiles were generated at the Arthur and Rochelle Beller Cancer Genomics Center at the Dana-Farber Cancer Institute. E.S. Martin is supported by National Cancer Institute 5T32CA09361. G. Tonon is supported by the International Myeloma Foundation, Brian D. Novis Research Award, and by a Special Fellow Award from the Leukemia and Lymphoma Society. A.C. Kimmelman is a recipient of the Leonard B. Holman Research Pathway fellowship. R.A. DePinho is an American Cancer Society Research Professor.

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.

References


Common and Distinct Genomic Events in Sporadic Colorectal Cancer and Diverse Cancer Types


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/67/22/10736

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2007/11/12/67.22.10736.DC1

Cited articles
This article cites 49 articles, 20 of which you can access for free at:
http://cancerres.aacrjournals.org/content/67/22/10736.full.html#ref-list-1

Citing articles
This article has been cited by 11 HighWire-hosted articles. Access the articles at:
/content/67/22/10736.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.