Gastrointestinal Adenocarcinomas of the Esophagus, Stomach, and Colon Exhibit Distinct Patterns of Genome Instability and Oncogenesis

Austin M. Dulak, Steven E. Schumacher, Jasper van Lieshout, Yu Imamura, Cameron Fox, Byoungyong Shim, Alex H. Ramos, Gordon Saksena, Sylvan C. Baca, Jose Baselga, Adam J. Bass, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215. Phone: 617-632-2477; Fax: 617-582-9830; E-mail: Rameen_beroukhim@dfci.harvard.edu

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

A.M. Dulak and S.E. Schumacher contributed equally to this article.

R. Beroukhim and A.J. Bass contributed equally to this article.

Corresponding Authors: Adam J. Bass, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215. Phone: 617-632-2477; Fax: 617-582-9830; E-mail: Adam.Bass@dfci.harvard.edu; and Rameen Beroukhim, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215. E-mail: Rameen_beroukhim@dfci.harvard.edu

doi: 10.1158/0008-5472.CAN-11-3893

©2012 American Association for Cancer Research.

Abstract

A more detailed understanding of the somatic genetic events that drive gastrointestinal adenocarcinomas is necessary to improve diagnosis and therapy. Using data from high-density genomic profiling arrays, we conducted an analysis of somatic copy-number aberrations in 486 gastrointestinal adenocarcinomas including 296 esophageal and gastric cancers. Focal amplifications were substantially more prevalent in gastric/esophageal adenocarcinomas than colorectal tumors. We identified 64 regions of significant recurrent amplification and deletion, some shared and others unique to the adenocarcinoma types examined. Amplified genes were noted in 37% of gastric/esophageal tumors, including in therapeutically targetable kinases such as ERBB2, FGFR1, FGFR2, EGFR, and MET, suggesting the potential use of genomic amplifications as biomarkers to guide therapy of gastric and esophageal cancers where targeted therapeutics have been less developed compared with colorectal cancers. Amplified loci implicated genes with known involvement in carcinogenesis but also pointed to regions harboring novel cancer genes, including a recurrent deletion found in 15% of esophageal tumors where the H3K27me3 transcription factor subunit RUNX1 was implicated, including by functional experiments in tissue culture. Together, our results defined genomic features that were common and distinct to various gut-derived adenocarcinomas, potentially informing novel opportunities for targeted therapeutic interventions. Cancer Res; 72(17); 4383–93. © 2012 AACR.

Introduction

Colorectal, gastric, and esophageal adenocarcinomas collectively account for approximately 180,000 cancer diagnoses and 76,500 deaths each year in the United States and approximately 1.3 million deaths worldwide (1, 2). A better understanding of the somatic genetics of these diseases is a prerequisite for earlier diagnosis and more effective treatment. Colorectal cancer (CRC) genomes have been studied extensively (3, 4); the value of this information is realized by persuasive evidence that KRAS and BRAF mutations in CRC predict lack of response to cetuximab (5, 6). Gastric cancer (GC) and especially esophageal adenocarcinoma (EA) has been subjected to fewer large-scale studies (7–9).

Cancers of the esophagus and stomach commonly arise in a background of intestinal metaplasia, but develop within distinct luminal environments. Nevertheless, they often are treated with identical chemotherapy, and many clinical trials combine patients with these 2 diseases (10, 11). Although the process of intestinal metaplasia preceding GC and EA suggests that these tumors may resemble adenocarcinomas arising from the intestine, they show distinct clinical behavior from CRC. It is therefore important to define the similarities and differences among digestive tract adenocarcinomas at the
genomic and molecular levels. Such a comparison can inform both mechanistic studies and strategies for biomarker-driven therapy.

Two challenges exist in the somatic genetic analysis of cancer: (i) distinguishing "driver" alterations that contribute to tumor development, maintenance, or proliferation from random "passenger" alterations that do not contribute to the neoplastic process, and (ii) identifying the specific genes that mediate tumor progression. Both challenges must be confronted in analysis of somatic copy-number alterations (SCNA) as tumors often harbor many such alterations, each of which often encompasses up to thousands of genes. The study of SCNAs has been greatly enhanced by high-density genomic arrays allowing resolution of individual SCNA boundaries and the ability to study large numbers of tumors. Statistical analysis of SCNAs across many samples can identify regions altered more frequently than expected by chance and also pinpoint the most likely culprit genes in these regions. Pooling data from different but related cancer types can increase both statistical power and the ability to resolve specific gene targets of SCNAs. Given the related origins and documented shared copy-number characteristics of gut-derived adenocarcinomas (12), we hypothesized that evaluating genomic events across these tumors will increase our power to identify common genes active in gut adenocarcinomas and also help uncover differences. Here, we report the largest analysis of SCNAs across gut adenocarcinoma genomes and systematically compare significantly recurrent structural genetic alterations in tumors from distinct regions of the gut. We find multiple known and novel recurrent alterations, including region-specific and shared events.

Materials and Methods

All samples were fresh frozen primary resections from patients not treated with prior chemotherapy or radiation. All cases had diagnoses confirmed by pathologic review and only cases with estimated carcinoma content >70% were selected (Supplementary Table S1). The sample set was not enriched for other features. Tumors annotated as having originated from the gastric–esophageal junction were assigned to the EA collection. DNA was extracted (Supplementary Table S1), quantified with Picogreen dye and hybridized to (215 samples) GeneChip Human Mapping StyI 250K arrays (Affymetrix) or (271 samples) Genome-Wide Human SNP Array 6.0 (SNP6.0; Affymetrix) genomic profiling arrays, according to the manufacturer’s instructions. Data from each of the 2 array platforms were independently normalized and segmented using all data present upon each of the 2 platforms (12, 13). Regions of known germline copy-number polymorphisms were then removed as previously described (14). Human genome build hg18 was used, and raw data files have been deposited at the Gene Expression Omnibus (GSE36460).

Significantly recurrent SCNAs were identified using GISTIC 2.0. (15) All data from each array were used to generate SCNA profiles for each tumor. To enable probe bound GISTIC analyses across data from 2 array platforms, the segmented data from each sample was remapped to the 196,800 probes shared by the 2 platforms. In some cases, this remapping modified the position of the probe bounding the transition between 2 copy-number segments. In these cases, the boundaries were remapped to the nearest probes in the joint set. To remove potentially spurious SCNAs, segments defined by fewer than 9 shared probes were removed (16). Additional details are described in Supplementary Methods.

Results

A comprehensive copy-number dataset from digestive tract adenocarcinomas

We analyzed a cohort of 363 new and 123 publically available genome array profiles from primary untreated gut adenocarcinomas including EA (186), GC (110), and CRC (190; Supplementary Table S1). We determined genome-wide copy-number profiles using either 250K StyI (238,000 probes) or SNP6.0 genome arrays (1.8 million probes; Supplementary Table S1). Copy-number alterations were identified using the full complement of data from each array type. To enable the analysis across platforms, the segmented copy-number data from each sample were remapped to the 196,800 probes common to both arrays. We found no evidence of bias introduced by pooling samples from the 2 platforms (Supplementary Note S1).

Chromosomal instability increases from lower to upper gut adenocarcinomas

Across this set of adenocarcinoma SCNA profiles, visual inspection of the segmented data showed alterations in nearly every part of the genome, and variations in the amount of genomic disruption between different cancers and cancer types (Fig. 1A). To compare levels of genomic disruption, we separately evaluated the frequencies of arm-level (comprising half or more of a chromosome arm) and focal SCNAs in each cancer type. Focal alterations were noted to occur throughout the chromosomes, but showed some predilection for the regions closer to the centromeres and telomeres of each chromosome (Supplementary Fig. S1).

The median number of arm-level gains varied little between types (Fig. 1B), but was significantly increased in focal amplifications in EA and GC compared with CRC (Fig. 1C). Some tumors, particularly among CRC and GC types, showed little apparent genomic disruption, potentially attributable to microsatellite instability or stromal contamination (Supplementary Note S2). The enhanced rate of focal amplification in EA and GC compared with CRC remained valid when these genomically quiet tumors were removed from analysis (Supplementary Fig. S2). Upper gut adenocarcinomas exhibited an even greater excess of higher-level, multicopy focal amplifications (Fig. 1D), which remained after we accounted for possible contaminating noncancer DNA through use of sample-specific thresholds for defining these events (Supplementary Fig. S3). The higher rates of focal genomic amplifications in EA and GC relative to CRC suggest that underlying mechanisms of genomic instability may differ between upper and lower gastrointestinal adenocarcinomas and that genomic...
amplification may be a more common means of oncogene activation in GC/EA.

The difference in rates of focal amplifications between upper and lower gastrointestinal adenocarcinomas did not hold for deletions. Fewer arm-level deletions were seen in GC than other gut adenocarcinomas (Fig. 1B). There was a modest increase in focal deletions in EA compared with GC and CRC (Fig. 1B), a trend that persisted when cases without arm-level SCNAs were excluded (Supplementary Fig. S2). EA and CRC exhibit similar rates of multicopy deletions that may represent homozygous deletions, rates that are significantly higher than in GC (Fig. 1D). This pattern persisted after exclusion of “quiet” samples and use of sample-specific thresholds to identify the multicopy deletions (Supplementary Fig. S3). The discordant bias toward focal amplifications but not deletions in upper gastrointestinal cancers suggests that the mechanisms and selective pressures leading to amplification may differ from deletion.

Recurrent amplifications and deletions across the pooled dataset

**Arm-level events.** We next carried out a GISTIC 2.0 analysis to define significantly recurrent SCNAs, starting with arm-level SCNAs. Arm-level amplifications of chromosomes 7p, 8q, 20p, and 20q recurred significantly across all 3 cancer types (Supplementary Fig. S4A). Events restricted to subtypes included 1q gains in GC and 13q gains in CRC and GC. These significant arm-level gains have been observed previously (17–23). Arm-level deletions were more variable across tumor types. In GC, deletions of 4p and 4q alone were significant reflecting a lower degree of arm-level losses detected earlier in GC (Supplementary Fig. S4B), but these and other deletions (8p, 18p, and 18q) were also detected in EA and CRC. Loss of arms 8p, 14q, and 15q were of higher significance in CRC, whereas loss of 5q, 9p, and 21q were particularly significant among EAs. Loss of 17p (containing...
TP53 was significant in EA and CRC, but not GC. Unique significant losses of 9p and 21q in EA are notable because these arms respectively contain the known and putative tumor suppressors CDKN2A and RUNXI, both targets of focal deletion in EA.

These results are in accordance with previously published data. We compared the frequency of alteration for each cytoband in each cancer type to frequencies determined across 998 CRC, 741 GC, and 71 EA in the Progenetix database curated from cCGH and aCGH data (24, 25). The cytoband-level data are similar for all cancer types, with the exception of lower frequencies of 17q gain in our GC and EA samples (Supplementary Fig. S5). Notably, the low frequency of deletions in GC data compiled by Progenetix mirrored our data suggesting that result is not unique to the samples in our collection.

**Focal alterations.** We next evaluated focal SCNAs across all 486 tumors and identified 33 regions subject to significant (q < 0.01) focal amplification (Fig. 2 and Table 1). Thirteen of these regions contain known oncogenes, including 4 genes involved in cell-cycle regulation (CCNE1, CCND1, CDK6, and MYC) and 7 members of tyrosine kinase/MAPK signaling pathways (EGFR, KRAS, MET, ERBB2, FGR1, FGFR2, and IGFIR). Twenty significant peaks contained no established oncogenes, suggesting potential presence of novel genes or noncoding transcripts that promote intestinal metaplasia and/or gastrointestinal carcinogenesis.

The 18q11.2 amplification peak is the second most significant peak after KRAS and contains only the endodermal transcription factor GATA6. Coupled with functional data suggesting a role for GATA6 in esophageal carcinogenesis (26, 27), these results implicate GATA6 as an important contributor to gastrointestinal neoplasia. The sixth most significant peak in the composite dataset, located at 8p23.1, contains 4 genes, including the related transcription factor GATA4, a candidate target noted previously (28).
We also identified 30 regions of significant focal deletion (Fig. 2 and Table 2). Eight of these regions include genes such as *FHIT* and *WWOX*, with exons spread over large genomic loci (in excess of 1 Mb). Prior studies suggest that such regions often lie in 'fragile sites' or areas of low gene density where deletion may be tolerated, and may not harbor functional tumor suppressors (12, 29). An additional 8 regions contain the known tumor suppressors *CDKN2A*, *SMAD4*, *PTEN*, *APC*, *RUNX1*, *ARID1A*, and *ATM* and the putative tumor suppressor *PARD3B* (30). Fourteen regions did not contain known tumor suppressors or large-footprint genes, but could contain novel factors whose loss contributes to intestinal metaplasia or cancer. Our analysis would not have detected regions of loss of heterozygosity that did not lead to copy-number loss.

The combined analysis across 3 tumor types enabled identification of less common SCNAs. Five amplification peaks were significant in the composite set but not in any individual cancer type. One of these, 13q22.1, contains only 2 genes, including the proliferative transcription factor *KLF5*. The combined dataset also enabled more precise identification of the likely targets of focal SCNAs. For example, the 1p36.11 deletion narrowed from 89 genes in the EA set to only 11 genes in the combined dataset, including the
these novel events and amplifications at 1q42.3. These findings may suggest cooperativity between these novel events and amplification of CCNE1, or reflect subsets of tumors that for other reasons tend to share alterations in these regions.

**Comparison of focal alterations across gut adenocarcinoma and other cancer types**

**Comparison of focal amplifications.** We analyzed focal alterations in each tumor type separately and identified 5, 14, and 25 amplification peaks in CRC, GC, and EA, respectively (Fig. 3A and Supplementary Fig. S6 and Table S2). Highlighting the similarities and differences among digestive tract cancers, only 3 amplifications were significant in all tumor types: 8q24.21 (containing MYC, 17q12 (containing ERBB2), and 18q11.2 (containing GATA6). Among these peaks, ERBB2 is amplified more commonly in esophageal (17%) and gastric (13%) than in colorectal (6%) tumors (Supplementary Fig. S7).

Two amplification peaks were restricted to CRCs. One amplicon contains the RTK, FGFR1, not previously reported on April 20, 2017. © 2012 American Association for Cancer Research. cancerres.aacrjournals.org Downloaded from cancerres.aacrjournals.org on April 20, 2017. © 2012 American Association for Cancer Research.

---

**Table 2.** Significantly deleted focal SCNAS in a cohort of esophageal, gastric, and colorectal adenocarcinomas (q-value <0.01)

<table>
<thead>
<tr>
<th>Cytoband</th>
<th>Residual q-value</th>
<th>Peak boundaries, Mb</th>
<th>Number of genes in peak</th>
<th>Candidate target(s)</th>
<th>Gut adenocarcinoma types represented</th>
</tr>
</thead>
<tbody>
<tr>
<td>3p14.2</td>
<td>1.207E–313</td>
<td>58.98–61.52</td>
<td>1</td>
<td>FHT</td>
<td>E, G, C</td>
</tr>
<tr>
<td>16q23.1</td>
<td>2.96E–151</td>
<td>76.69–78.21</td>
<td>2</td>
<td>WWOX</td>
<td>E, G, C</td>
</tr>
<tr>
<td>9p21.3</td>
<td>5.08E–56</td>
<td>21.98–22.01</td>
<td>1</td>
<td>CDKN2A</td>
<td>E</td>
</tr>
<tr>
<td>5q11.2</td>
<td>9.09E–38</td>
<td>58.30–59.84</td>
<td>3</td>
<td>PDE4D</td>
<td>E, G, C</td>
</tr>
<tr>
<td>16p13.3</td>
<td>8.41E–36</td>
<td>50.63–77.09</td>
<td>2</td>
<td>A2BP1</td>
<td>C</td>
</tr>
<tr>
<td>6p25.3</td>
<td>2.09E–28</td>
<td>1.54–2.57</td>
<td>2</td>
<td></td>
<td>E, G</td>
</tr>
<tr>
<td>4q22.1</td>
<td>2.60E–27</td>
<td>91.37–93.49</td>
<td>2</td>
<td>FAM190A</td>
<td>E, G, C</td>
</tr>
<tr>
<td>6q26</td>
<td>3.61E–21</td>
<td>161.61–163.13</td>
<td>1</td>
<td>PARK2</td>
<td>E, G, C</td>
</tr>
<tr>
<td>9p24.1</td>
<td>1.32E–14</td>
<td>7.77–12.71</td>
<td>3</td>
<td>PTPRD</td>
<td>E, G</td>
</tr>
<tr>
<td>18q12.2</td>
<td>7.52E–14</td>
<td>46.77–46.97</td>
<td>1</td>
<td>SMAD4</td>
<td>E, C</td>
</tr>
<tr>
<td>1p36.11</td>
<td>1.36E–11</td>
<td>26.77–27.31</td>
<td>11</td>
<td>ARID1A</td>
<td>E, C</td>
</tr>
<tr>
<td>21q11.2</td>
<td>1.61E–06</td>
<td>1.00–1.64</td>
<td>17</td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>1q4</td>
<td>2.64E–06</td>
<td>24.39–24.48</td>
<td>1</td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>8p23.3</td>
<td>2.64E–06</td>
<td>1.00–1.71</td>
<td>7</td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>10q23.31</td>
<td>2.64E–06</td>
<td>89.44–90.07</td>
<td>4</td>
<td>PTEN</td>
<td>C</td>
</tr>
<tr>
<td>21q22.12</td>
<td>5.76E–06</td>
<td>35.03–35.35</td>
<td>2</td>
<td>RUNX1</td>
<td>E</td>
</tr>
<tr>
<td>4q35.1</td>
<td>1.80E–05</td>
<td>179.06–187.23</td>
<td>31</td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>7q36.3</td>
<td>3.01E–05</td>
<td>156.49–158.82</td>
<td>12</td>
<td></td>
<td>E</td>
</tr>
<tr>
<td>5q22.2</td>
<td>5.22E–05</td>
<td>110.88–113.74</td>
<td>12</td>
<td>APC</td>
<td>C</td>
</tr>
<tr>
<td>1p13.2</td>
<td>8.53E–05</td>
<td>111.01–119.86</td>
<td>79</td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>2q33.2</td>
<td>1.15E–04</td>
<td>204.53–206.27</td>
<td>1</td>
<td>PARD3B</td>
<td>E</td>
</tr>
<tr>
<td>3p26.2</td>
<td>1.24E–04</td>
<td>4.32–4.51</td>
<td>1</td>
<td></td>
<td>None</td>
</tr>
<tr>
<td>4q35.2</td>
<td>4.15E–04</td>
<td>185.37–191.27</td>
<td>36</td>
<td>CASP3</td>
<td>C</td>
</tr>
<tr>
<td>17p12</td>
<td>1.38E–03</td>
<td>11.40–12.51</td>
<td>5</td>
<td></td>
<td>None</td>
</tr>
<tr>
<td>11q22.3</td>
<td>2.26E–03</td>
<td>103.49–112.70</td>
<td>62</td>
<td>ATM</td>
<td>E</td>
</tr>
<tr>
<td>5q15</td>
<td>3.85E–03</td>
<td>83.69–106.74</td>
<td>56</td>
<td></td>
<td>None</td>
</tr>
<tr>
<td>21q22.3</td>
<td>8.17E–03</td>
<td>38.21–46.94</td>
<td>116</td>
<td></td>
<td>None</td>
</tr>
<tr>
<td>12q21.2</td>
<td>8.88E–03</td>
<td>71.30–86.91</td>
<td>39</td>
<td></td>
<td>None</td>
</tr>
<tr>
<td>10q11.23</td>
<td>9.24E–03</td>
<td>52.31–53.77</td>
<td>5</td>
<td></td>
<td>C</td>
</tr>
</tbody>
</table>
in this disease, but noted to be overexpressed (33). The other unique peak is adjacent to the CRC oncogene \( CDK8 \) at 13q12.2 (34). Esophageal and gastric adenocarcinomas shared 7 amplifications, containing \( VEGFA, EGFR, GATA4, CCND1, MDM2, CCNE1, \) and \( KRAS \). The most significantly amplified gene across our dataset, \( KRAS \), showed a strong foregut preference. Only 5% of CRCs carried focal \( KRAS \) amplification, compared with 21% of upper GI cancers. Conversely, CRCs have substantially higher rates of \( KRAS \) mutation (Supplementary Fig. S8) showing how upper and lower gastrointestinal cancers show distinct ways of altering the same oncogene (35).

An additional 14 regions of amplification were specific to EA. Six of these peaks contained genes known to contribute to cancer (\( CDK6, MCL1, PRKCI, MYB, MET, \) and \( FGFR2 \)), although the others contain no previously described oncogenes. To evaluate how the large sample size enabled identification of relevant targets, we compared our analysis to the SCNA analyses of the largest previously published EA datasets, comprising 42 and 56 tumors (8, 36). Amplifications at \( CCNE1, MET, FGFR2, \) and \( MYB \) were not identified in previous datasets, but these genes have been noted to be overexpressed in EA (37–40). Many peaks that lack known oncogenes were also not noted in earlier reports. Although \( MET \) and \( CCNE1 \) amplifications had been detected previously by focused gene inquiry, \( MYB \) and \( FGFR2 \) amplifications were not noted in prior data. Our sample numbers also afforded greater resolution to identify targets of previously identified SCNAs. For example, an amplicon at 6p21.1 was reported in studies of 42 and 56 EAs to contain between 50 and 70 genes (8, 36). We narrowed this region to only 2 genes, including the vascular endothelial growth factor, \( VEGFA \), the target of the therapeutic antibody bevacizumab. A prior study of EA also identified a 94-gene region of amplification on 3q and attributed this event to \( PIK3CA \) (36). Our analysis narrows this peak to a region containing \( PRKCI \), \( > 5 \) Mb away from \( PIK3CA \).

The presence of \( FGFR2 \) amplifications in EA suggests a potential new therapeutic target for these tumors, similar to CRC (9). We confirmed the presence of \( FGFR2 \) amplification in individual EA cases through quantitative RT-PCR (Supplementary Fig. S9). These results indicate \( FGFR2 \) amplification may serve as a biomarker for the use of \( FGFR2 \)-directed therapy in EA in addition to GC.

Four amplification peaks were restricted to gastric adenocarcinoma. Among these, only 3q27.1, containing the RTK gene, \( EPHB3 \), has been suggested to play a role in cancer progression (41). Compared with prior studies, we noted all confirmed recurrent focal amplifications involving \( GATA4 \) and \( GATA6 \), but also identified novel peaks on 6p21.1 (\( VEGFA \)), 3q27.1 (\( EPHB3 \), 1p36.22, 12q15, and 1q42.3 (9).

The most significant focally amplified tyrosine kinases, \( ERBB2, EGFR, MET, FGFR1, \) and \( FGFR2 \), are known oncogenes and targets of therapeutic agents in current use or development. We detected amplifications involving one or more RTKs in 42% of EA, 28% of GC, and 14% of CRC samples (Fig. 4). Only 10% of tumors had concurrent amplifications of RTKs and \( KRAS \).

Finally, we compared amplifications in GI adenocarcinomas to those found in a study of 2,311 diverse cancers (12). Among the 33 focal amplifications in GI cancers, 42% overlapped with peak regions in other cancers (Supplementary Fig. S10), including 11 regions containing the known oncogenes \( MCL1, MYB, EGFR, FGFR1, MYC, CCND1, KRAS, MDM2, ERBB2, \) and \( CCNE1 \) (marked with asterisks in Fig. 3). Peak regions present in gut, but absent from nongut adenocarcinomas, encompassed genes encoding the tissue-specific transcription factors \( GATA4 \) and \( GATA6 \) and the known or putative oncogenes \( CDK6, FGFR2, MET, EPHB3, \) and \( EPHB6 \). Some of these genes are occasionally amplified in other cancer types, but did not show statistical significance. \( HMGA2 \) amplifications on chromosome 12 are notably absent in GI tumors, despite the presence of chromosome 12 amplification at \( MDM2 \) in both GI and non-GI carcinomas. Similarly, amplification of the \( G1-S \) cell-cycle–dependent kinase \( CDK6 \) is restricted to GI tumors, whereas \( CDK4 \) amplification is significant in all other tumors further suggesting that many cancer use similar pathways with multiple inputs to reach the same output (Supplementary Table S3).

**Comparison of focal deletions.** We identified 16, 8, and 21 peaks of significant deletion in CRC, GC, and EA, respectively (Fig. 3B and Supplementary Fig. S11 and Table S4). Six peaks were shared across the 3 cancer types, all containing genes spanning large genomic loci (including \( FHIT, WWOX, \) and \( MACROD2; \) refs. 12, 42).

Seven deletion peaks were unique to CRC, including 2 peaks that encompass known tumor suppressors (\( APC \) and \( PTEN \)), and 5 that do not. Among the latter, a previously unidentified...
peak at 10q25 contains the apoptosis effector caspase gene CASP7, suggesting a mechanism to evade apoptosis (Supplementary Table S4A).

Eleven deletion peaks were unique to EA whereas none were unique to GC (Fig. 3B). Among the EA peaks, 2 contain known esophageal tumor suppressors (CDKN2A and ATM; Supplementary Table S4B and Fig. S7D). One peak contains RUNX1, which we consider further below. Two peaks were seen in both esophageal and gastric tumors, including 1 large gene (PTPRD) and 1 region (6p25.3) with no known tumor suppressor.

Among the 30 deletion peaks in the composite tumor set (Table 2), 15 were also significant in the non-GI adenocarcinoma study (Supplementary Fig. S10). The common sites included 9 peaks encompassing large genes, 4 peaks containing known deleted tumor suppressors (APC, CDKN2A, ATM, and PTEN), and 2 peaks without either. One of the latter peaks contains the effector caspase gene CASP3, suggesting that deletion of an effector caspase, CASP3 or CASP7, may mark GI adenocarcinomas more generally and not CRC alone. Focal loss of RB1 and TP53 were identified in non-GI but not in digestive tract adenocarcinomas, although 17p deletion, containing TP53, was significant in CRC and EAC. The 15 deletion peaks restricted to GI adenocarcinomas include 13 that lack known tumor suppressors (Supplementary Table S5). The other 2 contain the known tumor suppressors SMAD4 and RUNX1.

**Functional significance of RUNX1 deletions in esophageal adenocarcinoma**

We observed highly focal RUNX1 deletions at 21q22.12 in 15% of EAs, also noted in a recent report (Fig. 5A; ref. 36). RUNX1 behaves as a tumor suppressor in leukemia, where translocations and mutations disrupt gene function (43). We, therefore, evaluated a possible tumor suppressor function for RUNX1 in EA by reintroducing it into the EA cell line OE33, which carries a focal RUNX1 deletion (Fig. 5A). We observed a 69% reduction in anchorage-independent growth relative to GFP-infected cells (Fig. 5B and C). As we did not possess an EA cell line without deletion at the RUNX1 locus, we ectopically expressed RUNX1 in A549 lung cancer cells, which have no focal RUNX1 deletion, to evaluate for generalized cellular toxicity because of overexpression of this gene. In contrast to OE33 cells, RUNX1 expression did not significantly affect A549 colony formation (Fig. 5B and C). These results are consistent with a potential role for RUNX1 as a tumor suppressor in EA.

**Discussion**

These data provide the most comprehensive, high-resolution analysis to date of SCNA patterns across the 3 most common forms of gut adenocarcinoma. Our cohort size enhanced the ability to detect significant SCNAS, and we identified several focused areas of recurrent genomic alteration pointing toward genes that may contribute to cancer. Notably, the genomes of EA and GC cancers contain alterations selected for their contributions to the process of both intestinal metaplasia and malignant transformation to adenocarcinoma.

We observed more focal amplifications in upper GI adenocarcinomas compared with CRC. Unlike CRC, EA, and GC emerge in a setting of bile and acid injury, which may generate DNA strand breaks and contribute to the high rates of SCNA (44). Alternatively, distinct DNA repair pathways or selection for differing stimuli may account for these differences. It is unclear why the enhanced rates of focal amplifications in EA/GC were not matched by a similarly increased rate of focal deletions nor why the rates of deletion in GC fell below those in EA or CRC. Although, the loci of the most significant amplifications peaks fell at known or plausible oncogenes, many focal deletions peaks lie in potential fragile sites. Thus, the mechanisms and selection pressures underlying deletions may differ from those responsible for amplification.

The GATA6 and GATA4 transcription factor genes lie in the second and sixth most significant amplification peaks across the full dataset, respectively. The developmental role of these...
transcription factors and selective amplification in GI cancers suggests that they may add to the growing number of lineage-survival transcription factor oncogenes (45). A parallel phenomenon is amplification of SOX2 in squamous esophageal and lung carcinomas (46), an event absent in EA.

A key clinical observation emerging from these data is that focially amplified RTKs were observed most prominently in EA and GC, suggesting that genomic amplifications will be more important biomarkers in upper gastrointestinal cancers than in CRC. Recent clinical trials reveal benefit when the HER2-directed antibody trastuzumab is combined with chemotherapy in treating ERBB2-amplified or overexpressing GC/EA (11). The presence of ERBB2 amplifications in 6% of CRCs suggests that ERBB2 (HER2)-directed therapy may benefit select patients with CRC (47). In addition, based on evidence that KRAS mutation negatively predicts cetuximab response, highly prevalent KRAS amplification in upper GI tumors may similarly impact clinical decisions. When KRAS is evaluated as a biomarker in upper GI cancers, it will be important to examine both SCNAs and point mutations. Our data also point to other genomic amplifications of clinical relevance. CCNE1 (48) and VEGFA focal amplifications may serve as a marker of response to targeted inhibitors. More broadly, these data point to the inclusion of gene amplifications to guide therapy in upper gastrointestinal adenocarcinomas.

Although these data suggest that many patients with EA or GC may benefit from treatment targeting an amplified RTK, it is unlikely that therapies directed against these targets alone will lead to durable responses. The progression-free survival provided by single-agent therapy with trastuzumab in GC/EA has been modest (11). The presence of complex SCNA profiles in these tumors suggest there could be co-occurring alterations that confer primary resistance (49) or enhanced genomic instability that speeds acquired resistance. In our cohort, the rates of co-occurrence between RTK-associated and other events did not significantly deviate from what would be expected for statistically independent events. However, co-occurrences were detected, and the rates with which individual pairs of events co-occur are likely to inform combinational treatment strategies. Although evaluation of RTK-targeted agents is needed, the genomic complexity and variability of these cancers suggests that combination inhibitor strategies will ultimately be essential.

Beyond these immediately clinically relevant targets, these data provide enhanced insight into specific genes responsible for different subtypes of gastrointestinal adenocarcinomas. Despite a shared intestinal origin, upper gastrointestinal cancers exhibit many distinct events from those seen in CRC, such as alteration of cell-cycle regulators (CCND1, CCNE1, CDK6, and CDKN2A). Moreover, as evidenced by the selective deletion of RUNX1 in EA, clear genomic differences exist between EA and GC. Although our analysis has helped identify loci of recurrent alteration and potential targets of these events, many recurrent regions of SCNA do not contain known oncogenes or tumor suppressors. One challenge moving forward will be to determine which of these regions harbor additional genes or noncoding elements that contribute toward intestinal metaplasia or transformation to cancer. High-resolution genome analyses of large sample numbers, as presented here, can guide future studies and inform the development of strategies to...
reverse the effects of the somatic genetic events that drive these cancers.

Disclosure of Potential Conflicts of Interest

R. Beroukhim, M. Meyerson, and R.A. Shvidrasani are consultants for Novartis. M. Meyerson is a founder of Foundation Medicine and Microgenomics, has commercial research grant from Novartis, is a consultant/advisory board member for Novartis and Foundation Medicine, and has patent royalties from Laboratory Corporation of America. J.D. Luketich has commercial research grant from Accuray. R.A. Shivdasani has commercial research grant from Novartis Oncology. R. Beroukhim has commercial research grant from Novartis Institutes for BioMedical Research and ownership interest (including patents) from AstraZeneca stock. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: A.M. Dulak, J. van Lieshout, J. Tabernero, M. Meyerson, D.G. Beer, R. Beroukhim, A.J. Bass
Development of methodology: J. van Lieshout, Y. Imamura, J. Luketich, R. Beroukhim, A.J. Bass
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. van Lieshout, C. Fox, S. Baca, J. Tabernero, J. Barretina, P. Enzinger, G. Corso, F. Rosvello, L. Lin, S. Bandia, J. Luketich, A. Pennathur, R.A. Shvidrasani, D.G. Beer, T.E. Godfried

References


Writing, review, and/or revision of the manuscript: A.M. Dulak, S. Schumacher, J. van Lieshout, B. Shim, J. Baselga, J. Tabernero, J. Barretina, P. Enzinger, G. Corso, S. Bandia, A. Pennathur, M. Meyerson, S. Ogino, R.A. Shvidrasani, D.G. Beer, T.E. Godfried, R. Beroukhim, A.J. Bass

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A.M. Dulak, J. van Lieshout, B. Shim, A. Ramos, F. Rosvello, J. Luketich, S. Ogino, A.J. Bass

Acknowledgments

The authors thank the members of the Broad Institute Biological Samples Platform and Genomic Analysis Platform for their work enabling this project and also thank the physicians and hospital staff whose efforts in collecting these samples is essential to this research.

Grant Support

This work was supported in part by the National Institute of Health - National Cancer Institute grant 3R01CA096666 (P.I. J.D. Luketich). A.J. Bass is supported by the National Cancer Institute (K08CA134931) and GI SPORE Developmental Project award—P50CA127003 and the DeGregorio Family Foundation. R. Beroukhim is supported by the National Cancer Institute (R01CA122853 and U54CA143798) and a V Foundation Scholarship. F. Rosvello is supported by a grant from Istituto Toscano Tumori.

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 December 1, 2011; revised May 21, 2012; accepted June 4, 2012; published OnlineFirst July 2, 2012.
Gastrointestinal Adenocarcinomas of the Esophagus, Stomach, and Colon Exhibit Distinct Patterns of Genome Instability and Oncogenesis

Austin M. Dulak, Steven E. Schumacher, Jasper van Lieshout, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-11-3893

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2012/07/02/0008-5472.CAN-11-3893.DC1

Cited articles
This article cites 47 articles, 18 of which you can access for free at:
http://cancerres.aacrjournals.org/content/72/17/4383.full.html#ref-list-1

Citing articles
This article has been cited by 4 HighWire-hosted articles. Access the articles at:
/content/72/17/4383.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.