Comparative Genomic Analysis of Esophageal Adenocarcinoma and Squamous Cell Carcinoma

Nishant Agrawal¹,2, Yuchen Jiao¹, Chetan Bettegowda¹,3, Susan M. Hutfless⁴, Yuxuan Wang¹, Stefan David⁴, Yulan Cheng⁴, William S. Twaddell⁶, Nyan L. Latt⁶, Eun J. Shin⁴, Li-Dong Wang⁷,⁸, Liang Wang⁷,⁹, Wancai Yang⁷,¹⁰, Victor E. Velculescu¹, Bert Vogelstein¹, Nickolas Papadopoulos¹, Kenneth W. Kinzler¹, and Stephen J. Meltzer⁴

ABSTRACT

Esophageal cancer ranks sixth in cancer death. To explore its genetic origins, we conducted exomic sequencing on 11 esophageal adenocarcinomas (EAC) and 12 esophageal squamous cell carcinomas (ESCC) from the United States. Interestingly, inactivating mutations of NOTCH1 were identified in 21% of ESCCs but not in EACs. There was a substantial disparity in the spectrum of mutations, with more indels in ESCCs, A:T>C:G transversions in EACs, and C:G>G:C transversions in ESCCs (P < 0.0001). Notably, NOTCH1 mutations were more frequent in North American ESCCs (11 of 53 cases) than in ESCCs from China (1 of 48 cases). A parallel analysis found that most mutations in EACs were already present in matched Barrett esophagus. These discoveries highlight key genetic differences between EACs and ESCCs and between American and Chinese ESCCs, and suggest that NOTCH1 is a tumor suppressor gene in the esophagus. Finally, we provide a genetic basis for the evolution of EACs from Barrett esophagus.

SIGNIFICANCE: This is the first genome-wide study of mutations in esophageal cancer. It identifies key genetic differences between EACs and ESCCs including general mutation spectra and NOTCH1 loss-of-function mutations specific to ESCCs, shows geographic disparities between North American and Chinese ESCCs, and shows that most mutations in EACs are already present in matched Barrett esophagus.

Cancer Discov; 2(10) ; 1–7. © 2012 AACR.

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

N. Agrawal, Y. Jiao, and C. Bettegowda share first authorship.

Corresponding Authors: Nishant Agrawal, The Johns Hopkins University School of Medicine, 601 North Caroline Street, JHOC 6, Baltimore, MD 21287. Phone: 410-955-8878; Fax: 410-955-0548; E-mail: nagrawal@jhmi.edu; Nickolas Papadopoulos (npapado1@jhmi.edu); Kenneth W. Kinzler (kinzler@jhmi.edu); and Stephen J. Meltzer (smeltzer@jhmi.edu).

do: 10.1158/2159-8290.CD-12-0189 ©2012 American Association for Cancer Research.
INTRODUCTION

Esophageal cancer is the sixth most common cause of cancer death and eighth in incidence worldwide, with almost 500,000 new cases and approximately 400,000 deaths in 2008 (1–3). The incidence and histologic subtypes of esophageal cancer exhibit considerable geographic variation. Overall, esophageal squamous cell carcinoma (ESCC) is the most frequent esophageal cancer subtype internationally, predominating in eastern Asia and parts of Africa. Tobacco and alcohol consumption are the major risk factors for ESCCs, but other environmental influences including nitrosamines, nutritional deficiencies, specific carcinogens, low socioeconomic status, limited intake of fruits and vegetables, and consumption of very hot beverages have been implicated in specific geographic regions (4–7). In contrast, esophageal adenocarcinoma (EAC) is the dominant subtype and one of the most rapidly increasing cancers in Western countries. Its increasing incidence has been associated with a corresponding increase in gastroesophageal reflux disease (GERD) and obesity (1, 8). Chronic GERD and its occasional development into Barrett esophagus are the major risk factors for EACs, along with tobacco and obesity (9–14). The 5-year survival rate of patients with esophageal cancer is poor (~15%), and most patients with esophageal cancer present with unresectable or metastatic disease (1, 15).

The molecular alterations underlying esophageal carcinogenesis have been studied in some depth. *TP53* point mutations occur in at least 50% of esophageal cancer cases (16–23). *TP53* mutations have also been detected in early stages of EAC and ESCC tumorigenesis as well as in benign Barrett esophagus mucosa (18, 21). A host of additional genes have been studied for mutation in esophageal cancer, but in most of these single-gene studies, very few mutations were identified. To our knowledge, a comprehensive evaluation of all coding regions for mutations has not yet been undertaken in esophageal cancer; thus, it is not yet known whether any previously unstudied genes are commonly mutated in these tumors. Furthermore, it has not been determined whether or not the mutational spectra of EACs and ESCCs differ. To address these unresolved issues, we conducted a comprehensive study of esophageal cancer exomes, comprising investigations of its 2 principal histologic subtypes, EACs and ESCCs.

RESULTS

Exomic Sequencing of EACs

DNA was purified from 11 tumors as well as matched non-neoplastic tissues and used to generate 22 libraries suitable for massively parallel sequencing. After capture of the coding sequences with a SureSelect Paired-End Version 2.0 Human Exome Kit (Agilent), the DNA was sequenced using an Illumina GAIIx instrument. The enrichment system included 38 Mb of protein-coding exons from the human genome, amounting to approximately 18,000 genes. The average distinct coverage of each base in the targeted region was 157-fold, and 95.3% of targeted bases were represented by at least 10 reads. Using stringent criteria for the analysis of these data, we identified 748 high-confidence nonsynonymous somatic mutations in 665 genes (Supplementary Table S1). The number of somatic mutations per tumor averaged 67 (range, 35–124; SD, ±28; Table 1). To confirm the specificity of our mutation-calling criteria, we evaluated 255 candidate mutations by Sanger sequencing and confirmed 215 (84%) of the mutations; 32 (13%) of the other candidate mutations could not be amplified by PCR because of unusually high guanine-cytosine content, difficulty in the design of unique primers, or other unknown factors preventing specific amplification and sequencing of the locus; the remaining 8 (3%) of the mutations were not present at levels detectable by Sanger sequencing.

**Exomic Sequencing of ESCCs**

We similarly determined the exomic sequences of 12 ESCCs; the average distinct coverage of each base in the targeted region was 304-fold, with 94.6% of targeted bases being represented by at least 10 reads. Using the same stringent criteria described above, we identified 997 high-confidence nonsynonymous somatic mutations in 884 genes (Supplementary Table S2). The number of somatic mutations per tumor averaged 83 (range, 48–144; SD, ±29). We evaluated 95 candidate mutations in ESCCs by Sanger sequencing and confirmed 83 (87%) of these; the remaining 12 (13%) could not be amplified by PCR for the reasons described above.

**Tumor Cell Purity**

The fraction of neoplastic cells in each specimen was estimated in 3 ways. First, representative cryosections of the blocks were examined by histopathology and only those portions of the blocks containing more than 50% neoplastic cells were used. Second, constitutional single-nucleotide polymorphisms (SNP) within the exomic regions sequenced were used to evaluate LOH in the tumors of each patient. Such losses can only be observed with high confidence if the fraction of neoplastic cells within the sample is high. In all cases, we could observe a substantial degree of LOH. The maximal fractional allelic loss (i.e., the chromosome exhibiting the highest degree of LOH) was consistent with the extent of neoplastic cell purity estimated by histopathology. Finally, we assumed that *TP53* alterations occurred relatively early during the neoplastic process.
Exome Sequencing of Esophageal Cancer

We did not observe a statistically significant difference between the number of somatic mutations in EACs versus ESCCs. The most common substitutions in both EACs (46%) and ESCCs (35%) were C>G>T transitions but with distinct spectra overall (Supplementary Table S3). We did, however, observe a statistically significant difference between the 2 tumor types in indels and transversions: A:T transversions were more common in EACs, whereas C>G>G:C transversions and indels were more frequent in ESCCs (P < 0.0001, Cochran–Mantel–Haenszel test). Although tobacco use is associated with a higher risk of ESCCs than of EACs, we did not observe a difference between these 2 tumor types in the C>G>A:T transversions that are typically associated with smoking. The mutational spectrum of ESCCs is similar to that of head and neck squamous cell carcinoma, in which our group also did not observe a smoking signature (24). These data suggest that the mutational effects of cigarette smoking and associated tobacco-derived carcinogens may be tumor type–dependent (25).

Comparative Analysis of Genes Mutated in EACs and ESCCs

The most commonly mutated gene in both EACs and ESCCs was TP53. Neither the incidence of TP53 mutations (73% vs. 92% for EACs vs. ESCCs, respectively) nor the type of TP53 mutation (missense vs. protein-truncating) differed significantly between the 2 tumor types. Thirty-eight genes were mutated in more than 1 of the 11 EACs studied, and TP53 was mutated in 8 EACs (Supplementary Table S1). Other than TP53, genes (or members of a related pathway) that were mutated in at least 3 of the 12 ESCCs comprised NOTCH1, NOTCH3, FBXW7, KIF16B, KIF21B, and MYCBP2 (Supplementary Table S2). To evaluate the incidence of mutations in these and closely related genes, we analyzed the sequences of TP53, NOTCH1, NOTCH2, NOTCH3, FBXW7, KIF16B, KIF21B, and MYCBP2 in 41 additional ESCCs and their corresponding normal tissues. In total, somatic mutations of TP53, NOTCH1, NOTCH2, NOTCH3, and FBXW7 were identified in 62%, 21%, 6%, 8%, and 6% of ESCCs, respectively (Supplementary Table S4). In general, these mutations were not mutually exclusive, although the degree of overlap was variable. The remaining 3 genes (KIF16B, KIF21B, and MYCBP2) were not mutated in any of the additional 41 tumors analyzed. We attempted to correlate NOTCH and TP53 with tumor stage. However, only 32 patients had stage information available, preventing meaningful interpretation of the results.

Comparative Analysis of North American versus Chinese ESCCs

Given the potential differences in risk factors and carcinogens between North American and Chinese ESCCs, we analyzed the complete coding sequences of TP53, NOTCH1, NOTCH2, NOTCH3, and FBXW7 in 48 Chinese ESCCs. As in our North American ESCC samples, the incidence of TP53 mutations was high (71%; Supplementary Table S5) and the fraction of mutant TP53 alleles was large (20%-90%), suggesting that the neoplastic cell content of the samples used for analysis was sufficient to identify mutations. The mutational spectrum of TP53 was not significantly different between North American and Chinese ESCCs (Supplementary Table S3), although our study only had 10% to 20% power to detect a statistically significant association. In Chinese ESCCs, the frequency of mutations was below 5% in all of the other 4 genes analyzed.

Comparative Analysis of Matched Barrett Esophagus and EAC Tissues

This genome-wide analysis of EACs provided an unprecedented opportunity to test their genome-wide relationship to Barrett esophagus epithelium, the presumed EAC precursor lesion. We were able to obtain matched Barrett esophagus mucosa from 2 of the 11 patients with EAC. DNA from the Barrett esophagus mucosa of patient ESO01T contained 65 of the 78 confirmed mutations present in this patient’s EAC. Similarly, DNA from the Barrett esophagus mucosa of patient ESO10T contained 31 of the 39 confirmed mutations present in this patient’s EAC. In particular, patient ESO10T had a TP53 mutation in both Barrett esophagus and EAC, whereas patient ESO01T did not have a TP53 mutation in either Barrett esophagus or EAC. These data suggest that the majority of the mutations present in the cancers were already present in their benign precursor lesions, providing very strong molecular evidence that EACs developed from Barrett esophagus epithelium in both of these patients. In addition, the data show that the advent of frank malignancy—that is, the ability to invade the underlying basement membrane of the esophageal mucosa—was associated with the accumulation of a relatively small number of additional mutations (Table 2). Although these additional mutations were not recurrent, it is possible that a subset of them was responsible for the invasive capacity of these EACs. In this regard, mutations present in EACs but not in matched Barrett esophagus mucosa are intriguing (Supplementary Table S6). Further functional studies are needed to evaluate the involvement of these particular mutations in tumor progression. An alternative hypothesis is that no additional driver mutations are necessary for progression of Barrett esophagus to EACs: for example, epigenetic events could be sufficient to cause the transition of Barrett esophagus into EACs.

DISCUSSION

Our study provides unequivocal evidence that NOTCH1 plays a tumor-suppressive role during ESCC development; we
observed 12 mutations, 8 of which were inactivating and predicted to result in loss of the majority of amino acids from the translated protein. The remaining 4 missense mutations were located in the N-terminal EGF-like ligand-binding domain. This finding is consistent with prior evidence indicating that in squamous cells (as opposed to other cell types), NOTCH1 signaling is growth-suppressive (26–28). For example, functional studies have shown that NOTCH genes suppress proliferation and promote differentiation of keratinocytes, the cell type that populates the normal keratinizing squamous epithelial lining (27–29). Moreover, loss of epidermal NOTCH1 promotes skin tumorigenesis by impacting the stromal microenvironment (30). Similarly, conditional NOTCH1-knockout (mice) develop cutaneous epithelial tumors, and transgenic mice expressing a pan-NOTCH inhibitor develop cutaneous squamous cell carcinomas (31–32). Nevertheless, a direct connection between NOTCH1 inactivation and human esophageal tumorigenesis had not been established before our study. A tumor-suppressive role for NOTCH1 in squamous cells is also supported by recent sequencing studies of related tumor types, such as squamous cell carcinomas of the head and neck, skin, and lung (24, 33–34). The development of skin cancers in patients treated with head and neck, skin, and lung (24, 33–34). The development of skin cancers in patients treated with or for degradation, including c-Myc and cyclin E (40–47); we observed inactivating FBXW7 gene mutations relatively frequently in our ESCCs, including those which harbored NOTCH mutations. Thus, our data also support a tumor-suppressive role for FBXW7 in ESCCs but one that could function independently of the NOTCH pathway. Barrett esophagus is the obligate precursor lesion of EACs, and progression from Barrett esophagus to EAC involves a stepwise series of molecular events (20, 48). Our data in matched samples provide strong support for a progressive molecular model of advancement from Barrett esophagus to EACs, with fewer mutations occurring in Barrett esophagus than in matched EAC. Interestingly, most mutations in EACs were already present in corresponding benign Barrett esophagus; this finding agrees with previous studies suggesting that Barrett esophagus, although histologically benign, actually constitutes a molecularly advanced stage during the evolution of EACs (49–55). It also raises the possibility of distinct molecular grades within the histologic category of benign Barrett esophagus, emphasizing the need for a comprehensive exome study of EAC-associated Barrett esophagus versus Barrett esophagus from patients with non-EACs. Esophageal cancer exhibits striking geographic variability, suggesting diverse pathogenetic pathways and etiologies, including genetic and environmental factors. This variability was evident in our study: NOTCH1 mutations occurred in North American but not in Chinese ESCCs. It is possible that germline genetic variations specific to the Chinese population substitute for somatic mutations or that epigenetic changes specific to Chinese environments inactivate the NOTCH pathway. Either way, this difference points to distinct tumorigenic mechanisms that can be evaluated further in future studies (56–58). It is often assumed that cancers with identical histopathologies result from the same genetic changes. However, the current study supports the contention that the genetic constitution of tumors from one geographic region cannot necessarily be generalized to those from other parts of the world. If true, this contention has important ramifications for future drug development, personalized therapy, and clinical trials.

METHODS

Samples Evaluated in Each Phase of the Study

For the initial massively parallel sequencing phase, 23 fresh-frozen primary tumors (12 ESCCs and 11 EACs) were evaluated at all coding exon positions by the SureSelect capture approach. For this study, only nonsynonymous mutations were considered. From these data, 255 high-quality mutations (for EACs) and 95 (for ESCCs) were chosen for validation by Sanger sequencing of the mutated genes in the same 23 tumors. These 255 and 95 high-quality genes were chosen for validation of mutation calling as follows: in EACs, the 8 TP53 mutations, all 117 mutations found in ESO01T and ESO10T, plus 130 randomly selected genes from other samples, were subjected to Sanger sequencing; in ESCCs, all 28 mutations in TPS3, NOTCH1, NOTCH3, FBXW7, KIF16B, KIF21B, and MYCBP2, plus 67 randomly chosen genes, were queried with Sanger sequencing to confirm our mutation calling. After this methodologic validation step, a set of 8 genes (TP53, NOTCH1, NOTCH2, NOTCH3, FBXW7, KIF16B, KIF21B, and MYCBP2: the only genes mutated in at least 3 tumors in the ESCC discovery screen) was chosen for “scale-up” Sanger sequencing of all coding exons in a larger, separate cohort comprising 41 fresh-frozen North American ESCCs. Because we were aware of preliminary findings from a parallel exome sequencing study being conducted in a larger cohort of EACs (A. Bass, personal communication), we did not conduct scale-up sequencing in any additional EACs. An additional cohort of 48 fresh-frozen Chinese ESCCs was also examined by Sanger sequencing of all coding exons in TP53, NOTCH1, NOTCH2, NOTCH3, and FBXW7. Finally, in the 2 patients from whom adequate high-quality DNA was available from matched Barrett esophagus epithelium, Sanger sequencing of all 78 genes that were confirmed as mutated in ESO01T and all 39 genes confirmed as mutated in ESO10T was conducted in the matching benign Barrett esophagus tissues. Although it would have been preferable to study multiple anatomic locations of Barrett esophagus within each patient, this was not possible in these cases because biopsy material from only one site was available from each patient.

Patient Characteristics and Preparation of Clinical Samples

Patient characteristics are detailed in Supplementary Table S7. Fresh-frozen resected tumor and matched blood were obtained from patients treated under an Institutional Review Board protocol at the Johns Hopkins Hospital (Baltimore, MD), University of Maryland (Baltimore, MD), and the First Affiliated Hospital of Zhengzhou University (Zhengzhou, China). Tumor tissue was analyzed by frozen fresh-frozen resected tumor and matched blood were obtained from patients treated under an Institutional Review Board protocol at the Johns Hopkins Hospital (Baltimore, MD), University of Maryland (Baltimore, MD), and the First Affiliated Hospital of Zhengzhou University (Zhengzhou, China). Tumor tissue was analyzed by frozen
5 μL of T4 polynucleotide kinase. All reagents used for this step and those described below were from New England Biolabs (NEB) unless otherwise specified. The 100 μL end-repair mixture was incubated at 20°C for 30 minutes, purified by a PCR purification kit (catalog number 28104, Qagen) and eluted with 32 μL of elution buffer (EB). (ii) To A-tail, all 32 μL of end-repaired DNA was mixed with 5 μL of 10× buffer (NEB buffer 2), 10 μL of 1 mmol/L dATP, and 3 μL of Klenow (exo-). The 50 μL mixture was incubated at 37°C for 30 minutes before DNA was purified with a MinElute PCR purification kit (catalog number 28004, Qagen). Purified DNA was eluted with 12.5 μL of 70°C EB and obtained with 10 μL of EB. (iv) For adapter ligation, 10 μL of A-tailed DNA was mixed with 10 μL of PE-adaptor (Illumina), 25 μL of 2× rapid ligation buffer, and 5 μL of Rapid Ligate. The ligation mixture was incubated at room temperature or 20°C for 15 minutes. (v) To purify adaptor-ligated DNA, 50 μL of ligation mixture from step (iv) was mixed with 200 μL of NT buffer from NucleoSpin Extract II kit (catalog number 636972, Clontech) and loaded into NucleoSpin column. The column was centrifuged at 14,000 × g in a desktop centrifuge for 1 minute, washed once with 600 μL of wash buffer (NT3 from Clontech), and centrifuged again for 2 minutes to dry completely. DNA was eluted in 50 μL of elution buffer included in the kit. (vi) To obtain an amplified library, 10 PCRs of 25 μL each were set up, each including 13.25 μL of H₂O, 5 μL of ×5 Phusion HF buffer, 0.5 μL of a dNTP mix containing 10 mmol/L of each dNTP, 0.5 μL of Illumina PE primer #1, 0.5 μL of Illumina PE primer #2, 0.25 μL of Hotstart Phusion polymerase, and 5 μL of the DNA from step (v). The PCR program used was 98°C for 1 minute; 6 cycles of 98°C for 20 seconds, 65°C for 30 seconds, 72°C for 30 seconds; and 72°C for 5 minutes. To purify the PCR product, 250 μL PCR mixture (from the 10 PCR reactions) was mixed with 500 μL NT buffer from a NucleoSpin Extract II kit and purified as described in step (v). Library DNA was eluted at 70°C warm elution buffer, and the DNA concentration was estimated by absorption at 260 nm.

**Exome and Targeted Subgenomic DNA Capture**

Human exome capture was conducted following a protocol from Agilent’s SureSelect Paired-End Version 2.0 Human Exome Kit (Agilent) with the following modifications. (i) A hybridization mixture was prepared containing 25 μL of SureSelect Hyb #1, 1 μL of SureSelect Hyb #2, 10 μL of SureSelect Hyb #3, and 13 μL of SureSelect Hyb #4. (ii) 3.4 μL (0.5 pg) of PE-library DNA described above, 2.5 μL of SureSelect Block #1, 2.5 μL of SureSelect Block #2, and 0.6 μL of Block #3 was loaded into one well in a 384-well Diamond PCR plate (catalog number AB-1111, Thermo-Scientific), sealed with microAmp clear adhesive film (catalog number 4306311; ABI), and placed in a GeneAmp PCR system 9700 thermocycler (Life Sciences Inc.) for 5 minutes at 95°C and then held at 65°C (with the heated lid on). (iii) Twenty-five to 30 μL of hybridization buffer from step (i) was heated for at least 5 minutes at 65°C in another sealed plate with heated lid on. (iv) Five microliters of SureSelect Oligo Capture Library, 1 μL of nucleic acid-free water, and 1 μL of diluted RNase Block (prepared by diluting RNase Block 1:1 with nucleic acid-free water) were mixed and heated at 65°C for 2 minutes in another sealed 384-well plate. (v) While keeping all reactions at 65°C, 13 μL of hybridization buffer from step (iii) was added to the 7 μL of the SureSelect Capture Library Mix from step (iv) and then the entire contents (9 μL) of the library from step (ii). The mixture was slowly pipetted up and down 8 to 10 times. (vi) The 384-well plate was sealed tightly and the hybridization mixture was incubated for 24 hours at 65°C with a heated lid.

After hybridization, 5 steps were conducted to recover and amplify the captured DNA library: (i) Magnetic beads for recovering captured DNA were suspended in 50 μL of Dynal MyOne Streptavidin C1 magnetic beads (catalog number 650.02, Invitrogen DYNAL) was placed in a 1.5 mL microfuge tube and vigorously resuspended on a vortex mixer. Beads were washed 3 times by adding 200 μL of SureSelect binding buffer, mixing on a vortex for 5 seconds, and then removing the supernatant after placing the tubes in a Dynal magnetic separator. After the third wash, beads were resuspended in 200 μL of SureSelect binding buffer. (ii) To bind captured DNA, the entire hybridization mixture described above (29 μL) was transferred directly from the thermocycler to the bead solution and mixed gently; the hybridization mix/bead solution was incubated in an Eppendorf thermomixer at 850 rpm for 30 minutes at room temperature. (iii) To wash the beads, the supernatant was removed from beads after applying a Dynal magnetic separator, and the beads were resuspended in 500 μL SureSelect wash buffer #1 by mixing on vortex mixer for 5 seconds and incubated for 15 minutes at room temperature. Wash buffer #1 was then removed from beads after magnetic separation. The beads were further washed 3 times, each with 500 μL prewarmed SureSelect wash buffer #2 after incubation at 65°C for 10 minutes. After the final wash, SureSelect wash buffer #2 was completely removed. (iv) To elute captured DNA, the beads were suspended in 50 μL SureSelect EB, vortex-mixed and incubated for 10 minutes at room temperature. The supernatant was removed after magnetic separation, collected in a new 1.5-mL microcentrifuge tube, and mixed with 50 μL of SureSelect neutralization buffer. DNA was purified with a Qagen MinElute column and eluted in 17 μL of 0.7-μL EB to obtain 15 μL of captured DNA library. (v) The captured DNA library was amplified in the following way: 15 PCR reactions each containing 9.5 μL of H₂O, 3 μL of 5× Phusion HF buffer, 0.3 μL of 10 mmol/L dNTP, 0.75 μL of dimethyl sulfoxide, 0.15 μL of Illumina PE primer #1, 0.15 μL of Illumina PE primer #2, 0.15 μL of Hotstart Phusion polymerase, and 1 μL of captured exome library were set up. The PCR program used was 98°C for 30 seconds; 14 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds; and 72°C for 5 minutes. To purify PCR products, 225 μL of PCR mixture (from 15 PCR reactions) was mixed with 450 μL of NT buffer from NucleoSpin Extract II kit and purified as described above. The final library DNA was eluted with 30 μL of 0.7-μL elution buffer and DNA concentration was estimated by optical density (OD)₂₆₀ measurement.

**Somatic Mutation Identification by Massively Parallel Sequencing**

Captured DNA libraries were sequenced with the Illumina GAIIx Genome Analyzer, yielding 150 (2 × 75) base pairs from the final library fragments. Sequencing reads were analyzed and aligned to human genome hg18 with the Eland algorithm in CASAVA 1.6 software (Illumina). A mismatched base was identified as a mutation only when (i) it was identified by more than 3 distinct tags; (ii) the number of distinct tags containing a particular mismatched base was at least 15% of the total distinct tags; and (iii) it was not present in >0.5% of the tags in the matched normal sample. SNP search databases included the NCBI database (59).

**Evaluation of Genes in Additional Tumors and Matched Normal Controls**

For the TP53, NOTCH1, NOTCH2, NOTCH3, FBXW7, KIF16B, KIF21B, and MYCBP2 genes that were mutated in at least 3 tumors in the ESCC discovery screen, the coding region was sequenced in 41 additional American ESCCs and matched controls. The coding regions of TP53, NOTCH1, NOTCH2, NOTCH3, and FBXW7 were sequenced in 48 Chinese ESCCs and matched controls. PCR amplification and Sanger sequencing were conducted following protocols described previously, using the primers listed in Supplementary Table S8 (60).

**Evaluation of Matched Barrett Esophagus**

The confirmed mutations in EAC samples ES001T and ES010T were sequenced in matched Barrett esophagus epithelium. PCR amplification and Sanger sequencing were conducted as described (60).

**Statistics**

Differences between EACs and ESCCs in the number of somatic mutations, type of specific mutations (TP53 and at least 1 NOTCH family member mutation), and mutation spectra were compared. The
total number of mutations and specific mutations between groups were compared using Cochran–Mantel–Haenszel tests for general association. The mutation spectra were compared using a continuity adjusted $\chi^2$ test to prevent overestimation of statistical significance. To examine whether there was a global trend for one subtype to have more spectra mutations of any type, a Cochran–Mantel–Haenszel test stratified by spectra was conducted. Differences between U.S. and Chinese ESCCs in type of mutations and predictors of mutations used the same tests as the comparisons between cancer subtypes. A post hoc power calculation was conducted to understand how well our study was powered to examine the relationship between mutations and region based on the prevalence of the mutation in the Chinese population and odds ratio of the mutation between the U.S. and Chinese population for a $P \leq 0.05$.

We had hoped to examine the relationship between smoking and specific mutations among the U.S. patients. Unfortunately, 7 of the 8 patients with ESCCs with reliable information were smokers, which made correlative comparisons difficult. To examine the relationship between NOTCH mutation and tumor stage, we created a logistic regression model of stage III or IV tumors compared with stage I or II tumors. Analyses were conducted in SAS 9.2 (Cary, North Carolina).

Disclosure of Potential Conflicts of Interest

Under agreements between the Johns Hopkins University, Genzyme, Exact Sciences, Inostics, Qiagen, Invitrogen, and Personal Genome Diagnostics, N. Papadopoulos, B. Vogelstein, K.W. Kinzler, and V.E. Velculescu are co-founders of Inostics and Personal Genome Diagnostics, are members of their Scientific Advisory Boards, and own Inostics and Personal Genome Diagnostics stock, which is subject to certain restrictions under Johns Hopkins University policy. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict-of-interest policies. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: N. Agrawal, C. Bettegowda, V.E. Velculescu, B. Vogelstein, N. Papadopoulos, K.W. Kinzler, S.J. Meltzer


Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): N. Agrawal, Y. Jiao, C. Bettegowda, S. David, Y. Cheng, W.S. Twaddell, N.L. Latt, E.J. Shinn, L.-D. Wang, L. Wang, W. Yang, B. Vogelstein, N. Papadopoulos, K.W. Kinzler, S.J. Meltzer


Writing, review, and/or revision of the manuscript: N. Agrawal, Y. Jiao, C. Bettegowda, S.M. Huttlless, W.S. Twaddell, E.J. Shinn, L. Wang, B. Vogelstein, N. Papadopoulos, K.W. Kinzler, S.J. Meltzer

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): N. Agrawal, C. Bettegowda, N.L. Latt, N. Papadopoulos, K.W. Kinzler, S.J. Meltzer

Study supervision: N. Agrawal, V.E. Velculescu, B. Vogelstein, N. Papadopoulos, K.W. Kinzler

Genomic sequencing: N. Agrawal, Y. Jiao, C. Bettegowda, V.E. Velculescu, N. Papadopoulos, K.W. Kinzler

Acknowledgments

The authors thank our patients for their courage and generosity and J. Peck, N. Silliman, L. Dobbyn, and J. Schaeffer for expert technical assistance.

Grant Support

This work was supported by the NIH grants R01CA102057, CA57345, CA121113, CA146799, CA133012, and DK087454, as well as an AACR Stand Up To Cancer Dream Team Translational Cancer Research Grant, the Virginia and D.K. Ludwig Fund for Cancer Research, China 863 High-Tech Key Projects (2012AA02A503, 2012AA02A209, and 2012AA02A201), and Innovation Scientists and Technicians Troop Construction Projects of Henan Province, China (3047).

Received April 26, 2012; revised July 31, 2012; accepted August 2, 2012; published OnlineFirst August 9, 2012.

REFERENCES

Exome Sequencing of Esophageal Cancer

45. Moberg KH, Bell DW, Wahner DC, Haber DA, Harinath IK. Archipel- lago regulates Cyclin E levels in Drosophila and is mutated in human cancer cell lines. Nature 2001;413:311–6.