Published in final edited form as:

Cancer Discov. 2012 October ; 2(10): 899–905. doi:10.1158/2159-8290.CD-12-0189.

Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma

Nishant Agrawal^{1,2}, Yuchen Jiao¹, Chetan Bettgowda^{1,3}, Susan M. Hutfless⁴, Yuxuan Wang¹, Stefan David⁵, Yulan Cheng⁴, William S. Twaddell⁶, Nyan L. Latt⁵, Eun J. Shin⁴, Li-Dong Wang^{7,8}, Liang Wang^{7,9}, Wancai Yang^{7,10}, Victor E. Velculescu¹, Bert Vogelstein¹, Nickolas Papadopoulos¹, Kenneth W. Kinzler¹, and Stephen J. Meltzer⁴

¹Ludwig Center for Cancer Genetics and Therapeutics and Howard Hughes Medical Institutions, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

²Department of Otolaryngology-Head and Neck Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

³Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.

⁴Division of Gastroenterology and Hepatology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

⁵Department of Internal Medicine, Greater Baltimore Medical Center, Baltimore, MD, USA.

⁶Department of Pathology, University of Maryland School of Medicine, Baltimore, MD, USA.

⁷Department of Pathology, Xinxiang Medical University, Henan 453003, China.

⁸Henan Key Laboratory for Esophageal Cancer Research of the First Affiliated Hospital, Zhengzhou University, Zhengzhou, Henan 450052, China.

⁹Department of Pathology, Medical College of Wisconsin, Milwaukee, Wisconsin, USA.

¹⁰Department of Pathology, University of Illinois at Chicago, Chicago, Illinois, USA.

Abstract

Esophageal cancer (EC) ranks sixth in cancer death. To explore its genetic origins, we performed exomic sequencing on 11 adenocarcinomas (EAC) and 12 squamous cell carcinomas (ESCCs) 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

Corresponding Authors: Nishant Agrawal (nagrawal@jhmi.edu), Nickolas Papadopoulos (npapado1@jhmi.edu), Kenneth W. Kinzler (kinzler@jhmi.edu) or Stephen J. Meltzer (smeltzer@jhmi.edu), The Johns Hopkins University School of Medicine, 1503 East Jefferson Street, Baltimore, MD 21287, U.S.A. Phone: 410-502-6071; Fax: 410-502-1329. Nishant Agrawal, Yuchen Jiao and Chetan Bettgowda share first authorship.

Disclosure of Potential Conflicts of Interest: Under agreements between the Johns Hopkins University, Genzyme, Exact Sciences, Inostics, Qiagen, Invitrogen and Personal Genome Diagnostics, NP, BV, KWK and VEV are entitled to a share of the royalties received by the University on sales of products related to genes and technologies described in this manuscript. NP, BV, KWK and VEV 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.

Author Contributions

NA, KWK, SJM, NP and BV designed the study. WY, LDW, LW, SJM, YC, SD, WST, NLL, and EJS collected and analyzed the EC samples. NA, YJ, CB, VEV, NP and KWK performed genomic sequencing. NA, YJ, CB, YW, VEV, NP, KWK and BV analyzed the genetic data; NA, SMH, SJM, KWK and BV wrote draft manuscripts. All authors contributed to the final version of the paper.

($p < 0.0001$). 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's esophagus (BE). These discoveries highlight key genetic differences between EAC and ESCC, American and Chinese ESCC, and suggest that *NOTCH1* is a tumor suppressor gene in the esophagus. Finally, we provide a genetic basis for the evolution of EACs from BE.

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 EC exhibit considerable geographic variation. Overall, esophageal squamous cell carcinoma (ESCC) is the most frequent EC subtype internationally, predominating in eastern Asia and parts of Africa. Tobacco and alcohol consumption are the major risk factors for ESCC, 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 rising incidence has been associated with a corresponding rise in gastroesophageal reflux disease (GERD) and obesity (1, 8). Chronic GERD and its occasional development into Barrett's esophagus (BE) are the major risk factors for EAC, along with tobacco and obesity (9–14). The five-year survival of patients with EC is poor (~15%), and most EC patients 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 EC cases (16–23). *TP53* mutations have also been detected in early stages of EAC and ESCC tumorigenesis, as well as in benign BE mucosa (18, 21). A host of additional genes has been studied for mutation in EC, 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 EC; 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 EAC and ESCC differ. To address these unresolved issues, we performed a comprehensive study of EC exomes, comprising investigations of its two principal histologic subtypes, EAC and ESCC.

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, Santa Clara, California), the DNA was sequenced using an Illumina GAIIx instrument. The enrichment system included 38Mb of protein-coding exons from the human genome, amounting to ~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 ten reads. Using stringent criteria for the analysis of these data, we identified 734 high-confidence non-synonymous somatic mutations in 665 genes (Supplementary Table 1). The number of somatic mutations per tumor averaged 67 (range from 35 to 124 and $SD \pm 28$) (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 GC 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 non-synonymous somatic mutations in 884 genes (Supplementary Table 2). The number of somatic mutations per tumor averaged 83 (range, 48 to 144; $SD \pm 29$). We evaluated 95 candidate mutations in ESCC 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 three ways. First, representative cryosections of the blocks were examined by histopathology, and only those portions of the blocks containing >50% neoplastic cells were used. Second, constitutional single nucleotide polymorphisms (SNPs) within the exomic regions sequenced were used to evaluate losses of heterozygosity 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 (that is, 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 and calculated the mutant allelic fraction (defined as number of non-redundant reads containing the mutation divided by the total number of non-redundant reads) (18, 21). The allelic fractions varied from 20% to 62% in ESCCs and 20% to 67% in EACs - in reasonable concordance with the histopathologic estimates after taking LOH of chromosome 17p into account. All three assessments therefore supported the view that the DNA used for analysis was derived from samples with adequate neoplastic cell purity for effective mutation detection.

Comparative analysis of mutational spectra

We did not observe a statistically significant difference between the number of somatic mutations in EACs vs. ESCCs. The most common substitutions in both EACs (46%) and ESCCs (35%) were C:G>T:A transitions but with distinct spectra overall (Supplementary Table 3). We did, however, observe a statistically significant difference between the two tumor types in indels and transversions: A:T>C:G substitutions were more common in EACs, while 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 ESCC than of EAC, we did not observe a difference between these two tumor types in the C:G>A:T transversions that are typically associated with smoking. The mutational spectrum of ESCC is similar to 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 two tumor types. Thirty-eight genes were mutated in more than one of the eleven EACs studied, and *TP53* was mutated in eight EACs (Supplementary Table 1). Other than *TP53*, genes (or members of a related pathway) that were mutated in at least three of the 12 ESCCs comprised *NOTCH1*, *NOTCH3*, *FBXW7*, *KIF16B*, *KIF21B*, and *MYCBP2* (Supplementary Table 2). 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 4). 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 vs. 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 5) and the fraction of mutant *TP53* alleles was large (20% to 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 3), 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 four genes analyzed.

Comparative analysis of matched BE and EAC tissues

This genome-wide analysis of EACs provided an unprecedented opportunity to test their genome-wide relationship to BE epithelium, the presumed EAC precursor lesion. We were able to obtain matched BE mucosa from two of the 11 EAC patients. DNA from the BE mucosa of patient ESO01T contained 65 of the 78 confirmed mutations present in this patient's EAC. Similarly, DNA from the BE mucosa of patient ESO10T contained 31 of the 39 confirmed mutations present in this patient's EAC. In particular, ESO10T had a *TP53* mutation in both BE and EAC, while ESO01T did not have a *TP53* mutation in either BE 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 BE epithelium in both of these patients. Additionally, the data demonstrate that the advent of frank malignancy - *i.e.*, 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 BE mucosa are intriguing (Supplementary Table 6). 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 BE to EAC: for example, epigenetic events could be sufficient to cause the transition of BE into EAC.

DISCUSSION

Our study provides unequivocal evidence that *NOTCH1* plays a tumor-suppressive role during ESCC development; we observed 12 mutations, eight of which were inactivating and predicted to result in loss of the majority of amino acids from the translated protein. The remaining four missense mutations were located in the N-terminal epidermal growth factor-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-repressive (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 prior to 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 gamma-secretase inhibitors, which prevent NOTCH nuclear translocation, is consistent with the interpretation of these sequencing studies (35).

Components of the NOTCH signaling pathway have been reported to interact with p53 (36–39). However, mutations in *TP53* and *NOTCH* genes were not mutually exclusive in esophageal tumors we evaluated; some tumors had mutations in both genes. NOTCH pathway disruption has also been tied to *FBXW7* gene mutation, although FBXW7 also targets other cancer-related proteins 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 ESCC, but one that could function independently of the NOTCH pathway.

BE is the obligate precursor lesion of EAC, and progression from BE 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 BE to EAC, with fewer mutations occurring in BE than in matched EAC. Interestingly, most mutations in EACs were already present in corresponding benign BE: this finding agrees with previous studies suggesting that BE, although histologically benign, actually constitutes a molecularly advanced stage during the evolution of EAC (49–55). It also raises the possibility of distinct molecular grades within the histologic category of benign BE, emphasizing the need for a comprehensive exome study of EAC-associated BE vs. BE from non-EAC patients.

EC 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 regarding 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 represented by the SureSelect capture approach. For this study, only non-synonymous 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 EAC, 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 ESCC, all 28 mutations in *TP53*, *NOTCH1*, *NOTCH3*, *FBXW7*, *KIF16B*, *KIF21B* and *MYCBP2*, plus 67 randomly chosen genes, were queried with Sanger sequencing to confirm our mutation calling. After this methodological 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 carried out in a larger cohort of EACs (Adam Bass, personal communication), we did not perform 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 two patients from whom adequate high-quality DNA was available from matched BE epithelium, Sanger sequencing of all 78 genes that were confirmed as mutated in ESO01T and all 39 genes confirmed as mutated in ESO10T was performed in the matching benign BE tissues. Although it would have been preferable to study multiple anatomic locations of BE 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 7. Fresh-frozen resected tumor and matched blood were obtained from patients treated under an Institutional Review Board Protocol at the Johns Hopkins Hospital, University of Maryland and the First Affiliated Hospital of Zhengzhou University in China. Tumor tissue was analyzed by frozen section to assess neoplastic cellularity. Tumors were macrodissected to remove residual normal tissue and enhance neoplastic cellularity, as confirmed by multiple frozen sections.

Preparation of Illumina genomic DNA libraries

Genomic DNA libraries were prepared following Illumina's (Illumina, San Diego, CA) suggested protocol with the following modifications. (1) 3 micrograms (μg) of genomic DNA from tumor or normal cells in 100 microliters (μl) of TE was fragmented in a Covaris sonicator (Covaris, Woburn, MA) to a size of 100–500 bp. DNA was purified with a PCR purification kit (Cat # 28104, Qiagen, Valencia, CA) and eluted in 35 μl of elution buffer included in the kit. (2) Purified, fragmented DNA was mixed with 40 μl of H₂O, 10 μl of 10 × T4 ligase buffer with 10 mM ATP, 4 μl of 10 mM dNTP, 5 μl of T4 DNA polymerase, 1 μl of Klenow Polymerase, and 5 μl of T4 polynucleotide Kinase. All reagents used for this step and those described below were from New England Biolabs (NEB, Ipswich, MA) unless otherwise specified. The 100 μl end-repair mixture was incubated at 20°C for 30 min, purified by a PCR purification kit (Cat # 28104, Qiagen) and eluted with 32 μl of elution

buffer (EB). **(3)** To A-tail, all 32 μ l of end-repaired DNA was mixed with 5 μ l of 10 $\times$ Buffer (NEB buffer 2), 10 μ l of 1 mM dATP and 3 μ l of Klenow (exo-). The 50 μ l mixture was incubated at 37°C for 30 min before DNA was purified with a MinElute PCR purification kit (Cat # 28004, Qiagen). Purified DNA was eluted with 12.5 μ l of 70°C EB and obtained with 10 μ l of EB. **(4)** For adaptor ligation, 10 μ l of A-tailed DNA was mixed with 10 μ l of PE-adaptor (Illumina), 25 μ l of 2 $\times$ Rapid ligase buffer and 5 μ l of Rapid Ligase. The ligation mixture was incubated at room temperature (RT) or 20°C for 15 min. **(5)** To purify adaptor-ligated DNA, 50 μ l of ligation mixture from step (4) was mixed with 200 μ l of NT buffer from NucleoSpin Extract II kit (cat# 636972, Clontech, Mountain View, CA) and loaded into NucleoSpin column. The column was centrifuged at 14000 g in a desktop centrifuge for 1 min, washed once with 600 μ l of wash buffer (NT3 from Clontech), and centrifuged again for 2 min to dry completely. DNA was eluted in 50 μ l elution buffer included in the kit. **(6)** To obtain an amplified library, ten PCRs of 25 μ l each were set up, each including 13.25 μ l of H₂O, 5 μ l of 5 $\times$ Phusion HF buffer, 0.5 μ l of a dNTP mix containing 10 mM of each dNTP, 0.5 μ l of Illumina PE primer #1, 0.5 μ l of Illumina PE primer #2, 0.25 μ l of Hotstart Phusion polymerase, and 5 μ l of the DNA from step (5). The PCR program used was: 98°C 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 min. To purify the PCR product, 250 μ l PCR mixture (from the ten PCR reactions) was mixed with 500 μ l NT buffer from a NucleoSpin Extract II kit and purified as described in step (5). Library DNA was eluted with 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 performed following a protocol from Agilent's SureSelect Paired-End Version 2.0 Human Exome Kit (Agilent, Santa Clara, CA) with the following modifications. **(1)** A hybridization mixture was prepared containing 25 μ l of SureSelect Hyb # 1, 1 μ l of SureSelect Hyb # 2, 10 μ l of SureSelect Hyb # 3, and 13 μ l of SureSelect Hyb # 4. **(2)** 3.4 μ l (0.5 μ g) of the PE-library DNA described above, 2.5 μ l of SureSelect Block #1, 2.5 μ l of SureSelect Block #2 and 0.6 μ l of Block #3; was loaded into one well in a 384-well Diamond PCR plate (cat# AB-1111, Thermo-Scientific, Lafayette, CO), sealed with microAmp clear adhesive film (cat# 4306311; ABI, Carlsbad, CA) and placed in GeneAmp PCR system 9700 thermocycler (Life Sciences Inc., Carlsbad CA) for 5 minutes at 95°C, then held at 65°C (with the heated lid on). **(3)** 25–30 μ l of hybridization buffer from step (1) was heated for at least 5 minutes at 65°C in another sealed plate with heated lid on. **(4)** 5 μ l of SureSelect Oligo Capture Library, 1 μ l of nuclease-free water, and 1 μ l of diluted RNase Block (prepared by diluting RNase Block 1: 1 with nuclease-free water) were mixed and heated at 65°C for 2 minutes in another sealed 384-well plate. **(5)** While keeping all reactions at 65°C, 13 μ l of Hybridization Buffer from Step (3) was added to the 7 μ l of the SureSelect Capture Library Mix from Step (4) and then the entire contents (9 μ l) of the library from Step (2). The mixture was slowly pipetted up and down 8 to 10 times. **(6)** 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, five steps were performed to recover and amplify the captured DNA library: **(1)** Magnetic beads for recovering captured DNA: 50 μ l of Dynal MyOne Streptavidin C1 magnetic beads (Cat # 650.02, Invitrogen Dynal, AS Oslo, Norway) was placed in a 1.5 ml microfuge tube and vigorously resuspended on a vortex mixer. Beads were washed three times by adding 200 μ l of SureSelect Binding buffer, mixing on a vortex for five seconds and then removing the supernatant after placing the tubes in a Dynal magnetic separator. After the third wash, beads were resuspended in 200 μ l of SureSelect Binding buffer. **(2)** To bind captured DNA, the entire hybridization mixture described above (29 μ l) 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 850rpm for 30 minutes at room temperature. (3) To wash the beads, the supernatant was removed from beads after applying a Dynal magnetic separator and the beads was 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 three times, each with 500 µl pre-warmed SureSelect Wash Buffer #2 after incubation at 65°C for 10 minutes. After the final wash, SureSelect Wash Buffer #2 was completely removed. (4) To elute captured DNA, the beads were suspended in 50 µl SureSelect Elution Buffer, 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 Qiagen MinElute column and eluted in 17 µl of 70°C EB to obtain 15 µl of captured DNA library. (5) 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 mM dNTP, 0.75 µl of DMSO, 0.15 µl of Illumina PE primer #1, 0.15µl 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 min. To purify PCR products, 225µl 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 70°C elution buffer and DNA concentration was estimated by OD260 measurement.

Somatic Mutation Identification by Massively Parallel Sequencing

Captured DNA libraries were sequenced with the Illumina GAIIX Genome Analyzer, yielding 150 (2 X 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 three 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's database and the 1000 Genomes Project database (59–60).

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 performed following protocols described previously, using the primers listed in Supplementary Table 8 (61).

Evaluation of matched BE

The confirmed mutations in EAC samples ESO01T and ESO10T were sequenced in matched BE epithelium. PCR amplification and Sanger sequencing were performed as described in the previous paragraph (61).

Statistics

Differences between EAC and ESCC 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 Cochrane-Mantel Haenszel tests for general association. The mutation spectra were compared using a continuity adjusted chi-square test to prevent overestimation of statistical significance. To examine if 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 performed. Differences between U.S. and China ESCC in type of mutations and predictors of mutations used the same tests as the comparisons between cancer subtypes. A post-hoc power calculation was performed 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-value 0.05.

We had hoped to examine the relationship between smoking and specific mutations among the U.S. patients. Unfortunately, 7 of the 8 ESCC patients 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 3 or 4 tumors compared with stage 1 or 2 tumors. Analyses were performed in SAS 9.2 (Cary, North Carolina).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank our patients for their courage and generosity. We thank J. Ptak, N. Silliman, L. Dobbyn, and J. Schaeffer for expert technical assistance.

Grant Support

This work was supported by the National Institutes of Health grants RC2DE020957, 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).

REFERENCES

1. Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med*. 2003; 349:2241–2252. [PubMed: 14657432]
2. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer*. 2010; 127:2893–2917. [PubMed: 21351269]
3. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin*. 62:10–29. [PubMed: 22237781]
4. Engel LS, Chow WH, Vaughan TL, Gammon MD, Risch HA, Stanford JL, et al. Population attributable risks of esophageal and gastric cancers. *J Natl Cancer Inst*. 2003; 95:1404–1413. [PubMed: 13130116]
5. Tran GD, Sun XD, Abnet CC, Fan JH, Dawsey SM, Dong ZW, et al. Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int J Cancer*. 2005; 113:456–463. [PubMed: 15455378]
6. Hotchkiss JH. Preformed N-nitroso compounds in foods and beverages. *Cancer Surv*. 1989; 8:295–321. [PubMed: 2696582]
7. Brown LM, Hoover R, Silverman D, Baris D, Hayes R, Swanson GM, et al. Excess incidence of squamous cell esophageal cancer among US Black men: role of social class and other risk factors. *Am J Epidemiol*. 2001; 153:114–122. [PubMed: 11159155]

8. Simard EP, Ward EM, Siegel R, Jemal A. Cancers with increasing incidence trends in the United States: 1999 through 2008. *CA Cancer J Clin.* 2012
9. Siewert JR, Lordick F, Ott K, Stein HJ, Weber WA, Becker K, et al. Induction chemotherapy in Barrett cancer: influence on surgical risk and outcome. *Ann Surg.* 2007; 246:624–628. discussion 8–31. [PubMed: 17893499]
10. Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med.* 1999; 340:825–831. [PubMed: 10080844]
11. Chow WH, Blot WJ, Vaughan TL, Risch HA, Gammon MD, Stanford JL, et al. Body mass index and risk of adenocarcinomas of the esophagus and gastric cardia. *J Natl Cancer Inst.* 1998; 90:150–155. [PubMed: 9450576]
12. Chow WH, Finkle WD, McLaughlin JK, Frankl H, Ziel HK, Fraumeni JF Jr. The relation of gastroesophageal reflux disease and its treatment to adenocarcinomas of the esophagus and gastric cardia. *Jama.* 1995; 274:474–477. [PubMed: 7629956]
13. Naef AP, Savary M, Ozzello L. Columnar-lined lower esophagus: an acquired lesion with malignant predisposition. Report on 140 cases of Barrett's esophagus with 12 adenocarcinomas. *J Thorac Cardiovasc Surg.* 1975; 70:826–835. [PubMed: 1186274]
14. Pickens A, Orringer MB. Geographical distribution and racial disparity in esophageal cancer. *Ann Thorac Surg.* 2003; 76:S1367–S1369. [PubMed: 14530066]
15. Howlander N.; Noone, A.; Krapcho, M.; Neyman, N.; Aminou, R.; Waldron, W., et al. SEER Cancer Statistics Review, 1975– 2008. National Cancer Institute; 2011.
16. Hollstein MC, Metcalf RA, Welsh JA, Montesano R, Harris CC. Frequent mutation of the p53 gene in human esophageal cancer. *Proc Natl Acad Sci U S A.* 1990; 87:9958–9961. [PubMed: 2263646]
17. Hollstein MC, Peri L, Mandard AM, Welsh JA, Montesano R, Metcalf RA, et al. Genetic analysis of human esophageal tumors from two high incidence geographic areas: frequent p53 base substitutions and absence of ras mutations. *Cancer Res.* 1991; 51:4102–4106. [PubMed: 1855226]
18. Casson AG, Mukhopadhyay T, Cleary KR, Ro JY, Levin B, Roth JA. p53 gene mutations in Barrett's epithelium and esophageal cancer. *Cancer Res.* 1991; 51:4495–4499. [PubMed: 1868473]
19. Bennett WP, Hollstein MC, Hsu IC, Sidransky D, Lane DP, Vogelstein B, et al. Mutational spectra and immunohistochemical analyses of p53 in human cancers. *Chest.* 1992; 101:19S–20S. [PubMed: 1541189]
20. Neshat K, Sanchez CA, Galipeau PC, Levine DS, Reid BJ. Barrett's esophagus: the biology of neoplastic progression. *Gastroenterol Clin Biol.* 1994; 18:D71–D76. [PubMed: 8013790]
21. Campomenosi P, Conio M, Bogliolo M, Urbini S, Assereto P, Aprile A, et al. p53 is frequently mutated in Barrett's metaplasia of the intestinal type. *Cancer Epidemiol Biomarkers Prev.* 1996; 5:559–565. [PubMed: 8827361]
22. Gamielidien W, Victor TC, Mugwanya D, Stepien A, Gelderblom WC, Marasas WF, et al. p53 and p16/CDKN2 gene mutations in esophageal tumors from a high-incidence area in South Africa. *Int J Cancer.* 1998; 78:544–549. [PubMed: 9808520]
23. Schneider PM, Casson AG, Levin B, Garewal HS, Hoelscher AH, Becker K, et al. Mutations of p53 in Barrett's esophagus and Barrett's cancer: a prospective study of ninety-eight cases. *J Thorac Cardiovasc Surg.* 1996; 111:323–331. discussion 31–3. [PubMed: 8583805]
24. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, et al. Exome Sequencing of Head and Neck Squamous Cell Carcinoma Reveals Inactivating Mutations in NOTCH1. *Science.* 2011
25. Blackford A, Parmigiani G, Kensler TW, Wolfgang C, Jones S, Zhang X, et al. Genetic mutations associated with cigarette smoking in pancreatic cancer. *Cancer Res.* 2009; 69:3681–3688. [PubMed: 19351817]
26. Katoh M. Notch signaling in gastrointestinal tract (review). *Int J Oncol.* 2007; 30:247–251. [PubMed: 17143535]
27. Dotto GP. Notch tumor suppressor function. *Oncogene.* 2008; 27:5115–5123. [PubMed: 18758480]

28. Rangarajan A, Talora C, Okuyama R, Nicolas M, Mammucari C, Oh H, et al. Notch signaling is a direct determinant of keratinocyte growth arrest and entry into differentiation. *EMBO J.* 2001; 20:3427–3436. [PubMed: 11432830]
29. Ohashi S, Natsuzaka M, Naganuma S, Kagawa S, Kimura S, Itoh H, et al. A NOTCH3-mediated squamous cell differentiation program limits expansion of EMT-competent cells that express the ZEB transcription factors. *Cancer Res.* 2011; 71:6836–6847. [PubMed: 21890822]
30. Demehri S, Turkoz A, Kopan R. Epidermal Notch1 loss promotes skin tumorigenesis by impacting the stromal microenvironment. *Cancer Cell.* 2009; 16:55–66. [PubMed: 19573812]
31. Nicolas M, Wolfer A, Raj K, Kummer JA, Mill P, van Noort M, et al. Notch1 functions as a tumor suppressor in mouse skin. *Nat Genet.* 2003; 33:416–421. [PubMed: 12590261]
32. Proweller A, Tu L, Lepore JJ, Cheng L, Lu MM, Seykora J, et al. Impaired notch signaling promotes de novo squamous cell carcinoma formation. *Cancer Res.* 2006; 66:7438–7444. [PubMed: 16885339]
33. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science.* 2011; 333:1157–1160. [PubMed: 21798893]
34. Wang NJ, Sanborn Z, Arnett KL, Bayston LJ, Liao W, Proby CM, et al. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. *Proc Natl Acad Sci U S A.* 2011; 108:17761–17766. [PubMed: 22006338]
35. Extnance A. Alzheimer's failure raises questions about disease-modifying strategies. *Nat Rev Drug Discov.* 2010; 9:749–751. [PubMed: 20885394]
36. Dotto GP. Crosstalk of Notch with p53 and p63 in cancer growth control. *Nat Rev Cancer.* 2009; 9:587–595. [PubMed: 19609265]
37. Kim SB, Chae GW, Lee J, Park J, Tak H, Chung JH, et al. Activated Notch1 interacts with p53 to inhibit its phosphorylation and transactivation. *Cell Death Differ.* 2007; 14:982–991. [PubMed: 17186020]
38. Lefort K, Mandinova A, Ostano P, Kolev V, Calpini V, Kolfshoten I, et al. Notch1 is a p53 target gene involved in human keratinocyte tumor suppression through negative regulation of ROCK1/2 and MRCKalpha kinases. *Genes Dev.* 2007; 21:562–577. [PubMed: 17344417]
39. Okuyama R, Ogawa E, Nagoshi H, Yabuki M, Kurihara A, Terui T, et al. p53 homologue, p51/p63, maintains the immaturity of keratinocyte stem cells by inhibiting Notch1 activity. *Oncogene.* 2007; 26:4478–4488. [PubMed: 17237812]
40. Tetzlaff MT, Yu W, Li M, Zhang P, Finegold M, Mahon K, et al. Defective cardiovascular development and elevated cyclin E and Notch proteins in mice lacking the Fbw7 F-box protein. *Proc Natl Acad Sci U S A.* 2004; 101:3338–3345. [PubMed: 14766969]
41. Lewis HD, Leveridge M, Strack PR, Haldon CD, O'Neil J, Kim H, et al. Apoptosis in T cell acute lymphoblastic leukemia cells after cell cycle arrest induced by pharmacological inhibition of notch signaling. *Chem Biol.* 2007; 14:209–219. [PubMed: 17317574]
42. Ishikawa Y, Onoyama I, Nakayama KI, Nakayama K. Notch-dependent cell cycle arrest and apoptosis in mouse embryonic fibroblasts lacking Fbxw7. *Oncogene.* 2008; 27:6164–6174. [PubMed: 18641686]
43. Matsumoto A, Onoyama I, Sunabori T, Kageyama R, Okano H, Nakayama KI. Fbxw7-dependent degradation of Notch is required for control of "stemness" and neuronal-glia differentiation in neural stem cells. *J Biol Chem.* 2011; 286:13754–13764. [PubMed: 21349854]
44. Koepf DM, Schaefer LK, Ye X, Keyomarsi K, Chu C, Harper JW, et al. Phosphorylation-dependent ubiquitination of cyclin E by the SCFFbw7 ubiquitin ligase. *Science.* 2001; 294:173–177. [PubMed: 11533444]
45. Moberg KH, Bell DW, Wahrer DC, Haber DA, Hariharan IK. Archipelago regulates Cyclin E levels in *Drosophila* and is mutated in human cancer cell lines. *Nature.* 2001; 413:311–316. [PubMed: 11565033]
46. Strohmaier H, Spruck CH, Kaiser P, Won KA, Sangfelt O, Reed SI. Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line. *Nature.* 2001; 413:316–322. [PubMed: 11565034]

47. Yada M, Hatakeyama S, Kamura T, Nishiyama M, Tsunematsu R, Imaki H, et al. Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. *EMBO J.* 2004; 23:2116–2125. [PubMed: 15103331]
48. Neshat K, Sanchez CA, Galipeau PC, Cowan DS, Ramel S, Levine DS, et al. Barrett's esophagus: a model of human neoplastic progression. *Cold Spring Harb Symp Quant Biol.* 1994; 59:577–583. [PubMed: 7587115]
49. Raskind WH, Norwood T, Levine DS, Haggitt RC, Rabinovitch PS, Reid BJ. Persistent clonal areas and clonal expansion in Barrett's esophagus. *Cancer Res.* 1992; 52:2946–2950. [PubMed: 1581911]
50. Reid BJ, Sanchez CA, Blount PL, Levine DS. Barrett's esophagus: cell cycle abnormalities in advancing stages of neoplastic progression. *Gastroenterology.* 1993; 105:119–129. [PubMed: 8514029]
51. Barrett MT, Galipeau PC, Sanchez CA, Emond MJ, Reid BJ. Determination of the frequency of loss of heterozygosity in esophageal adenocarcinoma by cell sorting, whole genome amplification and microsatellite polymorphisms. *Oncogene.* 1996; 12:1873–1878. [PubMed: 8649847]
52. Eads CA, Lord RV, Kurumboor SK, Wickramasinghe K, Skinner ML, Long TI, et al. Fields of aberrant CpG island hypermethylation in Barrett's esophagus and associated adenocarcinoma. *Cancer Res.* 2000; 60:5021–5026. [PubMed: 11016622]
53. Eads CA, Lord RV, Wickramasinghe K, Long TI, Kurumboor SK, Bernstein L, et al. Epigenetic patterns in the progression of esophageal adenocarcinoma. *Cancer Res.* 2001; 61:3410–3418. [PubMed: 11309301]
54. Selaru FM, Zou T, Xu Y, Shustova V, Yin J, Mori Y, et al. Global gene expression profiling in Barrett's esophagus and esophageal cancer: a comparative analysis using cDNA microarrays. *Oncogene.* 2002; 21:475–478. [PubMed: 11821959]
55. Wang S, Zhan M, Yin J, Abraham JM, Mori Y, Sato F, et al. Transcriptional profiling suggests that Barrett's metaplasia is an early intermediate stage in esophageal adenocarcinogenesis. *Oncogene.* 2006; 25:3346–3356. [PubMed: 16449976]
56. Vogelstein B, Kinzler KW. Carcinogens leave fingerprints. *Nature.* 1992; 355:209–210. [PubMed: 1309937]
57. Hollstein MC, Wild CP, Bleicher F, Chutimataewin S, Harris CC, Srivatanakul P, et al. p53 mutations and aflatoxin B1 exposure in hepatocellular carcinoma patients from Thailand. *Int J Cancer.* 1993; 53:51–55. [PubMed: 8380058]
58. Hollstein M, Marion MJ, Lehman T, Welsh J, Harris CC, Martel-Planche G, et al. p53 mutations at A:T base pairs in angiosarcomas of vinyl chloride-exposed factory workers. *Carcinogenesis.* 1994; 15:1–3. [PubMed: 8293534]
59. <http://www.ncbi.nlm.nih.gov/projects/SNP/>.
60. <http://browser.1000genomes.org/index.html>.
61. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science.* 2006; 314:268–274. [PubMed: 16959974]

SIGNIFICANCE

This is the first genome-wide study of mutations in EC. It identifies key genetic differences between EAC and ESCC including general mutation spectra and *NOTCH1* loss-of-function mutations specific to ESCC, demonstrates geographic disparities between North American and Chinese ESCC, and shows that most mutations in EAC are already present in matched BE.

Table 1

Summary of sequence analysis of EC in the discovery screen

	Adenocarcinoma (%)	SCC (%)
Discovery samples	11	12
High confidence mutations	734	997
Sanger attempted	255	95
Sanger confirmed	215 (84)	83 (87)
Sanger failed	32 (13)	12 (13)
Sanger did not confirm	8 (3)	0
High confidence mutations / tumor	67	83
TP53	8 (73)	11 (92)
NOTCH1	0	4 (33)
NOTCH2	1 (9)	0
NOTCH3	1 (9)	3 (25)
FBXW7	0	2 (17)

Table 2

Comparison of confirmed mutations in BE and EAC

	ESO01T	ESO10T
Mutations in EAC	78	39
Mutations in BE	65	31
Mutations present only in EAC	13	8