Tracking the Genomic Evolution of Esophageal Adenocarcinoma through Neoadjuvant Chemotherapy

Nirupa Murugaesu1, Gareth A. Wilson2,3, Nicolai J. Birkbak2,3, Thomas B.K. Watkins2, Nicholas McGranahan2,4, Sacheen Kumar5, Nima Abbassi-Ghadi5, Max Salm2, Richard Mitter2, Stuart Horswell2, Andrew Rowan2, Benjamin Phillimore2, Jennifer Biggs2, Sharmin Begum2, Nik Matthews2, Daniel Hochhauser1, George B. Hanna5, and Charles Swanton2,3

ABSTRACT

Esophageal adenocarcinomas are associated with a dismal prognosis. Deciphering the evolutionary history of this disease may shed light on therapeutically tractable targets and reveal dynamic mutational processes during the disease course and following neoadjuvant chemotherapy (NAC). We exome sequenced 40 tumor regions from 8 patients with operable esophageal adenocarcinomas, before and after platinum-containing NAC. This revealed the evolutionary genomic landscape of esophageal adenocarcinomas with the presence of heterogeneous driver mutations, parallel evolution, early genome-doubling events, and an association between high intratumor heterogeneity and poor response to NAC. Multiregion sequencing demonstrated a significant reduction in thymine to guanine mutations within a CpTpT context when comparing early and late mutational processes and the presence of a platinum signature with enrichment of cytosine to adenine mutations within a CpC context following NAC. Esophageal adenocarcinomas are characterized by early chromosomal instability leading to amplifications containing targetable oncogenes persisting through chemotherapy, providing a rationale for future therapeutic approaches.

SIGNIFICANCE: This work illustrates dynamic mutational processes occurring during esophageal adenocarcinoma evolution and following selective pressures of platinum exposure, emphasizing the iatrogenic impact of therapy on cancer evolution. Identification of amplifications encoding targetable oncogenes maintained through NAC suggests the presence of stable vulnerabilities, unimpeded by cytotoxics, suitable for therapeutic intervention.

INTRODUCTION

Esophageal cancer is associated with dismal clinical outcome, with only modest improvement in survival over recent decades (1). In addition, the incidence of esophageal adenocarcinoma has increased markedly over the past two decades in Western countries (2).

Although sequencing studies have identified drivers in esophageal adenocarcinoma (3, 4), little is known regarding the clonal evolution of surgically operable disease and the impact of cytotoxic therapy upon the genomic landscape of residual tumor at surgery. Treatment with curative intent of resectable esophageal adenocarcinoma involves surgical resection with neoadjuvant chemotherapy (NAC) or chemoradiotherapy (5).

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

1University College London Cancer Institute, London, United Kingdom. 2The Francis Crick Institute, London, United Kingdom. 3University College London Cancer Institute, Cancer Research UK Lung Cancer Centre of Excellence, London, United Kingdom. 4Centre for Mathematics and Physics in the Life Science and Experimental Biology (CoMPLEX), University College London, London, United Kingdom. 5Department of Surgery and Cancer, Imperial College London, London, United Kingdom.

Note: N. Murugaesu, G.A. Wilson, N.J. Birkbak, T. Watkins, and N. McGranahan contributed equally to this article. S. Kumar and N. Abbassi-Ghadi contributed equally to this article. Corresponding Author: Charles Swanton, The Francis Crick Institute, 44 Lincoln’s Inn Fields, London WC2A 3LY, UK. Phone: 44-207-269-3463; Fax: 44-207-269-3094; E-mail: Charles.Swanton@crick.ac.uk doi: 10.1158/2159-8290.CD-15-0412 ©2015 American Association for Cancer Research.
NAC confers a survival advantage over surgery alone (6), but 50% to 60% of tumors are resistant to treatment (5). Whole-exome sequencing in esophageal adenocarcinoma identified 26 significantly mutated genes, and in another study high-density genomic profiling arrays identified focal somatic copy-number aberrations containing therapeutically targetable kinases, such as ERBB2, FGFR1, FGFR2, EGFR, and MET, in a cohort of esophageal adenocarcinomas. The timing of these events within the context of esophageal adenocarcinoma evolution is unknown (3, 7, 8).

Several sequencing studies have begun to unravel the evolutionary histories of human cancers (9, 10). Emerging evidence reveals that phylogenetic analysis of multiregion exome sequence datasets from a single tumor allows the discrimination of conserved early genetic events present at all sites of the tumor sampled, from later mutations present in only part of the tumor. Distinguishing early from later somatic events during tumor evolution allows the temporal order of mutational processes that occur during the disease course to be resolved (4, 11–14). Previous studies have shown that the evolution from Barrett’s esophagus to adenocarcinoma of the esophagus is dominated by loss of TP53, genome doubling, chromosomal instability (CIN), and a high frequency of chromothripsis events resulting in genomic diversity and an increase in the prevalence of focal amplifications and copy-number gains and losses (4, 7, 15–17). We obtained multiple tumor regions from 8 esophageal adenocarcinomas, acquired consecutively within a prospective tissue collection protocol, both at diagnosis and at surgery following platinum-containing neoadjuvant chemotherapy, to distinguish early from later somatic driver events and to decipher the genomic architecture of esophageal adenocarcinoma during disease evolution and following cytotoxic therapy.

RESULTS
Genomic Architecture and Diversity in Human Esophageal Adenocarcinoma

We performed multiregion exome sequencing (M-seq) on 8 tumors from patients with operable esophageal adenocarcinomas to assess tumor evolution both spatially (sequencing different regions of the primary tumor) and temporally (sequencing further regions of the tumor following combination chemotherapy). In total, 40 regions were sequenced. Tumor and germline regions were sequenced at median 90-fold coverage across the exonic capture regions (Supplementary Table S1) and somatic variants were identified. Orthogonal validation resulted in a validation rate of 96.1% (see Methods). The clinical characteristics of these 8 patients are shown in Supplementary Table S2 and included patients with stage IIB to stage IIC disease.

To assess the extent of spatial and temporal intratumor heterogeneity in esophageal adenocarcinoma, nonsilent somatic mutations were first classified as either ubiquitously detected, if detected in all tumor regions, or heterogeneous, if identified in one or more tumor regions, but not all. All tumors demonstrated spatial and temporal heterogeneity with a median of 55.63% of nonsilent mutations being heterogeneous (range, 12.6%–69.4%; Fig. 1A). The total number of nonsilent mutations identified in these tumors (median = 213) was significantly greater ($P = 0.0109$, Wilcoxon rank sum test) than the number of mutations identified if only a single tumor region had been sampled (median = 127.5), highlighting intratumor heterogeneity and the underestimation of mutational burden and clonal architecture from single samples in esophageal adenocarcinoma (Supplementary Fig. S1). The proportion of heterogeneous mutations in each tumor, both before and after chemotherapy, is shown in Fig. 1B. We assessed this further by generating an intratumor heterogeneity (ITH) index. This was calculated for each tumor as the mean of the proportion of heterogeneous mutations relative to the total number of mutations determined by pairwise comparison of all pre- and post-chemotherapy tumor regions. Intriguingly, we observed a strong correlation between the ITH index and response to NAC (Supplementary Fig. S2, Spearman rho = 0.93, $P = 0.015$).

Evaluation of Truncal and Branched Driver Point Mutations

Filtered nonsilent variants from each tumor region were used to construct phylogenetic trees using the parsimony ratchet method (ref. 18; Fig. 1A and Supplementary Fig. S3), with branch lengths reflecting the number of nonsilent mutations. To account for single-nucleotide variant (SNV) heterogeneity driven by copy-number events, we identified all mutations in genomic segments of copy-number variation across tumor regions and filtered those whose absence, or low variant allele frequency, could be explained by copy-number loss. For example, in EAC015, we identified a genomic segment harboring 16 mutations. This segment was deleted in region R1 but present in all other tumor regions. Thus, the heterogeneity in this case can be explained by a single copy-number loss event (Supplementary Fig. S4).

We then classified somatic SNVs in known cancer genes into category 1 to 3 driver events depending on evidence supporting driver mutation status (Supplementary Methods and Supplementary Table S3), with all other SNVs classified as category 4. We evaluated where on the phylogenetic tree these driver...

**Figure 1.** Intratumor heterogeneity of somatic mutations. A, heatmaps show the regional distribution of all nonsilent mutations; presence (blue), absence, potentially due to copy-number loss (light blue), or absence (gray) of each mutation is indicated for every tumor region. Column next to heatmap shows the distribution of mutations; mutation present in all tumor regions (blue), shared in more than one but not all regions (orange), in one pre-chemotherapy tumor region (red), and in one post-chemotherapy tumor region (green). Mutations are ordered by tumor driver category with categories 1 to 3 indicated in the right column in black, dark gray, and light gray, respectively (details in Supplementary Table S3). Total number of nonsilent mutations (n) is provided for each tumor. Phylogenetic trees generated by a parsimony ratchet approach (18) based on the distribution of all detected mutations are shown to the right of the heatmap; trunk and branch lengths are proportional to the number of nonsilent mutations acquired. Category 1, 2, and 3 driver mutations are indicated next to the trunk or with an arrow pointing to the branches where they were acquired. B, plots show the proportion of trunk and branch mutations in the pre- and post-chemotherapy tumor regions. NA, data not available for these tumors; only one pre-chemotherapy tumor region was available for analysis in EAC015, and tumors EAC014, EAC003, and EAC009 had responded well to treatment with low tumor purity levels in the post-chemotherapy regions, rendering analysis intractable.
Tracking the Genomic Evolution of Esophageal Adenocarcinoma

EAC005
n = 190

EAC015
n = 320

EAC001
n = 219

EAC006
n = 214

EAC014
n = 212

EAC017
n = 249

EAC003
n = 123

EAC009
n = 127

A

B

Mutation present
Mutation absent
Mutation absent (potentially driven by copy-number loss)
Trunk mutations
Shared branch mutations
Private branch mutations (pre-chemotherapy)
Private branch mutations (post-chemotherapy)

Mutation absent (potentially driven by copy-number loss)

Ubiquitous mutations
Heterogeneous mutations

0.0
0.5
1.0

Proportion of mutations prechemotherapy

Proportion of mutations postchemotherapy

Poor response
Intermediate response
Good response

Published OnlineFirst May 23, 2015; DOI: 10.1158/2159-8290.CD-15-0412
mutations were located. Overall, 47% (n = 46) of putative driver events (category 1–3) were found to occur as branch mutations in the 8 esophageal adenocarcinomas. Although there was a trend (P = 0.08) for category 1 driver events on the trunks of the phylogenetic trees, 24% (n = 4) of the category 1 events were located on the branches, occurring in 3 of the 8 tumors (EAC003, EAC014, and EAC017) with PIK3R1 in EAC017 as a potentially targetable mutation. Interestingly, these data support our recent findings that mutations in PIK3R1 have a tendency to be later, subclonal events across cancer types (14). Fifty-two percent of category 2 and category 3 drivers were detected in only a subset of regions. In addition, we identified 10 frameshift indels within annotated driver genes, of which six were located on the trunk, including two in TP53 (EAC006 and EAC015). We also identified 15 of the 26 putative drivers reported in esophageal adenocarcinoma (3), and mutations in 7 of these 15 drivers were found to occur on the branches (SMAD4, TLR4, SLC39A12, TLL1, EYS, NUAK1, and DOCK2; Supplementary Fig. S5). Their presence on the branches of the phylogenetic tree suggests that they may be more prevalent than expected based on single sample analyses in esophageal adenocarcinoma. Supporting this, we identified a significantly greater number of driver mutations using the M-seq approach compared with a single biopsy for each of these tumors (P = 0.041, Wilcoxon rank sum test; Supplementary Fig. S6).

Next, to further resolve the clonal architecture of each tumor, we integrated copy number, variant allele frequencies, and purity estimates to calculate the cancer cell fraction (CCF) of each mutation within each tumor region. We found evidence for at least one driver mutation (categories 1–3) exhibiting an illusion of clonality in all but one tumor (EAC009; Supplementary Fig. S5). Such mutations appear clonally dominant within one tumor region, but are not detectable using our sequencing approach in other tumor regions. For instance, in EAC001, a nonsilent mutation in SLC39A12 (pA44E), a gene that was identified as recurrently mutated in esophageal adenocarcinoma (3), was estimated to be present in 100% of cancer cells within region R4 but was undetectable in all other tumor regions. Similarly, in EAC014, a mutation in SMAD4 (pR361H) was present in all cancer cells within region R3 but was undetectable in the remaining tumor regions. Conversely, mutations in SLC39A12 (pT37T1) and TP53 (splice site) in EAC014 and EAC001, respectively, were both found to be clonal in all tumor regions (Fig. 2A–E).

More generally, consistent with previous reports (3, 4), mutations in TP53 were always identified as fully clonal, and, in every case, mutations in TP53 were accompanied by copy-neutral LOH, highlighting the importance of mutations in TP53 as early driver events. Likewise, mutation and complete loss of the wild-type allele was observed in all tumor regions for CDKN2A and SMAD4 in EAC006 and EAC015, respectively. These mutations likely occurred before truncal genome-doubling events (see Supplementary Methods). However, in EAC014, although mutation and LOH was also observed in SMAD4, this was exclusive to region R3 and, similarly, for EAC017, mutation and LOH was observed in PIK3R1, but only in region R4.

In two of the eight tumors, we observed evidence for parallel evolution. In EAC005, multiple distinct mutations in NOTCH1 were identified in separate tumor regions (Fig. 1A; refs. 19, 20), whereas in EAC001, multiple distinct mutations in GNPATAB were detected in different tumor regions (Figs. 1A and 2A). These data imply the existence of convergent selection pressures at later stages of tumor evolution. The clonal architecture of each tumor region also sheds light on the spatial relationships between tumor regions. Certain regions, such as regions R2 and R4 in EAC001, did not share any subclonal mutations, potentially indicative of geographic barriers existing between these regions, fostering subclonal diversification (Fig. 2A). Conversely, other tumor regions, including regions R1 and R2 in EAC001, shared multiple subclonal mutations, consistent with intermingling subclones between these tumor regions (Fig. 2C).

**Chromosomal Instability and Copy-Number Amplifications Are Early Drivers**

Copy-number analysis on the M-seq data revealed that all tumors showed extensive somatic copy-number aberrations (Supplementary Fig. S7). In addition, all tumors had evidence of ubiquitous TP53 disruption, either through mutation (7/8) or by amplification of MDM2 (EAC003), and all tumor regions showed evidence of genome doubling. This is consistent with previous reports describing early TP53 inactivation followed by CIN and genome doubling as an early defining event in the evolution of esophageal adenocarcinoma (4, 17). We also found evidence of early chromothripsis events in two tumors: on chromosome 1 in EAC003 and on chromosome 19 in EAC017 (Supplementary Fig. S8). Both events were ubiquitously found in all tumor regions. To further measure the degree of CIN in this cohort, we calculated the weighted genome integrity index (wGII) as a proxy for CIN (21). We found that all tumor regions in this study showed high levels of wGII (median, 0.53; range, 0.31–0.66), indicative of CIN (21).

A similar level was observed in esophageal adenocarcinoma tumors from The Cancer Genome Atlas (TCGA) dataset (Supplementary Fig. S9), providing further support for CIN as a defining feature of esophageal adenocarcinoma.

Next, we analyzed the extent of heterogeneity at the copy-number level. Across our cohort of 8 tumors, we identified 1,090 genomic segments of copy-number gain (≥1 copy number relative to ploidy) and 824 segments of copy-number loss (≤1 copy number relative to ploidy). Of these, on average only 24% and 13%, respectively, were ubiquitously identified in all tumor regions sequenced within a tumor, consistent with extensive CIN in these tumors. However, chromosomal amplifications (≥2× ploidy) were significantly more likely to be ubiquitously identified across tumor regions than either gains or losses (Fig. 3A). These data suggest that chromosomal amplifications may often represent early events, which are maintained through NAC (Fig. 3B). In further support of ubiquitously identified amplifications representing shared events, haplotype analysis confirmed that the same allele was likely amplified in the majority of cases (data not shown).

We then assessed whether the somatic copy-number variations in our cohort were similar to those previously identified by TCGA using GISTIC2.0 (22). We found that 15 chromosomal segments previously identified in the TCGA cohort were amplified in at least one sample in our cohort, and 46 chromosomal segments exhibiting copy-number loss overlapped with TCGA recurrent deletions (Fig. 3B and C). With regard to amplifications, these were mostly ubiquitously identified and maintained within all tumor regions before
Tracking the Genomic Evolution of Esophageal Adenocarcinoma

Figure 2. Cancer cell fraction comparisons for different tumor regions. A–C, in each plot, the cancer cell fraction of all SNVs for one tumor region is plotted against another tumor region. The SNVs within each cluster are indicated. D–E, representative probability distributions over the cancer cell fraction for individual mutations are shown for specific tumors.

and after NAC and were often centered on specific oncogenes, such as MYC, CCND1, ERBB2, EGFR, and KRAS (Fig. 3B and D; Supplementary Table S4). Notably, 9 genes that were ubiquitously amplified in at least one tumor and maintained following NAC, including AKT2, AURKA, CCND1, CDK6, ERBB2, EGFR, and PIK3CA (Fig. 3D), have all been identified in the TARGET database (version 2) as potentially targetable copy-number aberrations (23).

Mutation Spectra of Truncal and Branch Mutations

Mutational processes may be dynamic both spatially and temporally within individual tumor types (11, 12). We
Figure 3. Intratumor heterogeneity of chromosomal alterations. A, plot showing the percentage of ubiquitous copy-number amplifications, gains, and losses (±SEM) for all tumors. B and C, heatmaps showing the distribution of potential tumor driver copy-number amplifications and deletions for each tumor region based on recurrently amplified and deleted chromosomal segments identified from TCGA esophageal cancer (ESCA) data. For each region, amplification was determined as ≥2× ploidy, and copy-number loss was determined as ≤1 copy number, relative to ploidy. If no putative driver genes were identified within the recurrent amplification or deletion, the nearest gene appears in square brackets as described by GISTIC2.0 (22). D, heatmap showing amplifications identified within the cohort of tumors containing a targetable oncogene as identified by the TARGET database (23) and whether these occur as ubiquitous (blue) or heterogeneous (orange) amplifications and if they occur as recurrent focal or arm-level amplifications based on TCGA ESCA data (22).
therefore examined the timing of mutational processes during esophageal adenocarcinoma evolution. We observed statistically significant changes in the mutation spectra over time when comparing early (trunk) versus late (branched) somatic mutations. In early mutations, we observed a strong enrichment for thymine to guanine (T>G) mutations within a CpTpT context (5 of 8 cases, \( P < 0.0002 \)). This mutation signature has previously been associated with Barrett’s esophagus and esophageal adenocarcinoma and may result from gastric acid reflux (3, 4, 17). When focusing on late mutations, we observed a relative decrease in the proportion of T>G mutations within a CpTpT context (Fig. 4A and

**Figure 4.** Temporal and spatial dissection of mutation spectra in esophageal adenocarcinoma samples. **A**, stacked bar plot showing the fraction of early mutations (trunk) and late mutations (branch) accounted for by each of the six mutation types in all M-seq samples and per case. The number of mutations analyzed is displayed on top of each bar. The difference between the spectra for trunk and branch mutations across all cases was assessed using a \( \chi^2 \) test. For specific mutation types, a Fisher exact test was used, and significant \( P \) values are shown. **B**, stacked bar plot showing the fraction of pre-chemotherapy mutations and post-chemotherapy mutations accounted for by each of the six mutation types in all M-seq samples and per case. The number of mutations analyzed is displayed on top of each bar. The difference between the spectra for trunk and branch mutations across all cases was assessed using a \( \chi^2 \) test. For specific mutation types, a Fisher exact test was used, and significant \( P \) values are shown. **C**, bar plot showing platinum signature enrichment OR for pre-chemotherapy (red bars) and post-chemotherapy (green bars) mutations for M-seq samples. The platinum signature encompasses cytosine to adenine (C>A) in CpC context (25). Values for 95% confidence intervals for the Fisher exact test are indicated. **D**, a model of tumor progression in esophageal adenocarcinoma. Inferring the evolutionary trajectories of esophageal adenocarcinoma tumors through NAC by M-seq, early and late mutational processes and the selective pressure of platinum treatment are identified.
the branches and the iatrogenic impact of platinum therapy port the fi ndings of the waning of the CpTpT signature in these tumors following NAC. Taken together, these data sup-
not observe a significant change in the mutational load of trend toward increased numbers of small indels in samples tion of enrichment at CpC sites. In addition, there was a A mutations, thereby preventing calcula-
to be unique to post-chemotherapy tumor regions with an acceptance of the phylogenetic trees. With the exception of PDGFRA, the remaining category 1 drivers were tumor-suppressor genes, such as TP53, CDK2NA, NFI, EXT2, DICER1, and CDH1, that are therapeutically difficult to target. However, for each patient within the cohort, there was at least one ubiquitously detected amplification, containing a potentially targetable driver oncogene persisting through NAC. This is supportive of a progression model of esophageal adenocarcinoma (Fig. 4D) where early CIN contributes to amplification of oncogenes early in tumor evolution, which may be selected for (26) and contribute to tumor progression, similar to our recent findings in papillary renal carcinoma (27). However, some of these amplifications may have occurred via breakage fusion bridges and chromothripsis, resulting in double-minute chromosomes that may be relatively resistant to targeted therapeutics (17).

In keeping with other studies, we detected T>G mutations at CpTpT sites (3, 4, 17). Our analysis confirms that this signature is enriched early in tumor evolution, occurring in all tumor regions, before the emergence of the most recent common ancestor. Interestingly, we observed a significant reduction in this mutational process in the late (branched) mutations with an associated relative increase in C>T mutations at CpG sites, consistent with findings in clear-cell renal cell carcinoma (11). This further highlights the benefit of multiregion sampling and provides evidence that distinct mutational processes operate during tumor initiation compared with those affecting mutational burden later in the disease course.

A number of putative driver mutations identified by Dulak and colleagues, SMAD4, TLR4, SLC39A12, TLL1, EYS, NUAK1, and DOCK2 (3), were identified subclonally in our cohort of tumors, suggesting these frequently occur later in tumor evolution. Many of these driver mutations resulted in an illusion of clonality, highlighting the importance of considering regional intratumoral heterogeneity in this tumor type to fully dissect the evolutionary history of these tumors and to more accurately identify fully clonal events which may serve as tractable therapeutic targets. Although M-seq captures more tumor diversity than single biopsies, with almost double the number of nonsi-
lent mutations detected compared with a single biopsy in this cohort, it still samples only a small fraction of the total tumor volume. Hence, this approach is still expected to underestimate the number of subclones present within a given tumor, limiting the resolution of the phylogenetic trees (11, 12).

Although it is difficult to discern the true extent of heterogeneity, using an M-seq approach we observed that tumors that were more heterogeneous as determined by an ITH index appeared to have a worse response to NAC in our cohort. However, with the limited number of cases in our cohort, our observations should be confirmed in an appropriately powered prospective cohort study, given the importance of defining predictive markers to determine response to chemotherapy in esophageal adenocarcinoma. Furthermore, the finding of a platinum scar after NAC highlights the need to understand
Tracking the Genomic Evolution of Esophageal Adenocarcinoma

the selective pressures of platinum chemotherapy on genetic diversity and the iatrogenic impact of treatment upon tumor evolution. Platinum preferentially cross-links G-G nucleotides possibly at random sites throughout the genome, and subsequent DNA repair using translesion DNA polymerases has a tendency to incorporate a wrong base, resulting in a single point mutation in a GpG or CpC context (28). The observation of an increase in C>A mutations in a CpC context in patients with residual disease following NAC, known to occur in model organisms following platinum exposure, further emphasizes the importance of discerning those patients who will not derive benefit from mutagenic cytotoxic chemotherapy.

Using M-seq we have identified considerable heterogeneity of nonsilent mutations and the presence of subclonal driver mutations, indicating that the incidence of driver mutations from single esophageal adenocarcinoma samples may be underestimated, similar to recent observations in other tumor types. We find that mutations in TP53 as well as mutations in other tumor-suppressor genes, such as CDKN2A, are early events, likely occurring before genome doubling, which was also found to be a ubiquitous event in all tumors. All tumors exhibit CIN and harbor considerable heterogeneity at the copy-number level. Despite this emerging chaos, many amplified oncogenes are ubiquitous throughout the tumor, are unperturbed by cytotoxics, and may represent vulnerabilities suitable for therapeutic intervention. Longitudinal studies in the setting of NAC are required to decipher the relationships between ITH and cytotoxic response and recurrence, which may support the identification of resistant disease in advance of therapy to limit the iatrogenic impact of cytotoxics on cancer evolution.

METHODS

See Supplementary Data for a full description of Methods.

Patient Cohort Description

Eight consecutive samples from patients diagnosed with esophageal adenocarcinoma who then went on to have NAC followed by definitive surgical resection (within a prospective tissue collection protocol) were obtained for sequencing. Samples were obtained with informed consent, the study was conducted in accordance with the Declaration of Helsinki, and ethics were approved by an Institutional Review Board for St. Mary’s Hospital, Imperial College Healthcare NHS Trust (REC Ref. no. 04/Q0403/119). Detailed clinical characteristics are provided in Supplementary Table S2. Treatment response to chemotherapy was classified as good, intermediate, or poor based on the prior evidence of the variant being a driver, with category 1 representing high confidence driver mutations (Supplementary Methods).

Identification and Classification of Driver Mutations

All nonsilent variants were compared against a list of potential driver genes (n = 598) and classified into category 1–4 depending on their occurrence before and after treatment with platinum chemotherapy. Variants were classed as post-treatment specific if they were absent from all regions extracted from pre-chemotherapy.

Disclosure of Potential Conflicts of Interest

D. Hochhauser reports receiving commercial research support from Merck-Serono. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions

Conception and design: N. Murugasu, S. Kumar, N. Abbassi-Ghadi, D. Hochhauser, G.B. Hanna, C. Swanton

which is hosted by the European Bioinformatics Institute (Accession No. EGAS00001001254). BWAMEM (bwa-0.7.7, ref. 30) was used to align raw paired-end reads (100 bp) to the full hg19 genomic assembly, and variants were called using VarScan2 somatic (v2.3.6; ref. 31).

Orthogonal validation was performed on all identified 685 nonsilent mutations from 3 tumors (EAC001, EAC003, and EAC005). These were manually reviewed and selected for deep sequencing (median coverage depth = 445) on the Ion Torrent Personal Genome Machine sequencer (Life Technologies), resulting in a validation rate of 96.1%.

Copy-Number Analysis

Processed sample exome SNP and copy-number data from paired tumor–normal samples were generated using VarScan2 (v2.3.6) and further processed using the Sequenza R package 2.1.1 (32) to provide segmented copy-number data and cellularity and ploidy estimates for all samples based on the exome sequence data. Processed copy-number data for each sample were divided by the sample mean ploidy and log transformed. Gain and loss were defined as log2(2.5/2) and log2(1.5/2), respectively. Copy-number heterogeneity was defined using minimum consistent regions (Supplementary Fig. S11). Amplification was defined as log2(4/2). Genome doubling was determined as previously described (33).

Cancer Cell Fraction Estimation and Cluster Analysis

The cancer cell fraction and mutation copy number of each mutation were estimated by integrating Sequenza-derived integer copy-number and tumor purity estimates with the variant allele frequency as outlined in Lohr and colleagues (34) and Landau and colleagues (35). SNVs were filtered to remove those whose absence, or low CCF values, may be driven by copy-number events. In total, across all tumors, 100 mutations were filtered as being driven by copy-number change.

Phylogenetic Tree Construction

All nonsilent mutations that passed validation (EAC001, EAC003, and EAC005) or further filtering (EAC006, EAC009, EAC014, EAC015, and EAC017) were considered for the purpose of determining phylogenetic trees. Trees were built using binary presence/absence matrices built from the regional distribution of variants within the tumor. The R Bioconductor package phangorn (1.99–7; ref. 36) was utilized to perform the parsimony ratchet method (18), generating unrooted genetic trees. Branch lengths were determined using the acctran function.

Temporal Dissection of Mutations

For each M-seq tumor, we classified each mutation as “early” or “late” based on whether it was located on the trunk or branch of the phylogenetic tree. All truncal mutations were classified as “early” and any branch mutations as “late.” The variants were also split according to their occurrence before and after treatment with platinum chemotherapy. Variants were classed as post-treatment specific if they were absent from all regions extracted from pre-chemotherapy.

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): N. Murugaesu, T.B.K. Watkins, S. Kumar, N. Abbassi-Ghadi, B. Phillimore, J. Biggs, S. Begum, N. Matthews, D. Hochhauser, G.B. Hanna, C. Swanton


Writing, review, and/or revision of the manuscript: N. Murugaesu, G.A. Wilson, S. Kumar, N. Abbassi-Ghadi, A. Rowan, B. Phillimore, J. Biggs, S. Begum, N. Matthews, C. Swanton

Study supervision: G.B. Hanna, C. Swanton

Other (dissection of tumor subclonal architecture): N. McGranahan

Acknowledgments

The results published here are in part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and the National Human Genome Research Institute. The data were retrieved through database of Genotypes and Phenotypes (dbGaP) authorization (Accession No. phs000178.v5.p5). Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at http://cancergenome.nih.gov/.

Grant Support

C. Swanton is a senior Cancer Research UK clinical research fellow and is funded by Cancer Research UK, the Rosette Trust, EU FP7 (projects PREDICT and RESPONSIFY, ID: 259303), the Prostate Cancer Foundation, the Breast Cancer Research Foundation, and the European Research Council (THESEUS). N. Murugaesu is a clinical lecturer for and is funded by the National Institute for Health Research. This research was funded by Academy of Medical Sciences and supported by researchers at the National Institute for Health Research University College London Hospitals and Imperial College London Biomedical Research Centre.

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 April 7, 2015; revised May 18, 2015; accepted May 19, 2015; published OnlineFirst May 23, 2015.

REFERENCES

Tracking the Genomic Evolution of Esophageal Adenocarcinoma


Tracking the Genomic Evolution of Esophageal Adenocarcinoma through Neoadjuvant Chemotherapy

Nirupa Murugaesu, Gareth A. Wilson, Nicolai J. Birkbak, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/2159-8290.CD-15-0412

Supplementary Material
Access the most recent supplemental material at:
http://cancerdiscovery.aacrjournals.org/content/suppl/2015/05/23/2159-8290.CD-15-0412.DC1

Cited articles
This article cites 33 articles, 15 of which you can access for free at:
http://cancerdiscovery.aacrjournals.org/content/5/8/821.full#ref-list-1

Citing articles
This article has been cited by 17 HighWire-hosted articles. Access the articles at:
http://cancerdiscovery.aacrjournals.org/content/5/8/821.full#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.