



Published in final edited form as:

*Nat Genet.* 2016 December ; 48(12): 1500–1507. doi:10.1038/ng.3683.

## Spatial intratumor heterogeneity of genetic, epigenetic alterations and temporal clonal evolution in esophageal squamous cell carcinoma

Jia-Jie Hao<sup>1,10</sup>, De-Chen Lin<sup>2,3,10,11</sup>, Huy Q. Dinh<sup>4,10</sup>, Anand Mayakonda<sup>5,10</sup>, Yan-Yi Jiang<sup>5,10</sup>, Chen Chang<sup>1</sup>, Ye Jiang<sup>1</sup>, Chen-Chen Lu<sup>1</sup>, Zhi-Zhou Shi<sup>6</sup>, Xin Xu<sup>1</sup>, Yu Zhang<sup>1</sup>, Yan Cai<sup>1</sup>, Jin-Wu Wang<sup>7</sup>, Qi-Min Zhan<sup>1</sup>, Wen-Qiang Wei<sup>8,11</sup>, Benjamin P. Berman<sup>4,11</sup>, Ming-Rong Wang<sup>1,11</sup>, and H. Phillip Koeffler<sup>2,5,9</sup>

<sup>1</sup>State Key Laboratory of Molecular Oncology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China

<sup>2</sup>Cedars-Sinai Medical Center, Division of Hematology/Oncology, UCLA School of Medicine, Los Angeles, USA

<sup>3</sup>Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Medical Research Center, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China

<sup>4</sup>Center for Bioinformatics and Functional Genomics, Biomedical Sciences, Cedars-Sinai Medical Center, UCLA School of Medicine, Los Angeles, USA

<sup>5</sup>Cancer Science Institute of Singapore, National University of Singapore, Singapore

<sup>6</sup>Faculty of Medicine, Kunming University of Science and Technology, Kunming, Yunnan, China

<sup>7</sup>Department of Pathology, Linzhou Cancer Hospital, Henan, China

<sup>8</sup>Department of Cancer Epidemiology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>11</sup>Correspondence should be addressed to M-R.W. ([wangmr2015@126.com](mailto:wangmr2015@126.com)), B.P.B. ([Benjamin.Berman@csmc.edu](mailto:Benjamin.Berman@csmc.edu)), D-C.L. ([dchlin11@gmail.com](mailto:dchlin11@gmail.com)) or W-Q W. ([weiqw2006@126.com](mailto:weiqw2006@126.com)).

<sup>10</sup>These authors contributed equally to this work.

### URLs

BWA-MEM, <http://arxiv.org/abs/1303.3997v2>.

fpFilter perl script, <https://github.com/ckandoth/variant-filter>.

Bam-readcount, <https://github.com/genome/bam-readcount>.

PHYLIP, <http://evolution.genetics.washington.edu/phylip.html>.

**Accession codes.** Digital sequencing and HM450 Bead array files have been deposited into Sequence Read Archive (SRP072112) and Gene Expression Omnibus (GSE79366), respectively.

### AUTHOR CONTRIBUTIONS

M.-R.W., D.-C.L., B.P.B. and H.P.K. conceived and designed the experiments. J.-J.H., D.-C.L., H.Q.D., W.-Q.W. B.P.B., M.-R.W., and H.P.K. wrote the manuscript. J.-J.H., D.-C.L., Y.J., C.C., C.-C.L., X.X., Y.C. performed the experiments. J.-J.H., H.Q.D., A.M., B.P.B., and Z.-Z.S. performed statistical analysis. J.-J.H., D.-C.L., H.Q.D., Y.-Y.J., B.P.B. and H.P.K. analyzed the data. X.X. contributed reagents. W.-Q.W. contributed materials. J.-W.W. and J.-J.H. read the H&E slides. D.-C.L., Y.Z., Q.-M.Z. and H.P.K. jointly supervised research.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

<sup>9</sup>National University Cancer Institute, National University Hospital Singapore, Singapore

## Abstract

Esophageal squamous cell carcinoma (ESCC) is among the most common malignancies, but little is known about its spatial intratumor heterogeneity (ITH) and temporal clonal evolutionary processes. To address this, we performed multiregion whole-exome sequencing on 51 tumor regions from 13 ESCCs, and multiregion global methylation profiling on three of these 13 cases. We found an average of 35.8% heterogeneous somatic mutations with strong evidence of ITH. Half of driver mutations located on the branches targeted oncogenes, including *PIK3CA*, *NFE2L2*, *MTOR*, etc. By contrast, the majority of truncal and clonal driver mutations occurred in tumor suppressor genes, including *TP53*, *KMT2D*, *ZNF750*, etc. Interestingly, the phyloepigenetic trees robustly recapitulated the topologic structures of the phylogenetic ones, indicating the possible relationship between genetic and epigenetic alterations. Our integrated investigations of the spatial ITH and clonal evolution provide an important molecular foundation for enhanced understanding of the tumorigenesis and progression of ESCC.

---

Esophageal carcinoma is among the most common human cancers, causing over 400,000 deaths worldwide annually<sup>1,2</sup>. The highest-risk areas are located in Eastern Asia, as well as Eastern and Southern Africa; and the most prevalent type is esophageal squamous cell carcinoma (ESCC)<sup>1,2</sup>. The five-year survival rates of ESCC patients undergoing surgery is below 30%, because a large proportion of the tumors are unresectable or have already metastasized before diagnosis<sup>3</sup>.

Recently, several large-scale genomic studies including ours have characterized ESCC genomes with hundreds of somatic mutations and copy number alterations, and have identified significantly mutated genes, including *TP53*, *PIK3CA*, *ZNF750*, etc.<sup>4-9</sup>. The APOBEC signature is a predominant mutational spectrum, and contributes to the mutagenic processes of ESCCs<sup>6,8</sup>. However, the genomic alterations identified in all of these studies were obtained using only single samples representing individual cases, and little is known about the spatial intratumor heterogeneity (ITH) and the temporal clonal evolutionary processes of mutational spectrum in ESCC. Moreover, although alterations in DNA methylation have been observed in ESCC, the ITH of these epigenetic changes is still unknown, and whether it correlates with the genetic architecture remains unexplored.

Precise understanding of both the genomic and epigenomic architecture of primary ESCC tumors is crucial for developing personalized patient treatment and molecular-based biomarkers<sup>10</sup>. Furthermore, an integrated investigation of the genomic and epigenomic evolutionary trajectory of ESCC may also reveal new insights into the relationship between the genome and epigenome. In the present study, we address these critical issues through integrative molecular approaches, including multiregional whole-exome sequencing (M-WES), global methylation profiling, as well as phylogenetic and phyloepigenetic tree construction.

## RESULTS

### Spatial ITH of ESCC

M-WES was performed on the genomic DNA from 13 primary ESCC patients, and the clinical-pathological parameters of these patients were listed in Supplementary Table 1. In total, 51 tumor regions and 13 matched morphologically normal esophageal tissues (four tumor regions and one matched normal tissue per case, with the exception of ESCC04, which only had three tumor regions) were sequenced, with the mean coverage depth of 150×. A total of 1,610 non-silent somatic mutations (affecting 1,427 genes) and 568 silent mutations were identified, with a validation rate of 90% (Supplementary Tables 2–3).

To explore ITH and genomic evolution of ESCC, phylogenetic trees were constructed based on somatic mutations (both silent and non-silent) identified from each tumor region. The trunk, shared branch and private branch of the tree represented mutations in all tumor regions, in some but not all regions, and only in one region, respectively. As shown in Fig. 1a and Supplementary Fig. 1, phylogenetic trees varied extensively among different cases, and all of the 13 ESCCs showed evidence of spatial ITH, with an average of 35.8% (780/2,178; range, 8.0%–60.9%) of somatic variants having spatial heterogeneity.

Characterization of the relative timing of mutations affecting driver genes with possible biological relevance is essential for revealing the evolutionary processes of the cancer genome, as well as further improving precision medicine strategies. To address this, we identified potential driver mutations according to recent large-scale ESCC sequencing data<sup>4–8</sup>, COSMIC gene census<sup>11</sup> and Pan-cancer analysis<sup>12</sup>; this was followed by tracing them within the phylogenetic trees (See Methods). Overall, driver mutations were significantly more enriched in the trunks than were passenger mutations (77.8% vs. 63.8%;  $P = 0.023$ ; Fig. 1b). This indicates that drivers are mutated as relatively early events during the evolutionary process of the tumors, which is in accordance with previous findings in other tumor types<sup>13</sup>. We next separated putative driver mutations into those occurring either in oncogenes or tumor suppressor genes (TSGs). Importantly, half of the driver mutations (50.0%) that mapped to the branches were within oncogenes, including *PIK3CA*, *KIT*, *NFE2L2*, *MTOR* and *FAM135B*. In comparison, only 22.4% of driver mutations located on the trunks affected oncogenes, and the rest were in TSGs. For example, *TP53* mutations were present in twelve of the thirteen cases, and were truncal in all of the mutated cases, in agreement with recent reports<sup>14,15</sup>. It is worthwhile to note that potentially actionable mutations such as those targeting *PIK3CA* and *MTOR* tended to be oncogenic branch events. These findings highlight the extra caution needed when considering inhibiting these mutants in ESCC, given previous studies showing that suppressing subclonal drivers led to growth acceleration of non-mutated subpopulations<sup>16</sup>.

### Clonal status of putative driver mutations

We next investigated the clonal status of somatic mutations within individual regions. Cancer cell fraction (CCF) in each tumor region was calculated as described previously through integrative analysis of local copy number, variant allele frequency (VAF) and tumor cell purity<sup>16,17</sup>. Several driver mutations were subclonal and possibly occurred as late events

in ESCC, including *MTOR*, *KEAP1*, *PTPRB* and *FAM135B*. In contrast, cancer genes on the trunks, such as *TP53*, *NOTCH1*, *CREBBP*, *KMT2D* and *ZNF750*, were predominantly mutated in a fully clonal manner (Fig. 2), further verifying our earlier phylogenetic tree analysis showing that these mutations were possibly early lesions during ESCC development. Of particularly noteworthy distinction, a number of driver variants detected as clonal within some individual tumor regions, were absent in others from the same individual, producing an “illusion” of clonal dominance. For example, a *PIK3CA* hotspot mutation (M1043I) was undetectable in tumor region T2 and T3 in case ESCC13 but was clonally dominant in the other two regions. Likewise, a hotspot mutation in *KIT* gene (E601K) was present in 100% tumor cells in regions T1 and T3 in case ESCC08, yet was absent in the rest of the tumor regions. Such clonal dominance was also observed in *NFE2L2* in case ESCC12. Our results suggest that driver mutations can have mixed and complex intratumoral clonal status in ESCC, and that current single-sampling approach may misinterpret these critical genomic lesions because of the “illusion” of clonal dominance. We further investigated all the non-silent variants within genes and related pathways that have potential targeting approaches. As shown in Supplementary Fig. 2, mutations affecting members in PI3K/MTOR pathway, *KIT*, *AURKA* and *CCND2* were always late events (branched/subclonal). By contrast, variants in *ERBB4*, *FGFR2*, *BRCA2*, *ATM* and *TP53*, were mutated as early events (truncal/clonal), suggesting their potentials as candidate actionable targets for ESCC.

### ITH of copy number alterations (CNA)

We next analyzed the ITH at the copy number level (Supplementary Table 4). First, recurrent copy number alterations which involve important cancer genes in ESCC were identified based on our previous results<sup>6</sup>, and we confirmed that the present cohort harbored these recurrent CNAs with similar frequencies (Supplementary Fig. 3). Although CNAs were generally more similar within cases than between different cases, we found extensive CNA ITH, with 90% (9/10) of all recurrent CNAs being spatially heterogeneous. For example in ESCC08, chr7p11.2 amplification (encompassing *EGFR*) was observed in regions T1 and T4, but not in regions T2 and T3. Similarly, deletions of chr9p21.3 (harboring *CDKN2A/B*) were ubiquitous in some cases but also occurred as heterogeneous aberrations in other samples. The only driver CNA found as consistently ubiquitous was the copy number gain of 11q13, which encompassed a number of oncogenes including *CCND1*, *ANO1*<sup>18–20</sup> and *CTTN*<sup>21,22</sup>, highlighting the importance of this aberration as a founder genomic lesion in the development of ESCC. These results suggest that similar as somatic mutations, CNAs also show significant spatial ITH, concordant with the observations in several other types of cancers<sup>23–25</sup>.

The within-patient mutational rate (mean = 168) was higher than the within-region rate (mean = 139, Supplementary Table 5), highlighting the improved resolution of our multi-biopsy approach for genomic interrogation. Particularly, in the case of branched cancer genes, current M-WES approach markedly increased the sensitivity of the detection rate (Table 1). For example, *ATR* and *TSC1* mutations, which were detected in only 2% of tumor regions (in agreement with previous results), occurred in 7.7% of cases. In addition, the proportion of subclonal mutations detected in each tumor region was much lower than that

in each case (Table 2 and Supplementary Fig. 4). These results again signify that analyzing sequencing data obtained from a single biopsy will likely underestimate the prevalence of the mutations, especially for those acquired late in the mutational process<sup>24</sup>.

### Temporal dissecting of mutational spectra and signatures

To determine the temporal dynamics of the mutagenic processes in ESCC, the mutational spectrum of both the trunks and branches was analyzed using deconstructSigs<sup>26</sup>, which identifies the linear combination of pre-defined signatures that most accurately reconstructs the mutational profile of a single tumor sample. As shown in Fig. 3a, the overall mutational spectra were similar between trunk and branch mutations, with very strong enrichment of Signature 1 substitutions (associated with age), and more subtle but enriched representation of APOBEC-associated Signature 2 and Signature 13 substitutions (C>G and C>T in the TpCpW context). We next calculated the contributions of individual mutational signatures to each tumor (Fig. 3b), and identified several signatures within the tumors tested, including Signature 1 (Age), Signatures 2 and 13 (APOBEC), and Signatures 6 and 15 (DNA mismatch repair), in agreement with previous results in esophageal squamous and other squamous-type cancers<sup>6,8,27</sup>. Interestingly, we noticed that a number of tumors displayed a prominent decrease of the relative contribution of Signature 1 in the branch compared with trunk mutations, albeit without obtaining statistical significance due to the relatively small number of tumors analyzed. In some of these cases, we also observed an increase of signatures associated with DNA damage (including Signatures 3 and 15) in the branch mutations (such as ESCC10 and ESCC12, shown in Figs. 3c–d and Supplementary Fig. 5). To interpret these temporal differences of mutational spectra within the same tumor will require further investigations, but the data indicate that various mutational processes might play important roles in subclonal diversification during the progression of ESCC.

### ITH of DNA methylation in ESCC

As with other cancers, epigenetic abnormalities have been associated with the development and pathogenesis of ESCC<sup>28–30</sup>. To decipher ESCC ITH at the epigenetic level and its potential relationship with subclonal gene mutations, the genome-wide methylation levels of fourteen M-WES-profiled tumor and normal tissues from three ESCC cases (ESCC01, ESCC03 and ESCC05) were profiled using the Illumina HumanMethylation450 (HM450) Bead array. We first identified CpG probes that showed significant differences between the tumor regions and normal tissues from the same patient (except for ESCC01, which did not have a matched normal tissue), then divided these differentially methylated probes into those with “shared” changes (i.e. consistent within all tumor regions from the same case), and “private” changes (those present in one or more of the regions, but not all). We used the probes with private changes to infer tumor evolution and constructed phyloepigenetic trees for each case based on the Euclidean pairwise distances of methylation profiles<sup>31,32</sup> (See Methods). Topological similarities were tested between phyloepigenetic and phylogenetic trees in all three cases based on Robinson-Foulds (RF) distance for unrooted trees<sup>33</sup> (Fig. 4a). Notably, in accordance with a recent report on glioma<sup>32</sup>, the RF distances (zeros for all 3 cases) suggested high concordance between genetic and epigenetic tree topologies in all three cases (see Methods). Since the distinction between the private and shared methylation changes was cutoff dependent, we further tested four different probe-selection cutoffs, and

noted that the phyloepigenetic trees were robust to the cutoff and showed highly similar topological structures (Supplementary Fig. 6). Moreover, to alleviate the confounding effects from non-tumor DNA contamination, two different methods were performed to account for and mitigate the potential influence from immune cells (the major source of non-cancer cells in these samples, Supplementary Fig. 7), and again, similar results were observed between the trees using uncorrected methylation values, and trees using either correction method (Supplementary Fig. 8, see Methods). These findings suggest the possible relationship between genomic and epigenomic alterations during the clonal evolution of ESCC cells, and are indicative of the presence of multiple epigenetically distinct, subclonal cell populations, as recently observed in prostate cancer<sup>31</sup>, glioma<sup>32</sup> and hepatocellular carcinoma [unpublished data, D-C. L., A. M., H. Q. D, Pinbo Huang, Lehang Lin, et al.].

We observed that a number of TSGs, including *EPHA7*<sup>34,35</sup>, *PCDH10*<sup>36,37</sup>, *DOK1*<sup>38,39</sup>, etc., were hypermethylated at their promoters within some but not all regions of the same case, indicating that their expression might be differentially suppressed in different tumor regions. Notably, some TSGs were both mutated and acquired promoter hypermethylation, such as *ASXL1* and *EPHA7*. Interestingly, *ASXL1* was subject to both truncal/clonal mutation and shared hypermethylation at its promoter, suggesting that *ASXL1* was disrupted early during both the genomic and epigenomic evolutionary processes.

To explore the potential biological significance of DNA methylation ITH in ESCC, we next sought to determine whether the differentially methylated DNA CpG loci in each case were enriched at particular functional genomic categories. We first divided CpG probes into those where tumor methylation was higher than the adjacent normal tissues (hyper-methylated) or lower (hypo-methylated). Shared probes were selected for their relatively consistent changes in different tumor regions (Supplementary Fig. 9), while the rest (private probes) exhibited prominent differences between the tumor regions (Fig. 4b) and reflected the extensive ITH seen in the phyloepigenetic trees. We next compared shared vs. private probes by assigning them to various relevant functional genomic categories including CpG Islands (CGIs), CGI Shores, promoters and enhancers, etc., and compared them to the background frequencies of these categories based on all probes on the array (Fig. 4c). As expected, shared CpG sites showed several methylation patterns commonly seen across cancer types<sup>40,41</sup>, including hypermethylated probes strongly enriched within CGI promoter regions, and depleted in both long-range partially methylated domains (PMDs) and enhancer regions (after removing CGIs). Shared hypomethylated probes showed an inverse distribution, i.e., markedly depleted in CGI promoters while enriched in PMDs as well as enhancer regions (Fig. 4c). Strikingly, private CpG sites for the most part resembled the distribution patterns of their shared counterparts (Fig. 4c). In light of the known contribution of tumor-specific methylation to cancer biology<sup>42,43</sup>, our results suggest that intratumoral methylation heterogeneity might play a role in the subclonal diversification of ESCC tumors. In support of this, GO analysis of the genes with privately-hypermethylated promoters showed that they were significantly enriched in cancer-related processes, including cell proliferation, differentiation, migration, adhesion and transcriptional regulation (Fig. 4d). In addition, we noticed that privately-hypermethylated probes were even more enriched in CGI Shores than shared-hypermethylated ones (Fig. 4c). Given the prior observations that i) cancer-specific differentially methylated regions occur more frequently within CGI shores than within

CGIs<sup>44,45</sup>, and ii) CGI Shore methylation correlates with the expression of the associated genes<sup>44</sup>, our observations further suggest the potential involvement of heterogeneity of DNA methylation in the evolutionary biology of ESCC cells.

## DISCUSSION

ESCC is one of the most common malignancies, with relatively low overall five-year survival rates. The main cause leading to unfavorable prognosis of ESCC patients is the lack of effective therapies. Currently, none of the targeted therapies have been established for clinical management of ESCC<sup>46</sup>. Hundreds of genomic alterations, including somatic mutations and copy number alterations, have recently been identified in ESCC<sup>4-9</sup>, but these data have not been translated into clinical applications. In addition, the genomic and epigenomic ITH and clonal evolution of ESCC tumors have not yet been characterized. In light of the evidence that ITH is the major cause of drug resistance and treatment failure<sup>47</sup>, deciphering the genomic diversity and clonal evolution of ESCC tumors will provide both a theoretical and translational basis for identifying new targets and designing personalized medicine strategies.

In the present study, the genomic ITH of 13 ESCC cases, as well as the epigenetic ITH of three of these individuals, were investigated through a variety of molecular approaches, and concordant tumor evolutionary trajectories were found as inferred by both their DNA mutations and methylation. A very recent study of two ESCC cases reported that the ITH rate for somatic mutations was approximate 90%<sup>48</sup>, whereas the rates in our study were much lower, with an average of 35.8%. The discrepancy may well be due to the differences of sequencing depths between the two studies (50× V.S. 150×). Although the true extent of ITH is difficult to define, high sequencing coverage in our study offers an improved resolution to decipher the spatial heterogeneity and clonal evolution of ESCC.

Although phylogeny analysis based on M-WES is not able to resolve completely the true temporal ordering of all the somatic variants, we calculated that an average of 93.5% (range from 87.8% to 97.7%) of somatic mutations were compatible with the present phylogenetic trees (Supplementary Fig. 1). For example, in case ESCC13, 282 out of 294 variants (95.9%) were compatible with the evolution model based on the topological structure of the phylogenetic tree; and only 12 mutations, including *PIK3CA*, were incompatible with the phylogenetic tree (Supplementary Table 6). Therefore, the phylogeny method correctly resolves the temporal order of the vast majority of the somatic mutations. Moreover, the evolutionary models inferred from the M-WES-based phylogeny are strongly supported by our DNA-methylation phylogeny in all three cases (Fig. 4a). Hence, this reconstruction of the phylogenetic topologies, from a completely independent epigenomic event, strongly reinforces the validity of these evolutionary models.

Resolving the clonal status of driver mutations will help to distinguish early from late events, and targeting clonally dominant driver mutations (early events) conceivably represents an optimized therapeutic strategy<sup>10,49</sup>. In this study, despite driver mutations having a tendency to be truncal/clonal compared with passenger mutations, approximate 40% of driver mutations were branched or subclonal. This observation suggests that these

driver mutations were relatively late events during tumor evolution, and contributed to the emergence of distinct subclonal expansions after the founding clones were established. Notable examples included *KIT*, and members of the PI3K/MTOR pathway (*PIK3CA* and *MTOR*) and NFE2L2 pathway (*NFE2L2* and *KEAP1*). These examples, most of which are oncogenes, were frequently mutated as late events in ESCC. Furthermore, evidence of “parallel evolution” was noted in some cases. For example, ESCC13 contained branch *PIK3CA* mutations derived in two spatially separated tumor regions, both harboring the *M1043I* variant, which is a hotspot mutation. Similar parallel evolution was also observed in *NFE2L2* mutations in ESCC12. Interestingly, *PIK3CA*, *KIT* and *NFE2L2* mutations were fully clonal in some tumor regions but were completely absent in others, showing an “illusion” of clonal dominance. In addition, the number of within-patient mutations was higher than the within-region mutations. These results strongly suggest that the prevalence of these driver events, and the rate of sub-clonality overall, are likely underestimated when using a single biopsy to represent an individual patient.

Although ESCC DNA methylation alterations have been profiled using single-sampling approaches, their intratumoral diversity and the relationship to genetic lesions remain unknown. In the present study, we found a number of TSGs with private hypermethylation at the promoters, some of which have been associated with either tumorigenesis or progression of other cancer types, such as *EPHA7*<sup>34,50</sup>, *ABCB4*<sup>51</sup>, *PCDH10*<sup>52,53</sup> and *DOK1*<sup>38,39</sup>. This finding suggests that these genes might be differentially inactivated in different tumor regions from same individuals. We revealed profound epigenetic ITH in ESCC through global methylation analysis. Importantly, subclonal evolutions inferred from DNA methylation closely recapitulated phylogenetic trees, indicating the possible relationship between genetic and epigenetic alterations in ESCC. Therefore, integrative analysis of both phylogenetic and phyloepigenetic trees may generate an enhanced understanding of clonal architecture, and reveal the basis for subclonal epigenetic driver events. These features of epigenetic and genetic ITH revealed by our study may have important implications in ESCC biology.

## ONLINE METHODS

### Patients and specimens

Tissue samples from 13 ESCC patients, including primary esophageal tumors and matched morphologically normal esophageal epithelial margins, were collected in Linzhou Esophagus Cancer Hospital, Henan province, China. All the samples used in this study were residual specimens collected after diagnosis sampling. All the patients received no treatment before surgery, and signed separate informed consent forms for the sampling and molecular analyses. We also considered the clinic-pathological parameters when selecting these ESCC patients, including gender, pathological tumor (pT) stage, regional lymph node metastasis, and tumor differentiation, to avoid bias towards particular pathological characteristics (Supplementary Table 1). Specifically, the male/female ratio in the current cohort was similar to that reported in the latest publication<sup>54</sup>. The number of patients with relatively early (pT1b/T2) and late (pT3) tumor stage were five and eight, respectively. The status of lymph node metastasis (negative, n = 4; positive, n = 9), as well as tumor differentiation (G1,



n = 1; G2, n = 6; G2/3, n = 2; G3, n = 4) were also taken into account. This study has been approved by the Ethics Committee/IRB of Cancer Hospital/Institute, Peking Union Medical College and Chinese Academy of Medical Sciences (Approval No. NCC2013-066). The collection and publication of Chinese human genetic data used in the present study has been approved by the Ministry of Science and Technology. In 12 out of 13 cases, four spatially separated tumor specimens were obtained from each individual, with each section at least 0.5 cm away from the others. In the case of ESCC04, three tumor regions were sampled. We carefully reviewed the hematoxylin and eosin (H&E) slides of each tumor region before subjecting them to WES analysis, in order to make sure that the tumor cell content of the selected regions were comparable and were at least greater than 60% (Representative H&E photos in Supplementary Fig. 10).

### Multiregional whole-exome sequencing (M-WES)

For each individual, genomic DNA (gDNA) of cells from different tumor regions and one matched normal epithelial tissue at the surgical margins were sequenced. Genomic DNA was extracted using Qiagen DNeasy Blood & Tissue Kit according to the manufacturer's instructions. For cases ESCC01 and ESCC02, whole-exome capture of gDNA was performed by the Beijing Genome Institute (BGI), using the BGI Exome Enrichment Kit, and massively parallel sequencing of captured gDNA was performed and analyzed by BGI using the Complete Genomics platform. For the 11 other cases, the Agilent SureSelect Human All Exon v4 (51 Mb) kit was used for whole-exome capture of gDNA, and the captured DNA was sequenced by BGI using the Illumina HiSeq4000 sequencing platform, with 150 base pair paired-end sequencing.

### Alignment of sequencing reads and somatic variant detection

150 base pair paired-end fastq files were aligned to the human reference genome (build hg19) using bwa-mem aligner in default mode (URL). Alignments were then filtered for duplicate reads using Sambler<sup>55</sup>, and bam files were indel realigned and base quality scores were recalibrated according to GATK best practices<sup>56</sup>.

Somatic variants were detected using VarScan<sup>257</sup>. Tumor and matched normal pileup files were generated using the samtools "mpileup" command and fed into the VarScan "somatic" command<sup>58</sup>. Reference genome positions covered at least by 10 reads in normal and 14 reads in tumor samples were considered for variant calling. Variants with VAF less than 0.07 were discarded. Raw somatic variants were filtered using the VarScan "processSomatic" command with arguments --min-tumor-freq 0.07, --max-normal-freq 0.02 and -p-value set to 0.05. These high quality somatic variants were filtered for false positives using fpFilter perl script (URL). These filtered variants were annotated with annovar<sup>59</sup> and filtered against dbSNP135 database for commonly occurring Single Nucleotide Polymorphisms (SNPs)<sup>59</sup>. Disease associated variants annotated in ClinVar database and COSMIC database were retained.

### Phylogenetic tree construction

For mutations that have been detected from at least one tumor region, a method described by Stachler *et al.*<sup>60</sup> was used to increase the sensitivity of detecting these mutations in other

tumor regions from the same individual with low VAF. In brief, Bam-Readcount (URL) was used to obtain read counts for unique somatic variants across all tumor regions. A variant was considered as absent if either its VAF was less than 0.02 or the reads were fewer than 3. The VAFs across all tumor regions of each individual were then used to generate a binary table. Phylogenetic trees were constructed based on the binary tables using Discrete Character Parsimony, implemented in PHYLogeny Inference Package (URL), with the matched morphologically normal epithelial margins as outgroup roots. Based on the calculated branch/trunk lengths inferred from mutation counts, the final trees were drawn manually.

### Cancer cell fraction (CCF) analysis

Copy number analysis from WES data was performed using ReCapSeg, which is implemented as part of the Genome Analysis Tool Kit (GATK v4). Briefly, read counts for each of the exome targets were extracted from all samples and were divided by the total number of reads to generate proportional coverage. A panel of normal controls<sup>14</sup> was created by using proportional coverage from all of the normal samples. Each of the tumor samples was compared to PoN, followed by tangent normalization. These normalized coverage profiles were then segmented using circular binary segmentation<sup>61</sup>. Variants on the sex chromosomes (X and Y) were excluded from this analysis.

Tumor cellularity was determined based on the VAF and segmented copy number data using ABSOLUTE<sup>62</sup>, in order to determine the cancer cell fraction (CCF) of each mutation, as was previously described by McGranahan *et al.*<sup>13</sup>. The clonal status was defined according to the confidence interval (CI) of the CCF. Mutations were classified as subclonal if the upper bound of their 95% CI was less than 1.

### Identification of putative driver mutations

We first identified putative cancer driver genes based on recent large-scale ESCC sequencing data<sup>4-8</sup>, COSMIC cancer gene census (Aug 2015)<sup>11</sup> and Pan-cancer analysis<sup>12</sup>. Next, non-silent variants in these genes were evaluated, and putative driver mutations were identified if they met one of the following requirements: i) Either the exact mutation, the same mutation site or at least three mutations located within 15 bp around the variant were found in COSMIC; ii) If the candidate gene was remarked to be recessive in COSMIC, and the variant was predicted to be deleterious, including stopgain, frameshift and splicing mutation and, had a SIFT score < 0.05<sup>63</sup> or a Polyphen score > 0.995<sup>64,65</sup>.

### Mutational signature analysis

Both silent and non-silent somatic mutations were classified as either truncal or branch as described earlier, and the mutational signatures of these variants were separately generated. We performed a multiple regression approach, deconstructSigs<sup>26</sup>, to extract the signatures based on Wellcome Trust Sanger Institute Mutational Signature Framework<sup>27</sup>, and to quantify statistically the contribution of each signature for each tumor.

## DNA methylation analysis and construction of phyloepigenetic tree

DNA methylation profiles of 12 tumor regions and 2 matched normal esophageal epithelial tissues from three M-WES-examined ESCC cases (ESCC01, ESCC03 and ESCC05) were generated using the Illumina Infinium HumanMethylation450K platform (Illumina, San Diego, CA) by the University of Southern California Norris Comprehensive Cancer Center Genomics Core. We performed basic data processing of the HM450k data using many of the same processing steps that we have performed previously for TCGA data analysis, which is based on the Methylumi R package<sup>66</sup> with several additional quality control steps. Probes with the detection p-value greater than 0.01 in any of the samples were removed, as were probes overlapping with dbSNP SNPs, and probes on the X or Y chromosomes.

For intratumoral analysis, we defined a probe as “private” if the difference in beta values between any single pair of tumor regions was at least 0.3, and defined a probe as “shared” if the differences in beta values between all pairs of tumor regions were less than 0.1. Only private probes were used for construction of phyloepigenetic trees. For each tumor, pairwise Euclidean distances were calculated between all tumor regions using the complete set of private probes.

The phyloepigenetic tree was constructed from these pairwise distances, using the Minimal Evolution method implemented by the *fastme.bal* function in the R package *ape*<sup>67</sup>. Different probe-selection cutoffs for calling private and shared probes produced similar results, with only minimal changes in case ESCC01 (at the cutoff of 0.5) and ESCC05 (at the cutoff of 0.2, Supplementary Fig. 6). The topological comparison of phylogenetic vs. phyloepigenetic and other tree pairs was performed using *RF.dist* function in the CRAN R package *phangorn*. The comparison in the case ESCC01 was done based on only tumor samples due to the lack of matched normal in DNA methylation data (for visualization purpose in Fig. 4a, we used normal samples from the other two cases as the root.)

To mitigate confounding effects of non-cancer cells in phyloepigenetic tree reconstruction, we performed additional bioinformatic analyses, as described below.

The major source of nonmalignant DNA contamination in the esophageal tumor is from immune cells<sup>68</sup>, which has been shown by TCGA to be the case for most solid tumors<sup>62,69</sup>. We confirmed this by review of all of our methylation-profiled samples through immunostaining of the leukocyte common antigen (LCA)/CD45 (representative immunohistochemical photos in Supplementary Fig. 7), which is a common marker of the immune cells and widely used in distinguishing the infiltration of immune cells<sup>70–73</sup>. To precisely determine the extent of immune cell contamination, we estimated the fraction of leukocyte cells in each sample using profiles of immune-specific methylation probes<sup>74</sup>, as described previously<sup>62,69</sup>. Using this method, we noted that case ESCC01 was highly pure (estimated immune-cell fraction 7.1%, ranging from 1% to 14% in various tumor regions) and case ESCC03 and ESCC05 contained an average of 20% and 32% immune cells, respectively (Supplementary Table 7).

We re-calculated each phyloepigenetic tree using one of two methods to model the mixture of cancer and immune cells within the samples:

- i. As demonstrated in several TCGA marker papers, performing an analysis using only the subset of Infinium probes that are unmethylated in purified leukocyte cells, and dichotomizing/binarizing the tumor beta value of these probes with a minimum beta value cutoff, could minimize the influence of contaminating leukocytes<sup>75–77</sup>. We used HM450k profiles from purified leukocyte populations<sup>74</sup>, and selected probes with a maximum beta value less than 0.3 across all leukocyte samples. We then binarized the tumor beta values as 1 if they were  $> 0.3$  and 0 otherwise. We computed pairwise distances between binarized values using the Jaccard index (*dist* function in R package), and performed tree construction using these pairwise distances as described above. The resulting trees are labeled “Dichotomized” in Supplementary Fig. 8.
- ii. In an independent approach, we modeled the tumor beta value as a linear combination of DNA from a mixture of cancer and leukocyte cells. The mixing ratio was estimated for each sample based on methylation of leukocyte-specific probes, as described above and in previously<sup>62,69</sup>. For each probe, we used the fixed mixing ratio, the average beta value of the probe in purified leukocytes<sup>74</sup>, and our measured beta value in the tumor, to estimate the methylation beta value of the cancer cells alone. This method was used to reconstruct phylogenetic trees for each case, and the resulting trees are labeled “Immune cell adjusted” in Supplementary Fig. 8.

Trees for both methods (i) and (ii) were compared to original trees using the RF method described above, and RF values are listed in Supplementary Fig. 8.

### Determining the genomic context of shared vs. private methylation patterns

Shared vs. Private probes were determined based on heterogeneity between different tumor regions, as described above. These were further divided “hypermethylated” and “hypomethylated”, based on comparisons between the tumor methylation and that of adjacent normal tissue. For hypermethylated probes within a specific case, we selected all probes with a methylation beta value  $< 0.3$  in the adjacent normal sample (or a maximum beta value of two other normal samples  $< 0.3$  for ESCC01) and a mean beta value across all tumor regions that was at least 0.3 above from mean of normal sample(s). Similarly for “hypomethylation”, we selected those probes with  $\geq 0.6$  in the normal, and at least below 0.3 for the tumor mean from mean of normals. For ESCC01, which had no matched normal, we averaged the beta values from the other two normals. Hyper- and hypomethylated probe sets are shown in Fig. 4b–d, and Supplementary Fig. 9. For the enrichment analysis in Fig. 4c, promoters were defined as 1.5 kbp up/down stream of RefSeq TSS, CpG islands were taken from the HMM-defined set<sup>78</sup>, shores and enhancers were defined using standard Illumina 450K annotation manifest. Partially Methylated Domains (PMD) were called using the Roadmap<sup>79</sup> normal esophagus sample (ID: E079), using an HMM-based segmentation method<sup>80</sup>. Enrichment/Depletion p-values for the enrichment of private vs. shared probes within each genomic context were computed based on a hypergeometric test based, where

null model frequencies were calculated based on all probes present on the array (shown as “Background” in Fig. 4c).

### Immunohistochemistry (IHC)

Formalin-fixed and paraffin-embedded tissue slides were deparaffinized using xylene, rehydrated using xylene and ethanol, and then immersed in 3% hydrogen peroxide solution for 10 min, heated in citrate at 95°C for 25 min, and cooled at room temperature for 60 min. The slides were incubated overnight at 4°C with the leukocyte common antigen (LCA)/CD45 antibody (Cell Marque, 145M-96, USA; diluted at 1:100), and visualized using PV-9000 Polymer Detection System following the manufacturer’s instructions (Beijing Zhongshan Golden Bridge Biotechnology Co. Ltd., China). Counterstaining was carried out with hematoxylin.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We thank Dr. Hui Shen (Van Andel Institute), Dr. Dan Weisenberger (University of Southern California) as well as Dr. Anand D Jeyasekharan (The Singapore Gastric Cancer Consortium) for their kind help on analysis and discussion. This work was funded by the Singapore Ministry of Health’s National Medical Research Council (NMRC) under its Singapore Translational Research (STaR) Investigator Award to H.P.K., NMRC Individual Research Grant (NMRC/1311/2011) and the NMRC Centre Grant awarded to National University Cancer Institute of Singapore, the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centres of Excellence initiatives to H.P.K. D.-C.L. was supported by American Society of Hematology Fellow Scholar Award, National Natural Science Foundation of China (81672786) and National Center for Advancing Translational Sciences UCLA CTSI Grant UL1TR000124. M.-R.W was supported by National Natural Science Foundation of China (81330052, 81520108023 and 81321091). Y.Z. was supported by Beijing Natural Science Foundation (7151008). This study was partially supported by a generous donation from the Melamed family, and NIH/NCI grant 1U01CA184826 as well as institutional support from the Samuel Oschin Comprehensive Cancer Institute to B.P.B and H.Q.D.

### REFERENCES

1. Torre LA, et al. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65:87–108. [PubMed: 25651787]
2. Ferlay J, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer.* 2015; 136:E359–E386. [PubMed: 25220842]
3. Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med.* 2003; 349:2241–2252. [PubMed: 14657432]
4. Agrawal N, et al. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discov.* 2012; 2:899–905. [PubMed: 22877736]
5. Song Y, et al. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature.* 2014; 509:91–95. [PubMed: 24670651]
6. Lin DC, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nat Genet.* 2014; 46:467–473. [PubMed: 24686850]
7. Gao YB, et al. Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet.* 2014; 46:1097–1102. [PubMed: 25151357]
8. Zhang L, et al. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am J Hum Genet.* 2015; 96:597–611. [PubMed: 25839328]
9. Cheng C, et al. Whole-Genome Sequencing Reveals Diverse Models of Structural Variations in Esophageal Squamous Cell Carcinoma. *Am J Hum Genet.* 2016; 98:256–274. [PubMed: 26833333]

10. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell*. 2015; 27:15–26. [PubMed: 25584892]
11. Futreal PA, et al. A census of human cancer genes. *Nat Rev Cancer*. 2004; 4:177–183. [PubMed: 14993899]
12. Lawrence MS, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014; 505:495–501. [PubMed: 24390350]
13. McGranahan N, et al. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med*. 2015; 7:283ra54.
14. Durinck S, et al. Temporal dissection of tumorigenesis in primary cancers. *Cancer Discov*. 2011; 1:137–143. [PubMed: 21984974]
15. Shah SP, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012; 486:395–399. [PubMed: 22495314]
16. Lohr JG, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*. 2014; 25:91–101. [PubMed: 24434212]
17. Landau DA, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013; 152:714–726. [PubMed: 23415222]
18. Shi ZZ, et al. Consistent and differential genetic aberrations between esophageal dysplasia and squamous cell carcinoma detected by array comparative genomic hybridization. *Clin Cancer Res*. 2013; 19:5867–5878. [PubMed: 24009147]
19. Shang L, et al. ANO1 protein as a potential biomarker for esophageal cancer prognosis and precancerous lesion development prediction. *Oncotarget*. 2016; 7:24374–24382. [PubMed: 27016410]
20. Britschgi A, et al. Calcium-activated chloride channel ANO1 promotes breast cancer progression by activating EGFR and CAMK signaling. *Proc Natl Acad Sci U S A*. 2013; 110:E1026–E1034. [PubMed: 23431153]
21. Luo ML, et al. Amplification and overexpression of CTTN (EMS1) contribute to the metastasis of esophageal squamous cell carcinoma by promoting cell migration and anoikis resistance. *Cancer Res*. 2006; 66:11690–11699. [PubMed: 17178864]
22. Lu P, et al. Genome-wide gene expression profile analyses identify CTTN as a potential prognostic marker in esophageal cancer. *PLoS One*. 2014; 9:e88918. [PubMed: 24551190]
23. de Bruin EC, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science*. 2014; 346:251–256. [PubMed: 25301630]
24. Gerlinger M, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet*. 2014; 46:225–233. [PubMed: 24487277]
25. Murugaesu N, et al. Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov*. 2015; 5:821–831. [PubMed: 26003801]
26. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol*. 2016; 17:31. [PubMed: 26899170]
27. Alexandrov LB, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500:415–421. [PubMed: 23945592]
28. Toh Y, Egashira A, Yamamoto M. Epigenetic alterations and their clinical implications in esophageal squamous cell carcinoma. *Gen Thorac Cardiovasc Surg*. 2013; 61:262–269. [PubMed: 23529258]
29. Agarwal R, et al. Epigenomic program of Barrett's-associated neoplastic progression reveals possible involvement of insulin signaling pathways. *Endocr Relat Cancer*. 2012; 19:L5–L9. [PubMed: 22194443]
30. Alvarez H, et al. Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis. *PLoS Genet*. 2011; 7:e1001356. [PubMed: 21483804]
31. Brocks D, et al. Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell Rep*. 2014; 8:798–806. [PubMed: 25066126]

32. Mazor T, et al. DNA Methylation and Somatic Mutations Converge on the Cell Cycle and Define Similar Evolutionary Histories in Brain Tumors. *Cancer Cell*. 2015; 28:307–317. [PubMed: 26373278]
33. Robinson DF, Foulds LR. Comparison of phylogenetic trees. *Math Biosci*. 1981; 53:131–147.
34. Oricchio E, et al. The Eph-receptor A7 is a soluble tumor suppressor for follicular lymphoma. *Cell*. 2011; 147:554–564. [PubMed: 22036564]
35. Lopez-Nieva P, et al. EPHA7, a new target gene for 6q deletion in T-cell lymphoblastic lymphomas. *Carcinogenesis*. 2012; 33:452–458. [PubMed: 22114070]
36. Yu J, et al. Methylation of protocadherin 10, a novel tumor suppressor, is associated with poor prognosis in patients with gastric cancer. *Gastroenterology*. 2009; 136:640.e1–651.e1. [PubMed: 19084528]
37. Zhao Y, et al. A novel wnt regulatory axis in endometrioid endometrial cancer. *Cancer Res*. 2014; 74:5103–5117. [PubMed: 25085246]
38. Saulnier A, et al. Inactivation of the putative suppressor gene DOK1 by promoter hypermethylation in primary human cancers. *Int J Cancer*. 2012; 130:2484–2494. [PubMed: 21796618]
39. Mercier PL, et al. Characterization of DOK1, a candidate tumor suppressor gene, in epithelial ovarian cancer. *Mol Oncol*. 2011; 5:438–453. [PubMed: 21856257]
40. Bergman Y, Cedar H. DNA methylation dynamics in health and disease. *Nat Struct Mol Biol*. 2013; 20:274–281. [PubMed: 23463312]
41. Quante T, Bird A. Do short, frequent DNA sequence motifs mould the epigenome? *Nat Rev Mol Cell Biol*. 2016; 17:257–262. [PubMed: 26837845]
42. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer*. 2011; 11:726–734. [PubMed: 21941284]
43. Lay FD, et al. The role of DNA methylation in directing the functional organization of the cancer epigenome. *Genome Res*. 2015; 25:467–477. [PubMed: 25747664]
44. Irizarry RA, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet*. 2009; 41:178–186. [PubMed: 19151715]
45. Doi A, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet*. 2009; 41:1350–1353. [PubMed: 19881528]
46. Gharwan H, Groninger H. Kinase inhibitors and monoclonal antibodies in oncology: clinical implications. *Nat Rev Clin Oncol*. 2016; 13:209–227. [PubMed: 26718105]
47. Gerlinger M, et al. Cancer: evolution within a lifetime. *Annu Rev Genet*. 2014; 48:215–236. [PubMed: 25292359]
48. Cao W, et al. Multiple region whole-exome sequencing reveals dramatically evolving intratumor genomic heterogeneity in esophageal squamous cell carcinoma. *Oncogenesis*. 2015; 4:e175. [PubMed: 26619400]
49. Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C. Intratumor heterogeneity: seeing the wood for the trees. *Sci Transl Med*. 2012; 4:127ps10.
50. Wang J, et al. Downregulation of EphA7 by hypermethylation in colorectal cancer. *Oncogene*. 2005; 24:5637–5647. [PubMed: 16007213]
51. Kiehl S, et al. ABCB4 is frequently epigenetically silenced in human cancers and inhibits tumor growth. *Sci Rep*. 2014; 4:6899. [PubMed: 25367630]
52. Jao TM, et al. Protocadherin 10 suppresses tumorigenesis and metastasis in colorectal cancer and its genetic loss predicts adverse prognosis. *Int J Cancer*. 2014; 135:2593–2603. [PubMed: 24740680]
53. Narayan G, et al. PCDH10 promoter hypermethylation is frequent in most histologic subtypes of mature lymphoid malignancies and occurs early in lymphomagenesis. *Genes Chromosomes Cancer*. 2013; 52:1030–1041. [PubMed: 23929756]

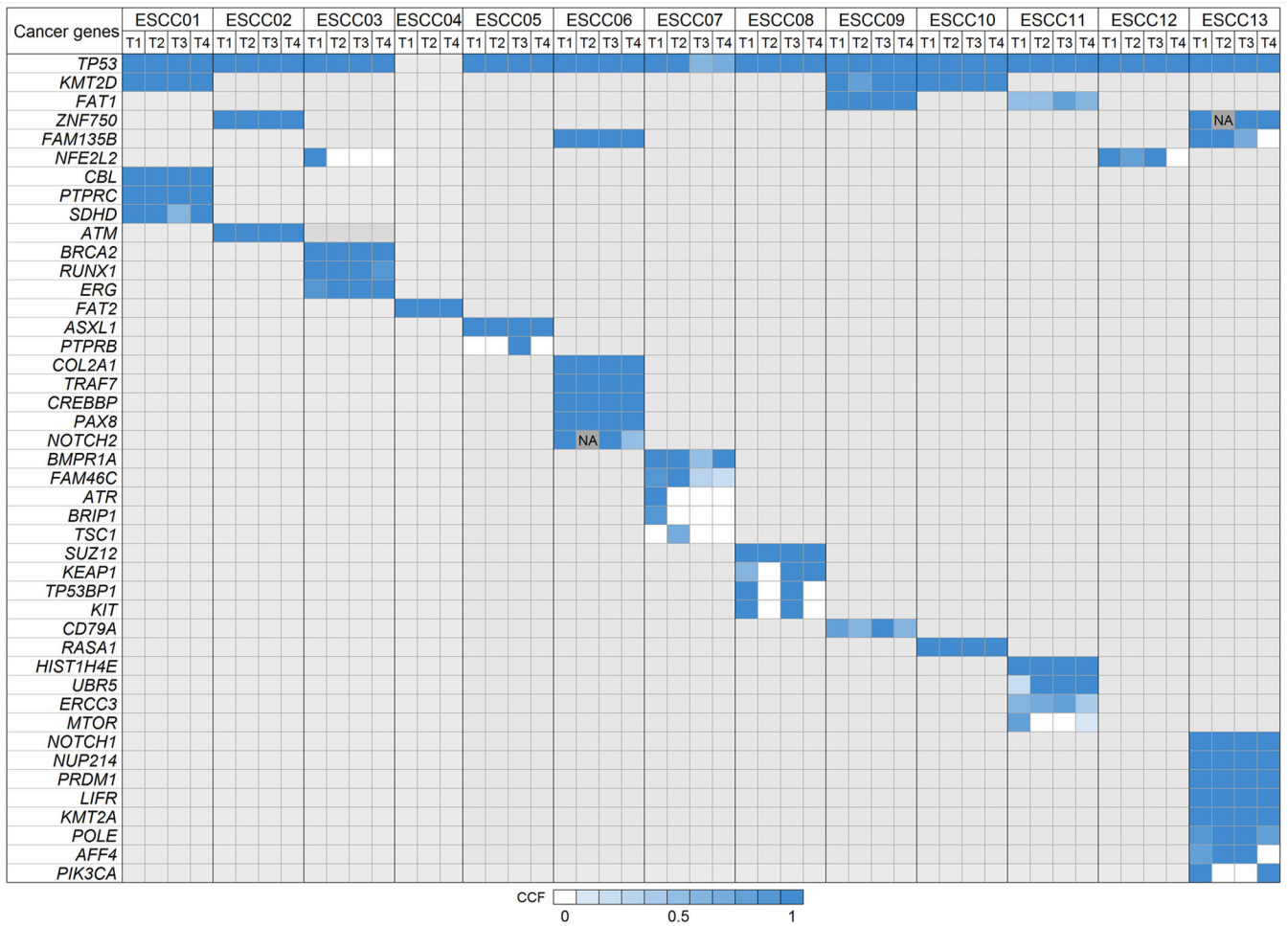
**METHODS-ONLY REFERENCES**

54. Chen W, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016; 66:115–132. [PubMed: 26808342]
55. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics.* 2014; 30:2503–2505. [PubMed: 24812344]
56. Van der Auwera GA, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013; 43:11.10.1–11.10.33. [PubMed: 25431634]
57. Koboldt DC, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 2012; 22:568–576. [PubMed: 22300766]
58. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25:2078–2079. [PubMed: 19505943]
59. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38:e164. [PubMed: 20601685]
60. Stachler MD, et al. Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat Genet.* 2015; 47:1047–1055. [PubMed: 26192918]
61. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics.* 2004; 5:557–572. [PubMed: 15475419]
62. Carter SL, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol.* 2012; 30:413–421. [PubMed: 22544022]
63. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009; 4:1073–1081. [PubMed: 19561590]
64. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7:248–249. [PubMed: 20354512]
65. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013; Chapter 7(Unit 7):20.
66. Triche TJ Jr, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* 2013; 41:e90. [PubMed: 23476028]
67. Desper R, Gascuel O. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comput Biol.* 2002; 9:687–705. [PubMed: 12487758]
68. Takahashi T, et al. Estimation of the fraction of cancer cells in a tumor DNA sample using DNA methylation. *PLoS One.* 2013; 8:e82302. [PubMed: 24312652]
69. Zack TI, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet.* 2013; 45:1134–1140. [PubMed: 24071852]
70. Pages F, et al. In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *J Clin Oncol.* 2009; 27:5944–5951. [PubMed: 19858404]
71. de Miranda NF, et al. Infiltration of Lynch colorectal cancers by activated immune cells associates with early staging of the primary tumor and absence of lymph node metastases. *Clin Cancer Res.* 2012; 18:1237–1245. [PubMed: 22261803]
72. Punt S, et al. Whole-transcriptome analysis of flow-sorted cervical cancer samples reveals that B cell expressed TCL1A is correlated with improved survival. *Oncotarget.* 2015; 6:38681–38694. [PubMed: 26299617]
73. Gorter A, Prins F, van Diepen M, Punt S, van der Burg SH. The tumor area occupied by Tbet+ cells in deeply invading cervical cancer predicts clinical outcome. *J Transl Med.* 2015; 13:295. [PubMed: 26357849]
74. Reinius LE, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS One.* 2012; 7:e41361. [PubMed: 22848472]
75. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature.* 2014; 513:202–209. [PubMed: 25079317]



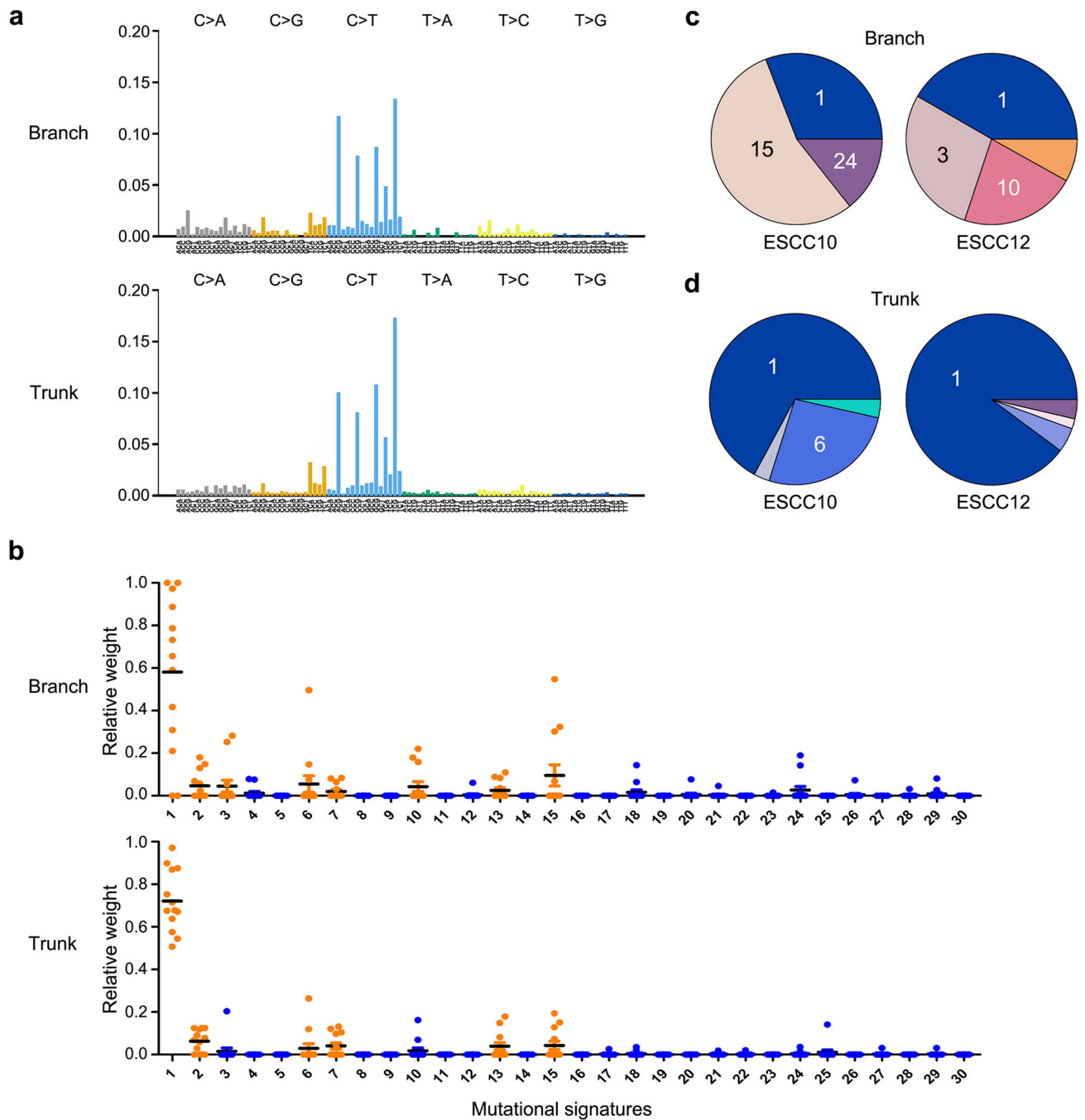
76. Ceccarelli M, et al. Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell*. 2016; 164:550–563. [PubMed: 26824661]
77. Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell*. 2015; 163:1011–1025. [PubMed: 26544944]
78. Wu H, Caffo B, Jaffee HA, Irizarry RA, Feinberg AP. Redefining CpG islands using hidden Markov models. *Biostatistics*. 2010; 11:499–514. [PubMed: 20212320]
79. Roadmap Epigenomics C. *et al*. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518:317–330. [PubMed: 25693563]
80. Song Q, et al. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One*. 2013; 8:e81148. [PubMed: 24324667]





**Fig. 2. Clonal status of putative driver mutations in ESCC tumors**

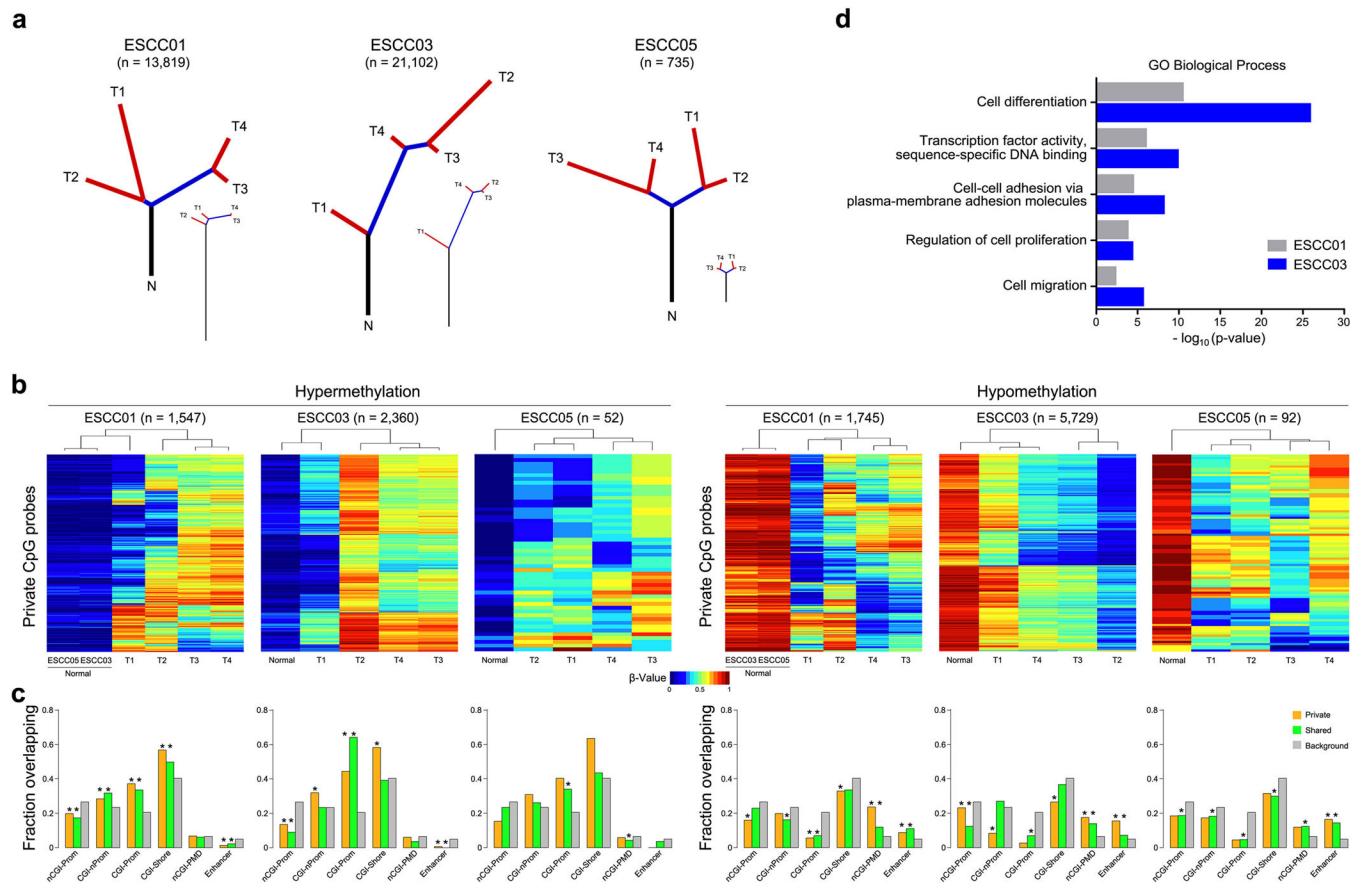
A heatmap displayed the cancer cell fraction (CCF) of driver mutations in each region of the ESCC tumors. Genomic regions with no segmentation data available were shown as NA.



**Fig. 3. Temporal dissection of mutational signatures in ESCC tumors**

(a) The 96 trinucleotide mutational spectrum of truncal (Bottom panel) and branched (Top panel) mutations across all regions was inferred by deconstructSigs. (b) Dot plots displayed the contributions of individual mutational signatures to individual cases, with each dot representing one case. Signatures 1–30 were based on the Wellcome Trust Sanger Institute COSMIC Mutational Signature Framework. Inferred signatures included: Signature 1 (associated with age), Signatures 2 and 13 (associated with APOBEC), Signatures 6 and 15 (associated with DNA mismatch repair), Signature 3 (associated with DNA double-strand

break-repair), Signature 7 (associated with UV exposure in squamous cancer). The bars represent the mean values. (c, d) Piecharts displayed the truncal and branch mutational signatures in cases ESCC10 and ESCC12, and only signatures with contributions over 10% were indicated.



#### Fig. 4. Epigenetic ITH in ESCC

(a) Phyloepigenetic trees of three ESCC cases. Lengths of trunks and branches were inferred using a phylogenetic approach, based on Euclidean distances between different tumor regions using private probes (see Methods). The total number of probes ( $n$ ) was provided above each tree. For comparison, phylogenetic trees from Fig. 1 were reproduced below each phyloepigenetic tree. (b) Heatmaps showed the beta values of private probes for each case, separated into hyper- and hypo-methylation. (c) Overlap between each probe set from panel (b), and a variety of functional genomic contexts: non-CpG Island Promoters (nCGI-Prom), non-Promoter CpG Islands (CGI-nProm), CpG Island Promoters (CGI-Prom), CpG Island Shores (CGI-Shore), Partially Methylated Domains excluding CpG Islands (nCGI-PMD) and enhancers. Overlapping frequencies of private probes from panel (b) were shown in yellow, shared probes (Supplementary Fig. 9) in green, and gray showed the frequency for the entire set of probes on the array. The hypergeometric test ( $* = P < 10^{-5}$ ) was used to compare the frequency of each private and shared probe set category to that of array background (see Methods). (d) Enriched GO biological processes for the genes associated with privately hypermethylated promoters in ESCC01 and ESCC03 (case ESCC05 was excluded due to the lack of sufficient privately hypermethylated promoters).

**Table 1**

Prevalence of non-silent mutations in ESCC (within-patient versus within-region)

Cancer gene	Prevalence (number of patients with mutations) in previous studies*	Within-region prevalence (number of regions with mutations) n = 51 regions	Within-patient prevalence (number of patients with mutations) n = 13 cases	Within-patient/within-region
<i>TP53</i>	78.9% (430)	94.1% (48)	92.3% (12)	0.98
<i>KMT2D</i>	13.8% (63)	23.5% (12)	23.1% (3)	0.98
<i>NOTCH1</i>	12.8% (70)	21.6% (11)	23.1% (3)	1.07
<i>FAT1</i>	11.2% (51)	15.7% (8)	15.4% (2)	0.98
<i>ZNF750</i>	5.7% (26)	15.7% (8)	15.4% (2)	0.98
<i>FAM135B</i>	6.4% (29)	13.7% (7)	15.4% (2)	1.12
<i>NFE2L2</i>	5.7% (26)	7.8% (4)	15.4% (2)	1.97
<i>PTPRB</i>	2.9% (13)	7.8% (4)	15.4% (2)	1.97
<i>ATM</i>	1.8% (8)	7.8% (4)	7.7% (1)	0.98
<i>BRCA2</i>	3.1% (14)	7.8% (4)	7.7% (1)	0.98
<i>CREBBP</i>	4.2% (19)	7.8% (4)	7.7% (1)	0.98
<i>KMT2A</i>	1.1% (5)	7.8% (4)	7.7% (1)	0.98
<i>NOTCH2</i>	3.3% (18)	7.8% (4)	7.7% (1)	0.98
<i>FAT2</i>	6.4% (29)	5.9% (3)	7.7% (1)	1.31
<i>KEAP1</i>	1.8% (8)	5.9% (3)	7.7% (1)	1.31
<i>MTOR</i>	1.1% (5)	3.9% (2)	7.7% (1)	1.96
<i>TP53BP1</i>	0.9% (4)	3.9% (2)	7.7% (1)	1.96
<i>KIT</i>	0.7% (3)	3.9% (2)	7.7% (1)	1.96
<i>PIK3CA</i>	9.0% (41)	3.9% (2)	7.7% (1)	1.96
<i>ATR</i>	1.1% (5)	2.0% (1)	7.7% (1)	3.92
<i>BRIP1</i>	0.9% (4)	2.0% (1)	7.7% (1)	3.92
<i>TSC1</i>	1.1% (5)	2.0% (1)	7.7% (1)	3.92

\* Summary of published data from Agrawal et al. (Ref. 4), Song et al. (Ref. 5), Lin et al. (Ref. 6), Gao et al. (Ref. 7), and Zhang et al. (Ref. 8). The total number of cases is 545 for *TP53*, *NOTCH1* and *NOTCH2* mutations, and is 456 for the rest gene mutations.

The last column showed the fold change when the prevalence was analyzed using individual cases instead of individual tumor regions.

Table 2

## Prevalence of subclonal mutations in ESCC

Case	Within-region prevalence				Within-patient prevalence
	T1	T2	T3	T4	
ESCC01	10.1%	16.1%	26.7%	15.4%	40.0%
ESCC02	14.7%	8.2%	10.4%	14.9%	20.5%
ESCC03	13.6%	7.2%	8.4%	24.1%	33.2%
ESCC04	10.7%	5.8%	NA	1.2%	13.3%
ESCC05	27.3%	21.4%	3.6%	33.3%	48.8%
ESCC06	6.9%	28.3%	5.4%	6.1%	33.3%
ESCC07	6.1%	21.1%	92.4%	61.1%	86.1%
ESCC08	11.6%	12.7%	15.4%	16.2%	31.7%
ESCC09	30.4%	41.3%	5.7%	20.0%	56.5%
ESCC10	21.2%	2.0%	3.1%	6.1%	27.0%
ESCC11	42.3%	35.5%	36.0%	41.7%	66.4%
ESCC12	1.4%	38.6%	6.1%	46.3%	62.5%
ESCC13	29.5%	3.0%	14.4%	29.5%	50.0%

Note: The within-patient prevalence was derived through dividing the number of subclonal mutations by the number of total mutations in each patients.