

ORIGINAL ARTICLE

Multiple region whole-exome sequencing reveals dramatically evolving intratumor genomic heterogeneity in esophageal squamous cell carcinoma

W Cao^{1,8}, W Wu^{1,2,7,8}, M Yan³, F Tian¹, C Ma¹, Q Zhang¹, X Li¹, P Han¹, Z Liu⁴, J Gu⁵ and FG Biddle⁶

Cancer is a disease of genome instability and genomic alterations; now, genomic heterogeneity is rapidly emerging as a defining feature of cancer, both within and between tumors. Motivation for our pilot study of tumor heterogeneity in esophageal squamous cell carcinoma (ESCC) is that it is not well studied, but the highest incidences of esophageal cancers are found in China and ESCC is the most common type. We profiled the mutations and changes in copy number that were identified by whole-exome sequencing and array-based comparative genomic hybridization in multiple regions within an ESCC from two patients. The average mutational heterogeneity rate was 90% in all regions of the individual tumors in each patient; most somatic point mutations were nonsynonymous substitutions, small Indels occurred in untranslated regions of genes, and copy number alterations varied among multiple regions of a tumor. Independent Sanger sequencing technology confirmed selected gene mutations with more than 88% concordance. Phylogenetic analysis of the somatic mutation frequency demonstrated that multiple, genomically heterogeneous divergent clones evolve and co-exist within a primary ESCC and metastatic subclones result from the dispersal and adaptation of an initially non-metastatic parental clone. Therefore, a single-region sampling will not reflect the evolving architecture of a genomically heterogeneous landscape of mutations in ESCC tumors and the divergent complexity of this genomic heterogeneity among patients will complicate any promise of a simple genetic or epigenetic diagnostic signature in ESCC. We conclude that any potential for informative biomarker discovery in ESCC and targeted personalized therapies will require a deeper understanding of the functional biology of the ontogeny and phylogeny of the tumor heterogeneity.

Oncogenesis (2015) 4, e175; doi:10.1038/oncsis.2015.34; published online 30 November 2015

INTRODUCTION

Cancer is a disease of genome instability and a resulting accumulation of genetic and epigenetic alteration. Global cancer genome projects have contributed to the identification and molecular classification of hundreds of cancer genes and the genomic architecture. With a deepening of the sequencing of the cancer genome, genomic heterogeneity is rapidly emerging as a defining feature of cancer, not only between tumors, but also within tumors. Accumulating evidence for intratumor heterogeneity^{1–6} is showing that tumors evolve through a process of branched evolution with genetically distinct subclones, which lead to tumor recurrence, drug resistance and metastatic potential.^{7,8}

Esophageal squamous cell carcinoma (ESCC) is the most common histological subtype of esophageal cancer in South-Eastern and Central Asia, particularly in China.⁹ Large-scale genome sequencing of ESCC has identified known frequently

mutated genes, such as TP53, and other previously unrecognized mutated genes,^{10–12} but intratumor heterogeneity in ESCC has not been well studied. Therefore, to establish the framework for a comprehensive and systematic whole-genome analysis of ESCC, we conducted a pilot study in ESCC patients, using multiple-region, whole-exome sequencing and array-based comparative genomic hybridization (aCGH), applied to 11 tumor regions from two surgically resected ESCCs, including metastatic lymph nodes. We show here that each tumor region has substantial genomic heterogeneity with its own unique profile of mutations and copy number alterations. Each tumor region was characterized for non-silent mutations and a branching evolution of cancer development was inferred. Functional analysis revealed common as well as unique, actionable and druggable mutated genes in the landscape of intratumor genomic heterogeneity in ESCC. Therefore, we will suggest that targeting the gene regulatory networks, which underlie the fitness landscape of ESCC, with a

¹Translational Medical Center, Zhengzhou Central Hospital Affiliated to Zhengzhou University, Zhengzhou, China; ²Department of Pathology and Laboratory Medicine, Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary Alberta, Canada; ³Medical School, Zhengzhou University, Zhengzhou, China; ⁴Linzhou Cancer Hospital, Linzhou, China; ⁵Science and Education Department, Health Bureau of Zhengzhou, Zhengzhou, China and ⁶Departments of Medical Genetics and Biological Sciences, University of Calgary, Calgary, Canada. Correspondence: Dr W Cao, Translational Medical Center, Zhengzhou Central Hospital Affiliated to Zhengzhou University, 195 Tongbai Road, Zhengzhou 450007, China.

E-mail: caoweiyou@hotmail.com

or Dr W Wu, Department of Pathology and Laboratory Medicine, Arnie Charbonneau Cancer Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada T2N 4N1.

E-mail: Wei.Wu@ucsf.edu

⁷Current address: Division of Hematology and Oncology, Department of Medicine, and Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA, USA.

⁸These authors contributed equally to this work.

Received 20 July 2015; revised 2 October 2015; accepted 8 October 2015

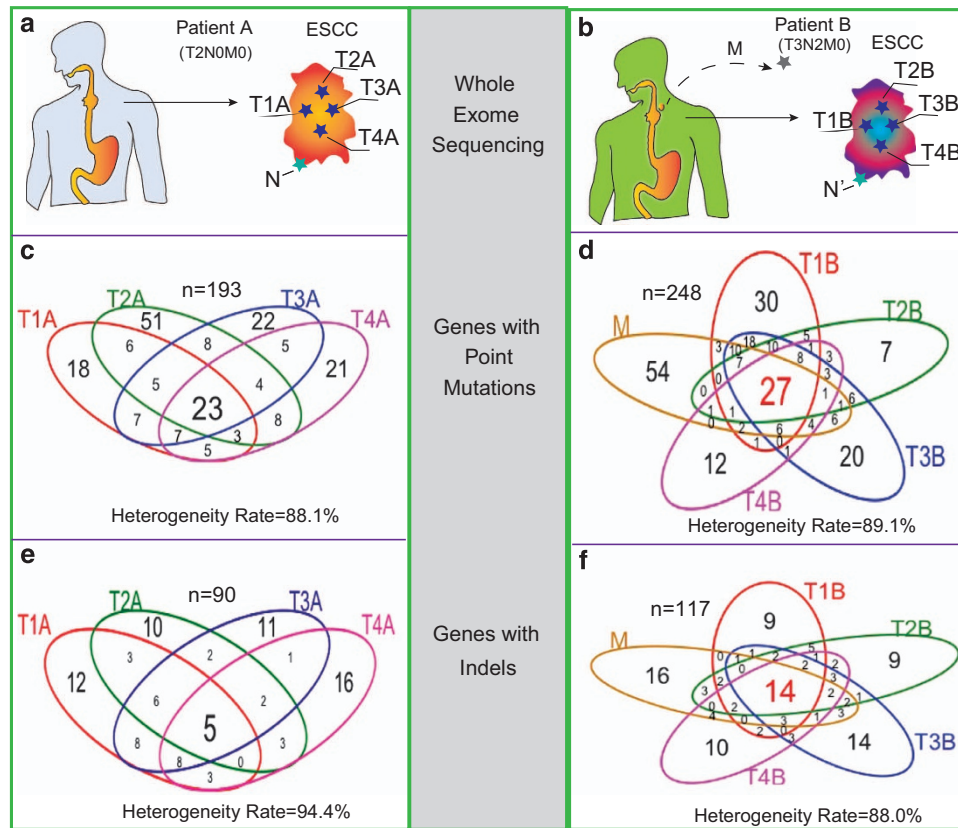


Figure 1. Intratumor heterogeneity of somatic mutations in ESCC. (**a, b**) Four spatially separated samples were obtained from a surgically resected ESCC as well as non-tumor tissue and associated metastatic lymph nodes were collected for multi-region exome sequencing and aCGH assay. (**c, d**) The Venn diagrams show the number of genes in each tumor region with SPMs from PtA and PtB. (**e, f**) The number of genes affected with insertions and deletions (Indels) in each region from PtA and PtB. The heterogeneity rate was computed with the total affected genes divided by the number of affected genes not shared by all tumor regions.

combinatorial approach rather than current canonical protocols may be necessary for optimal tumor treatment and control.

RESULTS

Mutational profile in multiple regions of ESCC

To characterize the extent of intratumor genomic heterogeneity in ESCC, we performed whole-exome sequencing on four separate regions within a primary ESCC sample and a non-tumor region from adjacent normal tissue from two patients (Pt), denoted as PtA and PtB (Figures 1a and b, Supplementary Table S1). PtA was diagnosed with ESCC without lymph node invasion, clinical stage was T2N0M0; PtB was diagnosed with advanced ESCC with lymph node metastasis, clinical stage was T3N2M0. Genomic aberrations in these samples were assessed for somatic point mutations (SPMs) and somatic small insertions or deletions (Indels). Annotation of the SPMs were categorized as nonsynonymous or missense mutations, synonymous or silent mutations, stop-gain or nonsense mutations, and splice site mutations. The non-silent mutations (nonsynonymous, stop-gain and splice site mutations) comprise more than 50% of the defined mutations in each tumor region (Supplementary Figures S1c and d). Indels occurred in exons of genes resulting in frameshift or non-frameshift mutations, as well as in the untranslated regions (5' or 3'-UTRs) and splice sites of genes. The majority of Indels were in UTRs (Supplementary Figures S1e and f). The predominant type of mutation in both patients is a C>T/G>A transition; the second most frequent mutation is an A>G/T>C transition in PtA, and a C>G/G>C transversion in PtB (Supplementary Figure S2). The difference in the mutation spectra between the two patients may

be due to a change in or a reflection of different mutational mechanisms during ESCC cancer development.^{12,13}

We then asked what genes are affected by the identified mutations across the tumor regions in each patient. Overall, only 23 of 193 genes with SPMs and 5 of 90 genes with Indels in PtA (Figures 1c and e) were found in all four regions, accounting for 12% and 7% common mutated genes with SPMs and Indels, respectively. Similarly, all tumor regions in PtB, including the metastasis, shared only 11% of the 248 mutated genes with SPMs and 12% of the 117 mutated genes with Indels (Figures 1d and f). Therefore, one measure of the degree of intratumor heterogeneity is the heterogeneity rate and it is defined as the number of mutated genes that are detected in one or more, but not all regions of a tumor. Both patients have similar average intratumor heterogeneity rates of approximately 90% for mutated genes with either SPMs or Indels.

Copy number alterations in multiple regions of ESCC

Copy number alterations in multiple regions of ESCC were surveyed using aCGH, and the amplifications and deletions of chromosomes were identified in each tumor region from PtA and PtB (Figure 2 and Supplementary Figures S3a and b). In PtA, 430 chromosomal segments were amplified or deleted in at least one region and among them, 71 (17%) genomic loci were amplified and 8 (2%) genomic loci were deleted in all four regions of the tumor (Figure 2a). The common amplified regions occurred in 2q, 3q, 8q, 10q, 11q, 14q and 18q, and the common deletions occurred in 3p21, 7q11.23, 10q24, 12q24, 19p13 and 19q13. All tumor regions did not share the majority of the chromosomal aberrations, indicating substantial intratumor chromosomal

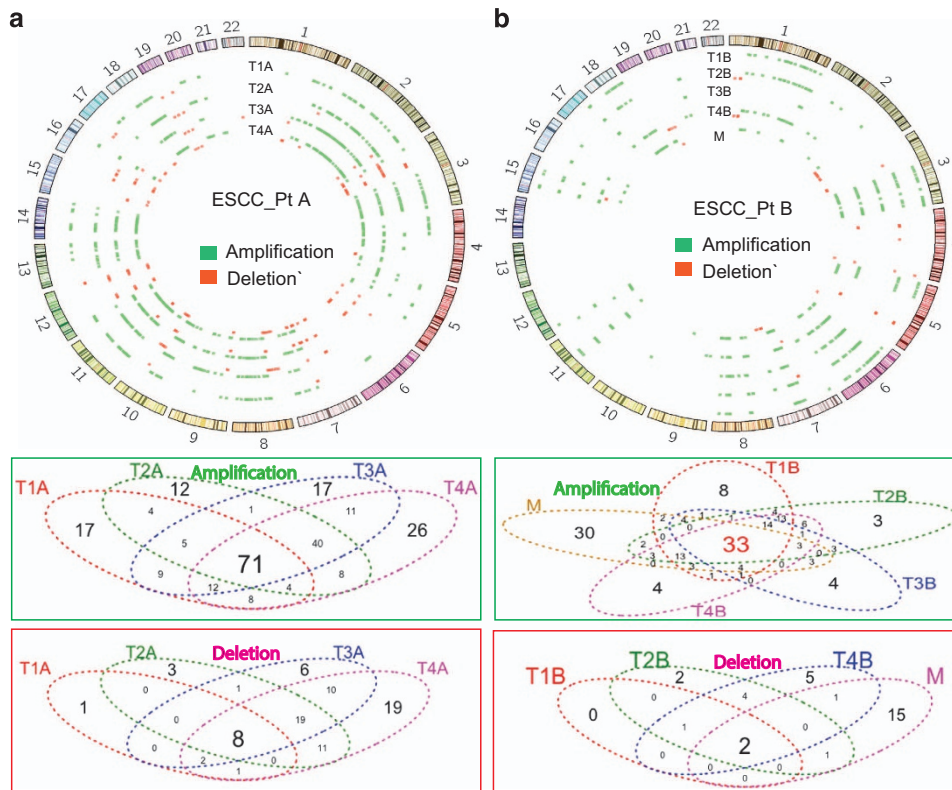


Figure 2. Intratumoral genomic copy number alterations of ESCC detected by aCGH. Genomic copy number alterations relative to non-tumor tissues in PtA (a) and in PtB (b) were plotted with Circos plot. Green color represents copy number amplification (Amp); Red color represents copy number deletion (Del), in each region of a given ESCC tissue. The total copy number alterations (Amp+Del): 403 in PtA, 262 in PtB. The Venn diagrams show the overlap of Amp or Del between each tumor region from PtA and PtB.

heterogeneity. In contrast, in PtB, only 262 chromosomal regions were amplified or deleted, in at least one region of the tumor (Figure 2b). There were 33 (13%) common amplified regions, located in 3p12, 3p11-13, 3q21.3-3q26.31, 6q, 7q, 8q and Xq, and two common deletions, which occurred at 4q35.2 and Yp11.31-Yq12, were shared only in tumor regions T1B, T2B, T4B and M (no deletion was detected in tumor region T3B). We note that the known frequent somatic CNVs, involving 3q26 (PRKCI gene amplification),¹⁴ 9p21 (SOX2 gene amplification)¹⁵ and 11q13.3 (CCTN gene amplification),¹⁶ were detected in this study. When taken together, heterogeneous amplifications and deletions were observed in multiple regions of each ESCC with an average heterogeneity rate of approximately 92%, which is very similar to the heterogeneity rate observed for SPMs and Indels.

Characterization of non-silent somatic mutations in multiple regions of ESCC

The nonsynonymous mutations, stop-gain mutations, splice site mutations and other mutations, including small Indels (frameshift and non-frameshift), generally change the protein function and we grouped them collectively as non-silent mutations. We note that the number of genes affected by mutations is relatively less than the total number of mutations, owing to the fact that multiple mutations may occur in the same gene. On the basis of this classification, there were 158 and 203 non-silent mutated genes (present in at least one tumor region) in PtA and PtB, respectively (Supplementary Tables S2 and S3), but only 16 genes were common to both patients (Figure 3a). We selected 51 gene mutations, including SPM and Indels, for independent validation by a Sanger sequencing approach, and 46 out of 51 gene mutations (88%) were confirmed (Supplementary Table S4).

To characterize intratumor heterogeneity of the non-silent mutated genes, we used the ‘trunk’ to represent ubiquitous mutations present in all regions of a tumor, the ‘branch’ to stand for heterogeneous mutations present in some, but not all regions of the tumor, and ‘private’ to correspond to mutations that are present in only one region of a tumor. We mapped these non-silent mutated genes across all spatially separated regions shown in the heatmap in Figures 3b and c. In PtA, we classified 158 non-silent mutated genes into 17 trunk gene mutations, 54 branch gene mutations and 87 private mutations. In PtB, 27 of 203 mutant genes were in the trunk, 76 mutant genes were located in the branch section and 100 genes were private mutations. Detection of private mutations in all tumor regions suggests that an ongoing, regional clonal evolution is occurring in PtA and PtB.

To determine whether these non-silent mutated genes are associated with cancer, we searched for our identified mutant genes in the COSMIC (Catalog Of Somatic Mutations In Cancer) database and other cancer-related studies.¹⁰⁻¹² There are 52 recognized cancer gene mutations across the tumor regions in PtA and 55 recognized cancer gene mutations across the tumor regions in PtB, and they are listed on the right side of the respective heatmaps in Figures 3b and c. In the trunk of the heatmap, 8 of the 17 mutated genes in PtA and 9 of the 27 mutated genes in PtB are known cancer-associated genes. PtB is a more advanced case with adjacent lymph node metastasis and we note that the TP53 (C85X, C85Y) mutation was detected in all tumor regions of PtB (but not PtA) and the metastatic region (M) of PtB has 18 cancer-associated gene mutations.

Tumor evolution in each patient was reconstructed from the regional mutation frequencies using a phylogenetic algorithm.¹⁷ A branching rather than a linear tumor evolutionary pattern was inferred for each patient and is illustrated in Figures 3d and e. Subclonal populations, with variable evolutionary distances

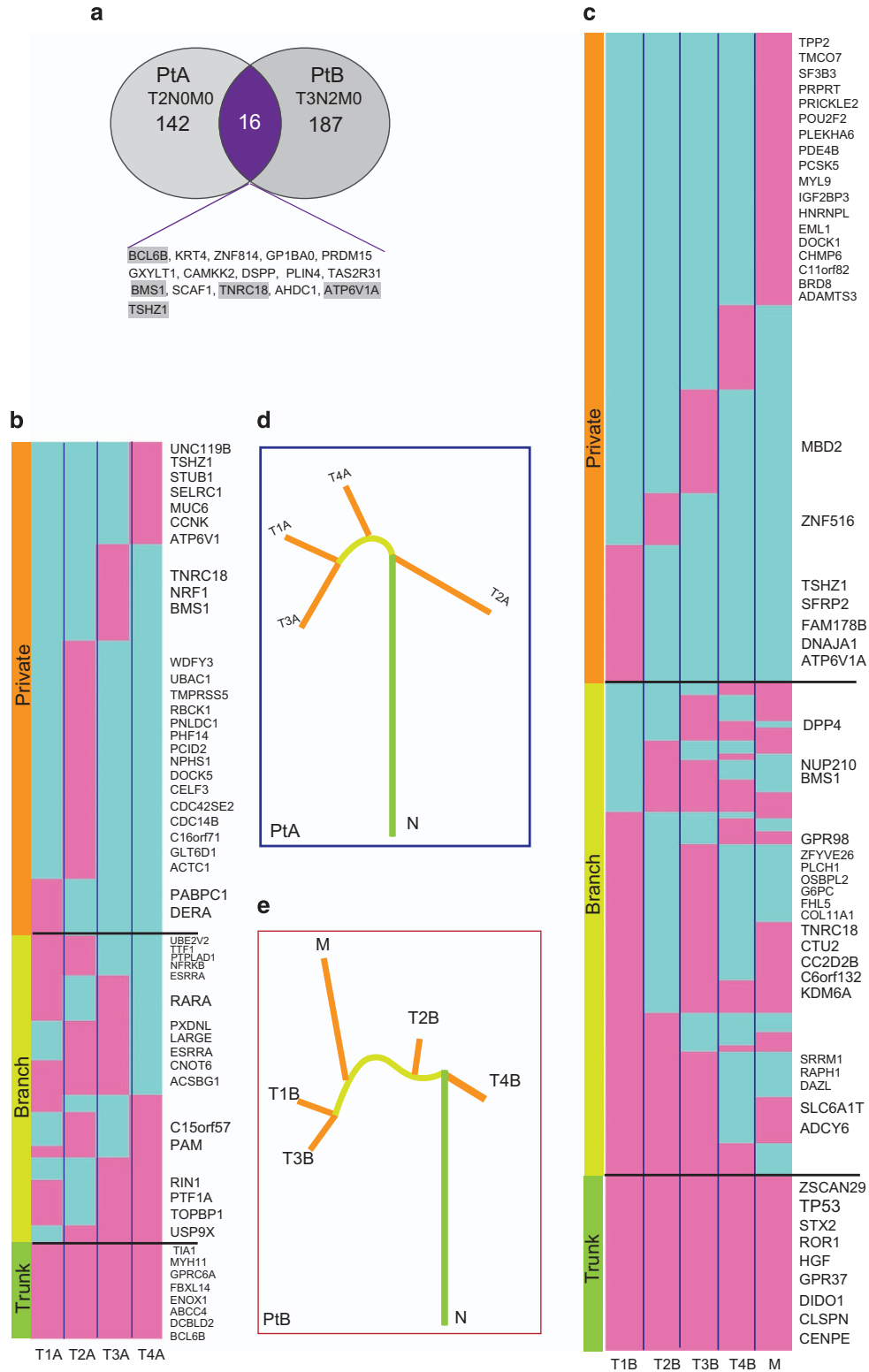


Figure 3. Characterization of non-silent mutations in multiple regional samples of ESCC. **(a)** Overlap of non-silent mutation affected genes between PtA ($n = 158$) and PtB ($n = 203$). **(b, c)** Heatmaps show the regional distributions of non-silent mutations in PtA and PtB; presence (pink), absence (light blue) is indicated for each tumor region. Gene mutations in all regions of a tumor shown with green bar, designated as 'Trunk'; gene mutations in some regions of a tumor shown with yellow bar ('Branch'); gene mutations in only one regions of a tumor shown with orange bar ('Private'), mutated cancer genes are listed on the right of the heatmaps. **(d, e)** Phylogenetic trees generated using the Ape and Phangorn packages in R based on the distribution of all detected mutations; trunk and branch lengths are proportional to the number of non-silent mutations in each tumor region.

Table 1. Pathways in each tumor region of ESCC			
	GO terms	P value	Associated genes
<i>PtA</i>			
T1A	Intracellular estrogen receptor signaling pathway	3.00E-04	CRIPAK, POU4F2, RARA
T2A	Secretory granule membrane	2.30E-03	ABCC4, PAM, PCSK4
	DNA packaging complex	1.60E-03	H1FOO, HIST1H3D, PXDNL, SMC4
T3A	Myosin complex	2.50E-03	MYH11, MYO18A, MYO18B
	Negative regulation of translation	1.80E-03	EIF2AK4, RARA, TIA1
T4A	O-linked glycosylation of mucins	9.80E-04	
	Termination of O-glycan biosynthesis	3.00E-04	MUC17, MUC5B, MUC6
	O-glycan processing	2.20E-03	
	Secretory granule membrane	1.60E-03	ABCC4, DBH, PAM
<i>PtB</i>			
T1B	Transport of mature transcript to cytoplasm	9.10E-03	NUP153, NXF1, SRRM1
	Transport of mature mRNA derived from an intron-containing transcript	1.40E-02	
	Melanoma	8.40E-03	HGF, PIK3R2, TP53
	ECM proteoglycans	1.20E-02	DSPP, MUSK, TNC
T2B	Melanoma	3.10E-03	HGF, PIK3R2, TP53
T3B	Ribosome biogenesis in eukaryotes	4.00E-03	BMS1, HEATR1, NXF1, REXO1
	ECM-receptor interaction	2.80E-03	COL11A1, GP1BA, THBS2, TNC
	Melanoma	7.50E-03	HGF, PIK3R2, TP53
T4B	NEP/NS2 interacts with the cellular export machinery	4.10E-04	NUP153, NUP210, XPO1
	Rev-mediated nuclear export of HIV RNA	3.00E-04	
	Export of viral ribonucleoproteins from nucleus	3.40E-04	
	Interactions of Rev with host cellular proteins	2.00E-04	
M	Prolactin signaling pathway	6.30E-03	CSN2, ESR1, PIK3R2
	Melanoma	1.20E-02	HGF, PIK3R2, TP53
	Muscle contraction	6.70E-03	MYL9, NEB, TPM2

Abbreviations: ESCC, esophageal squamous cell carcinoma; PtA, patient A; PtB, patient B.

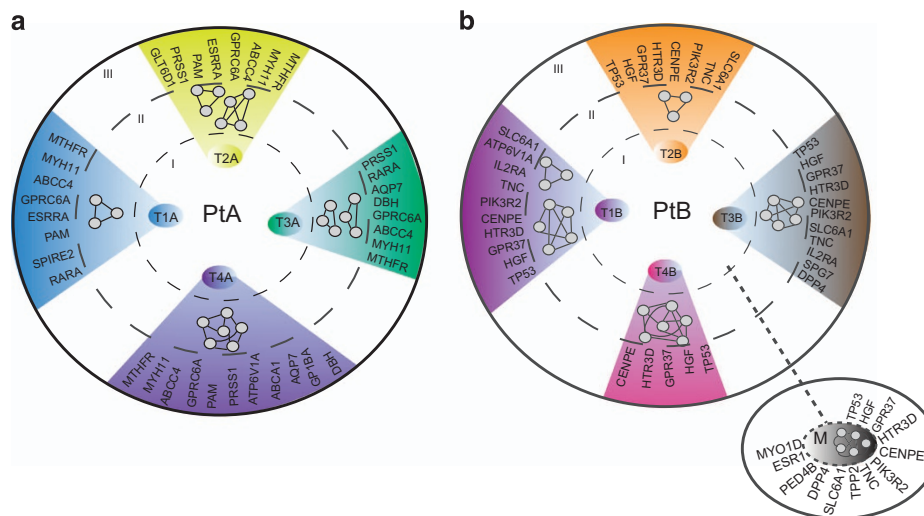


Figure 4. Functional intratumor genomic heterogeneity in ESCC. Non-silent mutations in each tumor region (circle I) were subjected to KEGG and Interactom database search using GlueGo algorithm. The networks of pathways and biological processes (summarized in Table 1) in each tumor region are illustrated conceptually as nodes and edges (circle II). The actionable and druggable genes in each tumor region were predicted with the Drug-Gene Interaction DATABASE (<http://dgidb.genome.wustl.edu/>) (circle III). Left panel: PtA, right panel: PtB; M, Metastasis.

between them, co-exist within a tumor and share trunk mutations. As well, the subclonal populations harbor their own private mutations, while continuing to maintain a stable phenotype or cancer property.

Characterization of mutations in UTRs from multiple regions of ESCC

We assessed SPMs and Indels in UTRs (3' and 5') that may change the regulatory elements of genes. UTRs contain binding sites for microRNAs and mutations in those UTRs may perturb interactions

between microRNAs and their target genes and, as a consequence, disrupt regulatory functions of the microRNAs. PtA had a total of 232 detected mutations in UTRs and PtB had a total of 366. We retrieved 1858 microRNA-binding sites that were predicted with the miRase algorithm and used them as a database to computationally assess the identified mutations in UTRs (Supplementary Figure S4). Two genes, CTDSPL2 (Chr15:44818173-44818173, an 'A' insertion) in region T3A and BCL7A (Chr12: 122498799-122498799, an 'A' insertion) in region T4A, were predicted to have mutations in 3'-UTR disrupting

microRNA-binding sites that are recognized by miR129/129-5p and miR204/211, respectively. The UTR mutations in PtB were not predicted to affect known microRNA-binding sites. Nevertheless, the majority of identified UTR mutations in both PtA and PtB appear to be of unknown significance.

Divergent complexity of functional intratumor genomic heterogeneity

Evolving intratumor mutational heterogeneity in ESCC reveals a complex divergence of misregulation of cancer-cell signaling pathways and the intratumor mutational heterogeneity may provide an informed guidance of combinatorial therapy. Most mutated genes do express and produce abnormal proteins and they may be components of different cancer pathways. Therefore, we made two functional assessments of the multiple tumor regions. First, we made functional predictions with GO terms in the KEGG and REACTOM search algorithms applied to the profiles of the non-silent mutations in the different tumor regions (Supplementary Figure S4). Functional heterogeneity in the different tumor regions suggested different regulatory signaling and cellular metabolic pathways (Table 1). The inferred gene regulatory networks of each tumor region from PtA and PtB, including the metastasis, are illustrated only symbolically with nodes and edges (circle II in Figures 4a and b). For example, MUC17, MUC5B and MUC6 gene mutations in tumor region T4A of PtA (Table 1) predict the perturbation of O-glycan biosynthesis and processing, and the GO terms in T4A differ from GO terms in other tumor regions. Similarly, mutations in the tumor regions from PtB suggest heterogeneous signaling pathways and biological processes, such as the HGF, PIK3R2 and TP53 mutations, which are known to result in cancer pathway activation. Mutations of genes CSN2, ESR1 and PIK3R2 may damage the prolactin-signaling pathway (Table 1). These data indicate that the divergent complexity of mutations, which is identified by genotyping of the separate tumor regions, may inevitably change the biological functionalities that are linked to intratumor heterogeneity.

Second, we used the well-established Drug-Gene Interaction Database (DGIDB) to predict actionable and druggable gene mutations. Drugs in this database are used clinically or they are currently in clinical trials. The results show that the patterns of potentially druggable mutated genes are heterogeneous in the different tumor regions. In PtA (circle III in Figure 4a), all tumor regions shared four common actionable mutated genes (MTFHR, MYH11, ABCC4 and GPRC6A); in PtB (circle III in Figure 4b), all tumor regions, including the metastasis, shared five druggable mutated genes (TP53, GPR37, HGF, CENPE and HTR3D). Perhaps more importantly, the tumor regions from each patient show an evolving divergent complexity of additional actionable and druggable mutated genes.

DISCUSSION

Single biopsy sampling for molecular analysis of tumors from patient cohorts is a standard analytical method and it is effective for diagnosis of single gene disorders. However, single tumor biopsy samples do not reveal the complexity of the genomic landscape in tumors with intratumor heterogeneity.^{1-3,5,18} Therefore, multiple regional sampling of tumors to estimate intratumor genomic heterogeneity has emerged as a non-canonical protocol to characterize actionable targets, make treatment decisions and manage chemo-resistance.⁷

In our pilot study, we assessed multiple regional samples of two primary ESCC tumors and an associated metastasis with whole-exome sequencing and aCGH. We found extensive divergent mutational profiles of SPMs, Indels, copy number amplifications and deletions in the spatially separated samples within each

tumor and metastasis with an average intratumor heterogeneity rate 90% (Figures 1 and 2). Further characterization of non-silent mutations in each tumor region showed only 17 (11%) out of 158 mutated genes in PtA (Figure 3b), and 27 (13%) of 203 in PtB (Figure 3c) to be common in all tumor regions ('trunk' mutations). The overall percentage of shared mutations between multiple tumor regions of ESCC in this study is dramatically lower than in other types of cancer such as clear cell renal cell carcinoma (30~35%)^{1,2} and high-grade serous ovarian cancers (52%).⁴ In other words, the rate of intratumor genomic heterogeneity is higher in ESCC (~90%) than in other types of tumor, including esophageal adenocarcinoma (56%).⁶ Higher intratumor heterogeneity index in esophageal adenocarcinoma is strongly correlated with poor responses to neoadjuvant chemotherapy.⁶ We suggest that our observed high intratumor heterogeneity rate in ESCC may contribute to its poor overall 5-year survival rates of 15-25%.¹⁹

To understand the biological significance of the non-silent mutation-affected genes that we detected in ESCC, we compared our mutations from PtA and PtB with the COSMIC database and other ESCC genome sequencing studies. Approximately one-third of our non-silent mutated genes are cancer-related. For example, the mutated genes common in all tumor regions from PtA, BCL6B,²⁰ MYH11,²¹ SUFU²² act as tumor suppressor genes, and mutations in BCL6B (Indel), MYH11(M816K) and SUFU(Y90S) could inactivate the gene function and contribute to tumorigenesis. Over expression of DCBLD2²³ may act as an oncogene interacting with the EGFR and PI3K/Akt signaling pathways; gain of function of DCBLD2 mutation (Indel) could participate with cellular proliferative pathways to promote tumor progression. In contrast, PtB is a more advanced case of ESCC with an adjacent lymph node metastasis and, as might be expected, more mutated genes were identified in PtB, but the intratumor genomic heterogeneity rate was the same as in PtA. We note that TP53 is a frequently and early mutated cancer gene,⁶ but it was detected as a trunk mutation only in PtB. CENPE is another trunk mutation gene in PtB and normally, it has a key role in the movement of chromosomes toward the metaphase plate during mitosis; the observed CENPE (Q1302E) gene mutation may disrupt its normal function and lead to chromosomal instability, but again, it only was found in PtB.

UTRs of genes contain microRNA regulatory-binding sites and mutations in these sequences could alter miRNA::mRNA interaction and significantly alter gene expression. These mutations in UTRs and the dysregulation of their functions are usually not reported in current studies of intratumor heterogeneity. In our computational search, we found multiple microRNA-binding sites mutated in target genes CTDSPL2 in region T3A and BCL7A in region T4A from PtA. These microRNAs are miR129-5p,^{24,25} miR204²⁶ and miR211,²⁷ which are putative tumor-suppressor microRNAs. These types of mutation in gene regulatory elements provide a hypothesis and a foundation for further experimental investigations.

Phylogenetic analysis is clearly a useful approach to visualize the complexity of intratumor heterogeneity as a tree structure in a given tumor because it summarizes the ontogeny of genomic similarity and divergence in all tumor regions. The present analysis unequivocally indicates that clonal subpopulations co-exist in the primary tumor and a metastatic subpopulation may be derived by additional mutations from initially non-metastatic parental clones in the primary tumor. This observation is supported by similar studies of intratumor heterogeneity in pancreatic cancer¹⁸ and clear cell renal cell cancer.¹

An intratumor heterogeneous genomic architecture may generate a microenvironment with divergent heterogeneous cellular signaling pathways and biological processes, which may harbor various actionable and druggable mutated genes. Therefore, we assessed the non-silent genes in each tumor region for potential regional specific functional pathways (Table 1) and

explored the potential for regional targeted therapy with small molecules (Figure 4). In PtA, no common pathways were identified, but there are four common actionable mutated genes, which are MHRF, MYH11, ABCC4 and GRPC6A. In PtB, a cancer pathway (melanoma) was identified in all tumor regions except T4B, and this pathway is composed of TP53, HGF and PIK3R2, which have existing targeted drugs. However, each tumor region also has its own uniqueness of cellular pathways and available targeted mutated genes. Thus, the functional intratumor heterogeneity presents a challenge to precision medicine. To achieve optimal tumor treatment and tumor control, it may be necessary to target clonally dominant truncal somatic events or adopt multiple targeted therapies in a combinatorial approach.²⁸

In summary, our observations, along with other studies, demonstrate that intratumor heterogeneity is a hallmark of ESCC, as it is for most cancers. Therefore, multi-regional sampling is a more informative approach for diagnosis, for potential biomarker discovery and for making treatment decisions. In addition, the evolving, divergent complexity of intratumor genomic and functional heterogeneity points to a necessity for unconventional strategies to understand the logic of the ESCC genomic system for reprogramming of its gene-regulatory network.

MATERIALS AND METHODS

Specimens

Written informed consent was obtained from two male patients before surgery and the Institutional Review Board for the use of human subjects at Zhengzhou Central Hospital, affiliated to Zhengzhou University, approved our study protocol. Multiregional specimens from primary tumors were obtained from two patients with ESCC who underwent surgical treatment and adjacent non-neoplastic tissues were obtained as references. One metastatic specimen was also obtained from one of the patients. Basic patient information is listed in Supplementary Table S1. Specimens were frozen in liquid nitrogen immediately after surgical resection. Neither patient had prior chemotherapy or radiotherapy, nor did they have any other serious disease. All ESCC tissues were histopathologically diagnosed by at least two independent senior pathologists.

Whole-exome sequencing

The genomic DNA from tumors of multiple regions, metastatic areas and matched adjacent normal tissues from the two ESCC patients were sheared by Bioruptor_NGS (Diagenode SA, Denville, NJ, USA) to produce fragment sizes of 200–300 bp. Samples of 100 ng of purified DNA were carried out through a process of end repair, phosphorylation and ligation to barcoded sequencing adapters. Ligated DNA was size-selected for lengths between 200 and 350 bp and subjected to exonic hybrid capture using SeqCap EZ Human Exome Library v3.0 (Roche Nimblegen, Madison, WI, USA). Each of the captured libraries were multiplexed and sequenced on multiple Illumina HiSeq 2000 (Illumina, San Diego, CA, USA) flow cells to average target exome coverage of 50× in neoplastic DNA and 60× in reference tissue using 76-bp paired-end reads.

Processing of sequencing data and detection of variation

After removing reads containing sequencing adapters and low-quality reads, high-quality reads were aligned to the NCBI reference genome (hg19) using Burrows-Wheeler Aligner 0.7.2 with default parameters. The SAMtools set of utilities was used to merge and remove duplicates. Local realignment and recalibration of base quality score was carried out with the GATK software package. SPMs (single nucleotide polymorphisms) and INDELS were predicted with GATK and variation file (*.vcf) was filtered and annotated with the ANNOVAR software tool. The sequence data have been deposited at the European Genome-Phenome Archive (EGA, <http://www.ebi.ac.uk/ega/>), which is hosted by the EMBL-EBI, under accession number EGAS 00001000965.

aCGH processing and analysis of copy number

Each genomic DNA (0.5 µg) was fluorescently labeled with the NimbleGen enzymatic labeling protocol which uses Cy3 and Cy5-labeled random nanomers (TriLink Biotechnologies, San Diego, CA, USA), a heat fragmentation step at 98 °C for 10 min, and amplification with Klenow

fragment 5'-3'-exo- (NEB, Ipswich, MA, USA). Five micrograms of each Cy5-labeled sample were co-hybridized with 5 µg of gender matched Cy3-labeled human male or female reference DNA (Promega, Madison, WI, USA) on Agilent SurePrint G3 Human Catalog CGH 8×60K (Design ID 021924, Agilent, Santa Clara, CA, USA) following the hybridization and washing conditions from the Agilent Oligonucleotide Array-based CGH for Genomic DNA Analysis Protocol v6.2. Arrays were scanned with the Agilent DNA Microarray Scanner at a 3-µm scan resolution, and quantified with Feature Extraction 11.0.1.1. CGH-processed signal was then uploaded into Partek Suite software (v6.6) where data were visualized and analyzed. The amplifications and deletions were called with default parameters except for changing of minimum genomic markers = 30 and signal/noise = 0.5 to increase the stringency. The aCGH data have been deposited in National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) and are accessible through (GEO) Series accession number GSE60625 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60625>).

Phylogenetic analysis

Phylogenetic analysis was conducted with published methods¹⁷ in R (V 2.15.0) using the Ape and Phangorn packages. Briefly, genetic distances were estimated among different regions under a generalized Kimura model. The initial phylogenetic trees were created using a neighbor-joining algorithm. Likelihood optimization was then used to obtain the maximum likelihood tree for each region. Finally, a non-parametric bootstrap on each maximum likelihood tree was conducted to estimate the support for individual branches ($n=1000$ iterations).

Functional analysis of non-silent mutations

The mutations (SPMs and Indels) in each tumor region obtained from whole-exome sequencing were further analyzed with the R project for statistical computing (version 3.1.2) and customized script (Supplementary Figure S3). The Cytoscape (v3.1.1) with GlueGo plugin was used for pathways analysis. The DGIDB (<http://dgidb.genome.wustl.edu/>) was used to predict actionable and druggable mutated genes.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (No: 81171992) and International Science and Technology Collaborative Project of Henan Province (No: 152102410088). We thank Anorroad Inc (Beijing, China) for providing sequencing services and primary data analysis. GenomeSky Inc provided help with bioinformatics and computational analysis. We especially thank Dr Andrew Hughes at the Center for Genome Sciences and Systems Biology, Washington University, St Louis, MO, for providing programming code for phylogenetic analysis.

REFERENCES

- Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012; **366**: 883–892.
- Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I *et al*. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* 2014; **46**: 225–233.
- Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC *et al*. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci USA* 2013; **110**: 4009–4014.
- Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A *et al*. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J Pathol* 2013; **231**: 21–34.
- Zhang XC, Xu C, Mitchell RM, Zhang B, Zhao D, Li Y *et al*. Tumor evolution and intratumor heterogeneity of an oropharyngeal squamous cell carcinoma revealed by whole-genome sequencing. *Neoplasia* 2013; **15**: 1371–1378.
- Murugaesu N, Wilson GA, Birkbak NJ, Watkins T, McGranahan N, Kumar S *et al*. Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov* 2015; **5**: 821–831.
- Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer* 2013; **108**: 479–485.

- 8 Seoane J, De Mattos-Arruda L. The challenge of intratumour heterogeneity in precision medicine. *J Intern Med* 2014; **276**: 41–51.
- 9 Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut* 2014; **64**: 381–387.
- 10 Song Y, Li L, Ou Y, Gao Z, Li E, Li X *et al*. Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* 2014; **509**: 91–95.
- 11 Lin DC, Hao JJ, Nagata Y, Xu L, Shang L, Meng X *et al*. Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nat Genet* 2014; **46**: 467–473.
- 12 Zhang L, Zhou Y, Cheng C, Cui H, Cheng L, Kong P *et al*. Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. *Am J Hum Genet* 2015; **96**: 597–611.
- 13 Gao YB, Chen ZL, Li JG, Hu XD, Shi XJ, Sun ZM *et al*. Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet* 2014; **46**: 1097–1102.
- 14 Yang YL, Chu JY, Luo ML, Wu YP, Zhang Y, Feng YB *et al*. Amplification of PRKCI, located in 3q26, is associated with lymph node metastasis in esophageal squamous cell carcinoma. *Genes Chromosomes Cancer* 2008; **47**: 127–136.
- 15 Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG *et al*. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet* 2009; **41**: 1238–1242.
- 16 Luo ML, Shen XM, Zhang Y, Wei F, Xu X, Cai Y *et al*. Amplification and over-expression 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.
- 17 Hughes AE, Magrini V, Demeter R, Miller CA, Fulton R, Fulton LL *et al*. Clonal architecture of secondary acute myeloid leukemia defined by single-cell sequencing. *PLoS Genet* 2014; **10**: e1004462.
- 18 Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B *et al*. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 2010; **467**: 1114–1117.
- 19 Pennathur A, Gibson MK, Jobe BA, Luketich JD. Oesophageal carcinoma. *Lancet* 2013; **381**: 400–412.
- 20 Xu L, Li X, Chu ES, Zhao G, Go MY, Tao Q *et al*. Epigenetic inactivation of BCL6B, a novel functional tumour suppressor for gastric cancer, is associated with poor survival. *Gut* 2012; **61**: 977–985.
- 21 Wang RJ, Wu P, Cai GX, Wang ZM, Xu Y, Peng JJ *et al*. Down-regulated MYH11 expression correlates with poor prognosis in stage II and III colorectal cancer. *Asian Pac J Cancer Prev* 2014; **15**: 7223–7228.
- 22 Liu X, Wang X, Du W, Chen L, Wang G, Cui Y *et al*. Suppressor of fused (Sufu) represses Gli1 transcription and nuclear accumulation, inhibits glioma cell proliferation, invasion and vasculogenic mimicry, improving glioma chemosensitivity and prognosis. *Oncotarget* 2014; **5**: 11681–11694.
- 23 Feng H, Lopez GY, Kim CK, Alvarez A, Duncan CG, Nishikawa R *et al*. EGFR phosphorylation of DCBLD2 recruits TRAF6 and stimulates AKT-promoted tumorigenesis. *J Clin Invest* 2014; **124**: 3741–3756.
- 24 Ma N, Chen F, Shen SL, Chen W, Chen LZ, Su Q *et al*. MicroRNA-129-5p inhibits hepatocellular carcinoma cell metastasis and invasion via targeting ETS1. *Biochem Biophys Res Commun* 2015; **461**: 618–623.
- 25 Tan G, Cao X, Dai Q, Zhang B, Huang J, Xiong S *et al*. A novel role for microRNA-129-5p in inhibiting ovarian cancer cell proliferation and survival via direct suppression of transcriptional co-activators YAP and TAZ. *Oncotarget* 2015; **6**: 8676–8686.
- 26 Butrym A, Rybka J, Baczynska D, Tukiendorf A, Kuliczkowski K, Mazur G. Low expression of microRNA-204 (miR-204) is associated with poor clinical outcome of acute myeloid leukemia (AML) patients. *J Exp Clin Cancer Res* 2015; **34**: 68.
- 27 Xia B, Yang S, Liu T, Lou G. miR-211 suppresses epithelial ovarian cancer proliferation and cell-cycle progression by targeting Cyclin D1 and CDK6. *Mol Cancer* 2015; **14**: 57.
- 28 Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C. Intratumor heterogeneity: seeing the wood for the trees. *Sci Transl Med* 2012; **4**: 127ps10.



Oncogenesis is an open-access journal published by Nature Publishing Group. This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Supplementary Information accompanies this paper on the *Oncogenesis* website (<http://www.nature.com/oncsis>).