2397

Mutation signatures and intratumor heterogeneity of esophageal squamous cell carcinoma in a Chinese cohort

2397

Yu Wang[1], Wenqing Yuan[2], Mengfei Liu[2], Jian Bai[3], Qingxuan Song[1], Amir Abliz[2], Zhen Liu[2], Jingjing Li[2], Changqing Zeng[3], Hong Cai[2], Yang Ke[2], Jun Li[1]

1. Department of Human Genetics, University of Michigan, Ann Arbor, MI. 2. Laboratory of Genetics, Key laboratory of Carcinogenesis and Translational Research, Peking University Cancer Hospital & Institute, Beijing, China . 3. Laboratory of Disease Genomics and Individualized Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Introduction

Esophageal squamous cell carcinoma (**ESCC**) is one of the most common and most aggressive cancers. Epidemiological features of ESCC are extremely complex, with strong geographic differentiation among world's populations. Rural **Anyang** in the Henan Province of China is a well-known high-incidence area, however the causal factors in this population remain elusive. We performed exome sequencing of 81 tumor-normal pairs, identified *TP53*, *ZNF750* and *NOTCH1* as significantly mutated genes, and observed highly recurrent aberrations in several other genes previously reported for ESCC (*PIK3CA*, *KMT2D*, *FAT2* and *FAT1*). The total set of ~7,000 single-nucleotide mutations revealed two main signatures: C>T transitions at NpCpG due to spontaneous deamination of 5-methyl-cytosine, and C>T and C>G mutations at TpCpW attributed to the APOBEC family of cytidine deaminases. The latter signature points to HPV infection as one of the potential mutagenic sources, even a possible etiological factor. To characterize intratumor heterogeneity we applied our newly developed method, *CHAT*, to estimate the clonal frequencies of copy number alternations and single nucleotide mutations in each tumor. Many tumors show a wide, sometimes multi-modal distribution of clonal frequencies, suggesting extensive within-tumor diversity due to coexistence of multiple clones. Survival analysis suggests that patients with complex multi-clonal tumor structure tend to have poorer prognosis. To better understand the patterns of growth, migration and metastatic potential among different cells within a tumor we performing exome sequencing to compare multiple regions in 10 ESCC as well as 2 esophageal neuroendocrine carcinoma tumors. For each, we analyzed 4-6 sectors of the tumor, 2-4 samples of adjacent normal tissue, and 1-2 nearby lymph nodes. In many ESCC's, each local-region sample still contains multiple clones, which are often shared with a distant region, suggesting extensive dispersal of the clonal populations and relatively slow sweep by the most dominant clone. We constructed "clone trees" to depict the most likely lineage relationship of the clones and the likely driver genes or pathways for each branch. Metastatic samples at lymph node often contain multiple clones, including those appearing in the early portions of the evolutionary tree, suggesting polyclonal seeding to the lymph nodes as well as invasion of early-stage tumor cells. Our analysis of spatial heterogeneity of molecular lesions in ESCC revealed likely temporal progression of tumorigenic events that may have driven the initiation, outgrowth, and metastasis.

Anyang Esophageal Cancer Cohort

Table 2. Selected characteristics of endoscopy participants in the baseline investigation, Anyang Esophageal Cancer Cohort Study (2006–2009).

Variables	No. of subjects (N = 8,112)	Percent
Village		
Chengbei	803	9.9
Hengcun	918	11.3
Malan	1,378	17.0
Shangzhuang	681	8.4
Shencun	654	8.1
Tunzhuang	1,389	17.1
Xiaopu	702	8.7
Xitaoyuan	409	5.0
Zhuzhao	1,178	14.5
Age (years)		
25–35	1,843	22.7
36–45	2,889	35.6
46–55	1,916	23.6
56–65	1,464	18.0
Mean age (Standard deviation)	44.3 (10.5)	
Gender		
Female	4,268	52.6
Male	3,844	47.4
Education		
Illiterate	959	11.8
Primary school	2,540	31.3
Junior high school	3,496	43.1
Senior high school or above	668	8.2
Unknown	449	5.5
Type of employment		
Farming	5,062	62.4
Non-farming	2,601	32.1
Unknown	449	5.5
Family history of esophageal cancer		
No	7,201	88.8
Yes	613	7.6
Unknown	298	3.7

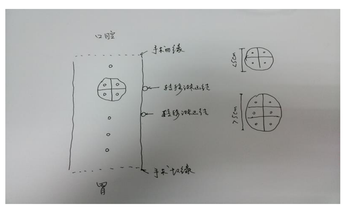


Figure 1. Epidemiological studies of the Anyang cancer cohort and planning of multi-region analysis of tumor heterogeneity.

Two dominant mutation signatures

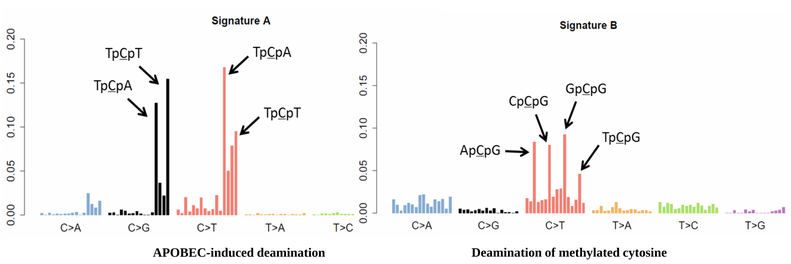
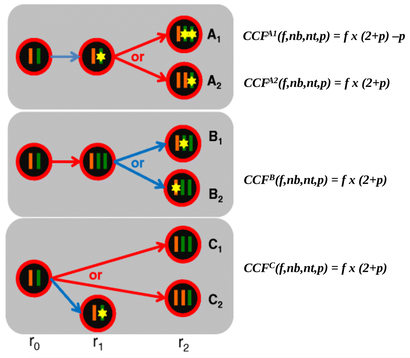


Figure 2. Somatic mutation signatures in ESCC. (A) APOBEC-induced deamination. **(B)** Deamination of methylated cytosine. The former suggest inflammation and possible HPV infection.

Clonal heterogeneity analysis tool *(CHAT)*



$CCF^{A1}(f,nb,nt,p) = f \times (2+p) - p$

$CCF^{A2}(f,nb,nt,p) = f \times (2+p)$

$CCF^{B}(f,nb,nt,p) = f \times (2+p)$

$CCF^{C}(f,nb,nt,p) = f \times (2+p)$

Figure 3. We developed a method, *CHAT*, to estimate the Cancer Cell Fraction (CCF) of each somatic SNV by adjusting its observed somatic allele frequency according to the background copy number status, while considering the sCNA clonality, the relative order of occurrence between the SNV and its associated sCNA, and their cis- or trans- relationship (Li and Li, 2014). Shown is an example of lineage scenarios for a mutation that fall in a region of heterozygous amplification.

Mutations in ESCC-related genes tend to have higher clonal freq. than those in other genes

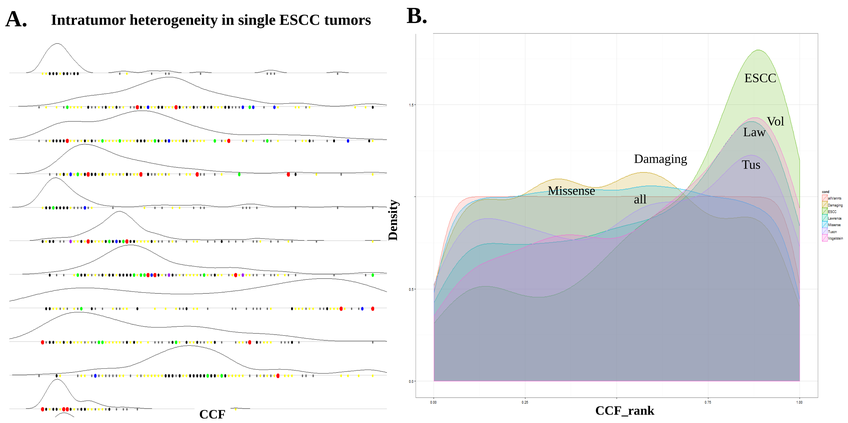


Figure 4. (A) Distribution of Cancer Cell Fractions (CCF) for mutations in 11 tumors (randomly selected from 81 ESCCs). The full range of CCF is (0-1) with lower values on the left. Mutations are annotated by predicted impact and grouped by color: Synonymous variants (grey); Missense non-damaging variants (yellow); Damaging variants predicted by CHASM (black) (Carter et al.,2009); Cancer-related genes were colored according to four published lists: Genes with high cancer probability (342) (green) (Davoli et al., 2013); Genes listed in pan-cancer research (240) (blue) (Lawrence et al., 2014); Genes listed by Vogelstein (pink) (144) (Vogelstein et al., 2013); ESCC related genes (red) (24) (Gao, 2014; Song, 2014; Lin, 2014; Zhang, 2015). **(B)** After re-scaling CCF values within each tumor to 0-100% and summing over 81 samples, we observe that ESCC related-genes tend to have mutations with higher CCFs.

Multi-clonal tumors tend to have poorer prognosis

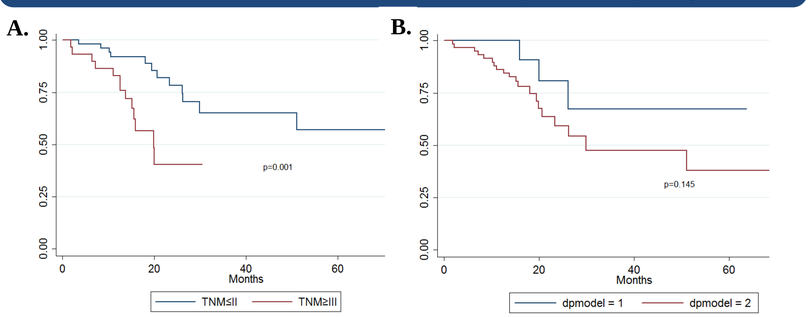


Figure 5. Survival time comparison of samples with high- and low TNM staging (A) and clonal heterogeneity (B). Shown are Kaplan-Meier survival curves (n=81).

Multi-region heterogeneity and clone evolution

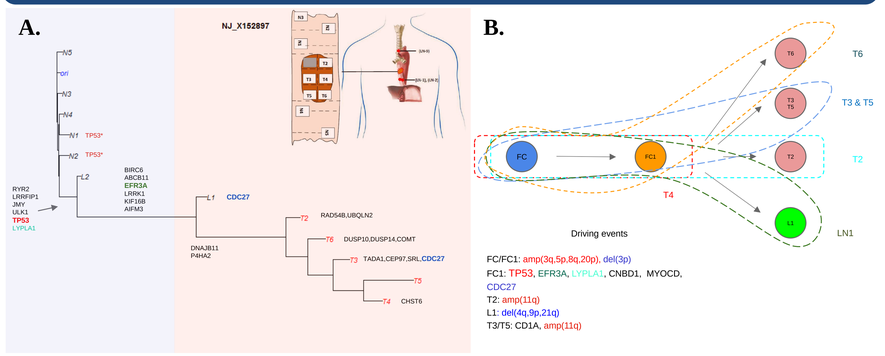


Figure 6. Intratumor heterogeneity as seen in a sample tree (A) and clone tree (B). We analyzed 4-6 sectors of each of 12 ESCC's, 2-4 adjacent normal tissue samples, and 1-2 nearby lymph node samples. For patient x152897, we used the mutation clonality values to construct phylogenetic relationship for samples or for clones, and infer the driver events along the evolution paths of individual regions. **(A)** in the sample tree *TP53* mutation is an early event (i.e., appearing in the trunk of the tree); and a series of mutations n other cancer-related genes occur in the branches. **(B)** Estimate subclones within all tumors and lymph node L1, and construct their evolutional relationship using the *SCHISM* method. Solid circles indicate the subclones. Dotted shapes indicates the samples. Possible driving events were marked below, and colored as before.

APOBEC-signature is enriched in "early" mutations

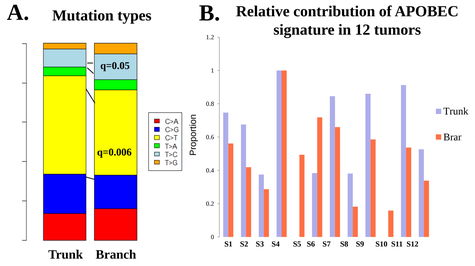


Figure 7. Differential loading of mutation signatures in trunk and branch mutations. We compared the subtype composition and mutation signature loadings between truncal and branch mutations. **(A)** The proportion of C>T transitions is lower in branch mutations. q values were Bonferroni-corrected p values. **(B)** Relative loadings of APOBEC-related mutations are higher in trunk mutations (blue) than branch mutations (red) in most samples

Conclusions

- By whole-exome sequencing of tumor-normal pairs (N=81) we identified significantly mutated genes in ESCC.
- We found two dominant somatic signatures: APOBEC-induced deamination and deamination of methylated cytosine. Since APOBEC activities are associated with exogenous viruses, the prominence of this signature suggests a role of HPV in ESCC etiology, consistent with our previous studies that detected HPV DNA in tumor samples and anti-HPV-E7 antibody in patient's blood.
- Known cancer-related genes and those reported for ESCC tend to have mutations with higher CCFs, consistent with their role in ESCC initiation.
- Many tumors show a multi-modal distribution of clonal frequencies, suggesting extensive within-tumor diversity due to coexistence of multiple clones. Survival analysis suggests that patients with complex multi-clonal tumor structure tend to have poorer prognosis.
- We further characterized multi-region heterogeneity and evolution paths in 12 ESCC tumors. Key mutations can be placed on the trunk or individual branches of the phylogenetic trees.
- Most samples in the phylogenetic tree contain multiple clones, making it necessary to infer clonal evolution before mapping to regional samples.
- ESCC seem to develop with extensive motility, resulting in multi-clone dispersal in distant regions.

References

- Li B, Li JZ. A general framework for analyzing tumor subclonality using SNP array and DNA sequencing data. Genome Biol, 2014. 14(9)
- Lawrence et al., Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature, 2013; 499(7457):214-8.
- Lawrence et al., Discovery and saturation analysis of cancer genes across 21 tumor types. Nature, 2014; 505(7484):495-501
- Vogelstein et al., Cancer Genome Landscapes. Science, 2013; 339(6127): 1546-1558
- Davoli et al., Cumulative haploinsufficiency and triplosensitivity driver aneupoloidy patterns to shape the cancer genome. Cell, 2013. 155(4):948-962
- Carter et al., Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. Cancer Res.2009: 69(16):6660-7.
- Gao et al., Genetic landscape of esophageal squamous cell carcinoma. Nat Genetic 2014. 46(10):1097-102
- Song et al., Identification of genomic alterations in oesophageal squamous cell cancer. Nature 2014. 509(7498):91-5
- Zhang et al., Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma. Am J Hum Genet. 2015; 96(4):597-611
- De-Chen Lin et al., Genomic and molecular characterization of esophageal squamous cell carcinoma. Nat Genetics 2014. 46(5):467-73