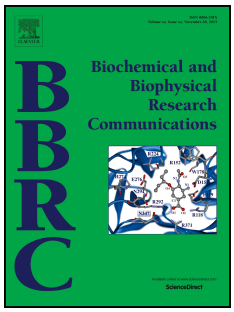
Accepted Manuscript



Noninvasive diagnosis and monitoring of mutations by deep sequencing of circulating tumor DNA in esophageal squamous cell carcinoma

Honglei Luo, Hong Li, Zhaoyang Hu, Hongjin Wu, Chenglin Liu, Ying Li, Xiaoyan Zhang, Ping Lin, Qiang Hou, Guohui Ding, Yan Wang, Shuang Li, Dongkai Wei, Feng Qiu, Yixue Li, Shixiu Wu

PII: S0006-291X(16)30201-7

DOI: 10.1016/j.bbrc.2016.02.011

Reference: YBBRC 35314

To appear in: *Biochemical and Biophysical Research Communications*

Received Date: 28 January 2016

Accepted Date: 4 February 2016

Please cite this article as: H. Luo, H. Li, Z. Hu, H. Wu, C. Liu, Y. Li, X. Zhang, P. Lin, Q. Hou, G. Ding, Y. Wang, S. Li, D. Wei, F. Qiu, Y. Li, S. Wu, Noninvasive diagnosis and monitoring of mutations by deep sequencing of circulating tumor DNA in esophageal squamous cell carcinoma, *Biochemical and Biophysical Research Communications* (2016), doi: 10.1016/j.bbrc.2016.02.011.

Noninvasive diagnosis and monitoring of mutations by deep sequencing of circulating tumor DNA in esophageal squamous cell carcinoma

Authors: Honglei Luo[2†], Hong Li[3†], Zhaoyang Hu[1†], Hongjin Wu[1], Chenglin Liu[4], Ying Li[1], Xiaoyan Zhang[1], Ping Lin[3], Qiang Hou[1], Guohui Ding[3], Yan Wang[1], Shuang Li[1], Dongkai Wei[5], Feng Qiu[5], Yixue Li[3*], Shixiu Wu[1*]

Affiliations:

[1]Hangzhou Cancer Institution, Hangzhou Cancer Hospital, Hangzhou, China.

[2]Department of Radiotherapy, Huai'an First People's Hospital, Huai'an, China.

[3]Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

[4]School of Life sciences and Biotechnology, Shanghai Jiaotong University, Shanghai, China

[5]Basepair Biotechnology, Co.,Ltd, Suzhou, China

***Corresponding Author:**

wushixiu@medmail.com.cn (SXW) or yxli @sibs.ac.cn (YXL).

Abstract

Circulating tumor DNA (ctDNA) is becoming an important biomarker in noninvasive diagnosis and monitoring of tumor dynamics. This study tested the feasibility of plasma ctDNA for the non-invasive analysis of tumor mutations in esophageal squamous cell carcinoma (ESCC) by sequencing of tumor, tumor-adjacent, and normal tissue, as well as pre-surgery and post-surgery plasma. Exome sequencing of eight patients identified between 29 and 134 somatic mutations in ESCCs, many of which were also determined in ctDNA. Comparison of pre-surgery and post-surgery plasma has shown that mutations had reduced frequency or disappeared after surgery treatment. We further evaluated the TruSight Cancer sequencing panel by using it to detect mutations in the plasma of three patients. Tumor mutations were only found in

one of them. To design a sequencing panel with improved targeting, we identified significantly mutated genes by meta-analysis of 532 ESCC genomes. Our results confirmed the well-known driver genes and found several uncharacterized genes. The new panel consisted of 90 recurrent genes, which theoretically achieved 94% and 75% of sensitivity when detecting at least 1 and 2 mutant genes in ESCC patients, respectively. Our results demonstrate the feasibility of using ctDNA to detect ESCCs and monitor treatment effect. The low-cost and sensitive target panel could facilitate clinical usage of ctDNA as a noninvasive biomarker.

1. Introduction

Esophageal cancer is the sixth leading cause of cancer-related death in the world. It includes two main types: esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EA) [1]. China is one of the countries with the highest incidence rate, with ESCC being the histopathological form in 90% cases [2]. Despite significant improvements in treatment modalities, the outcome of ESCC is dismal. The five-year survival rate ranges from 5% to 30%, based on the statistics of the National Cancer Institute. Patients with early disease have a better survival rate, but approximately 50% of patients already have a metastatic disease at diagnosis [3]. Thus, it is urgent to facilitate the development of reliable early screening methods or apparatuses for ESCC.

Circulating tumor DNA (ctDNA), the tumor-derived DNA in the plasma of cancer patients, has become an important biomarker in recent years. The concentration of ctDNA ranges from 0.1% to 10% in advanced cancers, and it is positively correlated with tumor stage [4] and tumor volume [5]. Since the amount of ctDNA in total plasma DNA is small, the detection and quantification of ctDNA is very challenging. Using digital polymerase chain reaction (PCR) based technology, a recent study of 640 cancer patients has found that ctDNA was detectable in >75% of cases of late-stage cancers (pancreatic, ovarian, colorectal, bladder, gastroesophageal, breast, melanoma, hepatocellular, and head and neck cancers), but in <50% of cases of primary cancers [4]. Newman et al. developed the new cancer-personalized profiling strategy by deep sequencing, which quantifies ctDNA using deep sequencing of recurrent tumor mutations or rearrangements to achieve a depth of coverage of around

10,000×. It detected ctDNA in 100% of patients with stages II–IV and in 50% of patients with stage I of non-small-cell lung cancers (NSCLCs) [5].

As a noninvasive method, ctDNA sequencing is valuable in many clinical applications, such as early detection, tumor progress monitoring, and resistance mutation identification. The analysis of 52 metastatic breast cancer patients revealed that ctDNA sequencing was able to identify a progressive disease much earlier than CT imaging [6]. Chan et al. reported a genome-wide investigation of ctDNA from four hepatocellular carcinoma patients [7]. Their results demonstrated that tumor-associated variations, in particular 8%-98% of copy number variants and 15%-94% of single-nucleotide variants (SNVs), could be found in pre-surgery plasma by whole-genome sequencing [7]. Murtaza et al. studied six patients with advanced cancers and followed them for 1–2 years after cancer therapy [8]. They compared the abundance of somatic mutations found in plasma before and after treatment, and identified mutations with significantly increased abundance after treatment, some of which are known to be responsible for therapy resistance [8].

Due to the lack of specific symptoms and poor diagnosis of early ESCC patients, ctDNA sequencing is a promising method for improving early detection of ESCC. To the best of our knowledge, no studies have characterized ctDNA in ESCC patients so far. Here, we used exome or targeted sequencing to detect somatic mutations in eleven ESCC patients and compared ctDNA from pre- and post-surgery plasma. To detect ctDNA with higher sensitivity, we used a bioinformatic approach to design a target-gene panel that would cover recurrently mutating genes in ESCC. Our studies will promote the clinical application of ctDNA in the detection and monitoring of ESCCs.

2. Methods

2.1 Patient and sample collection

This study recruited 11 patients from Huai'an First People's Hospital. Patients were diagnosed with ESCC between June 20, 2014 and January 20, 2015. Tumor, tumor-adjacent, and normal samples were obtained at the time of surgery. Adjacent samples were 0.5 cm away from tumors, while normal samples were 2 cm away from tumors. All non-tumor samples were validated by pathologists. Peripheral blood

samples were obtained at two points: the day before surgery and 7 days after surgery. This study was approved by the Hangzhou Cancer Hospital human research ethics committee and carried out in accordance with the approved guidelines. All patients signed the informed consent (IRB number 20131027025).

2.2 Sequencing and variants calling

Forty samples from eight patients were assigned to whole-exome sequencing. Fifteen samples from three patients were assigned to Illumina TruSight Cancer sequencing. Low-quality sequencing reads were removed and the remained reads were aligned to human genome (hg19). SNVs and indels were called by GATK based on GATK Best Practices. Somatic mutations were called by three softwares: varScan [9], muTect [10] and GATK. More details were described in the Supplementary methods.

2.3 Integrated analysis of ESCC genomes and ESCC target-gene panel

We collected gene mutation data of ESCCs from four public genome studies [11,12,13,14] and TCGA esophageal carcinoma project, downloaded mutation data of another 12 cancers from https://www.synapse.org/#!Synapse:syn1729383 [15]. We measured the significance of mutant genes by counting the number of patients with this gene (N_mutated), and then calculating the mutation recurrence rate (N_mutated/N_all) and mutation score (N_muated/Exon_length). To design a gene panel that could cover more ESCC patients using fewer genes, we selected genes using the following steps: 1) 18 ESCC driver genes that were reported by more than one large-scale ESCC genome study, 2) 55 potential ESCC driver genes that were reported only once, and 3) previously uncharacterized mutated genes that were found by integrated analysis of ESCC genomes.

3. Results

3.1 Overview

We recruited eleven ESCC patients whose tumors were removed by surgical treatment (Table S1). Tumor (T), tumor-adjacent (A), and normal tissue (N), as well as pre-surgery (Pre) and post-surgery plasma (Post), were collected in case patients gave informed consent (Fig 1A). Fig 1B shows the representative image of Hematoxylin and Eosin (H&E) staining of normal tissue and cancer tissue from patient ESCC1. The staining of adjacent tissue is the same as that of normal tissue (data not shown). Forty samples from eight patients (ESCC1 to ESCC8) were subjected to exome sequencing, and fifteen samples from another three patients (ESCC9 to ESCC11) were subjected to Illumina TruSight Cancer sequencing.

3.2 Somatic mutations in ESCC samples

Sequencing achieved 36 to 256 million reads, with a median of 60 million. After quality control and duplication removal, the average depth of normal, tumor, and plasma samples were 40×, 17×, and 32×, respectively. To validate the quality of variant calling in some low-coverage samples, we used hierarchical clustering to illustrate the similarity of germline SNVs among samples (Fig 1C). Except the pre-surgery plasma of ESCC1 and ESCC3, samples from the same patient were clustered together. Therefore, our sequencing results can capture the major characteristics of tumor and plasma samples.

Since normal tissues are difficult to obtain, the adjacent tissues of tumors are conventionally used to replace normal samples. Here, we evaluated their similarity by comparing the genomes. We identified somatic mutations of tumors using matching adjacent tissues or normal tissues as controls, and only used high-confidence mutations that were called by more than one software. Mutations obtained from these two methods were called “T-A” (adjacent tissue as a control) or “T-N” (normal tissue as a control). Fig 2A shows the number of somatic mutations that were identified by both methods or by only one method. The median overlap coefficient between “T-A” and T-N” was 89%. This suggested that the genome of tumor-adjacent tissues was very similar to that of normal tissues. We identified 29 to 134 functional mutations in tumor tissues (Fig 2B). These mutations were mapped to six known ESCC driver genes: *TP53*, *NOTCH1*, *RB1*, *CREBBP*, *DOCK10* and *CHEDK2*. Six patients (75%)

had functional *TP53* mutations. Two patients (25%) had *NOTCH1* mutations.

3.3 Noninvasive identification of ESCC mutations in plasma ctDNA

We first compared paired plasma and normal samples to identify potential somatic mutations in the plasma. Then, we limited our analysis to somatic mutations that were also identified in matched tumors in order to decrease the false positive rate. Fig 2C shows the comparison of allele frequency (AF) in tumor and pre-surgery plasma. Some tumor-associated mutations were detected in pre-surgery plasma, while others were not observed in plasma. Such results were expected due to the low level of plasma ctDNA. The number of detectable mutations in plasma ranged from 5 to 39, having a lower AF in plasma compared to tumors.

To determine whether significantly detectable mutations correlate with surgery status, we compared the AF in pre-surgery and post-surgery plasma. Generally, mutation AF decreased after surgery (Fig 2D). Examples of such mutations are listed in Table S2. For example, ESCC7 had a *TP53 N131I* mutation in tumor (AF=0.529) and pre-surgery plasma (AF=0.013), but not in post-surgery plasma. The lowered frequency of tumor mutations after surgery demonstrated that tumor cells had been completely or mostly abrogated. A few tumor mutations were found in post-surgery but not in pre-surgery plasma. They could be somatic mutations that were missed in pre-surgery plasma, or germline mutations that were not called in post-surgery.

For the first time, we used exome sequencing to monitor plasma ctDNA in ESCC patients. All of the results confirmed that exome sequencing had the ability to find tumor-associated mutations in plasma. However, exome sequencing has many limitations in ctDNA detection, such as high cost, low sensitivity for a single mutation, and difficulty with distinguishing low-frequency mutations from sequencing errors. In recent years, large-scale genome studies have provided more knowledge on the gene mutations in ESCC genomes. Therefore, we deeply mined recurrent genes in ESCC and then explored the clinical usage of targeted sequencing in ctDNA detection.

3.4 Recurrent genes in ESCC genomes

We collected whole-genome or whole-exome sequencing datasets from the

literature and the TCGA project. In total, we obtained somatic mutations of 532 ESCCs and 4429 other types of tumors. To measure the roles of genes in ESCC, we calculated their mutation recurrence rate and mutation score (Fig 3A). Genes were classified into four levels based on their association with ESCC: driver genes previously reported in 1 (pink) or 2 or more papers (red), genes listed in the Cancer Gene Census [16] (green), and other genes (black). Most of the previously reported driver genes are more frequently mutated and have higher mutation scores. We also observed previously uncharacterized mutant genes, such as *PRDM9* and *RBM15*, which were mutated in 4.5% and 3.0% of ESCCs. PRDM9 has histone-methyltransferase activity and is involved in chromatin organization [17]. Histone modifier genes have been proven to be frequently mutated in the ESCC genome [11,12]. Therefore, PRDM9 may play an important role in ESCC tumorigenesis. RBM15 has a repressor function in several signaling pathways and has been associated with leukemia [18,19]. In summary, integrative analysis of large tumor cohorts comprehensively captures the molecular heterogeneity of ESCC.

The recent pan-cancer studies reveal that many genes are significantly mutated across multiple cancer types [15]. To examine whether ESCC-specific genes exist, we compared ESCCs with the 12 types of pooled tumors. Fig 3B shows 11 genes with a higher recurrence rate in ESCC (FDR <0.05, recurrence rate >2%) than the average of other cancers. Compared to the IntOGen mutational cancer drivers database [20], 8 of the 11 genes are drivers in at least one cancer type, so their recurrence rate in ESCC is high but not fully specific. *DNAH6* (4.9%), *VWDE* (2.4%), and *ADGB* (2.2%) have not been detected as mutational cancer drivers in the IntOGen database, but their low mutation score (~0.002) and unclear biological functions make them less interesting. Tumor suppressor *PTCH1* is a predicted driver in medulloblastoma (6.2%). We observed that *PTCH1* is mutated in 4.9% ESCCs (Fig 3C), significantly higher than the average in other cancers (FDR = 0.04). A recent study also reported that *PTCH1* harbored frequent inactivating mutations in ESCC genomes [13].

3.5 Clinical usage of target-gene sequencing for ctDNA detection

The Illumina TruSight Cancer sequencing panel was designed for targeting 94 genes associated with both common and rare cancers. It covered a part of ESCC

driver genes that are shared with other cancers, such as *TP53*, *CDKN2A*, *RB1*, and *PTEN*. We examined the mutation profiles of these 94 genes in 532 sequenced ESCCs. Fig 5A shows the distribution of tumors with different mutation numbers. There are 230, 156, 54, and 22 tumors with 1, 2, 3, and 4 mutant genes, respectively. However, mutations in 60 tumors (11.3%) are not covered by the 94 genes in the TruSight Cancer panel. To address the difference between exome sequencing and targeted sequencing, TruSight Cancer sequencing was performed on 15 samples from three ESCC patients (ESCC9 to ESCC11). We observed 2, 3, and 1 somatic mutation in the tumor samples of ESCC9, ESCC10, and ESCC11, respectively (Fig 5B). The *TP53 C176F* mutation (chr17:7578403:C>A) in ESCC9 was also found in pre-surgery plasma. However, other mutations were not detected in plasma.

Finally, to improve the sensitivity for tumor detection, we designed a new target-gene panel for ESCC using the integrative analysis of 532 ESCCs. Genes were selected based on their functional roles and recurrence rate in ESCC: 18 driver genes that were reported in more than one paper, 55 potential driver genes that were reported in one paper, *PTCH1*, whose recurrence rate was higher in ESCC than in other cancers (Fig 3BC), and 16 recurrent genes whose recurrence rate was >2% and mutation score was >0.004 (Fig 3A). Collectively, the ESCC gene panel targeted 90 genes, covering ~506 kb. We examined the target regions in 532 sequenced ESCCs and calculated the number of mutant genes covered per tumor (Fig 5C). When using only 18 driver genes, 90%, 53%, and 18% tumors harbored at least 1, 2, and 3 mutant genes, respectively. When using all 90 genes, these numbers increased to 94%, 75%, and 52%, respectively. Compared to the TruSight Cancer sequencing panel (89%, 45%, 16%), the new ESCC panel covered more patients and more mutations per patient, while using less genes. Sensitive detection of ESCC ctDNA could be achieved by target enrichment and deep sequencing. However, significantly mutated driver genes are often shared by multiple cancers (Fig 5D). Therefore, detecting ctDNA in plasma is not specific enough for ESCC diagnosis. Combining ctDNA with other ESCC-specific biomarkers would achieve a more accurate diagnosis.

4. Discussion

Most of the recent ctDNA studies aim to accurately detect low-level ctDNA in

blood. The ctDNA level can be evaluated by PCR-based assays, whole-exome, or whole-genome sequencing. The PCR-based target method is fast and cost-effective, but requires knowledge of tumor-associated mutations. Sequencing-based methods do not require prior knowledge and can theoretically cover all patients, but they are expensive and require higher computational abilities.

Here, we used both exome sequencing and Illumina TruSight Cancer sequencing to detect ctDNA in the plasma of ESCC patients. Somatic mutations in tumor samples were detected by both sequencing platforms. However, only a small fraction of mutations can be found in plasma. The concentration of ctDNA cannot be well estimated by our computational analysis (data not shown). The sensitivity of ctDNA detection by sequencing is limited by 1) the amount of input ctDNA, 2) the proportion of repeated reads generated in the PCR step, 3) the deep coverage of mutated genes, and 4) sequencing or variant calling errors. Deeper sequencing may identify more mutations in plasma, but the amount of DNA was not sufficient to perform more sequencings in our study. Enrichment of recurrent genes is another way to improve sensitivity, but there is no targeted sequencing panel available for ESCC.

To resolve this problem, we selected target genes by integrating the analysis of previously sequenced ESCC genomes. Apart from known driver genes, we found 16 unreported recurrent genes. Combining the already known and the newly identified recurrent genes, we obtained an ESCC panel consisting of 90 genes. In theory, this small target panel (0.016% of the human genome) could cover 1 mutations in 94% of patients, and 3 mutations in 52% patients. Thus it is promising for the detection and monitoring of ctDNA in ESCC patients. We also realized a limitation of this panel: it cannot distinguish ESCC from other cancers. This is a general problem when applying ctDNA to tumor diagnosis, not just for ESCC. We can improve the specificity of ESCC diagnosis by combining ctDNA detection with other blood biomarkers or clinical examination.

Furthermore, we also collected plasma samples before and after surgery to examine the change of ctDNA. Compared to pre-surgery plasma, we found the AFs of some mutations were lower or equaled zero in post-surgery plasma. Such results demonstrate that ctDNA is a valuable biomarker for tracking tumor status and evaluating treatment effect. The future of ESCC treatment is early and precise diagnosis. Accurate methods, such as noninvasive assays for ctDNA in order to monitor ESCC formation and progression, will be necessary for clinical management.

5. References

[1] A.K. Rustgi, H.B. El-Serag, Esophageal carcinoma, N Engl J Med 371 (2014) 2499-2509.

[2] L.A. Torre, F. Bray, R.L. Siegel, J. Ferlay, J. Lortet-Tieulent, A. Jemal, Global cancer statistics, 2012, CA Cancer J Clin 65 (2015) 87-108.

[3] P.C. Enzinger, D.H. Ilson, D.P. Kelsen, Chemotherapy in esophageal cancer, Semin Oncol 26 (1999) 12-20.

[4] C. Bettegowda, M. Sausen, R.J. Leary, I. Kinde, Y. Wang, N. Agrawal, B.R. Bartlett, H. Wang, B. Luber, R.M. Alani, E.S. Antonarakis, N.S. Azad, A. Bardelli, H. Brem, J.L. Cameron, C.C. Lee, L.A. Fecher, G.L. Gallia, P. Gibbs, D. Le, R.L. Giuntoli, M. Goggins, M.D. Hogarty, M. Holdhoff, S.M. Hong, Y. Jiao, H.H. Juhl, J.J. Kim, G. Siravegna, D.A. Laheru, C. Lauricella, M. Lim, E.J. Lipson, S.K. Marie, G.J. Netto, K.S. Oliner, A. Olivi, L. Olsson, G.J. Riggins, A. Sartore-Bianchi, K. Schmidt, M. Shih l, S.M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T.T. Harkins, S. Veronese, T.L. Wang, J.D. Weingart, C.L. Wolfgang, L.D. Wood, D. Xing, R.H. Hruban, J. Wu, P.J. Allen, C.M. Schmidt, M.A. Choti, V.E. Velculescu, K.W. Kinzler, B. Vogelstein, N. Papadopoulos, L.A. Diaz, Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies, Sci Transl Med 6 (2014) 224ra224.

[5] A.M. Newman, S.V. Bratman, J. To, J.F. Wynne, N.C. Eclov, L.A. Modlin, C.L. Liu, J.W. Neal, H.A. Wakelee, R.E. Merritt, J.B. Shrager, B.W. Loo, Jr., A.A. Alizadeh, M. Diehn, An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage, Nat Med 20 (2014) 548-554.

[6] S.J. Dawson, D.W. Tsui, M. Murtaza, H. Biggs, O.M. Rueda, S.F. Chin, M.J. Dunning, D. Gale, T. Forshew, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas, N. Rosenfeld, Analysis of circulating tumor DNA to monitor metastatic breast cancer, N Engl J Med 368 (2013) 1199-1209.

[7] K.C. Chan, P. Jiang, Y.W. Zheng, G.J. Liao, H. Sun, J. Wong, S.S. Siu, W.C. Chan, S.L. Chan, A.T. Chan, P.B. Lai, R.W. Chiu, Y.M. Lo, Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing, Clin Chem 59 (2013) 211-224.

[8] M. Murtaza, S.J. Dawson, D.W. Tsui, D. Gale, T. Forshew, A.M. Piskorz, C. Parkinson, S.F. Chin, Z. Kingsbury, A.S. Wong, F. Marass, S. Humphray, J. Hadfield, D. Bentley, T.M. Chin, J.D. Brenton, C. Caldas, N. Rosenfeld, Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA, Nature 497 (2013) 108-112.

[9] D.C. Koboldt, Q. Zhang, D.E. Larson, D. Shen, M.D. McLellan, L. Lin, C.A. Miller, E.R. Mardis, L. Ding, R.K. Wilson, VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing, Genome Res 22 (2012) 568-576.

[10] K. Cibulskis, M.S. Lawrence, S.L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E.S. Lander, G. Getz, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples, Nat Biotechnol 31 (2013) 213-219.

[11] Y. Song, L. Li, Y. Ou, Z. Gao, E. Li, X. Li, W. Zhang, J. Wang, L. Xu, Y. Zhou, X. Ma, L. Liu, Z. Zhao, X. Huang, J. Fan, L. Dong, G. Chen, L. Ma, J. Yang, L. Chen, M. He, M. Li, X. Zhuang, K. Huang, K.

Qiu, G. Yin, G. Guo, Q. Feng, P. Chen, Z. Wu, J. Wu, L. Ma, J. Zhao, L. Luo, M. Fu, B. Xu, B. Chen, Y. Li, T. Tong, M. Wang, Z. Liu, D. Lin, X. Zhang, H. Yang, J. Wang, Q. Zhan, Identification of genomic alterations in oesophageal squamous cell cancer, Nature 509 (2014) 91-95.

[12] Y.B. Gao, Z.L. Chen, J.G. Li, X.D. Hu, X.J. Shi, Z.M. Sun, F. Zhang, Z.R. Zhao, Z.T. Li, Z.Y. Liu, Y.D. Zhao, J. Sun, C.C. Zhou, R. Yao, S.Y. Wang, P. Wang, N. Sun, B.H. Zhang, J.S. Dong, Y. Yu, M. Luo, X.L. Feng, S.S. Shi, F. Zhou, F.W. Tan, B. Qiu, N. Li, K. Shao, L.J. Zhang, L.J. Zhang, Q. Xue, S.G. Gao, J. He, Genetic landscape of esophageal squamous cell carcinoma, Nat Genet 46 (2014) 1097-1102.

[13] L. Zhang, Y. Zhou, C. Cheng, H. Cui, L. Cheng, P. Kong, J. Wang, Y. Li, W. Chen, B. Song, F. Wang, Z. Jia, L. Li, Y. Li, B. Yang, J. Liu, R. Shi, Y. Bi, Y. Zhang, J. Wang, Z. Zhao, X. Hu, J. Yang, H. Li, Z. Gao, G. Chen, X. Huang, X. Yang, S. Wan, C. Chen, B. Li, Y. Tan, L. Chen, M. He, S. Xie, X. Li, X. Zhuang, M. Wang, Z. Xia, L. Luo, J. Ma, B. Dong, J. Zhao, Y. Song, Y. Ou, E. Li, L. Xu, J. Wang, Y. Xi, G. Li, E. Xu, J. Liang, X. Yang, J. Guo, X. Chen, Y. Zhang, Q. Li, L. Liu, Y. Li, X. Zhang, H. Yang, D. Lin, X. Cheng, Y. Guo, J. Wang, Q. Zhan, Y. Cui, Genomic analyses reveal mutational signatures and frequently altered genes in esophageal squamous cell carcinoma, Am J Hum Genet 96 (2015) 597-611.

[14] D.C. Lin, J.J. Hao, Y. Nagata, L. Xu, L. Shang, X. Meng, Y. Sato, Y. Okuno, A.M. Varela, L.W. Ding, M. Garg, L.Z. Liu, H. Yang, D. Yin, Z.Z. Shi, Y.Y. Jiang, W.Y. Gu, T. Gong, Y. Zhang, X. Xu, O. Kalid, S. Shacham, S. Ogawa, M.R. Wang, H.P. Koeffler, Genomic and molecular characterization of esophageal squamous cell carcinoma, Nat Genet 46 (2014) 467-473.

[15] C. Kandoth, M.D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J.F. McMichael, M.A. Wyczalkowski, M.D. Leiserson, C.A. Miller, J.S. Welch, M.J. Walter, M.C. Wendl, T.J. Ley, R.K. Wilson, B.J. Raphael, L. Ding, Mutational landscape and significance across 12 major cancer types, Nature 502 (2013) 333-339.

[16] P.A. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, M.R. Stratton, A census of human cancer genes, Nat Rev Cancer 4 (2004) 177-183.

[17] G.L. Jiang, S. Huang, The yin-yang of PR-domain family genes in tumorigenesis, Histol Histopathol 15 (2000) 109-117.

[18] Y. Yang, S. Wang, Y. Zhang, X. Zhu, Biological effects of decreasing RBM15 on chronic myelogenous leukemia cells, Leuk Lymphoma 53 (2012) 2237-2244.

[19] H.H. Hsiao, M.Y. Yang, Y.C. Liu, H.P. Hsiao, S.B. Tseng, M.C. Chao, T.C. Liu, S.F. Lin, RBM15-MKL1 (OTT-MAL) fusion transcript in an adult acute myeloid leukemia patient, Am J Hematol 79 (2005) 43-45.

[20] G. Gundem, C. Perez-Llamas, A. Jene-Sanz, A. Kedzierska, A. Islam, J. Deu-Pons, S.J. Furney, N. Lopez-Bigas, IntOGen: integration and data mining of multidimensional oncogenomic data, Nat Methods 7 (2010) 92-93.

6. Acknowledgments

We thank Kangping Xu in Base Pair Biotechnologies Co. Ltd. for the technical support. The study was funded by the Zhejiang province natural funds Z15H220001, Science of Technology Commission of Shanghai Municipality (15YF1414100), and National Key Technology Support Program (2013BAI01B09).

Author contributions statement

Conceived and designed the experiments: SXW, YXL. Performed the experiments: HLL, ZYH, DKW, FQ. Analyzed the data: HL, CLL, PL, GHD, ZYH. Contributed reagents/materials/analysis tools: ZYH, HJW, YL, XYZ, QH, YW, SL. Wrote the paper: HLL, HL, ZYH.

Figure Legend:

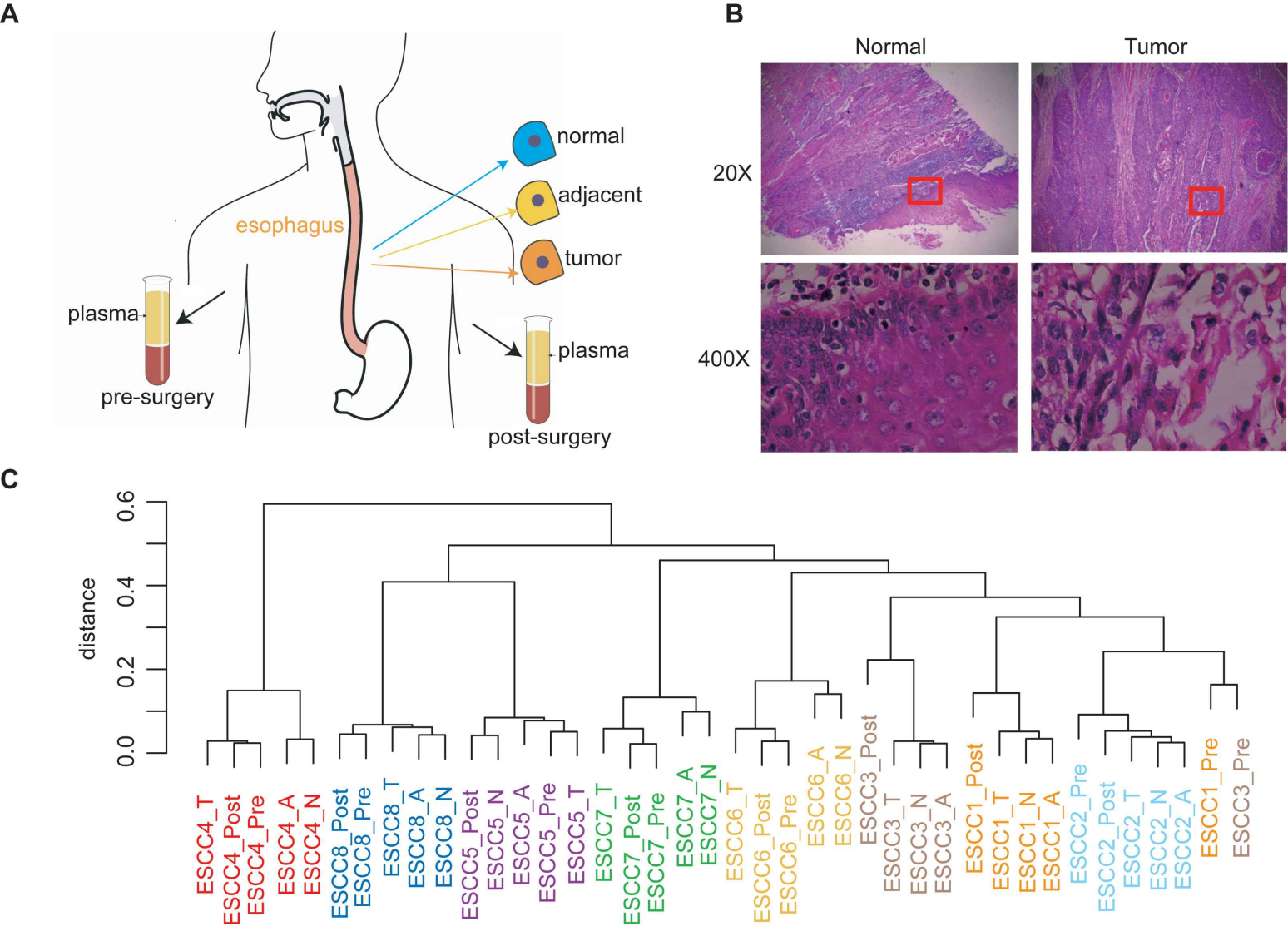
Fig 1. (A) Overview of the study design. (B) Hematoxylin- and eosin-stained normal and their paired tumor samples. (C) Cluster dendrogram based on the germline SNV overlap coefficient between samples.

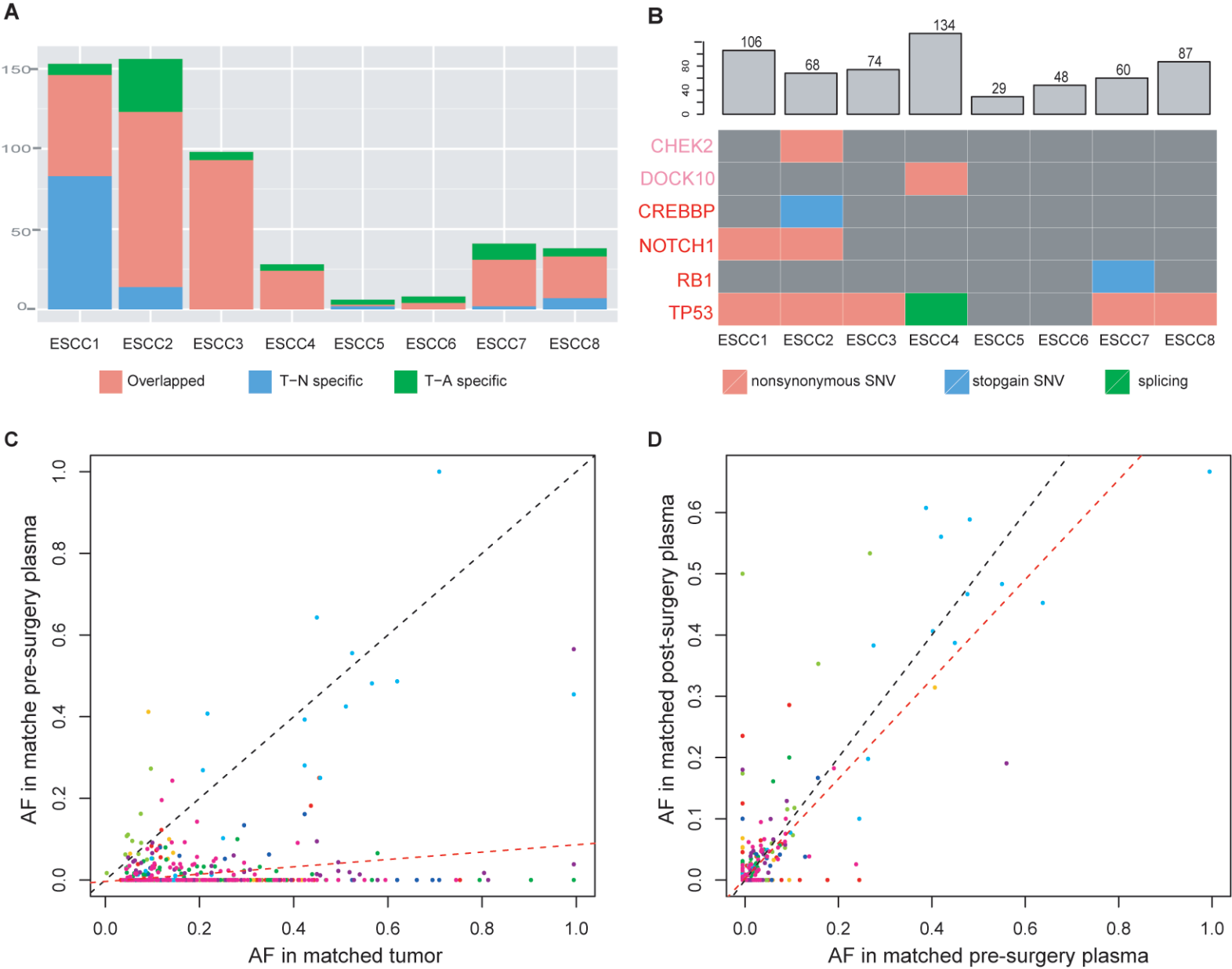
Fig 2. Somatic mutations in ESCC tissues and plasma. (A) Comparison of the number of tumor mutations by using adjacent tissues (T-A) or normal tissues (T-N) as control. (B) Distribution of functional mutations in known esophageal squamous cell carcinoma (ESCC) driver genes. Driver genes are labeled pink (reported only in one paper) or red (reported in more than one paper). (C) Comparison of mutation frequency in tumor and pre-surgery plasma. (D) Comparison of mutation frequency in pre-surgery and post-surgery plasma. Mutations from different patients are labeled in different colors. The red line is the fit-by linear model to represent real allele frequencies (AFs), while the black line represents equal AFs.

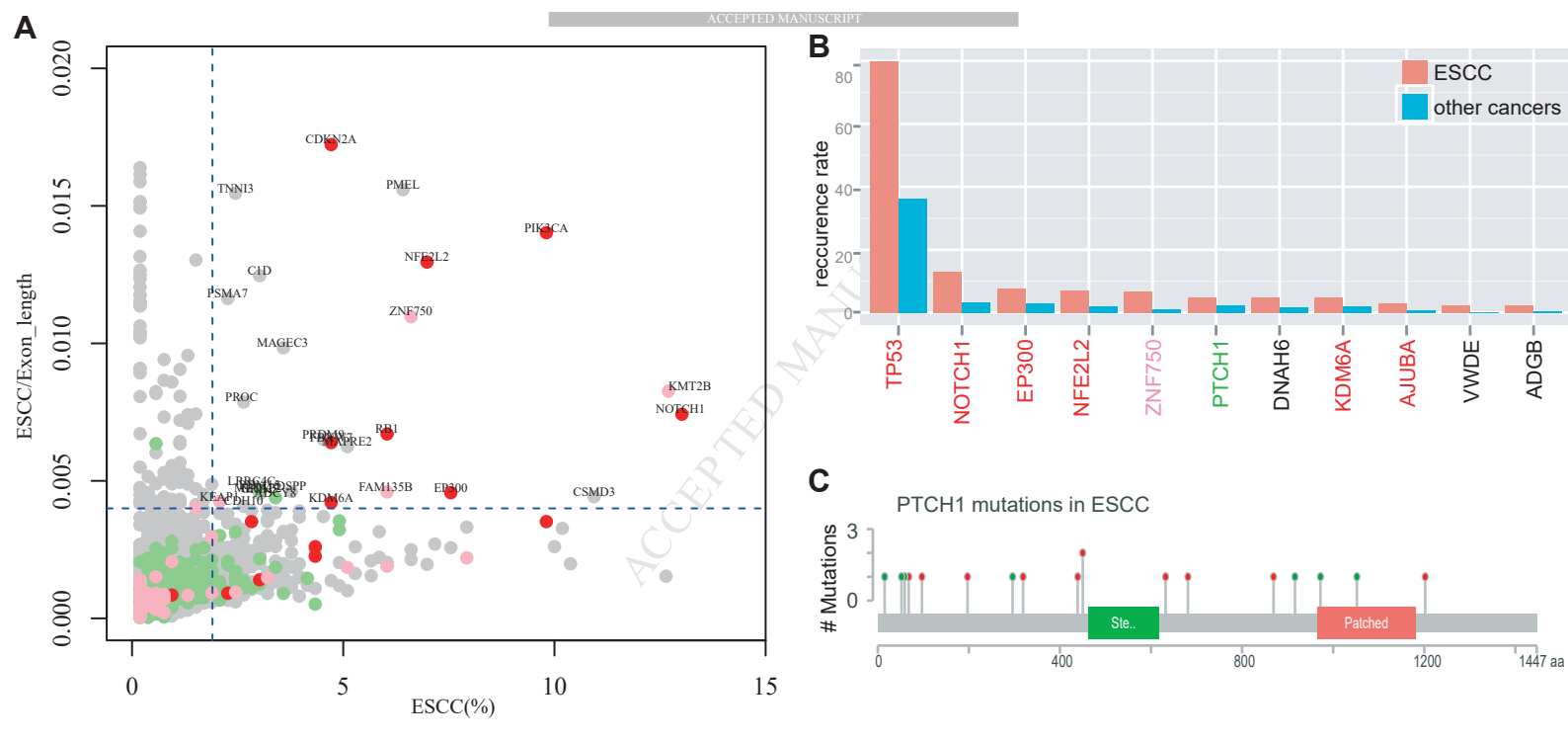
Fig 3. Integrated analysis of 532 ESCC genomes. (A) Recurrence rate and score of mutated genes. Known driver genes, having been reported in at least one paper, are labeled red or pink. Green points indicate other cancer genes collected in the Cancer Gene Census. Blue lines represent the cutoff of recurrent genes (recurrence rate of >2%, mutation score of >0.004). For a clearer display, genes having an extremely high recurrence rate or mutation score (such as *TP53*) were not plotted in this figure. (B) Genes with more mutations in ESCC than in most other cancers. (C) *PTCH1* mutation diagram in ESCC.

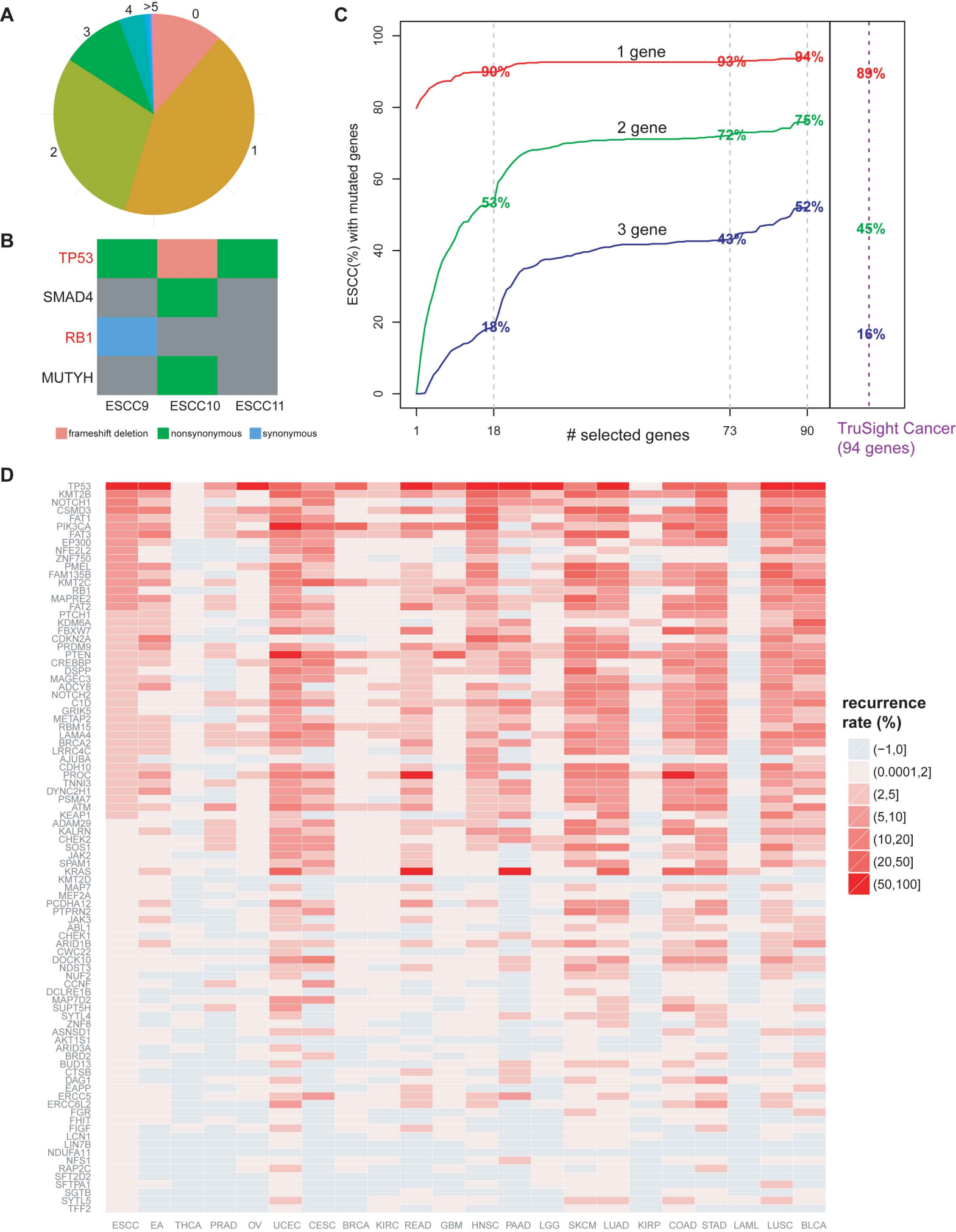
Fig 4 Target-gene panel for ESCC ctDNA identification. (A) Distribution of the number of mutated genes across all ESCC samples. Genes were selected from the

Illumina TruSight cancer panel. (B) Somatic mutations that were identified in three ESCCs by sequencing genes in the Illumina TruSight cancer panel. (C) Designing a new ESCC gene panel. Genes were gradually added based on their driver role and mutation score in ESCC. The percentage of patients with 1, 2, or 3 mutant genes was used to measure the sensitivity of the gene panel. (D) Mutation landscape of 90 selected genes in ESCC, EA and 12 other cancers.


A
normal
adjacent
tumor
esophagus
plasma
pre-surgery
plasma
post-surgery
B
Normal
Tumor
20X
400X
C
distance
0.0
0.2
0.4
0.6
ESCC4_T
ESCC4_Post
ESCC4_Pre
ESCC4_A
ESCC4_N
ESCC8_Post
ESCC8_Pre
ESCC8_T
ESCC8_A
ESCC8_N
ESCC5_Post
ESCC5_N
ESCC5_A
ESCC5_Pre
ESCC5_T
ESCC7_T
ESCC7_Post
ESCC7_Pre
ESCC7_A
ESCC7_N
ESCC6_T
ESCC6_Post
ESCC6_Pre
ESCC6_A
ESCC6_N
ESCC3_Post
ESCC3_T
ESCC3_N
ESCC3_A
ESCC1_Post
ESCC1_T
ESCC1_N
ESCC1_A
ESCC2_Pre
ESCC2_Post
ESCC2_T
ESCC2_N
ESCC2_A
ESCC1_Pre
ESCC3_Pre


A
150
100
50
0
ESCC1
ESCC2
ESCC3
ESCC4
ESCC5
ESCC6
ESCC7
ESCC8
Overlapped
T−N specific
T−A specific
B
106
68
74
134
29
48
60
87
0
40
80
CHEK2
DOCK10
CREBBP
NOTCH1
RB1
TP53
ESCC1
ESCC2
ESCC3
ESCC4
ESCC5
ESCC6
ESCC7
ESCC8
nonsynonymous SNV
stopgain SNV
splicing
C
AF in matche pre-surgery plasma
0.0
0.2
0.4
0.6
0.8
1.0
AF in matched tumor
D
AF in matched post-surgery plasma
0.0
0.1
0.2
0.3
0.4
0.5
0.6
AF in matched pre-surgery plasma




A
0
1
2
3
4
>5
B
TP53
SMAD4
RB1
MUTYH
ESCC9
ESCC10
ESCC11
frameshift deletion
nonsynonymous
synonymous
C
ESCC(%) with mutated genes
selected genes
1 gene
2 gene
3 gene
90%
93%
94%
89%
53%
72%
75%
45%
18%
43%
52%
16%
1
18
73
90
TruSight Cancer
(94 genes)
D
TP53
KMT2B
NOTCH1
CSMD3
FAT1
PIK3CA
FAT3
EP300
NFE2L2
ZNF750
PMEL
FAM135B
KMT2C
RB1
MAPRE2
FAT2
PTCH1
KDM6A
FBXW7
CDKN2A
PRDM9
PTEN
CREBBP
DSPP
MAGEC3
ADCY8
NOTCH2
C1D
GRIK5
METAP2
RBM15
LAMA4
BRCA2
LRRC4C
AJUBA
CDH10
PROC
TNNI3
DYNC2H1
PSMA7
ATM
KEAP1
ADAM29
KALRN
CHEK2
SOS1
JAK2
SPAM1
KRAS
KMT2D
MAP7
MEF2A
PCDHA12
PTPRN2
JAK3
ABL1
CHEK1
ARID1B
CWC22
DOCK10
NDST3
NUF2
CCNF
DCLRE1B
MAP7D2
SUPT5H
SYTL4
ZNF8
ASNSD1
AKT1S1
ARID3A
BRD2
BUD13
CTSB
DAG1
EAPP
ERCC5
ERCC6L2
FGR
FHIT
FIGF
LCN1
LIN7B
NDUFA11
NFS1
RAP2C
SFT2D2
SFTPA1
SGTB
SYTL5
TFF2
ESCC
EA
THCA
PRAD
OV
UCEC
CESC
BRCA
KIRC
READ
GBM
HNSC
PAAD
LGG
SKCM
LUAD
KIRP
COAD
STAD
LAML
LUSC
BLCA
recurrence
rate (%)
(−1,0]
(0.0001,2]
(2,5]
(5,10]
(10,20]
(20,50]
(50,100]