Research Paper

Low-abundance mutations in colorectal cancer patients and healthy adults

Yanfei Li[1,2], Zhengsheng Dai[3], Gang Huang[4], Yueling Jin[4], Zhongping Ning[1], Junwei Shen[1,5]

[1]Shanghai University of Medicine and Health Sciences Affiliated Zhoupu Hospital, Shanghai, China
[2]School of Medical Technology, Shanghai University of Medicine and Health Sciences, Shanghai, China
[3]Shanghai Pudong Hospital Affiliated to Fudan University, Shanghai, China
[4]Shanghai University of Medicine and Health Sciences, Shanghai, China
[5]Tongji University Affiliated Eastern Hospital, Shanghai, China

Correspondence to: Zhongping Ning, Junwei Shen; **email:** nzpwork@163.com, shenjunwei@tongji.edu.cn
Keywords: mutation, colorectal cancer, health, low-abundance
Received: September 7, 2019 **Accepted:** December 24, 2019 **Published:** January 12, 2020

ABSTRACT

Detecting low-abundance mutations is very important for cancer diagnosis and treatment. Here we describe an improved targeted sequencing analysis that dramatically increases sequencing depth. Seven colorectal cancer (CRC) patients and seven healthy adults were enrolled in this study. We examined genetic mutations in tissue samples from the central and peripheral regions of tumors from the CRC patients and in blood cells from the healthy adults. We observed that each CRC carried larger numbers of mutations more than previously estimated. These included numerous deletion mutations in the tumor tissue. While the cellular morphology in the surrounding normal colonic tissues was healthy, these cells also carried many mutations. Similarly, the blood cells from the healthy donors carried numerous mutations. These findings shed new light on the processes of tumorigenesis and aging, and also present a potentially effective method for detecting low-abundance mutations for cancer diagnosis and targeted treatments.

INTRODUCTION

Cancers result in part from the accumulation of genetic mutations, which are caused by multiple intrinsic factors, such as errors during cell proliferation and DNA repair, and extrinsic factors, such as UV light and aflatoxins [1–3]. For example, a guanine-to-cytosine transversion at amino acid 12 of *KRAS* is specifically associated with human neoplasms [4]. Similarly, mutation of *RB*, the first successfully cloned tumor suppressor gene, triggers the occurrence of retinoblastoma [5]. However, the apparent and direct causal associations between genetic mutations and cancer are infrequent; in fact, the roles of most mutations in cancer remain obscure. Although there are useful online mutation databases, such as COSMIC, it is very difficult to account for all the mutations that occur in all genes in cancer [6]. Furthermore, there may be large numbers of mutations within each gene. For instance, there are hundreds of mutations within *TP53* [7]. Additionally, compared to wild-type genes, mutant genes may acquire new functions, most of which have not been clearly identified [8, 9].

It has been hypothesized that tumorigenesis arises when normal stem cells mutate and transform into cancer stem cells [10]. After years of continuous proliferation, mutations may occur that are selected when they confer a fitness advantage for cellular adaptation and evolution [11, 12]. Ultimately, the stem cells become cancerous when they acquire critical malignant features. Although this hypothesis explains the heterogeneity of cancerous tissues, the precise steps in this process have not been fully illuminated.

Mutation detection is playing a more and more important role in the diagnosis, prevention and treatment of cancer.

For example, mutations in *KRAS* and *BRAF* are effective indicators for colorectal cancer (CRC) diagnosis and prognosis prediction [13]. Moreover, larotrectinib, a highly selective inhibitor that is widely used to treat cancer patients with mutations in Neuro Trophin Receptor Kinase (NTRK), was approved by the U.S. FDA last year [14]. Recently, a variety of effective methods have emerged for clinical mutation detection, including next-generation sequencing, third generation single-molecule sequencing, the amplification refractory mutation system (ARMS), and digital PCR [15–17]. FoundationOne CDx, for example, is a FDA-approved mutation detection kit based on targeted, massively parallel sequencing, which has been widely used in the diagnosis and treatment of cancer [18].

CRC is one of the most common and deadly cancers worldwide. Its incidence rate was ranked third in the United States in 2019 [19]. Due to the strong proliferative capacity and exposure to a complex environment, mutation rates are very high in CRC [1]. To investigate mutations in CRC, we have developed a modified method for targeted sequencing. By designing special primers, we have dramatically improved sequencing depth. In this report, we describe detection of mutant genes in cancer tissues from seven CRC patients and in blood cells from healthy donors and make comparisons between the two groups.

RESULTS

Clinical detection of seven CRC patients

To explore the tumor mutation burden (TMB) of CRC among all cancers, we analyzed TMB in The Cancer Genome Atlas (TCGA) dataset. CRC is associated with one of the highest TMBs in TCGA (Supplementary Figure 1). We also enrolled seven patients with CRC and collected their clinical data (Figure 1A, Supplementary Table 1). For tumor sample preparation, we divided the tissue into several pieces (Figure 1B). The results of hematoxylin and eosin staining showed that the cell boundaries within the tumor tissues were not clear. By contrast, cells within normal tissues maintained good cellular morphology (Figure 1C). Finally, tumor tissues were deeply immunostained for Ki67 and PCNA, which are indicative of the proliferative capacity of the cells and suggests strong cell division (Figure 1C and 1D). The PCNA protein was sharply upregulated in tumor tissues as well (Figure 1E and 1F).

Frequency of mutated genes in the seven CRC patients

To detect low-abundance mutations in CRC, we used high throughput sequencing with collected DNA samples (Figure 2A). We selected the targeted DNA sequences based on the mutation frequency and the gene's function (Figure 2B). The selected genes with a high mutation frequency and mutation hotspots were classified into five groups based to their function (Figure 2B). Because tumor sizes in CRC patients differ greatly, we used mutation data from T1 tumor tissue for further study. Interestingly, although we detected only 57 genes in total (27,976 total bases), 47 of them were mutated at least once in all patient samples, and the mutation probability in the tumors was very high (Figure 2C and Supplementary Figure 2). More than 20 mutant genes were detected in each patient sample. Moreover, the functions of the mutant genes varied widely in each patient (Supplementary Figure 2). Interestingly, the mutation frequency for several genes, including *CDH11* and *PTEN*, was quite high (Figure 2C). Because this may reflect Asian-specific genotypes rather than the mutations per se, we focused on genes with mutation frequencies <20%.

Mutation analyses of the selected genes in different samples from the seven CRC patients

Among the 57 genes tested, 5 were mutated in all seven patients, while mutations in 10 genes were not detected in the selected sequences at all (Figure 3A). Unique mutated genes were those detected in only one patient. The Venn diagram in Figure 3B shows that there the most unique mutated genes were detected in patient 7, while patient 7 carried the largest number of mutations as well (Figure 3C). We found that the frequency of deletion mutations was high in these tumor samples, especially in patient 5, whose deletion mutations exceeded 50% of all the mutations (Figure 3D). Analysis of the substitution mutations showed that A>G substitutions were the most common, followed by A>C and then C>T (Figure 3E). Comparison of the genomic changes in the central and peripheral areas of the T1 tumor samples revealed no significant differences in the mutant genes between the central and peripheral parts of the tumor samples (Figure 3F). We then compared mutant genes between matched normal and tumor tissues from the CRC patients. Intriguingly, although the morphologies of normal and tumor tissues markedly differed (Figure 1C), the differences in their mutated genes were not significant (Figure 3G).

Mutation detection in blood cells from seven healthy donors

Because numerous mutations were found in the normal tissues from CRC patients, we speculated that similar mutations may occur in other normal tissues, such as blood. We therefore enrolled seven matched healthy donors and collected their blood cells. Our subsequent

analysis revealed many mutations in these normal blood cells, with an average of about 20 mutant genes in each sample (Figure 4A). A large majority (93.5%) of the mutated genes in healthy donors were also detected in tumor tissues, suggesting that these mutations may be attributable to similar mechanisms, such as replication errors (Figure 4B). When we compared the mutated genes from the healthy donors and patients based on mutation frequency and gene function, we found that there was a large number of mutations in the healthy donors that were shared by CRC patients (Figure 4C). These results indicate that there may be many mutations in normal somatic cells in healthy people.

DISCUSSION

Cancerous tumors are heterogeneous tissues that contain a variety of cell types, including normal stem cells, supporting cells, immune cells, endothelial cells and diverse tumor cells with different mutations [20, 21]. Consequently, we need to be able to detect low-abundance mutations within tumor tissues, as some tumor cells are in low abundance at the early stages of cancer [21]. Detecting these low-abundance mutations is very important for cancer diagnosis and targeted treatment. Targeted sequencing has been an effective method for detecting low-abundance mutations in

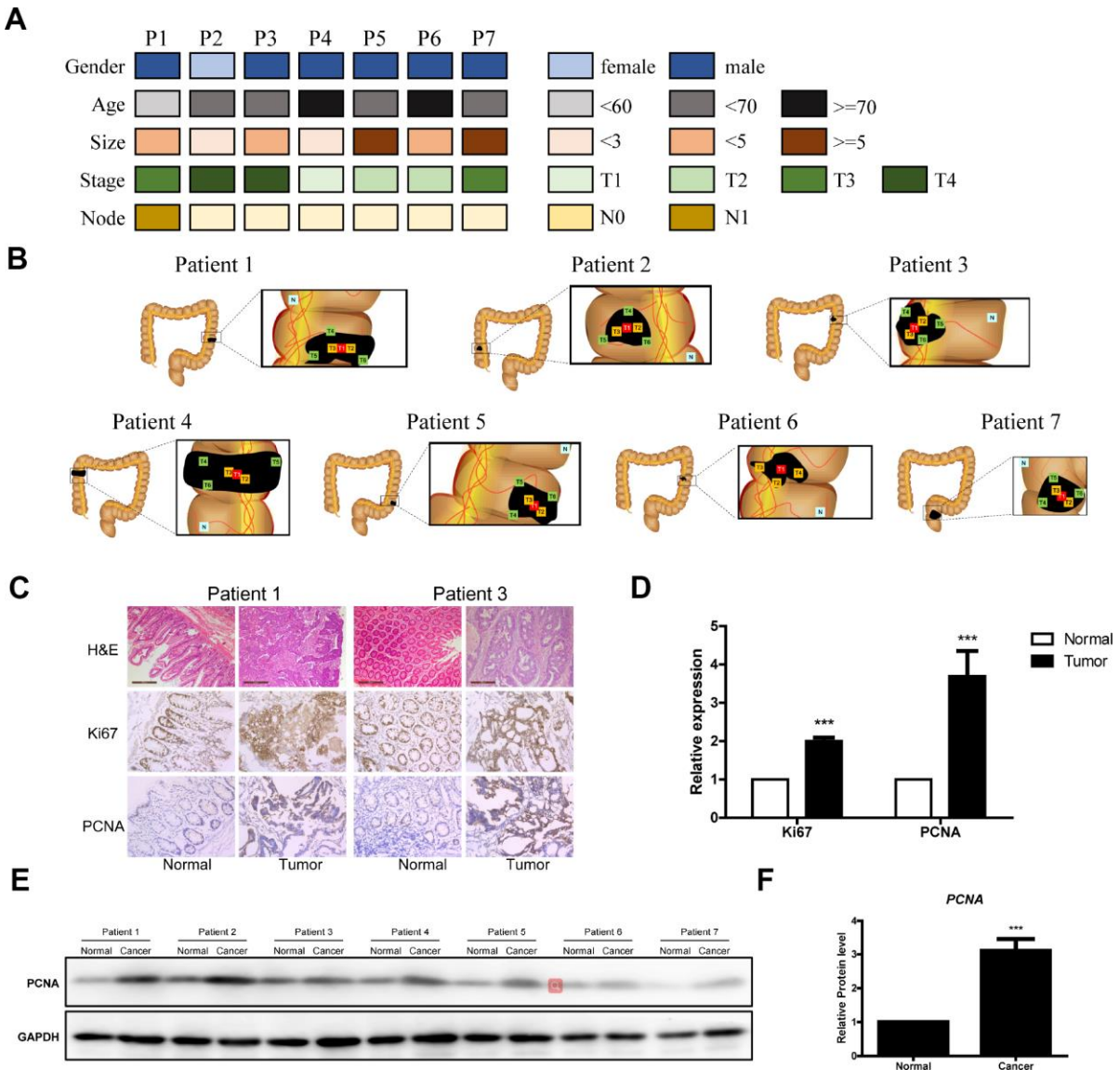


Figure 1. Clinical data of the seven CRC patients. (**A**) Clinical analyses of seven CRC patients. (**B**) Schematic diagram of the locations of the tumor tissues from the seven patients. (**C**) Representative images of H&E Staining, Ki67 immunostaining and PCNA immunostaining of samples from the seven patients. (**D**) Data analysis of Ki67 immunostaining and PCNA immunostaining of tumor tissue sections. (**E** and **F**) Western blot of PCNA in tumor tissue sections.

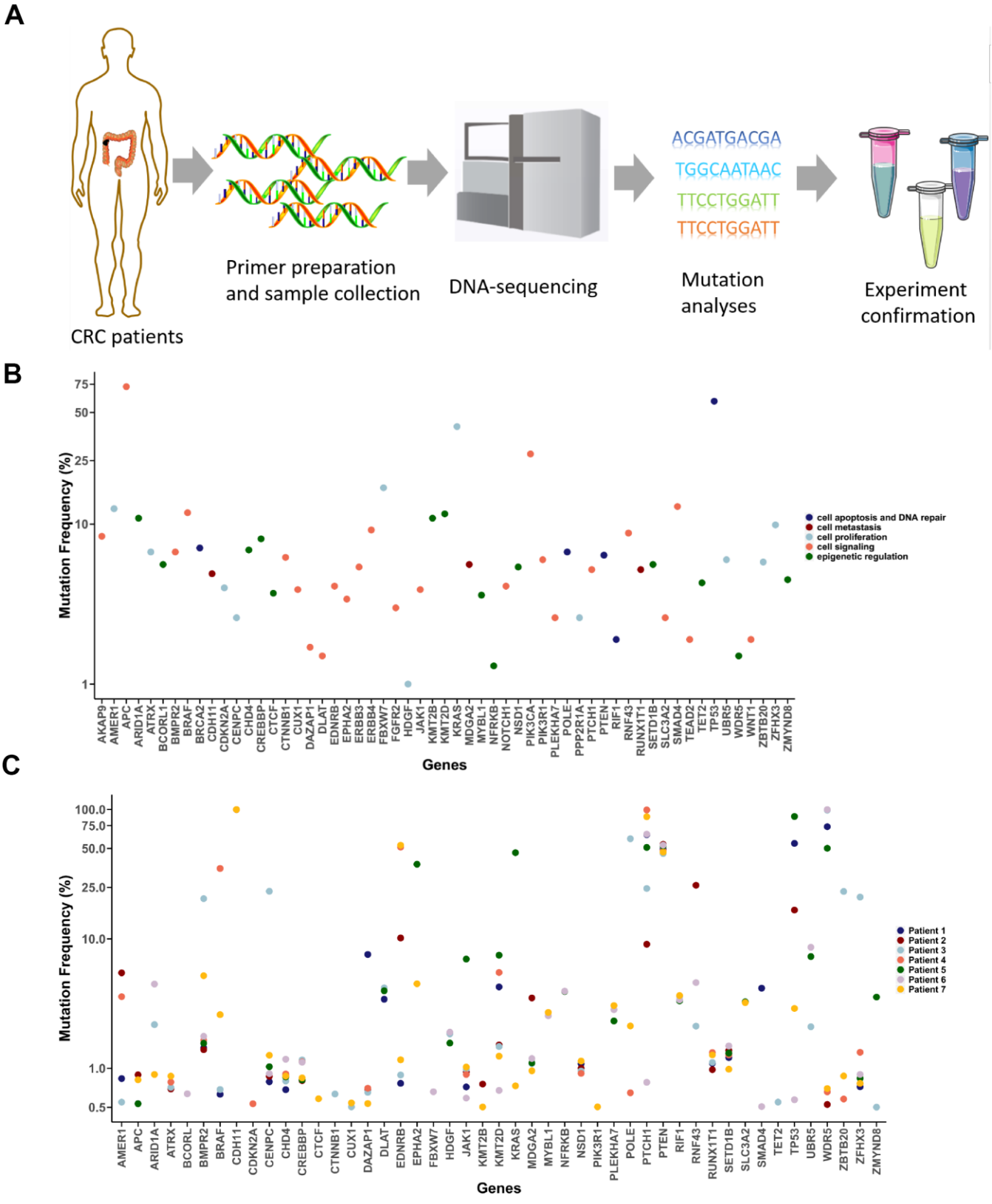


Figure 2. Detection of cancer-associated mutations in T1 cancer tissues. (**A**) Schematic diagram of the mutation detection method. (**B**) Selected CRC-associated genes. Mutation frequency data were acquired from the CRC project of TCGA. (**C**) Mutation detection in T1 tumor tissues from the seven CRC patients.

cancer studies and in clinical applications [22]. We improved upon this method and increased the depth of sequencing by at least 10-fold, greatly enhancing the detection sensitivity for low-abundance mutations. For example, *EGFR*, which is usually mutated within exons 19 and 21 is the gene targeted by Gefitinib in non-small cell lung cancer. [23, 24]. However, after 5 months of treatment, 50% of patients will carry the T790M (ACG>ATG) mutation and need further Osimertinib treatment [25]. We propose that these patients carrying the high-abundance *EGFR* mutation in exons 19 and 21, may also carry a low-abundance T790M EGFR mutation. With the method described herein, the low-abundance EGFR T790M mutation can be detected earlier. It would therefore be of interest to test this speculation through further experimentation and clinical exploration.

Mutation accumulation is thought to be one of the most important factors contributing to tumorigenesis [26]. While the results of numerous studies are consistent with that idea, the precise steps in the process remain unclear [27, 28]. It has been reported that cancers derive from a single cancer stem cell through mutation [29]. These cancer stem cells then continuously proliferate, and the mutations accumulate [30]. Over several years, tumor cells acquire new functions, such as anti-apoptosis, increased migration, and immune resistance [31]. Interestingly, we found that there were no marked differences in the mutations between normal tissue and CRC tissue. That is, normal tissues contained more mutations than we expected, which is consistent with the idea that that mutations accumulate in human adult stem cells throughout life [1, 32]. Based on our findings and those of others, we speculate that in the sea of

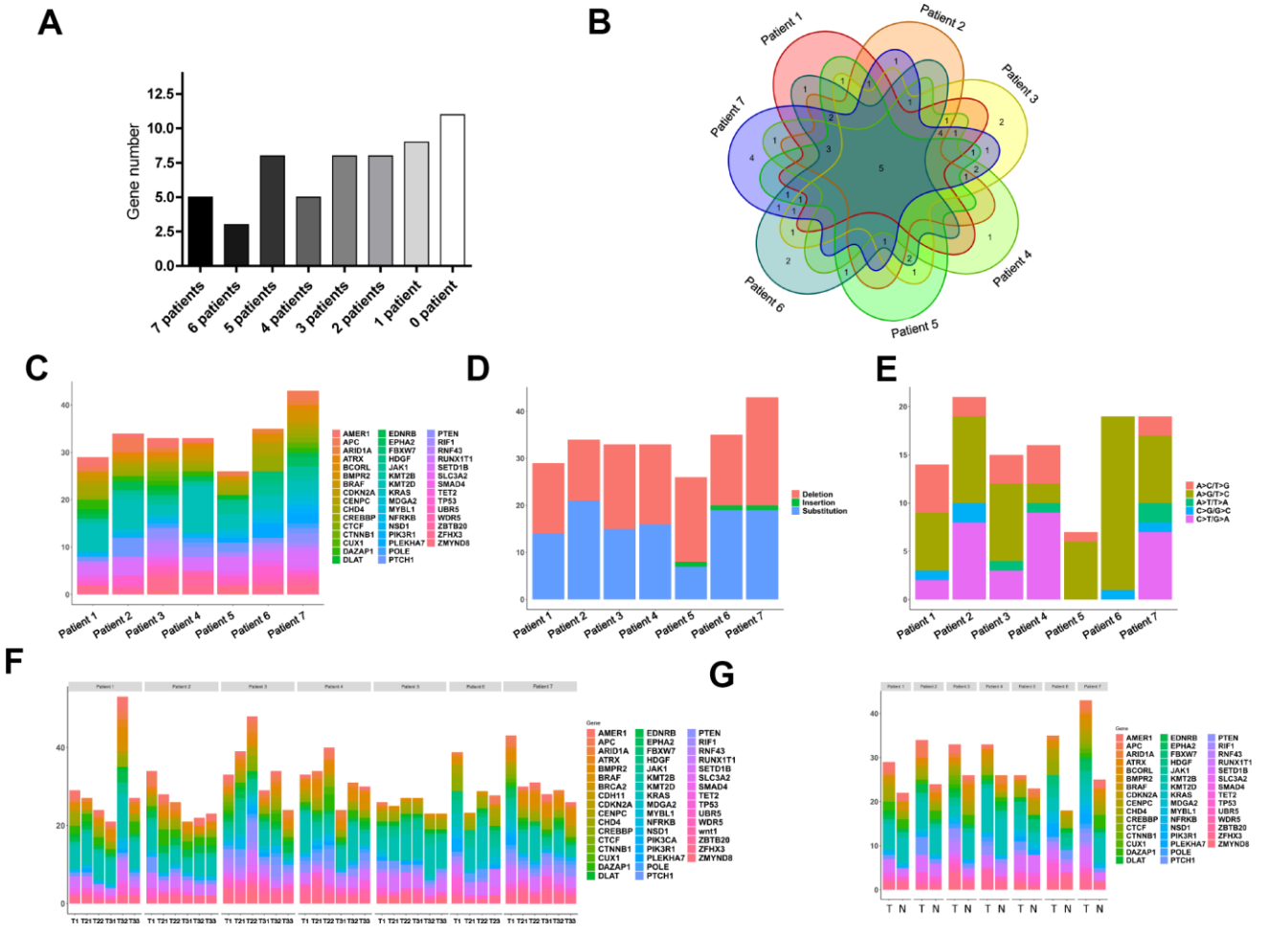


Figure 3. Widespread cancer-associated mutations in tumor and normal tissues. (**A**) Distribution of mutations in genes we selected in the seven CRC patients; the grayscale reflects the frequency of mutations. (**B**) Wayne diagram of the distribution of mutated genes in patients. (**C**) Distribution of mutated genes in the seven CRC patients. (**D**) Distribution of deletion and substitution mutations in the seven patients. (**E**) Distribution of subdivided substitutions in the seven patients. (**F**) Distributions of mutant genes in tissue samples from the central peripheral regions of the tumors in the seven CRC patients. (**G**) Distributions of mutant genes in samples from T1 tumors and normal colonic tissue from the seven CRC patients.

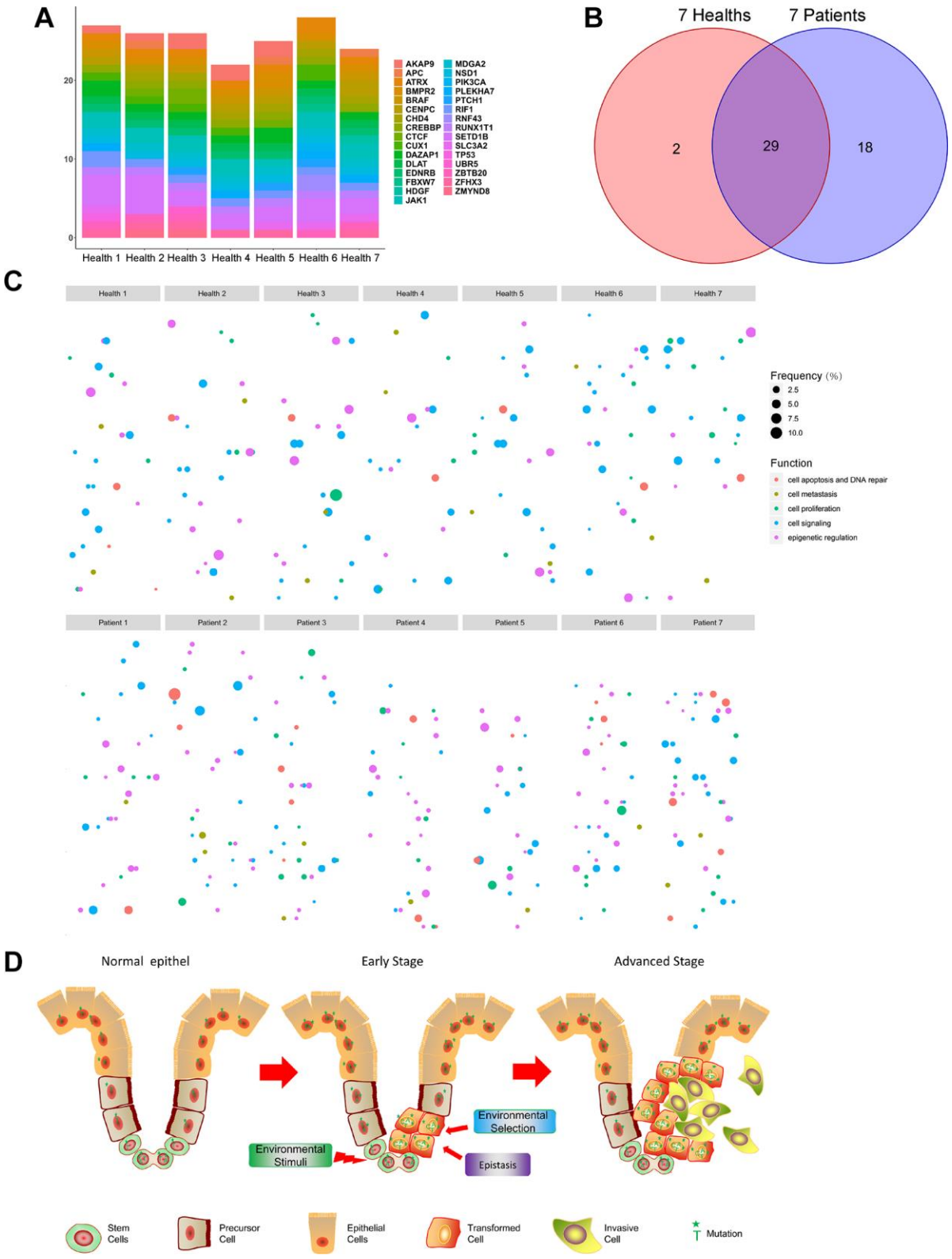


Figure 4. Detecting mutations in normal blood cells from healthy donors. (**A**) Distribution of selected mutant genes in the seven healthy donors. (**B**) Wayne diagram of the distribution of mutated genes in CRC patients and healthy donors. (**C**) Scatter plot of the mutated genes in CRC patients and healthy donors. (**D**) Schematic diagram. Stem cells in tissues proliferate over years and accumulate mutations. The majority of cells (including stem cells) carry mutant genes (Left). DNA replication and repair and environmental factors (UV, radiation, etc.) lead to the transformation of mutant stem cells into cancer stem cells, while environmental selection and epistasis influence mutations in the tumor (Middle). After continuous iterative selection, cancer stem cells eventually develop into malignant tumors (Right).

mutations, some cells suddenly obtain a decisive change, such as uncontrolled proliferation, due to environmental stimuli or an intrinsic mutation. The resultant malignant cells then begin to proliferate and quickly evolve (Figure 4D).

Aging is an eternal theme in medicine and health. Immune cells, secreted communication factors, and the shortening of telomeres partially explain this phenomenon [33, 34]. In addition, somatic tissues such as the normal human esophagus accumulate mutant clones with age [1]. Here we revealed that there are large numbers of mutations in the blood cells of healthy donors. Our present findings, together with earlier work [1], suggest that the accumulation of mutations in somatic cells with age is another important factor in aging, and it will be very interesting to identify the mechanisms underlying these mutations. Perturbation of DNA duplication and repair could cause accumulation of mutations [35, 36]. For instance, DNA lesions such as guanine N7 alkylation alter guanine hydrogen-bonding patterns in duplex DNA and inhibit DNA duplication [37]. These studies and our work partially explain the relationships among aging, mutations and tumors. However, the associations between mutations and other chronic diseases associated with aging, such as diabetes, cardiovascular disease, and neurodegenerative diseases, deserve further study.

Organ transplantation or regeneration is the ultimate treatment for many incurable diseases, such as liver, kidney, and heart failure [38, 39]. Induced pluripotent stem cell technology is thought to be a promising method for organ regeneration [40, 41]. However, unpredictable and uncontrollable carcinogenesis is a key problem with this method [42]. One possible reason is that the mixed cells with mosaic mutations may contain precancerous cells that can become malignant when exposed to the four iPS factors (Oct4, Sox2, Klf4 and c-Myc). It is therefore essential to detect mutations before iPS induction. Moreover, because the number of somatic cells carrying mutations increases with age, it may be necessary to store stem cells at a young age.

There are several limitations to this study. Though we have found the numbers of low-abundance mutations in CRC patients, the patient number is small, especially for various patients with different stages, different sexes and different age. Large-scale clinical research should be conducted to reveal more functions of low-abundance mutations in CRC. Besides clinical indicators, the backgrounds of the patients, such as ethnicity and eating habit, have important roles in the development of colorectal CRC [43]. It is of interest to study the roles of these background in low-abundance mutations. The tumorigenesis of CRC is a long-time period and illustrating the function of low-abundance mutations during the period is very important. Due to CRC's slow development and the capacity for early diagnosis, tissue samples such as colorectal polyps and precancerous lesions could be collected during different stages of the disease [44]. Thus, it is very valuable to unveil a hidden world of low-abundance mutation and CRC tumorigenesis through large-scale sequencing.

In summary, we used a new targeted sequencing method to detect mutations in CRC tissues and matched normal colonic tissues. This approach also enabled us to reveal that there are numerous mutations in blood cells from healthy donors. These findings not only shed new light on the mechanism of tumorigenesis, but also revealed the universality of somatic cell mutation with aging. This method has a very high sensitivity for low-abundance mutations and may have potential applications for the diagnosis and treatment of tumors.

MATERIALS AND METHODS

Patients and samples

For this study, seven CRC patients and seven healthy adults at Zhoupu Hospital (Shanghai, China) were enrolled between July 2017 and July 2018. All samples were obtained with written informed consent at study entry. All methods used in this study were approved by the Research Medical Ethics Committee of Zhoupu Hospital. The tumor tissues and the matched normal tissues were collected as shown in Figure 1B. The freshly collected tissue samples were placed in cryogenic storage tubes (Thermo Fisher Scientific) and immediately stored in liquid nitrogen (Thermo Fisher Scientific). Blood samples from healthy adults were collected into EDTA anticoagulant tubes and then placed in cryogenic storage tubes in liquid nitrogen as well. Genomic DNA was extracted from frozen tumor tissues using a standard protocol (FG304, Shanghai Finegene Biotech, Shanghai, China). The quantity and quality of the DNA was assessed using NanoDrop 2000 (Thermo Fisher Scientific).

Primer design for targeted sequence

To detect target regions, the reference regions of 57 genes are necessary for PCR [45]. The sequences of all these regions were downloaded from the UCSC Genome Browser (hg38; http://genome.ucsc.edu/). We designed the specific blunt hairpin primers of multiplex PCR as described previously [46, 47]. These primers contain two parts (Supplementary Figure 3). The first is the 3'-end containing about 20 bp, which constitutes the targeted region and has a melting temperature of 60-65°C (Supplementary Table 2). The second part is a

universal sequence (18 bp), which enables the PCR products to be amplified by adapter primers to construct the sequencing libraries. The sizes of PCR product are around 200 bp. We evaluated the specificity of these primers as previously described [48].

Adapter primers design

To construct the sequencing libraries with the different samples, we designed unique adapter primers containing Ion Torrent primer. These primers contained ten-base index sequences and universal sequences. All primers were synthesized by Biowing Applied (Shanghai, China). The adapter primers were HPLC purified and were supplied at standard desalting grade by Biowing Applied.

Multiplex PCR

To amplify the 200-bp target region, the PCR mixture (40 µl) contained 1× reaction buffer (HotStart DNA polymerase (1.2 U), 2 mM $MgCl_2$, 200 µM each dNTP and 2× KAPA2G Robust HotStart ReadyMix (Kapa Biosystems, USA), 1 µM each primer (), and 50 ng of genomic DNA. The PCR cycling protocol entailed: 95°C for 3 min and 20 cycles of 94°C for 15 s and 60°C for 4 min.

Library construction

For the library construction, the PCR mixture (10 µl) contained: 1× reaction buffer (2 mM $MgCl_2$, 0.4 U of KAPA2G Robust HotStart DNA polymerase), 200 µM each dNTP and 2× KAPA2G Robust HotStart ReadyMix (Kapa Biosystems, USA), 1 µM adapter primer, and 2 µl of template, which yielded the first round PCR products. The PCR cycling protocol entailed: 95°C for 3 min; 10 cycles of 94°C for 20 s, 65°C for 2 min, 72°C for 30 s; and 72°C for 5 min. The PCR products were purified using a TIANgel Midi Purification Kit (TIANGEN BIOTECH, Beijing, China).

Ion torrent PGM sequencing

The sequencing was performed by Biowing Applied. Briefly, the purified PCR products were sequenced on a PGM platform using commercially available protocols and were then processed on a OneTouch 2 instrument. After enrichment using a OneTouch 2 ES station, the products were sequenced using a 318 chip and an Ion PGMTM Sequencing 200 Kit v2 [47].

Sequencing data analysis

We separated all the sequencing reads according to their index combination information using a FASTX-Toolkit with the parameter that mismatch bases were less than 1 bp. We then trimmed out the index and adapter sequences using cutadapt software to generate target sequences for each sample. Thereafter, sequencing reads were mapped against the reference genome (hg38; http://genome.ucsc.edu/). For quality control, the filtration cutoff used a Phred-type quality score of Q20 (QPhred = 20) and the amplicons with > 5,000 sequencing depths were excluded in order to avoid extreme read depths [49, 50]. The detail of the sequencing such as the total number of reads were shown in Supplementary Table 3.

H&E Staining

Hematoxylin and eosin (H&E) staining was performed to evaluate the histopathological structure of CRC tumor tissue. The samples were fixed in formaldehyde (10%), embedded in paraffin, cut into 4-µm-thick sections, stained with hematoxylin solution for 5 min. This was following by 5 dips in acid ethanol (1% HCl in 70% ethanol). After rinsing the sections in distilled water, they were stained with eosin solution for 5 min, dehydrated through a graded alcohol series, clearing in xylene, and finally mounted in neutral gum. The samples were observed under a Nikon fluorescence microscope (Tokyo, Japan).

Immunohistochemical Staining

For immunohistochemistry, paraffin-embedded 4-µm sections first kept at 60°C for 24 h, deparaffinized with xylene and hydrating through an ethanol gradient (100%-70%), and incubated for 30 min in an antigen retrieval solution (Hangzhou Huabio Biotechnology Company; Hangzhou, China) and 3% H_2O_2. After being rinsed with water, the sections were incubated with the primary antibody (anti-Ki67 (1:50 dilution; ab197234, Abcam, Cambridge, UK) overnight at 4°C. They were then rinsed and incubated for 30 min with the secondary antibody (Hangzhou Huabio Biotechnology Company; Hangzhou, China) followed by 3,3′-diaminobenzidine (DAB) and hematoxylin. The immunostained sections were observed under a Nikon fluorescence microscope (Tokyo, Japan).

Western blot

The protein of PCNA in the normal and tumor tissues were analyzed by Western blot. Briefly, the tissues were treated with RIPA buffer (high) (Solarbio, Beijing, China) in ice for 10min and were lysed with pestle in ice for 5 min. Total lysates were centrifuged and the protein concentration of the supernatant was tested with a BCA protein assay kit (Beyondtime, Nantong, China). 20 µg of total proteins were separated by SDS-PAGE and

electro-blotted onto a PVDF membrane (Merck KGaA). The membranes were blocked with 3% BSA for 1 h at room temperature, and incubated with different primary antibodies: anti-PCNA (#ab18197; Abcam, Cambridge, UK) and anti-GAPDH (#ab181602; Abcam). After incubated with the appropriate secondary antibodies for 1 h at room temperature, the bands were visualized using an ECL assay kit (Amersham Pharmacia Biosciences) and observed with an LAS3000® Luminescent image analyzer.

AUTHOR CONTRIBUTIONS

Funding support: J. Shen, Z. Ning; Conception and design: J. Shen; Development of methodology: Y. Li, Z. Dai, J. Shen; Acquisition of data: Y. Li, G. Huang, Z. Ning; Analysis and interpretation of data: Y. Li, Y. Jin, J. Shen; Writing, review, and revision of the manuscript: Y. Li, J. Shen; Administrative, technical, or material support: Y. Li, Y. Jin.

ACKNOWLEDGEMENTS

We thank all patients and blood sample donors for participating in the clinical study at Zhoupu hospital.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

FUNDING

This study is supported by National Nature Science Foundation of China (11602295); Pandeng Plan, the Special Program for Collaborative Innovation and Bairenku project of Shanghai University of Medicine and Health Sciences (A1-2601-19-311308, SPCI-17-17-001, B1-0200-19-311145).

REFERENCES

1. Martincorena I, Fowler JC, Wabik A, Lawson AR, Abascal F, Hall MW, Cagan A, Murai K, Mahbubani K, Stratton MR, Fitzgerald RC, Handford PA, Campbell PJ, et al. Somatic mutant clones colonize the human esophagus with age. Science. 2018; 362:911–17. https://doi.org/10.1126/science.aau3879 PMID:30337457

2. Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, Dawson KJ, Ju YS, Iorio F, Tubio JMC, Koh CC, Georgakopoulos-Soares I, Rodríguez-Martín B, et al. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. Cell. 2019; 176:1282–1294.e20. https://doi.org/10.1016/j.cell.2019.02.012 PMID:30849372

3. Kucab JE, Zou X, Morganella S, Joel M, Nanda AS, Nagy E, Gomez C, Degasperi A, Harris R, Jackson SP, Arlt VM, Phillips DH, Nik-Zainal S. A Compendium of Mutational Signatures of Environmental Agents. Cell. 2019; 177:821–836.e16. https://doi.org/10.1016/j.cell.2019.03.001 PMID:30982602

4. Santos E, Martin-Zanca D, Reddy EP, Pierotti MA, Della Porta G, Barbacid M. Malignant activation of a K-ras oncogene in lung carcinoma but not in normal tissue of the same patient. Science. 1984; 223:661–64. https://doi.org/10.1126/science.6695174 PMID:6695174

5. Squire J, Goddard AD, Canton M, Becker A, Phillips RA, Gallie BL. Tumour induction by the retinoblastoma mutation is independent of N-myc expression. Nature. 1986; 322:555–57. https://doi.org/10.1038/322555a0 PMID:2426601

6. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, Stefancsik R, Harsha B, Kok CY, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017; 45:D777–83. https://doi.org/10.1093/nar/gkw1121 PMID:27899578

7. Soussi T, Wiman KG. TP53: an oncogene in disguise. Cell Death Differ. 2015; 22:1239–49. https://doi.org/10.1038/cdd.2015.53 PMID:26024390

8. Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, Yamamoto S, Tatsuno K, Katoh H, Watanabe Y, Ichimura T, Ushiku T, Funahashi S, et al. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. Nat Genet. 2014; 46:583–87. https://doi.org/10.1038/ng.2984 PMID:24816255

9. Zmajkovic J, Lundberg P, Nienhold R, Torgersen ML, Sundan A, Waage A, Skoda RC. A gain-of-function mutation in EPO in familial erythrocytosis. N Engl J Med. 2018; 378:924–30. https://doi.org/10.1056/NEJMoa1709064 PMID:29514032

10. Rycaj K, Tang DG. Cell-of-Origin of Cancer versus Cancer Stem Cells: assays and Interpretations. Cancer Res. 2015; 75:4003–11. https://doi.org/10.1158/0008-5472.CAN-15-0798 PMID:26292361

11. Bratulic S, Toll-Riera M, Wagner A. Mistranslation can enhance fitness through purging of deleterious mutations. Nat Commun. 2017; 8:15410. https://doi.org/10.1038/ncomms15410 PMID:28524864

12. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007; 8:610–18. https://doi.org/10.1038/nrg2146 PMID:17637733

13. De Roock W, Claes B, Bernasconi D, De Schutter J, Biesmans B, Fountzilas G, Kalogeras KT, Kotoula V, Papamichael D, Laurent-Puig P, Penault-Llorca F, Rougier P, Vincenzi B, et al. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. Lancet Oncol. 2010; 11:753–62. https://doi.org/10.1016/S1470-2045(10)70130-3 PMID:20619739

14. Scott LJ. Larotrectinib: First Global Approval. Drugs. 2019; 79:201–06. https://doi.org/10.1007/s40265-018-1044-x PMID:30635837

15. Segovia R, Tam AS, Stirling PC. Dissecting genetic and environmental mutation signatures with model organisms. Trends Genet. 2015; 31:465–74. https://doi.org/10.1016/j.tig.2015.04.001 PMID:25940384

16. Camunas-Soler J, Lee H, Hudgins L, Hintz SR, Blumenfeld YJ, El-Sayed YY, Quake SR. Noninvasive Prenatal Diagnosis of Single-Gene Disorders by Use of Droplet Digital PCR. Clin Chem. 2018; 64:336–45. https://doi.org/10.1373/clinchem.2017.278101 PMID:29097507

17. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. 1989; 17:2503–16. https://doi.org/10.1093/nar/17.7.2503 PMID:2785681

18. Ready N, Hellmann MD, Awad MM, Otterson GA, Gutierrez M, Gainor JF, Borghaei H, Jolivet J, Horn L, Mates M, Brahmer J, Rabinowitz I, Reddy PS, et al. First-Line Nivolumab Plus Ipilimumab in Advanced Non-Small-Cell Lung Cancer (CheckMate 568): Outcomes by Programmed Death Ligand 1 and Tumor Mutational Burden as Biomarkers. J Clin Oncol. 2019; 37:992–1000. https://doi.org/10.1200/JCO.18.01042 PMID:30785829

19. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019; 69:7–34. https://doi.org/10.3322/caac.21551 PMID:30620402

20. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017; 168:613–28. https://doi.org/10.1016/j.cell.2017.01.018 PMID:28187284

21. Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. Mol Cancer. 2017; 16:41. https://doi.org/10.1186/s12943-017-0600-4 PMID:28209166

22. Nikiforova MN, Wald AI, Roy S, Durso MB, Nikiforov YE. Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer. J Clin Endocrinol Metab. 2013; 98:E1852–60. https://doi.org/10.1210/jc.2013-2292 PMID:23979959

23. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004; 350:2129–39. https://doi.org/10.1056/NEJMoa040938 PMID:15118073

24. Arulananda S, Do H, Musafer A, Mitchell P, Dobrovic A, John T. Combination Osimertinib and Gefitinib in C797S and T790M EGFR-Mutated Non-Small Cell Lung Cancer. J Thorac Oncol. 2017; 12:1728–32. https://doi.org/10.1016/j.jtho.2017.08.006 PMID:28843359

25. Fujii A, Harada T, Iwama E, Ota K, Furuyama K, Ijichi K, Okamoto T, Okamoto I, Takayama K, Nakanishi Y. Hypermethylation of the CpG dinucleotide in epidermal growth factor receptor codon 790: implications for a mutational hotspot leading to the T790M mutation in non-small-cell lung cancer. Cancer Genet. 2015; 208:271–78. https://doi.org/10.1016/j.cancergen.2014.12.005 PMID:25682017

26. Fujii M, Shimokawa M, Date S, Takano A, Matano M, Nanki K, Ohta Y, Toshimitsu K, Nakazato Y, Kawasaki K, Uraoka T, Watanabe T, Kanai T, Sato T. A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. Cell Stem Cell. 2016; 18:827–38. https://doi.org/10.1016/j.stem.2016.04.003 PMID:27212702

27. Notta F, Hahn SA, Real FX. A genetic roadmap of pancreatic cancer: still evolving. Gut. 2017; 66:2170–78. https://doi.org/10.1136/gutjnl-2016-313317 PMID:28993418

28. Vogelstein B, Kinzler KW. The Path to Cancer—Three Strikes and You're Out. N Engl J Med. 2015; 373:1895–98.

https://doi.org/10.1056/NEJMp1508811 PMID:26559569

29. Akbari-Birgani S, Paranjothy T, Zuse A, Janikowski T, Cieślar-Pobuda A, Likus W, Urasińska E, Schweizer F, Ghavami S, Klonisch T, Łos MJ. Cancer stem cells, cancer-initiating cells and methods for their detection. Drug Discov Today. 2016; 21:836–42. https://doi.org/10.1016/j.drudis.2016.03.004 PMID:26976692

30. Magee JA, Piskounova E, Morrison SJ. Cancer stem cells: impact, heterogeneity, and uncertainty. Cancer Cell. 2012; 21:283–96. https://doi.org/10.1016/j.ccr.2012.03.003 PMID:22439924

31. Vlashi E, Pajonk F. Cancer stem cells, cancer cell plasticity and radiation therapy. Semin Cancer Biol. 2015; 31:28–35. https://doi.org/10.1016/j.semcancer.2014.07.001 PMID:25025713

32. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, Nijman IJ, Martincorena I, Mokry M, et al. Tissue-specific mutation accumulation in human adult stem cells during life. Nature. 2016; 538:260–64. https://doi.org/10.1038/nature19768 PMID:27698416

33. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013; 153:1194–217. https://doi.org/10.1016/j.cell.2013.05.039 PMID:23746838

34. Shay JW. Role of Telomeres and Telomerase in Aging and Cancer. Cancer Discov. 2016; 6:584–93. https://doi.org/10.1158/2159-8290.CD-16-0062 PMID:27029895

35. Deng W, Henriet S, Chourrout D. Prevalence of Mutation-Prone Microhomology-Mediated End Joining in a Chordate Lacking the c-NHEJ DNA Repair Pathway. Curr Biol. 2018; 28:3337–3341.e4. https://doi.org/10.1016/j.cub.2018.08.048 PMID:30293719

36. Kou Y, Koag MC, Lee S. Structural and Kinetic Studies of the Effect of Guanine N7 Alkylation and Metal Cofactors on DNA Replication. Biochemistry. 2018; 57:5105–16. https://doi.org/10.1021/acs.biochem.8b00331 PMID:29957995

37. Kou Y, Koag MC, Lee S. N7 methylation alters hydrogen-bonding patterns of guanine in duplex DNA. J Am Chem Soc. 2015; 137:14067–70. https://doi.org/10.1021/jacs.5b10172 PMID:26517568

38. Neuberger J, Trotter P, Stratton R. Organ transplantation rates in the UK. BMJ. 2017; 359:j5218. https://doi.org/10.1136/bmj.j5218 PMID:29133305

39. Rana A, Gruessner A, Agopian VG, Khalpey Z, Riaz IB, Kaplan B, Halazun KJ, Busuttil RW, Gruessner RW. Survival benefit of solid-organ transplant in the United States. JAMA Surg. 2015; 150:252–59. https://doi.org/10.1001/jamasurg.2014.2038 PMID:25629390

40. Hu C, Li L. In vitro culture of isolated primary hepatocytes and stem cell-derived hepatocyte-like cells for liver regeneration. Protein Cell. 2015; 6:562–74. https://doi.org/10.1007/s13238-015-0180-2 PMID:26088193

41. Takagi R, Ishimaru J, Sugawara A, Toyoshima KE, Ishida K, Ogawa M, Sakakibara K, Asakawa K, Kashiwakura A, Oshima M, Minamide R, Sato A, Yoshitake T, et al. Bioengineering a 3D integumentary organ system from iPS cells using an in vivo transplantation model. Sci Adv. 2016; 2:e1500887. https://doi.org/10.1126/sciadv.1500887 PMID:27051874

42. Nagoshi N, Okano H. iPSC-derived neural precursor cells: potential for cell transplantation therapy in spinal cord injury. Cell Mol Life Sci. 2018; 75:989–1000. https://doi.org/10.1007/s00018-017-2676-9 PMID:28993834

43. Song M, Garrett WS, Chan AT. Nutrients, foods, and colorectal cancer prevention. Gastroenterology. 2015; 148:1244–60.e16. https://doi.org/10.1053/j.gastro.2014.12.035 PMID:25575572

44. Williams C, DiLeo A, Niv Y, Gustafsson JA. Estrogen receptor beta as target for colorectal cancer prevention. Cancer Lett. 2016; 372:48–56. https://doi.org/10.1016/j.canlet.2015.12.009 PMID:26708506

45. Dahl F, Gullberg M, Stenberg J, Landegren U, Nilsson M. Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. Nucleic Acids Res. 2005; 33:e71. https://doi.org/10.1093/nar/gni070 PMID:15860768

46. Chen K, Dong SS, Wu N, Wu ZH, Zhou YX, Li K, Zhang F, Xiao JH. A novel multiplex fluorescent competitive PCR for copy number variation detection. Genomics. 2019; 111:1745–1751. https://doi.org/10.1016/j.ygeno.2018.11.029 PMID:30529537

47. Chen K, Zhou YX, Li K, Qi LX, Zhang QF, Wang MC, Xiao JH. A novel three-round multiplex PCR for SNP

genotyping with next generation sequencing. Anal Bioanal Chem. 2016; 408:4371–77. https://doi.org/10.1007/s00216-016-9536-6 PMID:27113460

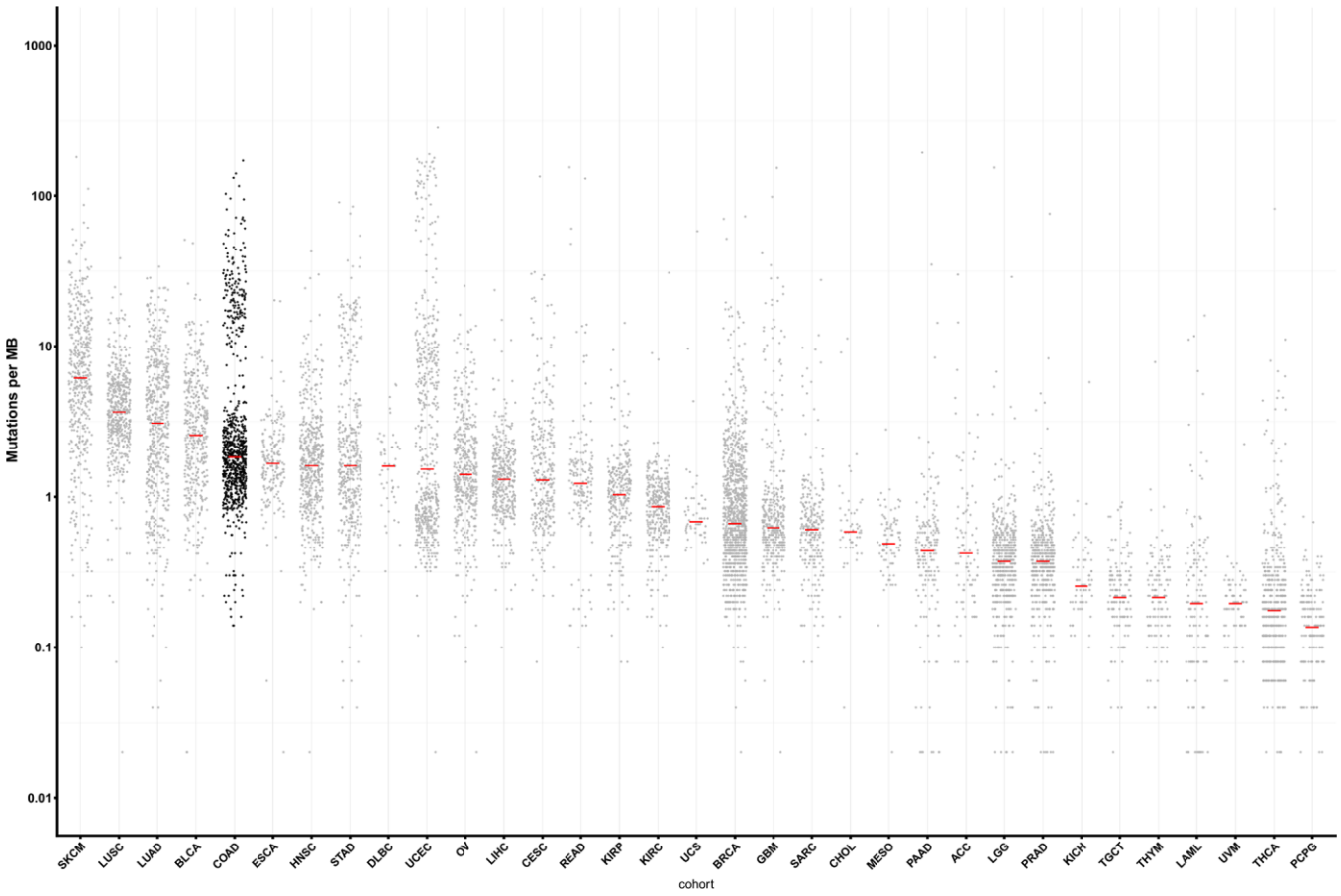
48. Chen K, Zhou YX, Li K, Xiao JH. Multiplex PCR with the blunt hairpin primers for next generation sequencing. Biotechnol Bioprocess Eng. 2017; 22:347–51. https://doi.org/10.1007/s12257-017-0133-0

49. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–98. https://doi.org/10.1038/ng.806 PMID:21478889

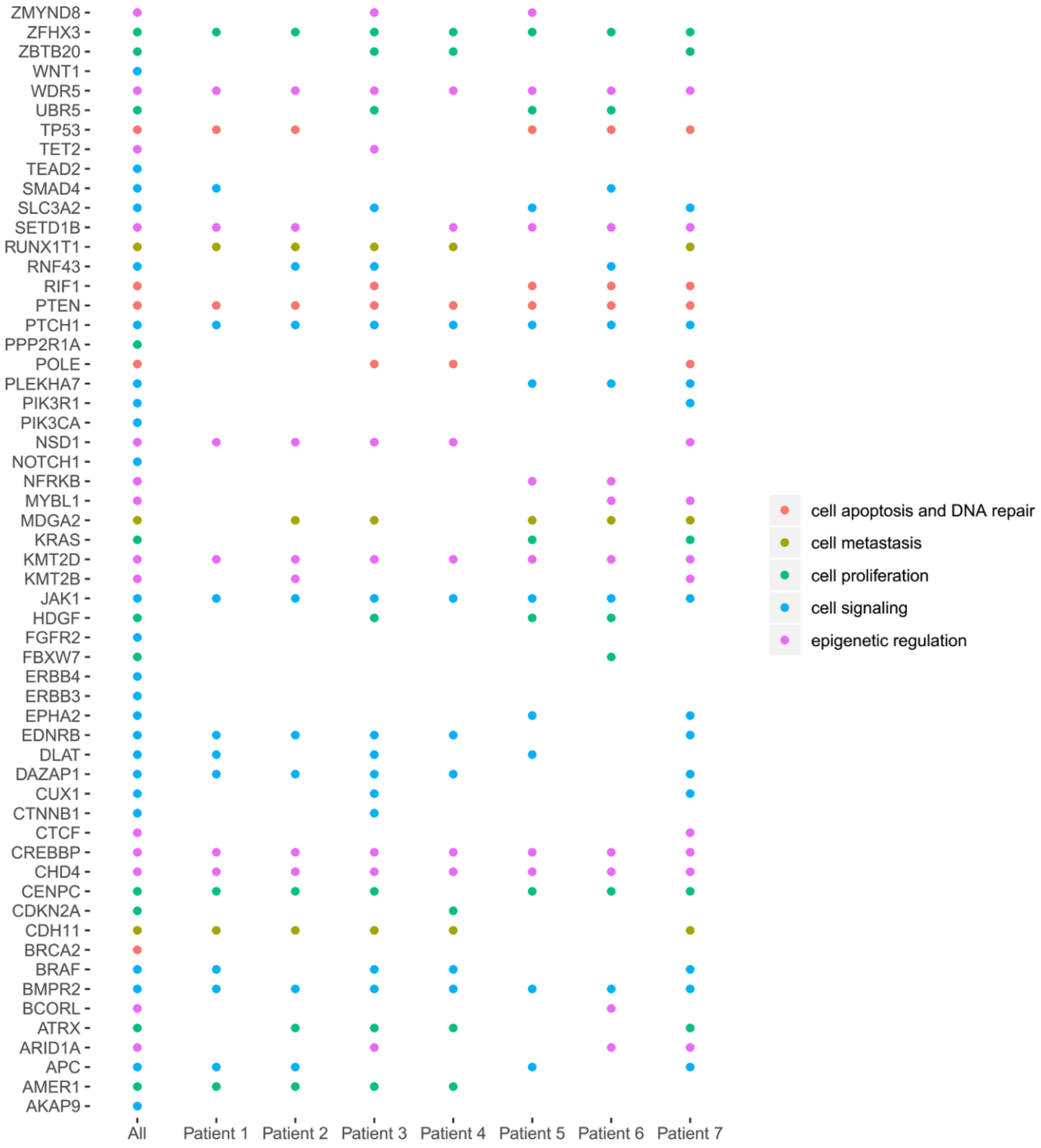
50. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011; 12:443–51. https://doi.org/10.1038/nrg2986 PMID:21587300

SUPPLEMENTARY MATERIALS

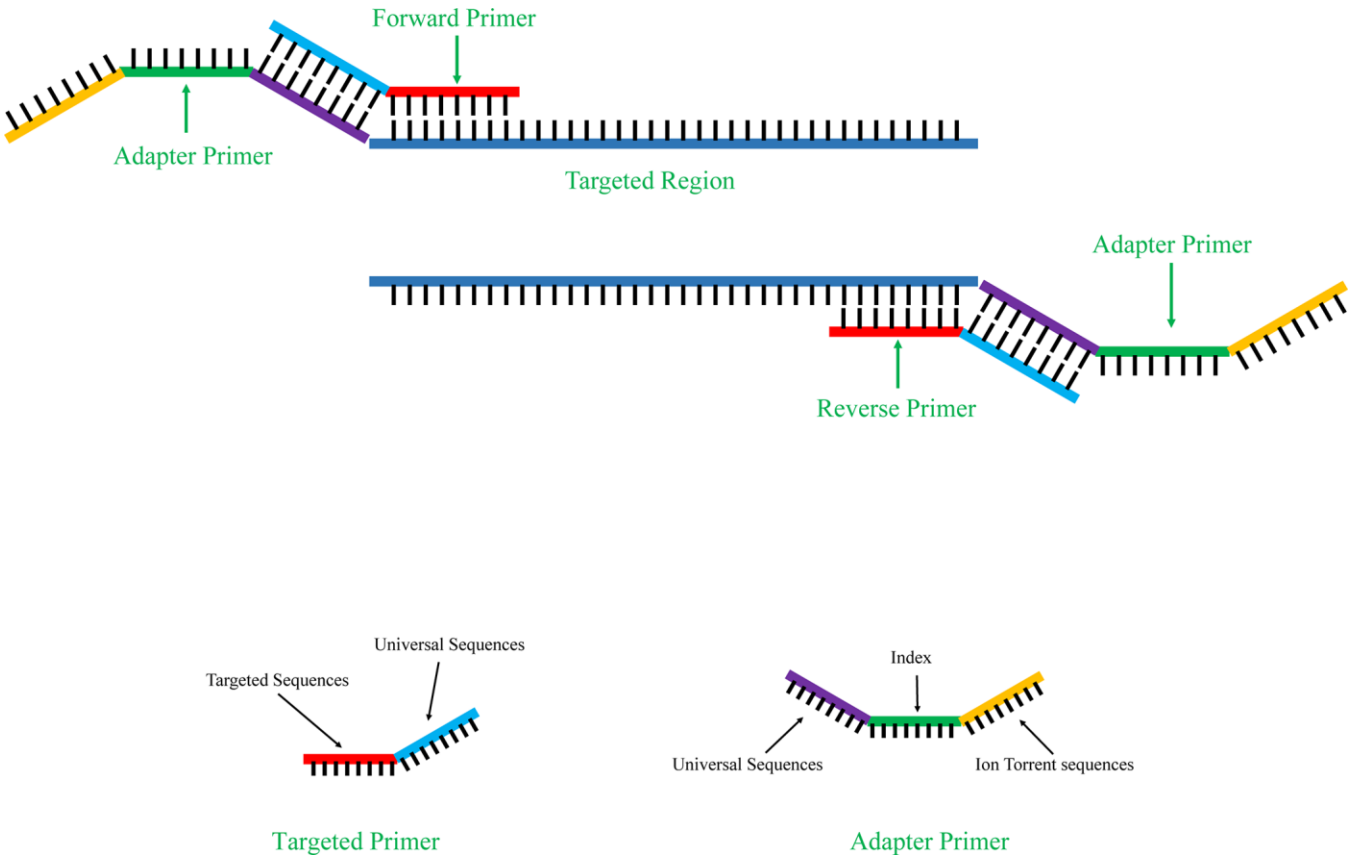
Supplementary Figures



Supplemental Figure 1. The nonsynonymous tumor mutation burden across 33 cancer types in TCGA. Each dot corresponds to a tumor sample. The log10 scaled vertical axis demonstrates the number of nonsynonymous mutations per megabase. The 33 cancer types are ordered on the horizontal axis according to their median numbers of nonsynonymous mutations per megabase, which were represented by the short red horizontal lines. The COAD lied in the fifth column.



Supplemental Figure 2. Mutation detection of T1 tumor tissues from the six COAD patients. All the selected genes lie in the first column. The mutant genes that detected in each patients lied in the second to seventh column. All the detected mutant genes lied in the eighth column.



Supplemental Figure 3. Schematic diagram of primer design. Two primers were design in this study: the targeted primer and the adapter primer, as shown in methods section. The sequence of the targeted primer were shown in Supplementary Table.

Supplementary Tables

Supplementary Table 1.

No.	Sex	Age	Tumor size	Stage	Lymph.Nodes.Positive	CT detection
201702011	female	65	1.5*1.5*0.5cm	III	1	1
201705053	male	52	4.5*4.5*0.5cm	IV	0	1
201701065	male	66	4.5*4.5*1.0cm	IV	0	1
201802060	male	73	1.5*1.0*0.5cm	I	0	1
201804011	male	62	5.0*4.0*3.0cm	II	0	1
201806006	male	81	4.0*1.0*0.8cm	II	0	1
201902005	male	64	6.5*3.5*1.5cm	III	0	1

Please browse Full Text version to see the data of Supplementary Table 2

Supplementary Table 2.

Supplementary Table 3.

Patients	sample	raw reads	raw data (bp)	Q30(%)	clean reads	clean data (bp)
Health 1	blood	2874958	431243700	98.60%	2834786	425217900
Health 2	blood	3555754	533363100	98.65%	3507699	526154850
Health 3	blood	3791302	568695300	98.55%	3736490	560473500
Health 4	blood	3465252	519787800	98.60%	3416689	512503350
Health 5	blood	2733798	410069700	98.66%	2697123	404568450
Health 6	blood	2374414	356162100	98.35%	2335284	350292600
Health 7	blood	2597040	389556000	98.51%	2558332	383749800
Patient 1	T1	8983730	1347559500	94.44%	8484669	1272700350
Patient 1	T21	3889436	583415400	94.61%	3679665	551949750
Patient 1	T22	6487360	973104000	95.23%	6177609	926641350
Patient 1	T31	5021792	753268800	95.15%	4778485	716772750
Patient 1	T32	5293576	794036400	95.10%	5034356	755153400
Patient 1	T33	7208058	1081208700	95.47%	6881547	1032232050
Patient 1	normal	10611972	1591795800	95.14%	10096302	1514445300
Patient 2	T1	12907630	1936144500	94.81%	12237581	1835637150
Patient 2	T21	643660	96549000	95.38%	613951	92092650
Patient 2	T22	6291702	943755300	94.59%	5951473	892720950
Patient 2	T31	5938210	890731500	95.01%	5641969	846295350
Patient 2	T32	12856498	1928474700	94.88%	12198651	1829797650
Patient 2	T33	11851226	1777683900	95.20%	11282304	1692345600
Patient 2	normal	26942700	4041405000	94.34%	25418950	3812842500
Patient 3	T1	6077548	911632200	96.13%	5842055	876308250
Patient 3	T21	9745696	1461854400	95.74%	9330659	1399598850
Patient 3	T22	4530982	679647300	95.81%	4341041	651156150
Patient 3	T31	4008834	601325100	94.92%	3805217	570782550
Patient 3	T32	6339228	950884200	95.03%	6024258	903638700

Patient 3	T33	13664850	2049727500	95.46%	13044423	1956663450
Patient 3	normal	7876790	1181518500	94.97%	7480867	1122130050
Patient 4	T1	37350466	5602569900	94.61%	35337186	5300577900
Patient 4	T21	30142032	4521304800	94.49%	28480857	4272128550
Patient 4	T22	12475154	1871273100	95.75%	11945368	1791805200
Patient 4	T31	15092174	2263826100	95.48%	14409462	2161419300
Patient 4	T32	10925402	1638810300	95.76%	10462200	1569330000
Patient 4	T33	10587636	1588145400	95.67%	10128681	1519302150
Patient 4	normal	29675188	4451278200	95.16%	28238376	4235756400
Patient 5	T1	2200670	330100500	95.30%	2097344	314601600
Patient 5	T21	5394972	809245800	95.81%	5169181	775377150
Patient 5	T22	2657258	398588700	95.28%	2531863	379779450
Patient 5	T31	3108924	466338600	95.41%	2966341	444951150
Patient 5	T32	7477154	1121573100	95.33%	7127832	1069174800
Patient 5	T33	4484220	672633000	95.21%	4269267	640390050
Patient 5	normal	2867686	430152900	95.51%	2739052	410857800
Patient 6	T1	1228996	184349400	95.62%	1175189	176278350
Patient 6	T21	4565322	684798300	95.42%	4356374	653456100
Patient 6	T22	2607780	391167000	95.78%	2497784	374667600
Patient 6	T23	1288430	193264500	95.64%	1232260	184839000
Patient 6	normal	1706826	256023900	95.97%	1637963	245694450
Patient 7	T1	6903882	1035582300	95.88%	6619402	992910300
Patient 7	T21	12933792	1940068800	95.92%	12405758	1860863700
Patient 7	T22	7628024	1144203600	96.39%	7352945	1102941750
Patient 7	T31	12530866	1879629900	95.76%	12000062	1800009300
Patient 7	T32	3809072	571360800	95.70%	3645240	546786000
Patient 7	T33	8493836	1274075400	95.63%	8122311	1218346650