Clonal Hematopoiesis in Lung Cancer and Its Impact on Liquid Biopsy Based on Targeted Deep Sequencing

Yang Xue
Sichuan Provincial People's Hospital: Sichuan Academy of Medical Sciences and Sichuan People's Hospital
Youyu Wang
Sichuan Provincial People's Hospital: Sichuan Academy of Medical Sciences and Sichuan People's Hospital
Yingmei Li
HaploX Biotechnology
Danni Liu
HaploX Biotechnology
Mengqi Shao
Sichuan Provincial People's Hospital: Sichuan Academy of Medical Sciences and Sichuan People's Hospital
Lei Liu
Sichuan Provincial People's Hospital: Sichuan Academy of Medical Sciences and Sichuan People's Hospital
Yida Liao
Sichuan Provincial People's Hospital: Sichuan Academy of Medical Sciences and Sichuan People's Hospital
Tanxiao Huang
HaploX Biotechnology
Jing Zhang
HaploX Biotechnology
Tianhao Mu
HaploX Biotechnology
Shifu Chen
HaploX Biotechnology
Gang Feng (steedvan@163.com)
Sichuan Provincial People's Hospital: Sichuan Academy of Medical Sciences and Sichuan People's Hospital https://orcid.org/0000-0001-9855-616X
Kegang Jia

Sichuan Provincial People's Hospital: Sichuan Academy of Medical Sciences and Sichuan People's Hospital

---

Research

Keywords: Clonal hematopoiesis, lung cancer, liquid biopsy

Posted Date: September 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-907910/v1

Abstract

Background: Blood based liquid biopsy has proved its potential in enormous clinical applications, such as cancer screening, diagnosis, treatment guidance, disease tracking and monitoring. In certain scenario (e.g., molecular residual disease), it requires the technique to be able to detect mutation with very low frequency (0.001% ~ 1%). The major hurdle of ultra-sensitive circulating tumor DNA sequencing is the high background noise of plasma cell-free DNA (cfDNA) and clonal hematopoiesis (CH). Here in this study, we investigated the prevalence of CH in lung cancer patients and its interference with liquid biopsy.

Methods: We retrospectively analyzed cfDNA and blood cell genomic DNA (gDNA) sequencing data sets (n=1261) from a group of Chinese lung cancer patients. Threshold (1%) and subthreshold (0.2%) for variant allele frequency were set and compared. We focused on 23 clonal hematopoiesis genes that were selected based on previous publications.

Results: CH mutations were detected in 27.68% of all the patients at the threshold and 62.01% at the subthreshold, and the detection rate increased with age. *DNMT3A* was the most frequently mutated CH gene, accounted for more than half of the CH mutations. The CH mutations had a higher detection rate in smokers (72%) than non-smokers (59.4%) at subthreshold. VAFs of CH mutations in cell-free DNA strongly correlated with their VAFs in gDNA (Pearson's R =0.92, $p<2.2x10^{-16}$), while tumor derived somatic mutations didn't have such correlation.

Conclusion: Our study showed that clonal hematopoiesis is very common in lung cancer patients, especially when examining low frequency mutations. Sequencing of gDNA at equivalent depth is very important to filter out CH mutation in cancer liquid biopsy.

Background

Blood based circulating tumor DNA (ctDNA) sequencing techniques (i.e. liquid biopsies)[1,2] have made great progress in cancer clinical applications[3,4], such as cancer screening[5], companion diagnosis[6] and treatment monitoring.[4,7,8] As part of the cell-free DNA (cfDNA) in plasma, the amount of ctDNA presents a large dynamic range among patients with different stages of cancer.[9,10] Despite the achievement liquid biopsy has made, there remain some limitations and challenges not to be ignored in this field related to the nature of cfDNA biology. One of the main issues is the interference with the undefined biological background cell-free DNA.[11,12] Previous studies have investigated those different sources of cfDNA and predicted the majority of cfDNA were from blood cells.[13,14] Besides genomic DNA fragments from blood cells, hematopoietic progenitor cells also accumulates mutations during aging, called clonal hematopoiesis (CH). [15–19] These CH mutations introduce false-positive mutation calling in liquid biopsy if sequenced or interpreted incorrectly, particularly when the tumor derived mutation frequency is very low.[15,20] As there's growing interest of using liquid biopsy in molecular residual disease (MRD) monitoring,[21–25] it usually requires ultra-deep sequencing and careful removing background noise. Previous studies

have identified dozens of common mutated CH genes,[18,26–29] and here in this research we summarized those literature and focused on 23 canonical CH related genes. We investigated the mutation landscape of these featured CH genes in Chinese lung cancer patients and the correlation with clinical characteristics. We sought to analyzed how the variant allele frequency (VAF) setting in the mutation calling and filtering affected ctDNA and CH derived mutation detection.

Methods

Sample processing

The authors retrospectively analyzed all non-small cell lung cancer (NSCLC) cases with both plasma cfDNA (n = 1261) and blood cell gDNA (n = 1049) sequencing data in a single laboratory from July 2019 to May 2021. Blood samples (n = 54) from healthy individuals were also sequenced and compared. In brief, 8 mL blood from each individual was collected in Streck BCT® and delivered to lab within 48 hours at ambient temperature. Plasma was then separated by sequential centrifugation at 1600g for 10 minutes and 16,000g for 10 minutes. Cell-free DNA was extracted by QIAamp Circulating Nucleic Acid Kit (QIAGEN, 55114), and blood cell gDNA was extracted by TIANamp Blood DNA Kit (TIANGEN, DP348-03). Extracted gDNA was fragmented using WGS Fragmentation Enzyme Mix (QIAGEN, Y9410), underwent size selection by KAPA Pure Beads (Roche, 7983298001), and was quantified by the Qubit 2.0 fluorometer (Thermo Fisher Scientific). Cell-free DNA was quantified by Qubit 2.0 fluorometer and 4200 TapeStation system (Agilent Technologies) without fragmentation. Sequencing libraries were prepared with the KAPA library prep kit (KAPA Biosystems) following the manufacturer's protocol. DNA was enriched by a 601-gene panel (HaploX, see supplementary tabular data Table S1 for full gene list) with HyperCap Target Enrichment Kit (Roche, 08286345001) following the manufacturer's protocol. DNA libraries were quantified by KAPA Library Quant kit (Roche,07960140001), and examined by 4200 TapeStation system, then sequenced using 150-bp paired-end runs on the NovaSeq 6000 system (Illumina).

Bioinformatic analysis

Data Processing. Raw sequencing data were preprocessed by fastp (v0.12.6)[30] for trimming of adapters and removal of low-quality (q < = 20) reads. Clean reads were aligned to the reference human genome (build hg19) using sentieon bwa-mem with default parameters. Duplicate reads were identified and marked with sentieon tools (v201808).[31] SNVs and Indels were called using VarScan2 (v2.3.8)[32] and annotated by ANNOVAR (v2018-04-16).[33]

CH alterations. Filter procedures were mainly following the previous publication[26], briefly,

1. Germline mutations with a ≥ 30% AF in both WBC gDNA and cfDNA were filtered out. Exceptions were frameshifting Indels or truncating SNVs occurring in the 23 canonical genes associated with CH.
2. Synonymous variants were not analyzed.

3. Variants covered by < 500 reads or with < 5 high-quality sequencing reads (mapqthres > 30, baseqthres > 30) in WBC gDNA or cfDNA were disregarded.
4. Only alterations covered by ≥ 2 forward and reverse strand reads were considered.
5. Remove mutations with low variant allele frequency (VAF threshold < 1%; VAF subthreshold < 0.2%) in WBC gDNA and cfDNA.
6. Remove variant sites with mutant allele frequency (MAF) > 0.01 in any one of the population variant database (esp6500siv2_all, 1000g2015aug_eas, ExAC_EAS,ExAC_All)
7. Mutations in repetitive sequences (simple tandem repeats, segmental duplications, and microsatellites) were removed.
8. We built an in-house database of cfDNA background noise from 54 healthy individuals. Variants were filtered if also present in the healthy group in genes other than the 23 canonical CH genes.

Blood tumor mutation burden (bTMB). The bTMB was calculated as the number of competent mutations divided by the length of the panel-covered genomic region (1.89 Mb). High bTMB (bTMB-H) was defined as bTMB ≥ 10 mut/Mb.

Microsatellite instability (MSI). MSI status were determined by VisualMSI.[34] Tumor tissue samples were divided into microsatellite stable (MSS), low MSI (MSI-L) and High MSI (MSI-H).

Statistical analysis. The consistency between two continuous variables was assessed using Pearson correlation analysis. The proportional compositions of two variables were compared using Chi-square tests. R (v3.5.1) was used for all statistical analyses. P values < 0.05 were considered statistically significant.

Results

Study overview and the sequencing depth

We retrospectively reviewed 1261 plasma cfDNA sequencing data sets with matched white blood cell (WBC) from 1046 lung cancer patients in a single sequencing center. Notice that 181 patients had multiple sampling. Plasma cell-free DNA was assigned to be sequenced with 20,000X raw depth (uncollapsed), while 10,000X for WBC gDNA (Fig. 1a). Patients' information was summarized in Fig. 1b. To the best of our knowledge, this is the first systemic CH study of large Chinese lung cancer population. Notice that the library preparation, targeted enrichment and sequencing processes inevitably introduced replicates and other unwanted reads, we plotted the uncollapsed and actual collapsed coverage for cfDNA and WBC gDNA (Fig. 1c and 1d). The median uncollapsed coverage was 17742X and 9173X for cfDNA and gDNA, and median collapsed coverage was 2757X and 2349X for cfDNA and gDNA respectively. Considering the amount of cfDNA varied a lot among samples, and all the extracted cfDNA was used for library prep, we analyzed how cfDNA input affected the mean coverage using Pearson's correlation (figure S1). The input cfDNA positively correlates with collapsed mean coverage.

Mutation landscape of featured CH genes in lung cancer patients and healthy controls.

Twenty-three featured CH genes were selected from several previous publications[18,26–29] : *ASXL1, ATM, CBL, CHEK2, CREBBP, DNMT3A, FGFR3, GNAS, IDH1, IDH2, JAK2, KMT2D, MED12, NF1, NOTCH1, RUNX1, SETD2, SF3B1, SRSF2, TET2, TNFAIP3, TP53, U2AF1*. The mutational landscape of all the cfDNA samples from lung cancer patients and 54 healthy individuals was shown in Fig. 2. To determine the appropriate cutoff for CH mutation calling in liquid biopsy, we set the VAF at threshold (1%) and subthreshold (0.2%). We detected CH variants at threshold in 349 (27.68%) of 1261 lung cancer patients, and 782 (62.01%) at subthreshold. The top mutated CH related gene was *DNMT3A*, followed by *TET2, ASXL1, TP53*. On the contrary, in 54 healthy controls, only 7 (12.96%) and 21 (38.89%) showed variants in CH related genes at threshold and subthreshold respectively. Besides these 23 reported CH genes, we used our pipeline (see methods) to filter top 30 potential blood cell derived mutated genes (figure S2) to see if there were other neglected CH genes. We can see that the most popular CH genes (*DNMT3A, TET2, ASXL1, TP53*, etc.) were present in both methods. Among all the 452 CH mutations detected at the threshold, *DNMT3A* accounted for nearly 60% of them, followed by *TET2, ASKL1, KMR2D* and *TP53*. While in all the 1513 CH mutations detected the subthreshold, *DNMT3A* accounted for 51.55% of them, followed by *TET2, U2AF1, CREBBP*, and *ASKL1* (figure S3).

Prevalence of Clonal hematopoiesis

We observed that CH related mutations correlated with increasing age (Fig. 3) in cancer patients, which was consistent with previous studies.[11,35,36] No CH mutations were detected in the age range from 20–29 with threshold and subthreshold. The highest detection frequency was in 80–89 range: 47.06% (Fig. 3a) with threshold and 83.32% (Fig. 3b) with subthreshold respectively. However, there was a drop in age range 90–99 with threshold (33.33%), but a similar detection level in 90–99 (83.33%) and 80–90 with subthreshold. This suggests with 1% VAF cutoff, some mutations were missed so that the trend of increasing detection frequency with increasing age deviated at the end (highlighted with red dashed square). When analyzing number of mutations detected in different age ranges, the proportion of multiple mutations (> = 2) also increased with age (Fig. 3c and 3d), and the trend was even more clear with the subthreshold. While in healthy individuals, we didn't observe such trend (Figure S4).

Source of mutations in cell-free DNA

To identify the source of a mutation from cfDNA sequencing and the requirement of sequencing depth, we first analyzed VAFs of all mutations (Fig. 4a) and CH related mutations detected by our pipeline (Fig. 4b) in different depth range with the subthreshold setting. It was clear to see that in order to reliably detect mutations with VAF as low as 0.2% (red dash line), the collapsed coverage needed to be above 2500X. This was also confined by the bioinformatic algorithms which requested at least 5 mutant reads covered the mutational site. We plotted all somatic mutations and CH related mutations with its VAF in cfDNA and in gDNA (Fig. 4c), and most CH mutations fell in the diagonal range (Pearson's R = 0.92, p <

$2.2x10^{-16}$), suggested their source of blood cells. Similar results can be seen with the threshold setting (figure S5). Here we presented one example of a patient (P53198) harboring *TP53* mutations (Fig. 4d and 4e). When calling mutation below 1% VAF, a new mutation of TP53_p.C275Y appeared. As it had similar VAF in gDNA sequencing, it was classified as CH mutation from blood cells.

Correlation of CH mutations and clinical characteristics

In order to investigate the CH mutations in different clinical settings, we compared proportion of CH mutation positive patients with respect to smoking status (Fig. 5a), bTMB status (Fig. 5b), cancer subtype (Fig. 5c), and MSI status (Fig. 5d). Among them, MSI status was calculated by samples with matched tumor sequencing data (n = 461). No significant difference was observed in these subpopulations except for smokers and non-smokers with subthreshold. Smokers had significantly ($p < 0.001$) higher CH detection rate than non-smokers, which can be explained by the chemical mutagenesis nature of cigarette.[37]

Discussion

By decreasing the VAF cutoff from 1–0.2%, the detection frequency in cancer population increased from 27.68–62.01%, and the number of CH mutations increased from 452 to 1513. More than double of the mutations above 1% fell into range 0.2% ~ 1%. In order to reliably distinguish mutation at 0.2% level, the collapsed sequencing depth for gDNA also needs to be above 2500X. As liquid biopsy moves into the application of molecular residual disease (MRD) testing, it requires the identification of mutation with VAF as low as 0.01%.[38] In this scenario, setting minimum VAF to 2% for CH[16,28] is apparently no longer suitable. For such ultra-sensitive detection, the significance of gDNA sequencing should not be neglected in liquid biopsy.

DNMT3A is the most popular mutated CH genes in cancer patients, not only in our cohort, but also in many other publications.[18,26–29] It was also reported to have a strong indication with hyperprogression after immunotherapy.[39] In real world, some patients with late stage cancer who have chance to receive immunotherapy could only provide blood samples for clinical examinations, thus clinicians need to pay extra attention when patients‘ blood are positive for this gene mutation. If the *DNMT3A* mutation is only a clonal hematopoiesis, then the patient may still benefit from immunotherapy. *TP53* is the most famous tumor suppressor gene and it’s often mutated in many cancer types. It is also one of the top 5 CH genes, making it extremely crucial in liquid biopsy. In our example in Fig. 4d and 4e, when decreasing the VAF cutoff from 1–0.2%, the CH derived mutation appeared, which was demonstrated by similar VAF in blood gDNA. This shows us the importance of requirement of equivalent sequencing depth for WBC gDNA.

We find that smoking is correlated with higher detection rate of CH mutations. There are a plenty of studies that proved the carcinogenesis and mutagenesis of cigarette. However, without knowing the detailed smoking history of the patients, we couldn’t estimate to what extent the smoking behavior affected CH mutations. With respect to other clinical features, we haven’t found any relation with bTMB,

MSI status or cancer subtypes. Recently, Dr. Pritchard and co-workers reported hematopoiesis in DNA repair genes was associated with prostate cancer plasma cell-free DNA testing interference.[40] It is also reported that inflammation[41] and treatment[42] might induce clonal hematopoiesis as well. The cause of hematopoiesis[27,35] is still under investigation and its clinical implication[28] is yet to be systemically studied.

Conclusions

Our results demonstrated that clonal hematopoiesis is very common in Chinese lung cancer patients and would interfere with blood-based ctDNA testing. Equivalent sequencing depth of gDNA could help distinguish tumor derived mutation from clonal hematopoiesis. While scientists are learning more about clonal hematopoiesis, liquid biopsy has been growing in cancer clinical applications. Physicians and molecular pathologist need carefully consider the effect of clonal hematopoiesis when interpreting the sequencing reports.

Abbreviations

cfDNA: cell-free DNA

ctDNA: circulating tumor DNA

CH: clonal hematopoiesis

VAF: variant allele frequency

NSCLC: non-small cell lung cancer

LUAD: lung adenocarcinoma

LUSC: lung squamous carcinoma

MRD: molecular residual disease

bTMB: blood tumor mutation burden

MSI: microsatellite instability

MSS: microsatellite stable

Declarations

Ethics approval and Consent to participate: All procedures performed in this study involving the patient were in accordance with the ethical standards of the ethical standards of the institutional research committee. Clinical samples were obtained through the Sichuan Provincial People's Hospital Institutional

Review Board-approved informed consent process, and only excess amount of tumor samples not required for pathological diagnosis was utilized in this study.

Consent for publication: The authors declare that they agree to submit the article for publication.

Availability of data and materials: Supplementary figures are available online through the journal website. The assembled mutational data from cfDNA, blood gDNA and tumors are provided as supplementary tabular data (Table S2-S4). Raw sequencing data cannot be deposited for legal and privacy reasons, as sequencing was performed for clinical purposes.

Competing interests: Yingmei Li, Danni Liu, Tanxiao Huang, Jing Zhang, Tianhao Mu and Shifu Chen are employed by the company, HaploX Biotechnology. All other authors declare that they have no competing interests.

Funding: We thank grants from the Science and Technology Innovation Committee of Shenzhen Municipality (KQJSCX20180330124428928; JSGG20180508152646606) and Development and Reform Commission of Shenzhen Municipality (XMHT20190104006; XMHT20200104013).

Authors' Contributions: Conceptualization-Gang Feng, Kegang Jia, Shifu Chen; Clinical sample and data collection- Yang Xue, Mengqi Shao, Yida Liao; Sequencing data analysis- Youyu Wang, Danni Liu, Lei Liu, Tanxiao Huang, Jing Zhang, Yingmei Li; Manuscript preparation- Yang Xue, Youyu Wang, Yingmei Li; Discussion and revisions – all authors.

Acknowledgements: We acknowledge the generosity of the patients for allowing us to use their clinical samples in this research. We thank Yuanyuan Liu and Xiaoni Zhang for technical support and helpful discussion.

References

1. Filipska M, Rosell R. Mutated circulating tumor DNA as a liquid biopsy in lung cancer detection and treatment. *Mol Oncol*. 2021:1-16. doi:10.1002/1878-0261.12983
2. Crowley E, Nicolantonio F Di, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the Blood. *Nature*. 2013;10(8):472-484. doi:10.1038/nrclinonc.2013.110
3. Deveson IW, Gong B, Lai K, et al. Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. *Nat Biotechnol*. 2021. doi:10.1038/s41587-021-00857-z
4. Cescon DW, Bratman S V., Chan SM, Siu LL. Circulating tumor DNA and liquid biopsy in oncology. *Nat Cancer*. 2020;1(3):276-290. doi:10.1038/s43018-020-0043-5
5. Chen M, Zhao H. Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Hum Genomics*. 2019;13(1):34. doi:10.1186/s40246-019-0220-8
6. Vaidyanathan R, Soon RH, Zhang P, Jiang K, Lim CT. Cancer diagnosis: from tumor to liquid biopsy and beyond. *Lab Chip*. 2019;19(1):11-34. doi:10.1039/c8lc00684a

7. Zviran A, Schulman RC, Shah M, et al. Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. *Nat Med*. 2020;26(7):1114-1124. doi:10.1038/s41591-020-0915-3
8. Beaver JA, Park BENH. Circulating cell-free DNA for molecular diagnostics and therapeutic monitoring. In: *Genomic Applications in Pathology*. ; 2015:85-97. doi:10.1007/978-1-4939-0727-4
9. Bettegowda C, Sausen M, Leary RJ, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 2014;6(224):224ra24. doi:10.1126/scitranslmed.3007094
10. Newman AM, Bratman S V, To J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med*. 2014;20(August 2013):548-554. doi:10.1038/nm.3519
11. Liu J, Chen X, Wang J, et al. Biological background of the genomic variations of cf-DNA in healthy individuals. *Ann Oncol*. 2019;30(3):464-470. doi:10.1093/annonc/mdy513
12. Keller L, Belloum Y, Wikman H, Pantel K. Clinical relevance of blood-based ctDNA analysis: mutation detection and beyond. *Br J Cancer*. 2021;124(2):345-358. doi:10.1038/s41416-020-01047-5
13. Moss J, Magenheim J, Neiman D, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-07466-6
14. Razavi P, Li BT, Brown DN, et al. High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. *Nat Med*. 2019;25(12):1928-1937. doi:10.1038/s41591-019-0652-7
15. Hu Y, Ulrich BC, Supplee J, et al. False-positive plasma genotyping due to clonal hematopoiesis. *Clin Cancer Res*. 2018;24(18):4437-4443. doi:10.1158/1078-0432.CCR-18-0143
16. Chan HT, Chin YM, Nakamura Y, Low SK. Clonal hematopoiesis in liquid biopsy: From biological noise to valuable clinical implications. *Cancers (Basel)*. 2020;12(8):1-18. doi:10.3390/cancers12082277
17. Park SJ, Bejar R. Clonal hematopoiesis in cancer. *Exp Hematol*. 2020;83:105-112. doi:10.1016/j.exphem.2020.02.001
18. Ptashkin RN, Mandelker DL, Coombs CC, et al. Prevalence of clonal hematopoiesis mutations in tumor-only clinical genomic profiling of solid tumors. *JAMA Oncol*. 2018;4(11):1589-1593. doi:10.1001/jamaoncol.2018.2297
19. Maras F, Stephens D, Ptashkin R, et al. Fragment Size Analysis may Distinguish Clonal Hematopoiesis from Tumor-Derived Mutations in Cell-free DNA. *Clin Chem*. 2020;66(4):616-618. doi:10.1080/13518040701205365
20. Abbosh C, Birkbak NJ, Swanton C. Early stage NSCLC — challenges to implementing ctDNA-based screening and MRD detection. *Nat Rev Clin Oncol*. 2018;15(9):577-586. doi:10.1038/s41571-018-0058-3
21. Abbosh C, Birkbak NJ, Wilson GA, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. 2017;545(7655):446-451. doi:10.1038/nature22364
22. Cohen JD, Douville C, Dudley JC, et al. Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands. *Nat Biotechnol*. 2021. doi:10.1038/s41587-021-00900-

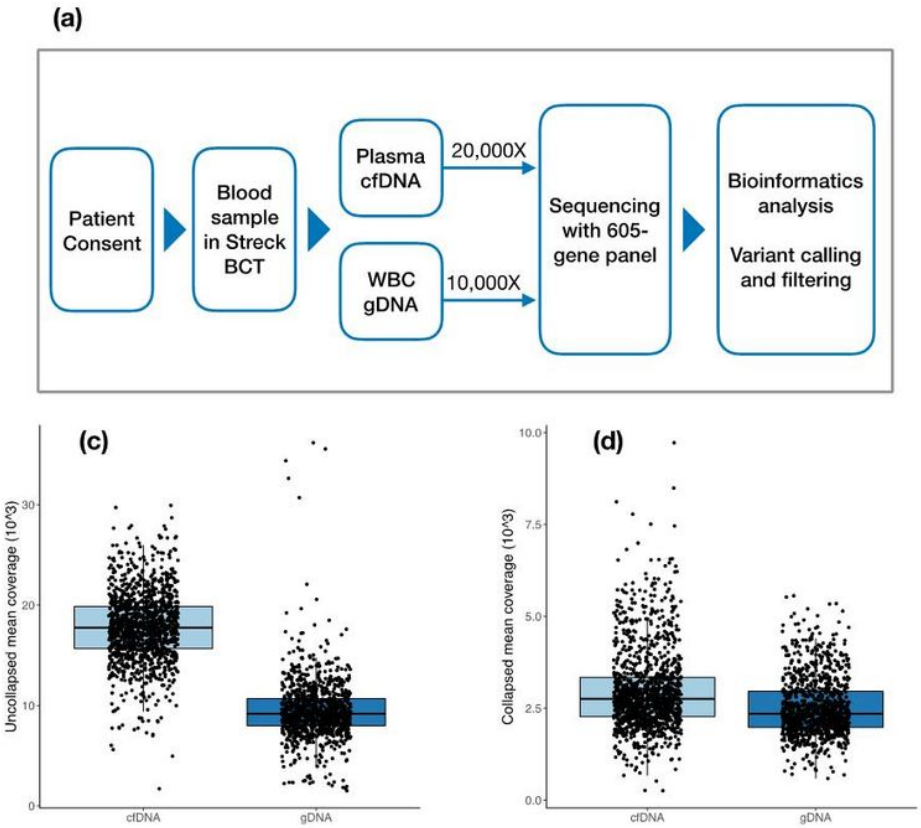
z

23. Garcia-Murillas I, Chopra N, Comino-Méndez I, et al. Assessment of Molecular Relapse Detection in Early-Stage Breast Cancer. *JAMA Oncol*. 2019;5(10):1473-1478. doi:10.1001/jamaoncol.2019.1838
24. Tie J, Cohen JD, Wang Y, et al. Circulating tumor dna analyses as markers of recurrence risk and benefit of adjuvant therapy for stage III colon cancer. *JAMA Oncol*. 2019;5(12):1710-1717. doi:10.1001/jamaoncol.2019.3616
25. Hrebien S, Citi V, Garcia-Murillas I, et al. Early ctDNA dynamics as a surrogate for progression-free survival in advanced breast cancer in the BEECH trial. *Ann Oncol*. 2019;30(6):945-952. doi:10.1093/annonc/mdz085
26. Zhang Y, Yao Y, Xu Y, et al. Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. *Nat Commun*. 2021;12(1):1-14. doi:10.1038/s41467-020-20162-8
27. Bick AG, Weinstock JS, Nandakumar SK, et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature*. 2020;586(7831):763-768. doi:10.1038/s41586-020-2819-2
28. Bolton KL, Ptashkin RN, Gao T, et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat Genet*. 2020;52(11):1219-1226. doi:10.1038/s41588-020-00710-0
29. Li Z, Huang W, Yin JC, et al. Comprehensive next-generation profiling of clonal hematopoiesis in cancer patients using paired tumor-blood sequencing for guiding personalized therapies. *Clin Transl Med*. 2020;10(7):1-5. doi:10.1002/ctm2.222
30. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890. doi:10.1093/bioinformatics/bty560
31. Freed D, Aldana R, Weber J, Edwards J. The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. *bioRxiv*. 2017:115717. doi:10.1101/115717
32. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568-576. doi:10.1101/gr.129684.111
33. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):1-7. doi:10.1093/nar/gkq603
34. Chen S, Qu H, Yang B, et al. Detect and visualize tumor microsatellite instability status from next-generation sequencing data by simulating PCR techniques. *J Clin Oncol*. 2019;37(15_suppl):e13052-e13052. doi:10.1200/JCO.2019.37.15_suppl.e13052
35. Jaiswal S, Fontanillas P, Flannick J, et al. Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N Engl J Med*. 2014;371(26):2488-2498. doi:10.1056/nejmoa1408617
36. Ococks E, Frankell AM, Masque Soler N, et al. Longitudinal tracking of 97 esophageal adenocarcinomas using liquid biopsy sampling. *Ann Oncol*. 2021;32(4):522-532. doi:10.1016/j.annonc.2020.12.010
37. De Flora S, D'Agostini F, Balansky R, et al. Modulation of cigarette smoke-related end-points in mutagenesis and carcinogenesis. *Mutat Res - Fundam Mol Mech Mutagen*. 2003;523-524:237-252.

doi:10.1016/S0027-5107(02)00340-8

38. Pantel K, Alix-Panabières C. Liquid biopsy and minimal residual disease — latest advances and implications for cure. *Nat Rev Clin Oncol*. 2019;16(7):409-424. doi:10.1038/s41571-019-0187-3
39. Kato S, Goodman A, Walavalkar V, Barkauskas DA, Sharabi A, Kurzrock R. Hyperprogressors after immunotherapy: Analysis of genomic alterations associated with accelerated growth rate. *Clin Cancer Res*. 2017;23(15):4242-4250. doi:10.1158/1078-0432.CCR-16-3133
40. Jensen K, Konnick EQ, Schweizer MT, et al. Association of Clonal Hematopoiesis in DNA Repair Genes with Prostate Cancer Plasma Cell-free DNA Testing Interference. *JAMA Oncol*. 2021;7(1):107-110. doi:10.1001/jamaoncol.2020.5161
41. Boettcher S, Manz MG. Regulation of Inflammation- and Infection-Driven Hematopoiesis. *Trends Immunol*. 2017;38(5):345-357. doi:https://doi.org/10.1016/j.it.2017.01.004
42. Lucas D, Scheiermann C, Chow A, et al. Chemotherapy-induced bone marrow nerve injury impairs hematopoietic regeneration. *Nat Med*. 2013;19(6):695-703. doi:10.1038/nm.3155

Figures



(b)

Parameter	Sample (%)
Age	median = 62 (21-92)
Gender	
Male	659 (52.26%)
Female	602 (47.74%)
Subtype	
LUAD	734 (58.21%)
LUSC	59 (4.68%)
Unknown	468 (37.11%)
Smoking history	
Never smoker	593 (47.03%)
Smoker	275 (21.81%)
Unknown	393 (31.16%)
bTMB	
bTMB-H	74 (5.87%)
bTMB-L	1187 (94.13%)
MSI (confirmed by tissue sequencing)	461
MSI-H	10 (2.17%)
MSI-L	21 (4.55%)
MSS	430 (93.28%)
Total	1049

Figure 1

Study overview. (a) Overview of research pipeline. (b) Clinical information summary of lung cancer patients. (c) Uncollapsed sequencing depth of cfDNA and gDNA for all lung cancer samples. (d) Collapsed sequencing depth of cfDNA and gDNA for all lung cancer samples.

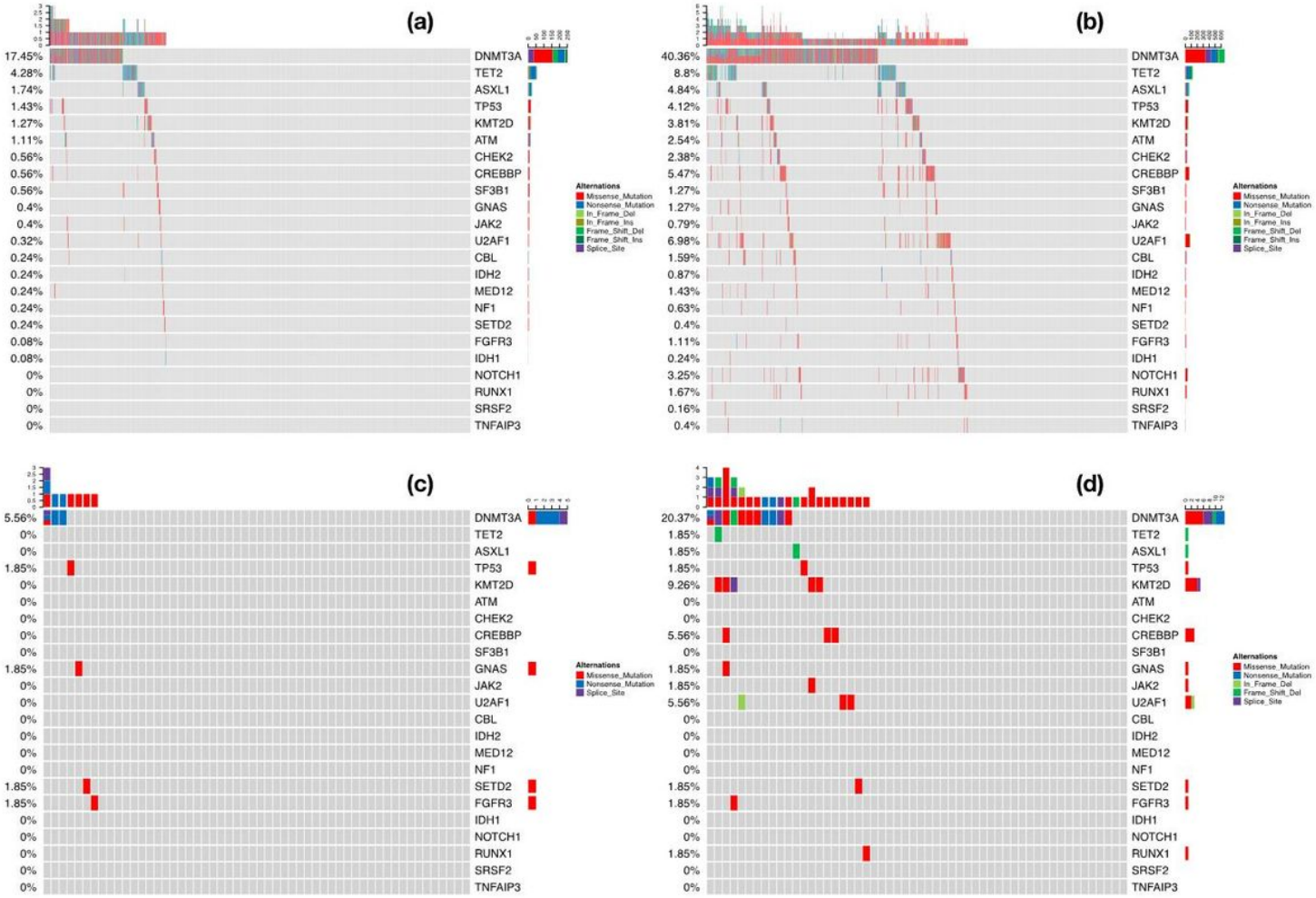


Figure 2

Mutation landscape of featured CH genes for (a) lung cancer patients with the threshold setting, (b) lung cancer patients with the subthreshold setting, (c) healthy individuals with the threshold setting, and (d) healthy individuals with the subthreshold setting.

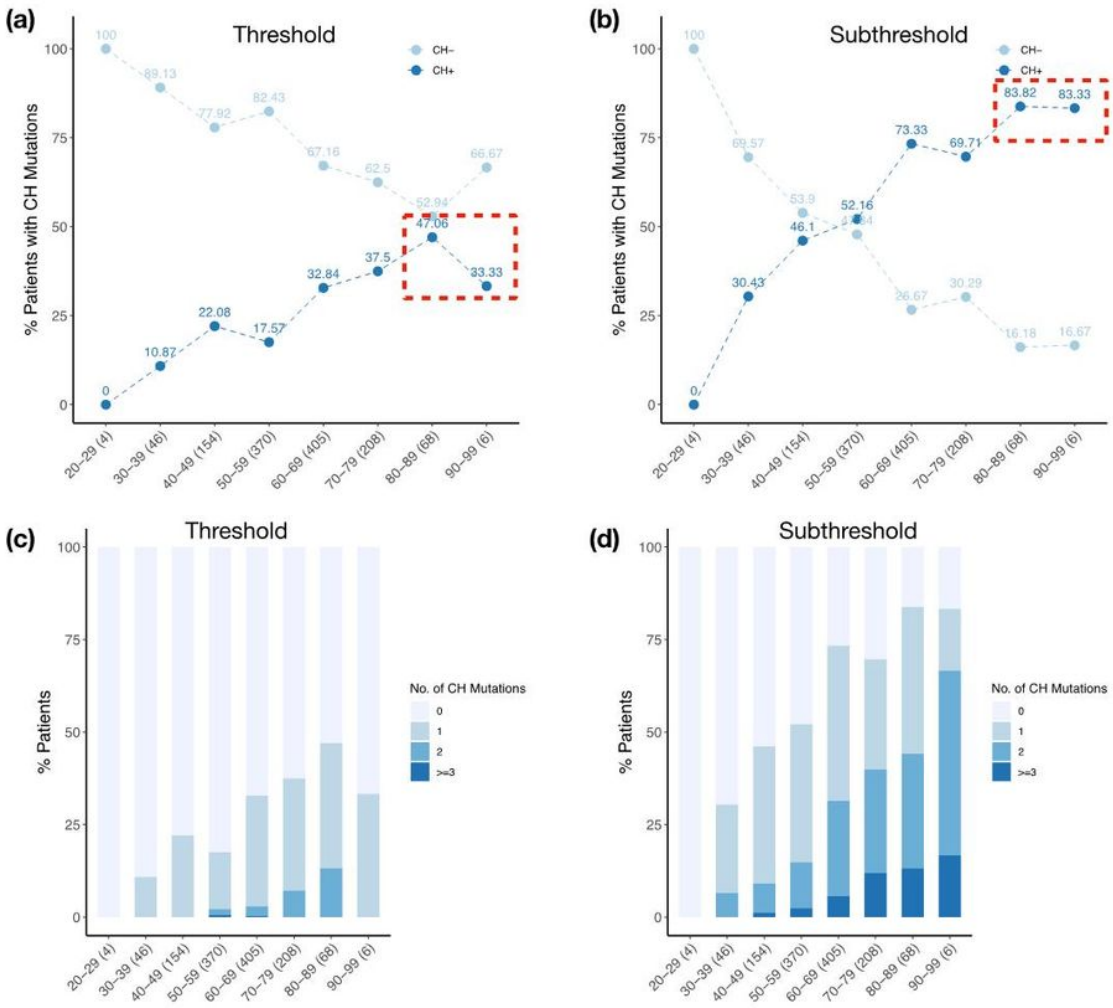


Figure 3

Prevalence of CH mutations in lung cancer patients. Proportion of CH mutation positive and negative patients in different age ranges with (a) the threshold and (b) the subthreshold setting. Group the patients with different number of detected CH mutations. Proportion of each group in different age ranges with (c) the threshold and (d) the subthreshold setting.

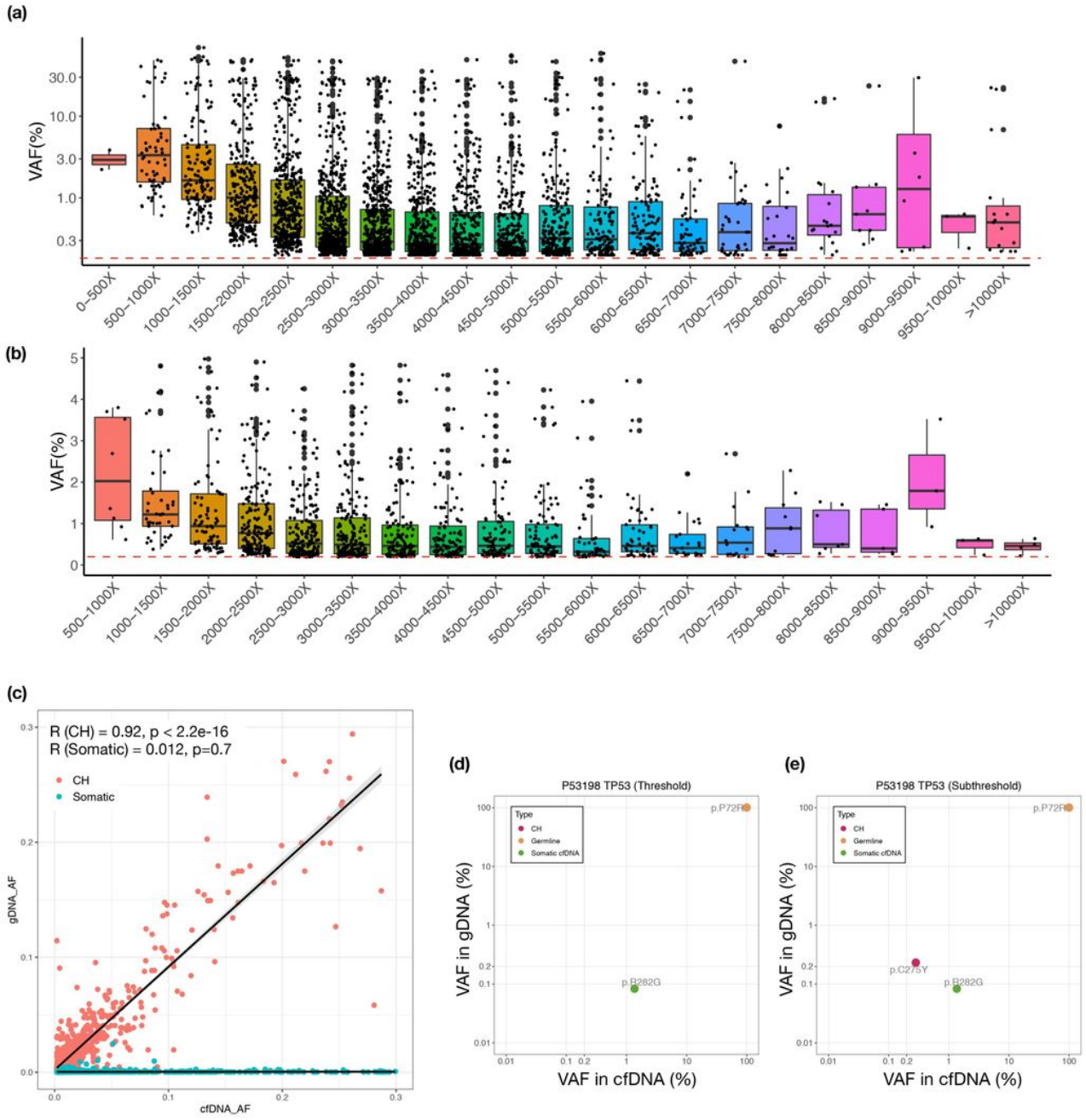


Figure 4

Sources and VAFs of mutations in cfDNA. (a) All mutations detected in cfDNA at different sequencing depth ranges; (b) all CH mutations detected in cfDNA at different sequencing ranges; (c) mutant allele frequencies in cfDNA and gDNA (Pearson's correlation); one example of a patient's mutation at (d) the threshold and (e) the subthreshold.

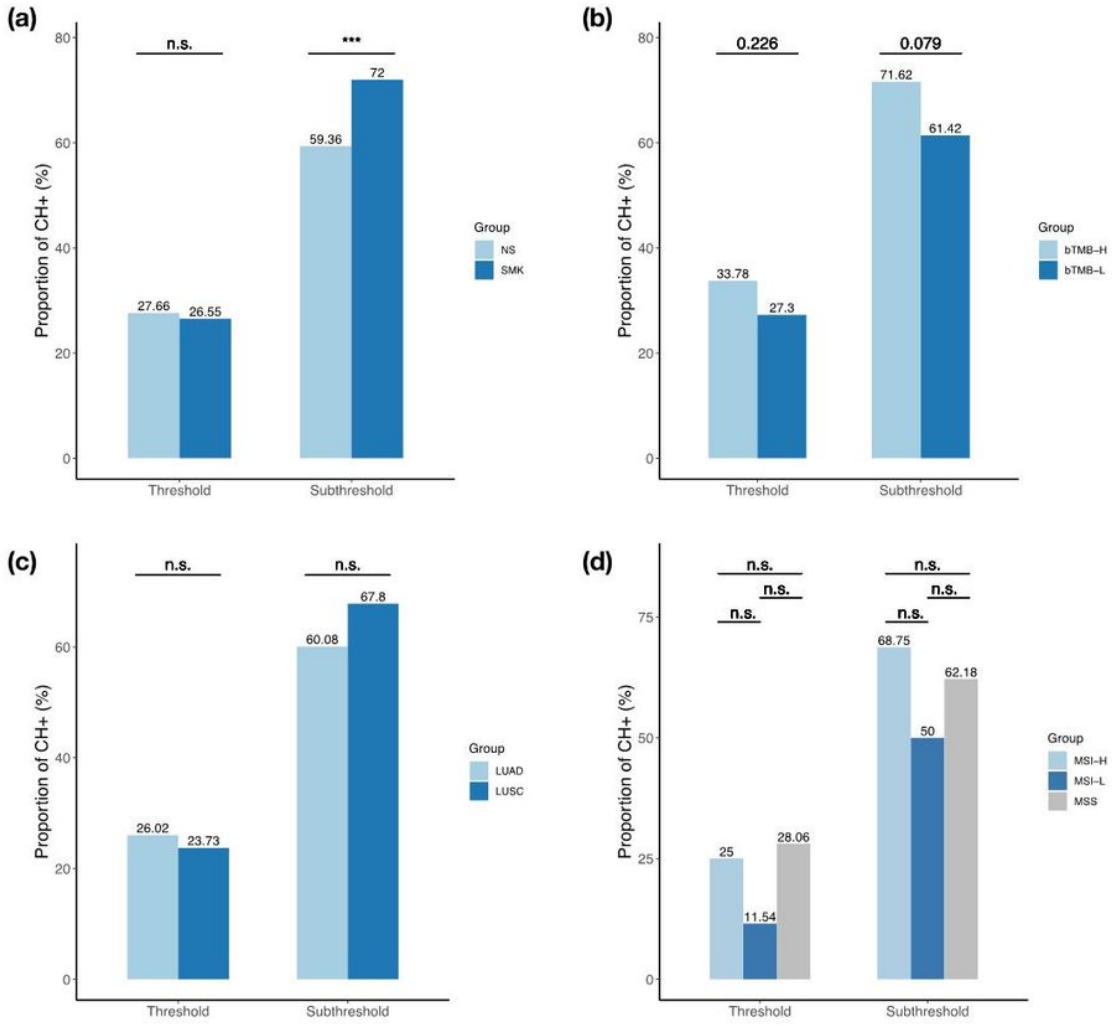


Figure 5

Correlation of CH mutations and clinical characteristics. CH mutation detection rates between (a) Smokers and non-smokers, (b) bTMB-H and bTMB-L patients, (c) lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), and (d) MSI-H, MSI-L and MSS patients. Chi-square p-value is shown in each plot (n.s.: not significant; ***: p-value < 0.001).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigures.docx
- SupplementarytabulardataTableS1Panelgenelist.csv
- SupplementarytabulardataTableS2AllCHthreshold.xlsx
- SupplementarytabulardataTableS3AllCHsubthreshold.xlsx

- SupplementarytabulardataTableS4Somatic.cfDNA.AllMut.xlsx