Biomedical & Translational Science

Original Article

Uncovering Novel Stage I Molecular Biomarkers for the Early Detection of Non-Small Cell Lung Cancer

Shu-Ti Lin*, Chandler Raj, Jun-Seong Ha

OncoDxRx, California, USA

Correspondence

Shu-Ti Lin

OncoDxRx, LLC, 150 N Santa Anita Ave., Suite 300, Arcadia, CA 91006, USA

Abstract

Although a number of driver biomarkers that are clinically relevant in the management of patients with non-small cell lung cancer (NSCLC) have been well documented, there is still about 25% unidentified genes and somatic mutations involved in the early onset of the disease. Most importantly, a clear distinction must be made between stage I vs. II and III of NSCLC (i.e., the early stages) in terms of biomarker discovery to refine inclusion criteria for screening programs and to benefit therapeutic decision in the neoadjuvant and adjuvant setting. There is much room for improvement in early-stage NSCLC outcome, identification of stage I-specific biomarkers herein could be the key to refine risk-stratification, prognosis and treatment of these patients. In this study, we aim to identify both tissue and circulating stage I biomarkers in NSCLC as the stepstone to improve clinical decision-making, especially regarding early detection, diagnostics and adjuvant therapy.

- Received Date: 09 Jun 2022
- Accepted Date: 10 Nov 2022
- Publication Date: 21 Nov 2022

Keywords

Biomarker, Stage I Lung Cancer, Early Detection, Somatic Mutation

Introduction

LLung cancer remains the most commonly diagnosed malignancy and the leading cause of cancer-related mortality worldwide, with non-small-cell lung cancer (NSCLC) represents over 80% of lung cancer cases [1,2]. National data showed over 50% of all lung cancer was diagnosed at a distant organ metastasis, with an overall five-year survival of 19.4% [3]. Given that lung cancer has such a poor survival profile, there is an urgently unmet need to discover new effective treatment, and most importantly, sensitive and specific technologies to detect lung cancer at an early stage to commence treatment as soon as possible. This leads to one of the major current challenges in oncology: the search for predictive genetic biomarkers of early-stage lung cancer.

Low-dose computed tomography (LDCT) is the current standard-of-care screening tool for NSCLC. Once a suspicious nodule is identified by this method, an invasive biopsy will be performed. However, the data from National Lung Screening Trial (NLST) has shown that LDCT decreased mortality by 20%, but alarmingly, this screening method had a 96.4% false positive rate [4], making it extremely non-specific and leading to many unnecessary invasive procedures and associated complications. Furthermore, annual LDCT screening also leads to increased radiation exposure. Due to the lack of sensitive and/or specific radiological screening methods, there has been much research effort focusing on non-invasive circulating genetic biomarkers. Today, utilizing liquid biopsy as a molecular test for patients with NSCLC, is included amongst a list of recommendations in the new guidelines for the College of American Pathologists (CAP), as well as the International Association for the Study of Lung Cancer (IASLC), and the Association for Molecular Pathology (AMP) [5].

Since circulating cell-free DNA (cfDNA) is mainly derived from apoptotic or dead tumor cells, they carry several characteristics that are informative and actionable regarding the tumor genetic contexts. First, since cancer is a disease of clonal evolution, the somatic mutations of malignant origin are likely to be present in clonally amplified cell populations [6]; therefore, the cancer-associated mutations are likely to be enriched at a relatively higher allele frequency than those background somatic mutations from non-tumor tissues in cfDNA. Second, the somatic mutations typically carry mutational signatures of tissue-specific tumorigenesis and/or DNA repair deficiency, profiling of the mutational signatures in cfDNA can help identify malignant tissue types [7]. Third, the neoplastic transformation typically involves genome instability which can lead to distinct mutational profiles. Fourth, cfDNA sequencing from plasma can provide orthogonal metagenomic signatures characteristics of tumor heterogeneity, adding yet another line of evidence for the detection of the disease. Taken

Citation: Lin ST, Raj C, Ha JS. Uncovering Novel Stage I Molecular Biomarkers for the Early Detection of Non-Small Cell Lung Cancer. Biomed Transl Sci. 2022; 2(4):1-6

together, in addition to the detection of oncogenic mutations, the emerging molecular signatures in cfDNA sequencing and inferred chromatin compositions can provide multi-faceted insights into tumor etiology and disease status, and guide imaging or molecular marker-based targeted follow-ups and treatment options.

A number of genomic technologies and bioinformatic methods have been developed for sensitive detection of somatic mutations and other molecular signatures from cfDNA sequencing. It is now possible to detect somatic mutations at an ultra-low allele frequency [8]. Moreover, integrative analysis of driver and passenger somatic mutations from cfDNA has been able to infer clonal and subclonal mutations, and also clonal architecture in the primary tumor [9] which opens up the possibility of tracking clonal dynamics in vivo in real-time. Collectively, these advances provide a rich toolbox to track cancer progression non-invasively from blood, even at an early stage of the disease and recurrence.

Different subtypes of lung cancer can secrete a different amount of cfDNA due to the degree of necrosis in early staged NSCLC, the utilization of cfDNA-based tests may be specific in detecting later stage NSCLC and their mutation status, and thus need to be combined with other biomarkers or imaging modalities to detect early-stage NSCLC [9]. The identification of patients with early-stage NSCLC who are most likely to benefit from adjuvant therapy is the key challenge to biomarker discovery. The role of gene signatures and the use of newer platforms such as RNA, methylation and protein signatures is being explored in patients with early-stage NSCLC. Despite a concerted effort, there is a lack of biomarkers with potential application in the management of early-stage NSCLC.

In this study, we performed tumor genomic profiling of stage I NSCLC using massively parallel sequencing to sensitively detect multiple types of mutations and identified potential biomarkers relevant to early-stage NSCLC. Data integration analysis from both tumor genomic DNA and longitudinal cfDNA not only identified novel oncogenic mutations, but also distinct mutational signatures for early- and late-stage NSCLC.

Materials and methods

Tumor Tissue gDNA and plasma cfDNA Isolation and Genomic Profiling

After routine pathology review for selection of stage I NSCLC patients, 5 x 10um unstained slides were cut from formalin-fixed paraffin-embedded (FFPE) tissues and used for DNA extraction and sequencing. Tissue genomic DNA and plasma cfDNA were extracted using QIAsymphony automated workstation (Qiagen, Hilden, Germany).

Mutational landscape was interrogated by next generation sequencing (NGS) on a 530-gene custom panel with 18 FFPE as well as 55 longitudinal plasma samples from 5 stage II-IV NSCLC patients. The sequencing panel was able to detect SNV (single nucleotide variant), indel (insertion/deletion), CNV (copy number variation), rearrangements and certain fusions. TMB (tumor mutational burden) and MSI (microsatellite instability) were also obtainable from this NGS panel. NGS assays were performed in a Clinical Laboratory Improvement Amendments-certified (CLIA) laboratory and followed standard protocols for library preparation and bioinformatic analysis.

Capture-Based Targeted DNA Sequencing

For NGS library construction, DNA Fragments between 200 and 400 bp were purified (Agencourt AMPure XP Kit, Beckman Coulter, CA, USA), followed by hybridization with capture probes baits, target region enrichment and library amplification. The quality and the size of the libraries were assessed with the high sensitivity DNA kit using Bioanalyzer 2100 (Agilent Technologies, CA, USA). Target capture was performed using a custom panel consisting of 530 cancer-related genes, spanning 1.70 megabases (Mb) of the human genome (OncoDxRx, CA, USA). Indexed samples were sequenced on Novaseq6000 (Illumina, Inc., USA) with paired-end reads and average sequencing depth of 1,000× for tissue samples and 20,000× for plasma samples.

Sequence data analysis

Sequence data were analyzed using optimized bioinformatics pipelines for somatic and germline variant calling and annotation. Briefly, the sequence data were mapped to the reference human genome (hg19) using Burrows-Wheeler Aligner v.0.7.10 [10]. Local alignment optimization, duplication marking and variant calling were performed using Genome Analysis Tool Kit v.3.2 [11], and VarScan v.2.4.3 [12]. Somatic mutations were filtered using the VarScan fpfilter pipeline, loci with depth less than 100 were filtered out. Base calling in plasma and tissue samples required at least eight supporting reads for SNV and two and five supporting reads for indel, respectively. Variants with population frequency over 0.1% in the ExAC, 1,000 Genomes, dbSNP or ESP6500SI-V2 databases were grouped as single nucleotide polymorphisms and excluded from further analysis. The remaining variants were annotated with ANNOVAR (2016-02-01 release) [13] and SnpEff v.3.6 [14]. Analysis of DNA translocation was performed using Factera v.1.4.3 [15]. CNV were analyzed based on the depth of coverage data of capture intervals. Coverage data were corrected against sequencing bias resulting from GC content and probe design. The average coverage of all captured regions was used to normalize the coverage of different samples to comparable scales. Copy number (CN) was calculated based on the ratio between the depth of coverage in tumor samples and average coverage of an adequate number ($n > 50$) of samples without copy number variations as references per capture interval. CNV is called if the coverage data of the gene region was quantitatively and statistically significant from its reference control. The limit of detection for CNV is 1.5 for CN deletion and 2.64 for CN amplifications.

Statistical analysis

Categorical variables were summarized as frequency or percentage and compared using Fisher's exact test or chi-square test. The statistical test was two-sided, and $P < 0.05$ was considered statistically significant. The P values were corrected using the error discovery rate (FDR or Q value) and multiple hypothesis testing was performed according to the Benjamin-Hochberg program.

Results

The clinical and pathological characteristics of the 18 NSCLC patients included in our study are shown in Table 1. The median age of the patients was 70.5 years. All of the patients were stage I at the time of diagnosis consisting of 78% of IA and 22% of IB cases. The distribution of molecular subtypes were 56% adenocarcinoma and 44% carcinoma. A majority (10/18, 55.6%) of the patients were histopathologically classified as T1N0Mx, and graded as well or moderately differentiated tumors.

Table 1. *Demographics and clinical characteristics of stage I NSCLC patient cohort in this study*

Parameter		Number (n=18)	Patient Percentage (%)
Gender	Male	9	50
	Female	9	50
Median Age		70.5	(56-83)
Race	White	11	61
	Black	2	11
	Undefined	5	28
	Adeocarcinoma	10	56
NSCLC	Carcinoma, Squamous cell	5	28
	Carcinoma, non-small cell	2	11
	Carcinoma, large cell	1	5
Overall Clinical Stage	Stage 1A	14	78
	Stage 1B	4	22

NSCLC: non-small-cell lung cancer

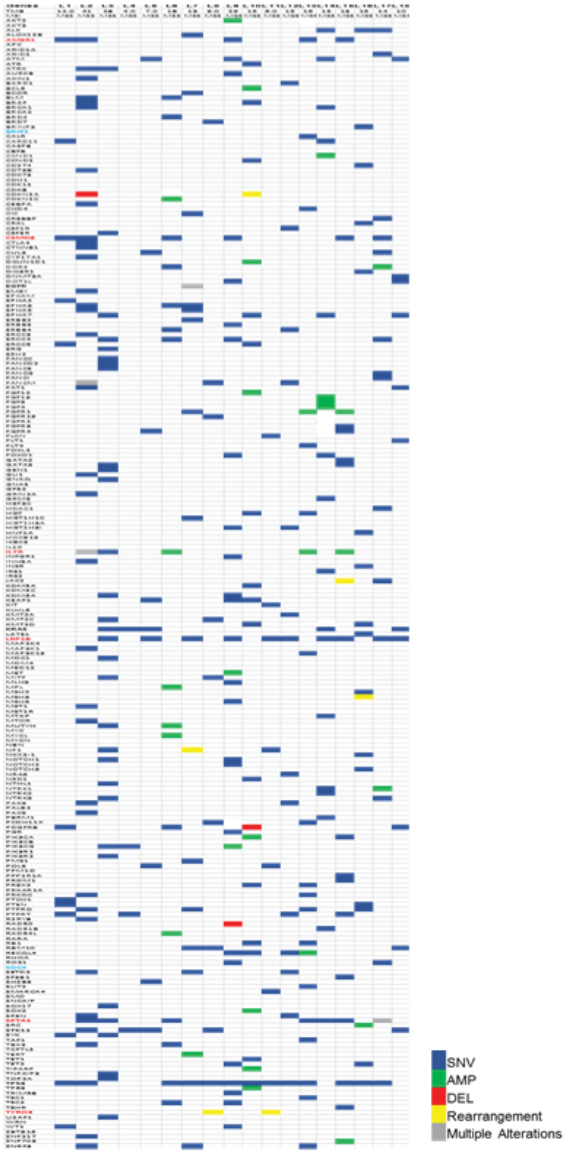

Figure 1. *Heatmap of NGS mutational profiling on 18 Stage I NSCLC tumor tissues. SNV, single nucleotide variant; AMP, amplification; DEL, deletion.*

Targeted sequencing was performed on tissue samples of 18 stage I NSCLC patients. A total of 617 somatic mutations in 234 genes were identified, all patients have at least one somatic mutation detected. Various mutation types were detected including 574 SNV, 30 CN amplification, 3 CN deletion, 6 rearrangements and 4 cases of multiple types of alterations. Figure 1 summarized the mutation landscape of stage I NSCLC tumors. TP53 (72%), LRP1B (56%), and SPTA1 (39%) were the most frequently mutated genes in our cohort. Other key genes that were mutated in ≥10% of the stage I patients included CSMD3 (33%), KRAS (28%), STK11 (28%), IL7R (28%), AMER1 (28%) and KEAP1 (17%). Most surprisingly, EGFR only displayed 5.6% mutation prevalence in stage I NSCLC tumors.

When we analyzed the mutation frequency of these candidate oncogenic drivers, using KRAS as a reference point, to identify potential stage I biomarkers, we found TP53, LRP1B, SPTA1, CSMD3, STK11, IL7R and AMER1 genes were mutated at frequencies greater than or comparable to KRAS (Figure 2), suggesting that these are stage I NSCLC biomarkers which have not been identified before. The lower prevalence of EGFR and KEAP1 mutations indicated these 2 genes could be post-stage I drivers. Other clinically relevant NSCLC drivers such as MET, ERBB2, BRAF, PIK3CA alterations, as well as ALK, ROS1, RET, NTRK fusions were also detected at much lower frequencies in our cohort.

Genomic profiling of cfDNA for diagnostic purposes, especially at an early stage of cancer is not trivial. Tumor DNA fraction in cfDNA in blood might be very low during the early non-invasive stage of the disease and can vary considerably, depending on the location of the tumor in the tissue contexts. Large cohorts are required to assess the cancer type-specific accuracy of cfDNA profiling in the early detection setting. We then performed further analysis to verify whether the stage I mutational features of tumor tissues can be recapitulated in plasma cfDNA samples. We tested the prevalence of these stage I aberrations in 5 stage II-IV NSCLC patients that have multiple longitudinal plasma samples available: 4, 12, 10, 17 and 12 serial blood samples were collected from each patient P1-5, respectively (Figure 3). In keeping with data from the stage I tumor set, frequent TP53 aberrations were found in 20% of plasma samples. In the same liquid biopsy cohort, EGFR mutations were dominated and detected in 55% of cases with the majority been exon 19 deletion, L858R and T790M, whereas only one cfDNA sample showed KRAS mutation (1.8%). We could not detect any alteration associated with two stage I drivers CSMD3 and STK11. Low mutation prevalence was also found in AMER1 (3.6%) and IL7R (1.8%), confirming both involved in early stage of cancer development. Frequencies of non-synonymous somatic mutations detected in KEAP1, SPTA1 and LRP1B genes were 20%, 11% and 7.3%, respectively (Figure 3).

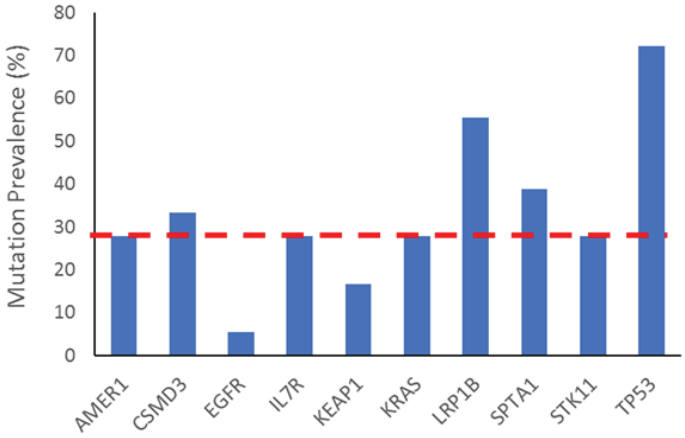

Figure 2. *Specific driver gene alterations identified in stage I NSCLC tumor tissues. The KRAS mutation prevalence served as a reference point (red dash line) for the relative dominance of each biomarker.*

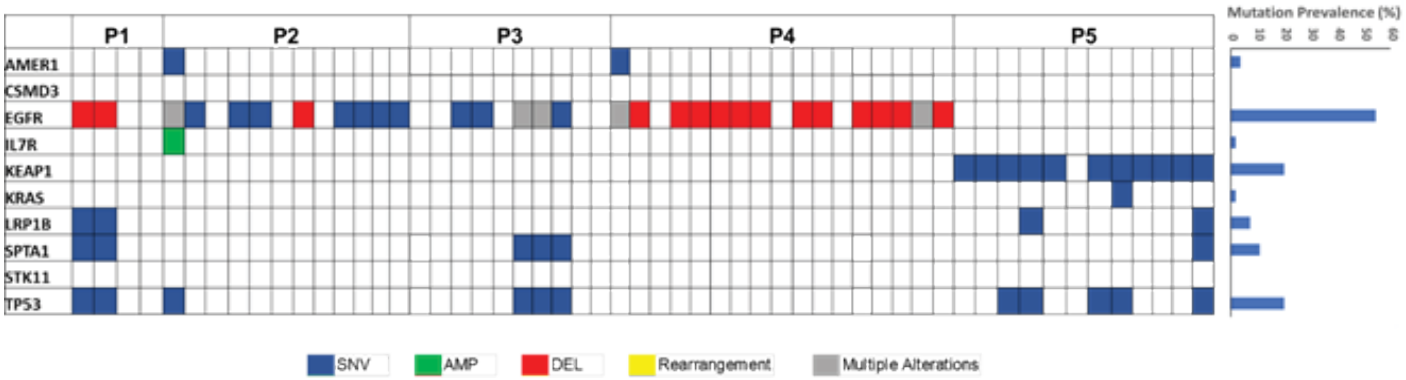

Figure 3. *Heatmap of 10 stage I biomarkers generated by NGS mutational profiling on 55 longitudinal plasma samples from 5 stage II-IV NSCLC patients (P1-P5). Mutation prevalence of each biomarker was also shown on the right side of heatmap. SNV, single nucleotide variant; AMP, amplification; DEL, deletion..*

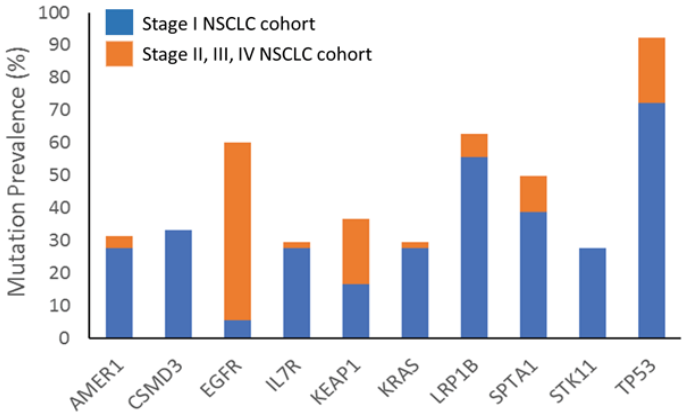

Figure 4. *Relative distribution of the 10 targeted molecular biomarkers in stage I and post-stage I NSCLC patients. .*

Lung cancer is a multistep process, characterized by the sequential accumulation of genetic and molecular abnormalities after carcinogen exposure, resulting in the clonal selection of tumor cells with uncontrolled and de-regulated cell proliferation. As shown in Figure 4, our NGS genomic profiling on both tumor tissue and plasma samples have identified novel stage I and post-stage I biomarkers in NSCLC patients, i.e., AMER1, CSMD3, IL7R, KRAS, LRP1B, SPTA1, STK11 and TP53 gene alterations were strongly associated with stage I cohort, whereas EGFR and KEAP1 were highly likely the post-stage I drivers. Our findings here was supported by the fact that there is consensus across international guidelines recommending EGFR, KRAS, BRAF, MET, ERBB2, ALK, ROS1, RET and NTRK biomarker testing for targeted therapies in advanced metastatic NSCLC patients. Unfortunately, lung cancer symptoms occur late in the disease, so the majority of patients were diagnosed late with advanced disease. Therefore, early detection via stage I biomarker discovery could be a valuable approach to detect the disease at an earliest possible, asymptomatic and potentially curable stage.

Discussion

In current clinical practice, LDCT, a standard-of-care procedure, is the only screening test performed for preventive, diagnostic, and follow-up purposes for patients undergoing radical stage I surgery. Screening high-risk population with LDCT can reduce lung cancer mortality but also causes false-positive results leading to unnecessary tests, invasive procedures and overdiagnosis. Discovery of novel stage I molecular biomarkers will bring additional crucial information to the standards of care in early detection, screening strategy and classifying subgroups of patients for adjuvant therapy. In this context, a panel of stage I biomarkers (with the assistance of AI and machine-learning algorithm) will help to establish a molecular classification of NSCLC, based not on the tumor histology, but rather on the genetic composition of each patient. We are thus striving to identify novel stage I-specific biomarkers for early detection and post-surgery MRD (molecular residual disease) of NSCLC.

The use of 530-gene NGS panel enabled and increased the detection of novel oncogenic drivers in stage I NSCLC patients. In our cohort, all of the stage I tumors showed at least one mutation across 530 cancer-related genes. TP53 (72%), LRP1B (56%), SPTA1 (39%), and CSMD3 (33%) were mutated in more than 30% of the cohort. In contrast, actionable biomarkers that have therapeutic value for NSCLC including genetic aberrations in EGFR, ROS1, ALK, MET and NTRK were detected at extremely low rate. These findings were further confirmed by NGS profiling of serial plasma samples from stage II-IV NSCLC patients. Overall, these observations strongly supported the authenticity of the 8 stage I biomarkers we identified and suggested that comprehensive genomic analysis of NSCLC is required to facilitate the discovery of novel biomarkers.

Among the 8 newly identified stage I NSCLC biomarkers, AMER1 is a gene located in chromosome X that codifies a highly conserved membrane protein that acts as scaffold for β-catenin degradation. In tumors, AMER1 is a negative regulator of the WNT/β-catenin pathway by promoting β-catenin ubiquitination and degradation. AMER1 maintains the integrity of cellular junctions by mediating the membrane localization of APC [16]. Though mutations in AMER1 have been observed in colorectal cancer, this gene has not received proper attention as a potential driver for NSCLC. IL7R is a protein-coding gene, which encodes the receptor of interleukin 7 (IL7). IL7R was significantly associated with infiltration of B cell, CD4+ T cell, DC cell, and it also affected the prognosis of patients with lung adenocarcinoma [17]. LRP1B gene (encodes the low-density lipoprotein receptor-related protein 1b) is a large gene located on chromosome 2q and is a member of the LDL receptor family. LRP1B deleterious alterations were associated with improved response rates and survival outcome when treated with immunotherapy [18]. SPTA1 is a gene that encodes an actin crosslinking and molecular scaffold protein that links the plasma membrane to the actin cytoskeleton. Patients with SPTA1 mutations had shorter TKI remission time than those without mutations [19]. Little is known about the function of CSMD3, a predictive membrane protein with a role in adhesion. Although mutations in CSMD3 were linked to a better prognosis in squamous cell carcinoma, a poor prognosis was found to be linked to mutations in TP53 or STK11 in adenocarcinoma [20].

Our genomic profiling analysis revealed the distinct mutation landscape between stage I and post-stage I cohorts, which could provide new insights into the molecular and clonal evolution between different stages of NSCLC and could shed light on early detection and novel therapeutic strategy. The mutation profiles we have observed from our cohort of stage II-IV NSCLC are

similar to those reported in other studies and are consistent with the current standard-of-care guidelines. Nevertheless, this study has some limitations. Our study only investigated the mutation landscape in 530 cancer-related genes, which might miss some novel regulators of NSCLC development. The small sample size of stage I tumor tissues also limited our analysis. Most importantly, we plan to expand our study together with AI and machine-learning algorithm, so we could determine the predictive value of these novel stage I biomarkers for early detection of NSCLC. Further research is needed to verify our findings in a multi-center study with a larger cohort.

It is our heartfelt hope that the patients will one day be screened primarily through a non-invasive molecular test, involving sample collection of their biological fluids. While liquid biopsy-based tracking of cancer progression is already integrated into the clinical management for cancer patients, leveraging it for non-invasive screening for early detection of pan-cancer risk at a low cost in a standardized setting will have important implications for clinical management of high-risk patient population and public health at large.

Conclusion

The identification and validation of stage I molecular biomarkers combining clinical and radiological data have the potential for great clinical impact on cancer early screening and detection. The development of these predictive models will not only improve risk stratification, diagnosis, and treatment of NSCLC patients, but also, provide a complete management plan integrating all patient characteristics. The clinical validation and implementation of predictive biomarkers for early detection of lung cancer has become the greatest challenge to date. The same holds true for our capacity to predict the efficacy and choose the modality of adjuvant treatments. We need a better understanding of the detailed molecular makeup of early-stage lung tumors especially at stage I and its correlation with outcomes and response to therapy.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
2. Lemjabbar-Alaoui H, Hassan OU, Yang YW, Buchanan P. Lung cancer: Biology and treatment options. Biochim Biophys Acta. 2015;1856(2):189-210.
3. Howlader N NA, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, et al. SEER Cancer Statistics Review, 1975-2017. USA: National Cancer Institute Bethesda, MD (2020).
4. National Lung Screening Trial Research Team, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395-409.
5. Rolfo C, Mack PC, Scagliotti GV, et al. Liquid Biopsy for Advanced Non-Small Cell Lung Cancer (NSCLC): A Statement Paper from the IASLC. J Thorac Oncol. 2018;13(9):1248-1268.
6. Podlaha O, Riester M, De S, Michor F. Evolution of the cancer genome. Trends Genet. 2012;28(4):155-163.
7. Gerstung M, Jolly C, Leshchiner I, et al. The evolutionary history of 2,658 cancers. Nature. 2020;578(7793):122-128..
8. Greytak SR, Engel KB, Parpart-Li S, et al. Harmonizing Cell-Free DNA Collection and Processing Practices through Evidence-Based Guidance. Clin Cancer Res. 2020;26(13):3104-3109.
9. Abbosh C, Birkbak NJ, Wilson GA, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature. 2017;545(7655):446-451.
10. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.
11. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-1303.
12. 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.
13. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
14. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80-92.
15. Newman AM, Bratman SV, Stehr H, et al. FACTERA: a practical method for the discovery of genomic rearrangements at breakpoint resolution. Bioinformatics. 2014;30(23):3390-3393.
16. Sanz-Pamplona R, Lopez-Doriga A, Paré-Brunet L, et al. Exome Sequencing Reveals AMER1 as a Frequently Mutated Gene in Colorectal Cancer. Clin Cancer Res. 2015;21(20):4709-4718.
17. Wang X, Chang S, Wang T, et al. IL7R Is Correlated With Immune Cell Infiltration in the Tumor Microenvironment of Lung Adenocarcinoma. Front Pharmacol. 2022;13:857289.
18. Brown LC, Tucker MD, Sedhom R, et al. LRP1B mutations are associated with favorable outcomes to immune checkpoint inhibitors across multiple cancer types. J Immunother Cancer. 2021;9(3):e001792.
19. Liu J, Xu WY, Ye M, Liu Z, Li C. Genetic Alteration Profiling of Chinese Lung Adenocarcinoma and Its Effect on Targeted Therapy Efficacy. Front Oncol. 2021;11:726547.
20. La Fleur L, Falk-Sörqvist E, Smeds P, et al. Mutation patterns in a population-based non-small cell lung cancer cohort and prognostic impact of concomitant mutations in KRAS and TP53 or STK11. Lung Cancer. 2019;130:50-58.