Changing the landscape of non-small cell lung cancer disparities

Joab O. Odera¹, Muthana Al Abo², Steven R. Patierno¹,², Jeffrey M. Clarke¹,²*, Jennifer A. Freedman¹,²*

¹Department of Medicine, Division of Medical Oncology, Duke University School of Medicine, Durham, NC, 27710, USA
²Duke Cancer Institute, Duke University School of Medicine, Durham, NC, 27710, USA
*These authors contributed equally to this work
*Author for correspondence: Email: jennifer.freedman@duke.edu

Abstract

In the United States, lung and bronchus cancers are the second most common types of cancer and are responsible for the largest number of deaths from cancer, with African Americans suffering disproportionately from lung and bronchus cancers. This disparity likely results from a complex interplay among social, psycho-social, lifestyle, environmental, health system, and biological determinants of health. Toward improving outcomes for lung cancer patients of all races and ethnicities and mitigating lung cancer disparities, in this commentary, we bring forward biological factors that contribute to lung cancer disparities, efforts to identify, functionally characterize, and modulate novel ancestry-related RNA splicing-related targets in lung cancer for precision intervention, and translational and clinical research needs to improve outcomes for lung cancer patients of all races and ethnicities and mitigate lung cancer disparities.

Keywords: Lung cancer, Non-small cell lung cancer, Lung squamous cell carcinoma, Lung adenocarcinoma, Disparities, Ancestry, RNA splicing, Biomarkers, Therapeutics, Immunotherapies

Introduction

In the United States, lung and bronchus cancers are the second most common types of cancer diagnosed, with an estimated 228,820 new cases in 2020, representing 12.7% of all newly diagnosed cancer cases [1]. Of these cases, 80-85% are non-small cell lung cancer (NSCLC), including adenocarcinoma, squamous cell carcinoma, and large cell carcinoma subtypes [2]. Lung and bronchus cancers are responsible for the largest number of deaths from cancer, accounting for an estimated 135,720 deaths in 2020, representing 22.4% of all cancer deaths [1]. Significant racial disparities are observed in lung cancer diagnoses and deaths, with African Americans (AAs) receiving a diagnosis of lung cancer three years younger on average than European Americans (EAs), and AAs having the highest number of deaths per 100,000 persons for lung and bronchus cancer [1,3]. Evidence suggests that cigarette smoking patterns and rates are not a primary driver of these disparities [4]. Therefore, there is an urgent need for research investigating the contribution to these disparities of other determinants of health, including additional lifestyle, environmental, social, psycho-social, health system, and biological factors.

In this commentary, we discuss biological factors that contribute to lung cancer disparities, while being keenly aware of the complex interplay of individual-, societal-, neighborhood-, and institutional-level determinants of cancer disparities and being avid proponents of the critical need for future research to address intersection among determinants of cancer disparities. Given the disproportionate burden of lung cancer within the AA community, the limited exploration of molecular mechanisms underlying lung cancer in patients of African ancestry to date, and the small number of available precision interventions, we highlight efforts to identify, functionally characterize, and modulate novel ancestry-related RNA splicing-related targets in NSCLC for precision intervention. We conclude by raising translational and clinical research needs to improve outcomes for NSCLC patients of all races and ethnicities and mitigate NSCLC disparities.

Importance of Precision Intervention in NSCLC

It is of utmost importance to use omics in translational and clinical NSCLC research to identify
genomic, epigenomic, proteomic, metabolomic, and immune-related targets and evaluate precision interventions based on such targets. A number of studies have identified TP53, EGFR, NTRK1, U2AF1, NTRK2, NTRK3, RET, MET, BRAF, ALK, ROS1, PIK3CA, and BRAF as the most commonly mutated genes in NSCLC [5]. Analyses of mutations in circulating tumor nucleic acids have also detected mutations in STK11, NRAS, and KRAS in lung squamous cell carcinoma. Comparison of smokers and non-smokers have found strong association between KRAS, STK11, NF1, KEAP1, and SMARCA4 mutations and current and former smokers. In never smokers; however, EGFR has been identified as the most common mutation [6]. Many of these commonly mutated drivers may be targeted using tyrosine kinase inhibitors [7,8]. Recent studies have reported that NSCLC patients having specific driver mutations, who received targeted therapy with corresponding tyrosine kinase inhibitors, had a better prognosis than those who received non-targeted therapy [9-11]. This work emphasizes the importance of identifying targets, developing precision interventions based on such targets, providing access to tumor profiling, and evaluating such precision interventions in NSCLC patients.

**Ancestry-related Differences in Driver Mutations and Aggregate RNA Expression in NSCLC**

Identification of targets, development of precision interventions based on such targets, tumor profiling, and evaluation of precision interventions must be done in NSCLC patients of all races and ethnicities, as ancestry-related genomic differences in NSCLC have been reported. For example, Lusk et al. compared driver mutation frequency using a targeted panel and genome-wide mutations using whole exome sequencing between AA and White NSCLC patients [12]. While the targeted panel revealed a lower driver mutation frequency in AAs compared with EAs, the whole exome sequencing revealed many mutations in such driver genes as well as in novel genes in AAs compared with EAs. Furthermore, Arauz et al. used whole exome sequencing to examine genome-wide mutations in NSCLC in AAs and reported that STK11 and RB1 mutations were higher in lung adenocarcinoma in AAs compared with lung adenocarcinoma in EAs in The Cancer Genome Atlas (TCGA) [13]. STK11 mutations were also higher in lung squamous cell carcinoma in AAs compared with lung squamous cell carcinoma in EAs in TCGA.

Ancestry-related genomic differences in NSCLC at the RNA-level have also been reported. For example, Mitchell et al. used mRNA and miRNA expression arrays to investigate transcriptomic differences in NSCLC in AAs compared with EAs [5]. Differential aggregate RNA expression in NSCLC in AAs was particularly enriched in stem cell and invasion pathways, whereas differential aggregate RNA expression in NSCLC in EAs was particularly enriched in cell cycle, mitosis, and proliferation pathways. Differences in aggregate RNA expression signatures in NSCLC between AAs and EAs translated to differences in predicted drug responses. Ancestry-related differences in miRNA expression were also reported, with seven miRNAs identified to be differentially expressed in NSCLC in AAs and ten miRNAs identified to be differentially expressed in NSCLC in EAs. Hypergeometric studies suggested that the ancestry-related miRNA expression signatures may drive the ancestry-related aggregate RNA expression patterns observed in the NSCLCs.

As the aforementioned studies exemplify, the limited exploration of molecular mechanisms underlying lung cancer in patients of African ancestry to date have largely focused on ancestry-related differences in driver mutations and aggregate RNA expression. These approaches have likely missed important ancestry-related drivers of NSCLC biological and clinical heterogeneity and have uncovered a limited number of targets that are therapeutically actionable. Therefore, there is an urgent need to explore additional molecular mechanisms underlying lung cancer in patients of African ancestry.

**RNA Splicing in NSCLC**

RNA splicing is a key step in gene expression in higher eukaryotes, being the physiological process that generates different mRNA variants and thus proteoforms from the same sequence of DNA [14]. Such proteoforms can have functions that are as distinct as proteins that are encoded by different genes [15]. Dysregulation of RNA splicing can result in disease, including cancer and, just as RNA splicing drives evolutionary biological diversity, RNA splicing is emerging as a major driver of tumor-related biological diversity [16,17]. Variation in cis-acting splicing elements, differential expression of trans-acting splicing factors, or mutation in genes encoding components of the RNA splicing machinery can all alter RNA splicing and result in disease, including cancer [16]. Recent studies from our laboratory and others have identified RNA splicing-related genetic and genomic variation in tumors, oncogenes dysregulated by RNA splicing, RNA splice variants driving race-related cancer aggressiveness and drug response, spliceosome-dependent transformation, and RNA splicing-related immunogenic epitopes in cancer [18,19].

Recent studies have begun to explore RNA splicing in NSCLC. For example, Zhuhong et al. have analyzed RNA splicing in lung adenocarcinoma in TCGA, with RNA splice variants identified playing a role in biological processes and associating with patient survival [20]. In addition, Li et al. have analyzed lung adenocarcinoma in TCGA and identified RNA binding proteins that were related to RNA splicing, distinguished lung adenocarcinoma, and associated with patient survival [21]. Moreover, Brooks et al. have analyzed lung adenocarcinoma in TCGA with mutations in the RNA splicing factor U2 Small Nuclear RNA Auxiliary Factor 1 (U2AF1) and identified alterations in RNA splicing that were associated with such mutations [22]. Furthermore, Vinayawanuattikun et al. have reported a higher number of mutations in genes encoding RNA splicing factors in formerly bronchiolo-alveolar carcinoma that had been reclassified as preinvasive/local invasive/invasive lesions compared with lung adenocarcinoma [23]. Finally, several studies have identified RNA splicing signatures having prognostic value in NSCLC [24-27]. Zhao et al. have reported that RNA splicing events associated with survival in NSCLC were enriched in epithelial-mesenchymal transition, cell adhesion, and cell cycle pathways [28]. In addition, the integrin A9B1 pathway was enriched in lung adenocarcinoma in males, MYC repression was enriched in lung adenocarcinoma in females, the VEGFR1/2 and FAK pathways were enriched in lung squamous cell carcinoma in males, and p53 binding was enriched in lung squamous cell carcinoma in females.

**RNA Splicing in NSCLC Disparities**

With the goal of further understanding molecular mechanisms underlying NSCLC disparities and providing an abundance of novel RNA splicing-related targets for development of new biomarkers and therapeutic agents for NSCLC in patients of African and European ancestry, our team conducted two studies. In the first
study, using ancestral genotyping and Affymetrix Clarion D arrays, we reported differences in RNA splicing and aggregate RNA expression in lung squamous cell carcinoma between patients of West African and European ancestry who had a history of smoking [29]. These differences were identified in genes previously reported in NSCLC as well as in novel genes. The observed number of differentially spliced genes exceeded the observed number of genes differing in aggregate RNA expression, supporting RNA splicing as a major driver of tumor-related biological diversity. Almost 18% of the differentially spliced genes and almost 11% of the genes differing in aggregate RNA expression had been previously reported to be drivers, oncogenes, and/or tumor suppressor genes. Among the differentially spliced genes and genes differing in aggregate RNA expression, biological processes enriched included metabolic process, biological regulation, and multicellular organismal process and, among DSGs, ion transport. Overlapping canonical pathways among differentially spliced genes included neuronal signaling pathways and among genes differing in aggregate RNA expression included cell metabolism involving biosynthesis. Differentially spliced genes were enriched in KRAS Signaling, UV Response, E2F Targets, Glycolysis, and Coagulation. A number of the differential RNA splicing events and the genes differing in aggregate RNA expression potentially associated with lung squamous cell carcinoma patient survival.

In the second study, using TCGA, we reported differences in alternative RNA splicing and transcription events between NSCLC from self-identified AA and White patients [30]. In both lung adenocarcinoma and lung squamous cell carcinoma, the most common race-related alternative RNA splicing, and transcription events were exon skipping and alternative promoter followed by alternative terminator followed by retained intron followed by mutually exclusive and alternative donor. Comparing race-related alternative RNA splicing and transcription events among NSCLC, breast cancer, colon cancer, and prostate cancer, the majority of the events were unique to lung adenocarcinoma or lung squamous cell carcinoma, but a number of the events overlapped with those in the other aforementioned cancers. Gene sets enriched among DSGs in alternative RNA splicing and transcription events in NSCLC included immune-related signaling and MTORC1 signaling. In addition, in lung adenocarcinoma, gene sets included adipogenesis and spermatogenesis and in lung squamous cell carcinoma, gene sets included G2M checkpoint, androgen response, apical junction/surface, epithelial mesenchymal transition, and UV response. A number of the race-related alternative RNA splicing and transcription events in NSCLC associated with patient survival. In lung squamous cell carcinoma, an alternative promoter event in MET associated with patient survival. MXI1 has been reported to function as an antagonist for MYC transcriptional activity and to play a role in lung cancer progression and radioresistance [31,32]. This event was identified to be associated with lung squamous cell carcinoma patient overall survival among AA patients, but not among White patients [30]. Additional examples of race-related alternative RNA splicing and transcription events associated with lung squamous cell carcinoma patient overall survival included an exon 3.1 alternative acceptor event in ARGLIP4 and an alternative promoter event in NPIPB4. The association between the level of these two events with lung squamous cell carcinoma patient overall survival was identified by two independent studies using TCGA SpliceSeq and TCGA clinical datasets [30,33]. Furthermore, Al Abo et al. examined the association between these events and patient survival among AA patients and White patients separately and found that the level of ARGLIP4 and NPIPB4 alternative RNA splicing and transcription events were associated with lung squamous cell carcinoma patient overall survival among White patients, but not among AA patients. Likewise, a number of race-related alternative RNA splicing and transcription events in lung adenocarcinoma were associated with patient survival, including an alternative promoter event in TRXAS1, whose low aggregate RNA expression has been reported in high grade breast cancer tumors and has been reported to be associated with poor prognosis [34]. In addition, an exon 2.2 alternative donor event in BANF1 was found to be associated with lung adenocarcinoma patient survival among AA patients, but not among White patients. BANF1 has been reported to play a crucial role in resetting oxidative stress induced by PARP1 activity [35].

Implications of Ancestry-related RNA Splicing for Precision Interventions in NSCLC

Biomarkers

RNA splice variants can serve as biomarkers for cancer detection. An example is b-variant CDKN1A Interacting Zinc Finger Protein 1 (CIZ1). CIZ1 is a nuclear matrix protein that cooperates with Cyclin A2 and Cyclin Dependent Kinase 2 (CDK2) to promote mammalian DNA replication [36]. It is also needed for recruiting Cyclin A and Cyclin E, and thus also plays a role in S-phase transition [37,38]. In vitro suppression of b-variant CIZ1 has been reported to inhibit NSCLC cell proliferation [39], b-variant CIZ1 has been shown to be present in lung tumors, but not adjacent tissue, and in plasma from lung cancer patients [39,40]. In addition, RNA splice variants can serve as biomarkers for drug response. Alterations in MET, which encodes a receptor tyrosine kinase that plays a role in cellular survival, embryogenesis, cellular migration, and cellular invasion, have been shown to drive oncogenesis, including NSCLC [41]. One class of MET alterations reported in NSCLC specimens affects cis-acting splicing elements controlling RNA splicing of exon 14 resulting in skipping of exon 14 and activation of MET [42-44]. Indeed, multiple therapies are now FDA approved for METex14 driven lung cancer including tepotinib and capmatinib [45-47]. Another example of an RNA splice variant that has potential to serve as a biomarker for drug response involves the Erib-B2 Receptor Tyrosine Kinase 2 (ERBB2) Ace16 variant. Recent work has shown its ability to transform immortalized epithelial cells and induce resistance to afatinib, a pan-ERBB family tyrosine kinase inhibitor [48]. Furthermore, this variant has also been shown recently to induce resistance to osimertinib, a third generation EGFR tyrosine kinase inhibitor [49]. And, as mentioned previously, several studies have identified RNA splicing signatures having prognostic value in NSCLC [24-28]. Similar to the aforementioned examples, ancestry-related RNA splice variants have potential to serve as biomarkers for detection, drug response, and/or survival in NSCLC patients of African and European ancestry.

Therapeutics

Splice-switching oligonucleotides (SSOs) comprise a promising therapeutic strategy to combat human disease by modulation of RNA splicing. Unlike standard RNA interference to inhibit gene expression, SSOs simultaneously limit production of pathogenic proteins and maintain/induce expression of protein variants with therapeutic value. SSOs modulate RNA splicing by binding to target pre-mRNAs and blocking access of the splicing machinery to
a particular splice site [50,51]. One example of a SSO, Nusinersen/Spinraza, is proving efficacious in clinical trials for treatment of spinal muscular atrophy [52]. Spinal muscular atrophy patients have a mutation in the Survival Motor Neuron 1 gene that interferes with RNA splicing and also in the duplicate Survival Motor Neuron 2 gene, resulting in production of a non-functional Survival Motor Neuron protein. Nusinersen/Spinraza binds to the Survival Motor Neuron pre-mRNA and induces inclusion of an exon, resulting in functional Spinal Motor Neuron protein. The phosphorothioate 2’-O-methylxoyethyl oligonucleotide Nusinersen/Spinraza has been evaluated in four clinical trials Phase 1-3 using intrathecal dosing and based on positive motor response in patients enrolled in such trials, Nusinersen/Spinraza received approval by the Food and Drug Administration (FDA). Another example of a SSO, but one which is being evaluated using systemic dosing, is Eteplirsen, which is proving efficacious in clinical trials for treatment of Duchenne muscular dystrophy [52]. Most Duchenne muscular dystrophy patients have mutations in the gene encoding Dystrophin. Such mutations cause premature translation termination and result in a truncated, unstable, non-functional Dystrophin protein. Eteplirsen binds to the Dystrophin pre-mRNA and induces skipping of exons containing such mutations, thus enabling translation to continue and resulting in an internally deleted, but partially functional Dystrophin protein. Initially, the phosphodiamidate morpholino oligomer Eteplirsen was evaluated by injection intramuscularly in the extensor digitorum brevis muscle of Duchenne muscular dystrophy patients and then long-term systemic dosing was evaluated in Duchenne muscular dystrophy patients. Based on the increase in Dystrophin in patients enrolled in such trials, Eteplirsen received accelerated approval by the FDA. In addition to using oligonucleotide-based approaches to modulate RNA splicing for therapeutic application, small molecule-based approaches are also showing promise. Similar to Nusinersen/Spinraza, Risdiplam is a small molecule that induces inclusion of an exon, resulting in functional Spinal Motor Neuron protein. The small molecule Risdiplam has been evaluated in two clinical trials Phase 2-3 using oral dosing and based on positive ability to sit independently and motor function, survival without permanent ventilation, and increased expression of functional SMN protein in patients enrolled in such trials, Risdiplam received approval by the FDA [53-55]. Such splice-switching oligonucleotide- and small molecule-based approaches to modulate RNA splicing for therapeutic application have potential to be expanded from neuromuscular diseases to cancer. Ancestry-related RNA splice variants have potential to be targeted by such approaches for therapeutic application to improve survival of NSCLC patients of African and European ancestry and mitigate NSCLC disparities.

Immunotherapies

Deriving antigens for vaccines and T cell therapies is a considerable challenge in precision oncology. Tumor-specific RNA splicing presents a large new class of RNA splicing associated potential neoantigens that could affect the immune response and could be exploited for cancer immunotherapy [56]. Using TCGA, The Genotype-Tissue Expression project, and Clinical Proteomic Tumor Analysis Consortium data, Kahles et al. have identified tumor-specific RNA splicing events that result in tumor-specific exon-exon neojunctions and neojunction and single nucleotide variant-derived peptides predicted to bind to major histocompatibility complex-1 that result in tumor-specific putative neoantigens [57]. In addition, Oka et al. have evaluated binding of RNA splice variants to HLA molecules and identified putative neoantigens in NSCLC [58]. Moreover, Oka et al. have identified aberrant RNA splice variants and peptides that are putative activators of T cell responses. Some such RNA splice variants are likely to be ancestry-related and have potential to serve as antigens for vaccines and T cell therapies in NSCLC patients of African and European ancestry.

Conclusions and Future Directions

Precision interventions to improve outcomes for NSCLC patients and mitigate NSCLC disparities are in part dependent on fully understanding molecular mechanisms driving NSCLC in patients of all races and ethnicities and developing biomarkers and therapeutics based on such mechanisms. As discussed herein, ancestry-related RNA splicing is emerging as one such mechanism and has the potential to aid in development of novel precision interventions for NSCLC patients of African and European ancestry. To meet the goal of improving outcomes for NSCLC patients and mitigating NSCLC disparities, larger numbers of annotated biospecimens from NSCLC patients of color need to be accessible for analyses and larger numbers of NSCLC patients of color need to be represented in clinical trials. Such cohorts will enable more robust analyses, allow analyses to take into account all clinically relevant parameters, empower analyses to intersect social, psycho-social, lifestyle, environmental, health system, and biological determinants of health, and permit evaluation of therapeutic regimens by ancestry. The clinical relevance of intersection between determinants of health has been established; for example, Wang et al. have reported association between smoking history and KRAS and EGFR mutations as well as tumor mutation burden in NSCLC [6]. In addition, although we have focused herein on DNA- and RNA-level molecular mechanisms driving NSCLC disparities, further study of ancestry-related epigenomic, proteomic, metabolomics, and immune-related mechanisms underlying NSCLC is needed. Moreover, although we have focused herein on ancestry-related biological determinants of health in NSCLC patients of African and European ancestry, study of ancestry-related determinants of health in NSCLC patients of all races and ethnicities are needed. Association between other ancestries and DNA mutation has been revealed; for example, Carrot-Zhang et al. have reported association between Native American ancestry and EGFR, KRAS, and STK11 mutations as well as tumor mutation burden in lung cancer [59]. Finally, it will be of utmost importance for translational and clinical research in NSCLC going forward to take into account both self-identified race and ethnicity and ancestry to enable evaluation of social, psycho-social, lifestyle, environmental, health system, and biological determinants of health that impact NSCLC outcomes [60].

Conflicts of Interest

Authors declare no conflicts of interest.

Funding

This work was partially supported by a NIH Feasibility Studies to Build Collaborative Partnerships in Cancer Research P20 Award 1P20-CA202925-01A1 to SRP Overall PI and JAF PI of Pilot Project One, a NIH Basic Research in Cancer Health Disparities R01 Award R01CA220314 to SRP PI and JAF Co-I, a Prostate Cancer Foundation Movember Challenge Award #18CHAL04 to SRP and JAF Multi-PI, a NIH Feasibility and Planning Studies for Development of Specialized Programs of Research Excellence.
Author Contributions Statement


References


