BIOL 4225
Christopher Thompson
Fall 2024

Somatic point mutation profiles for lung adenocarcinoma patients diagnosed at older vs. younger age

Objective

Lung cancer has proven difficult to treat, and it is the leading estimated cause of cancer death for both men and women[1]. Understanding the driver mutations of all cancers is beneficial for treatment discovery as we move through an age of more targeted cancer therapies that aim to combat these driver mutations. For example, EGFR, which is a known cancer gene in lung adenocarcinoma, is frequently caused by the “L858R” point mutation on exon 21[2]. Patients with this L858R mutation often respond well to tyrosine kinase inhibition, combating the effects of the mutation[2]. Using transcriptomic data on lung adenocarcinoma patients, it is possible to gain a better understanding of the driver mutations and develop new treatments targeting these mutations. Many factors play a role in the types and amounts of mutations on known oncogenes. Age, smoking status, whether a patient has diabetes are examples of such factors that can cause different mutations in the oncogenes that lead to lung adenocarcinoma. The objective of this study is to evaluate the differences in the amount and types of somatic point mutations present on seven known cancer driving genes (EGFR, KRAS, NRAS, PIK3CA, BRAF, CTNNB1, and MET) in lung adenocarcinoma patients based on age. We will take transcriptome sequences from both healthy and unhealthy tissue samples from 2 patients diagnosed with lung adenocarcinoma above the age of 60 years and 2 patients diagnosed with lung adenocarcinoma under the age of 60 years. We hypothesize that patients who were diagnosed above the age of 60 will harbor more somatic point mutations on cancer driving genes than those diagnosed under the age of 60.

BIOL 4225
Christopher Thompson
Fall 2024

Background

In 2012, a study titled “The transcriptional landscape and mutational profile of lung adenocarcinoma” by Seo et al. analyzed the transcriptome and whole exome of 76 lung adenocarcinoma patients[3]. 200 lung adenocarcinoma biopsies were collected from 200 different Korean lung cancer patients. Of the 200 biopsies collected, 87 samples had driver mutations that were not detected by screening, and of these, 77 underwent transcriptome and whole exome sequencing for further analysis[3]. The study discovered a number of possible new lung adenocarcinoma driver mutations including SNPs, short indel mutations, fusions, and splicing events[3]. The study did not do this cross analysis in relation to age of diagnosis and mutational profile. However, a more recent study, “Next-generation Sequencing Reveals Age-dependent Genetic Underpinnings in Lung adenocarcinoma” by Wu et al. evaluated 2025 Chinese lung cancer patients who underwent next-generation sequencing from 2014-2019[4]. The mutational profiles of lung cancer genes among the patients in the study were analyzed and organized based on age of the patient. This study found that 20 known somatic mutations of interest were reported at significantly higher rates for elderly patients (diagnosed after 50 years) than younger patients (diagnosed before 50 years) and found 14 unique mutations in the elderly population of the study not found in the younger population[4]. Understanding the mutational profile of cancer genes in elder populations is crucial as researchers try and answer questions related to treatment via targeted therapies and immunotherapies, especially given that over 40% of lung cancer patients are diagnosed over the age of 70 and that it is commonly found as a high-incidence malignant tumor in patients diagnosed over the age of 70[4]. Through this study, we hope to build on previous work to provide even more answers as to the type and frequency of known somatic point mutations present in patients diagnosed at an older age vs. younger age with the hope that better, more targeted treatment can result from this information.

BIOL 4225
Christopher Thompson
Fall 2024

Methods

The patient data from the 2012 study, “The transcriptional landscape and mutational profile of lung adenocarcinoma” by Seo et al. including patient demographic information, the unhealthy tissue transcriptomes, and adjacent healthy tissue transcriptome data for each patient can be found from the SRA database through NCBI. We downloaded the healthy and unhealthy transcriptome data for 2 patients above the age of 60 at diagnosis and 2 patients below the age of 60 at diagnosis. The patients utilized were LC_C1, LC_C7, LC_S14, and LC_S20) who’s SRA values for healthy and unhealthy transcriptome were found from NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE40419. Following this, the reads from each transcriptome were trimmed to remove bases with quality scores <10 and length =>80. Following trimming, the reads for the healthy and unhealthy tissue trimmed transcriptomes were mapped with and aligned to the GRCh38.p14 transcriptome assembly as a reference using BBMAP.sh. We then sorted and indexed the resulting .bam file and ran analysis using bcftools to identify SNPs and short indel variations on the known cancer driving genes of interest (EGFR, KRAS, NRAS, PIK3CA, BRAF, CTNNB1, and MET) for the healthy and unhealthy transcriptomes. To help distinguish the somatic point mutations vs. germline mutations for each patient, we filtered out the mutations that appeared on the gene of interest in the healthy and unhealthy tissue transcriptomes. Following this an organization of the mutational profiles on the genes of interest for each patient was done using the pyVCF library. This data was organized to create the visual representations of the data shown in the figures within the results.

BIOL 4225
Christopher Thompson
Fall 2024

Results

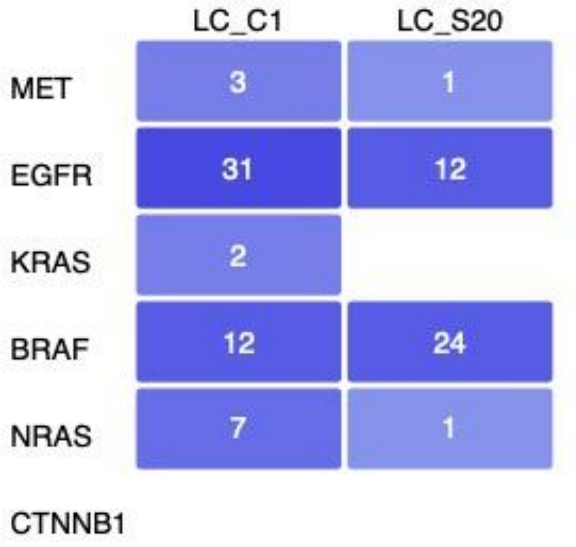

Figure 1A. Mutational profile heatmap for patients diagnosed under the age of 60 years

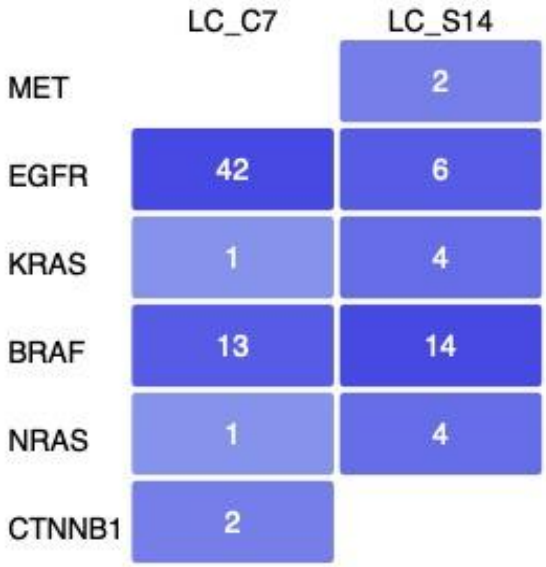

Figure 1B. Mutational profile heatmap for patients diagnosed after the age of 60 years

BIOL 4225
Christopher Thompson
Fall 2024

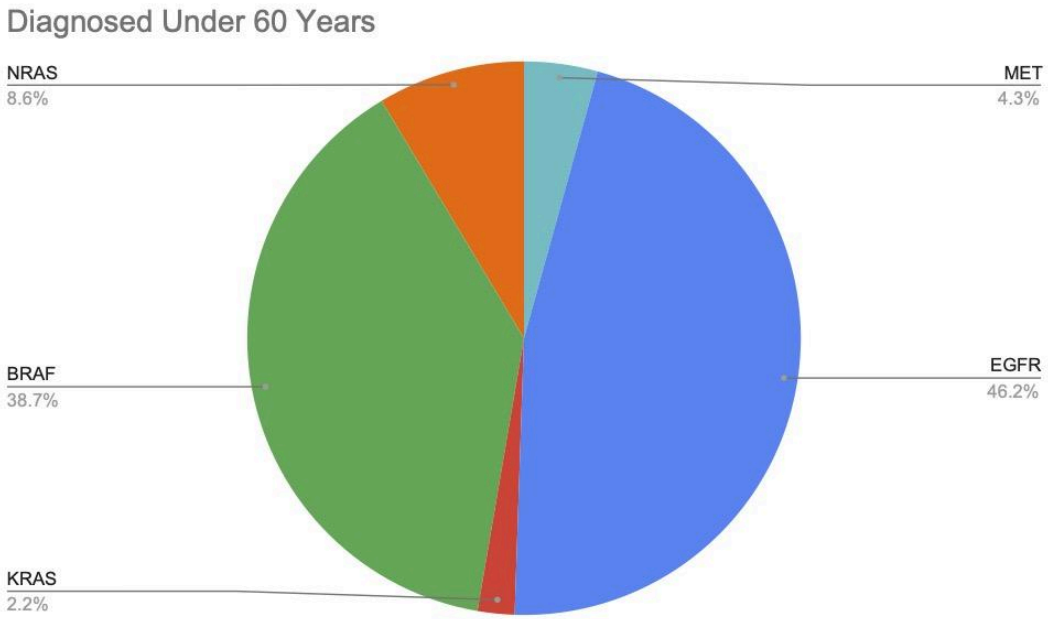

Figure 2A. Mutational profile pie chart for patients diagnosed under the age of 60

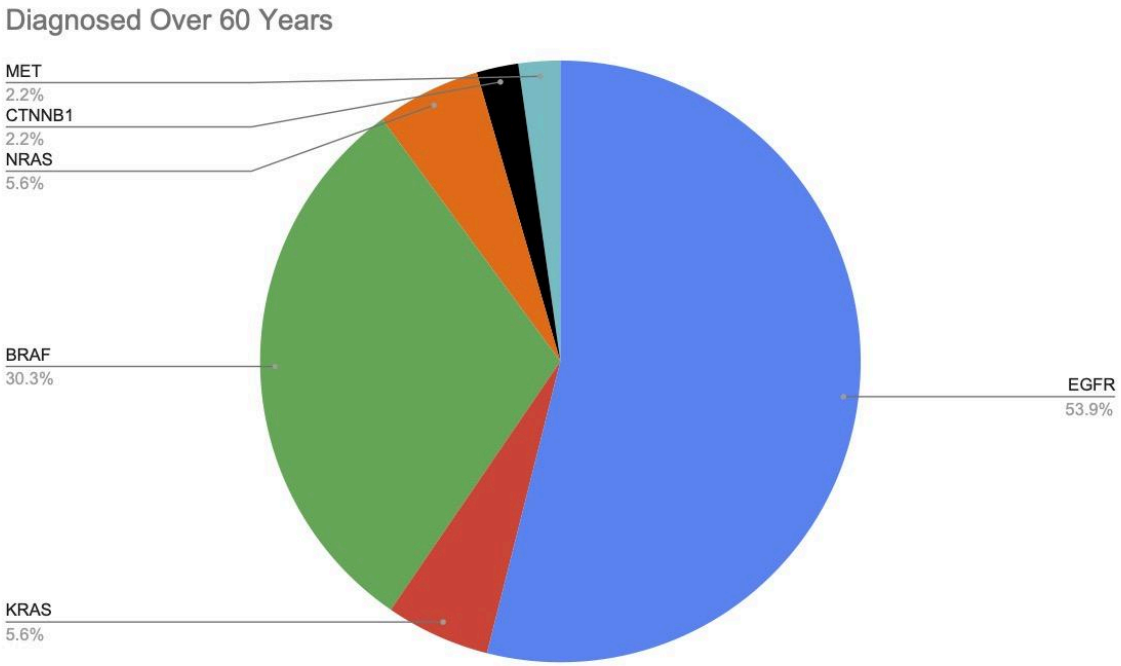

Figure 2B. Mutational profile pie chart for patients diagnosed over the age of 60

Figures 1A and 1B show the mutational profile for each patient in each patient in this study. The total number of mutations in patients diagnosed under the age of 60 was 93 in the 7 cancer driving genes selected, and the total number of mutations in patients diagnosed over the age of 60 was 89 in the 7 cancer genes of interest. Figures 2A and 2B outline the mutational profiles in pie charts, and more clearly displays the similarities and differences in the relative amounts and types of mutations in both groups (patients diagnosed under the age of 60 (Figure 2A) and patients diagnosed over the age of 60 (Figure 2B)). EGFR and BRAF were the genes with the highest number of mutations in both groups.

BIOL 4225
Christopher Thompson
Fall 2024

Discussion

The results of this study showed that patients diagnosed after the age of 60 had fewer total somatic point mutations than patients diagnosed before the age of 60 on 7 known cancer driving genes (EGFR, BRAF, KRAS, NRAS, CTNNB1, MET, and PIK3CA). This result is inconsistent with the hypothesis of the study, that patients diagnosed at an older age (above 60 years) would harbor more somatic point mutations on the 7 cancer-driving genes of interest than patients diagnosed at a younger age (under 60 years).

While the total number of somatic point mutations was less for patients diagnosed at older ages in this study, patients diagnosed after the age of 60 years had more genes with somatic point mutations than patients diagnosed below 60 years; the pie charts (figures 2A and 2B) show that patients diagnosed above the age of 60 had mutations on CTNNB1, and the CTNNB1 gene had no somatic point mutations for patients younger than 60 years. This could be an indication that patients diagnosed above the age of 60 may have more genes affected by somatic point mutations than those diagnosed at younger ages. This is consistent with the background information that lead to the hypothesis of this study, the fact that genetic mutations accumulate as people age, and this in theory would lead to more genes being affected by the increase in genetic mutations as aging proceeds.

The results of this study also shed light on the similarity in mutational profiles between the two groups (diagnosed over 60 years and diagnosed after 60 years). This is also shown most clearly in figures 2A and 2B. EGFR and BRAF harbored the vast majority of mutations for both groups.

There were a few drawbacks to the design and implementation of the study. Namely, the number of subjects in the study was low. Only 4 patients were selected; 2 in the diagnosed

above 60 years group and 2 in the diagnosed under 60 years group. Additionally, the patients were only chosen based on their age. Future studies would likely benefit from selecting patients with other factors being aligned such as smoking status, gender, and cancer stage. All three of these factors affect the amount and type of mutations present in lung adenocarcinoma.

Understanding the mutational landscape of lung adenocarcinoma, the amounts and types of mutations present in patients diagnosed at older versus younger ages, could lead to an improved understanding of the underlying mechanisms driving cancers at different ages. This understanding could lead to more targeted therapies to be developed to better treat lung adenocarcinoma diagnosed at different ages.

BIOL 4225
Christopher Thompson
Fall 2024

References

1. Dela Cruz, Charles S., Lynn T. Tanoue, and Richard A. Matthay. “Lung Cancer: Epidemiology, Etiology, and Prevention.” *Clinics in Chest Medicine* 32, no. 4 (December 2011): 10.1016/j.ccm.2011.09.001. https://doi.org/10.1016/j.ccm.2011.09.001.

2. Li, Allan R., Dhananjay Chitale, Gregory J. Riely, William Pao, Vincent A. Miller, Maureen F. Zakowski, Valerie Rusch, Mark G. Kris, and Marc Ladanyi. “EGFR Mutations in Lung Adenocarcinomas.” *The Journal of Molecular Diagnostics : JMD* 10, no. 3 (May 2008): 242–48. https://doi.org/10.2353/jmoldx.2008.070178.

3. Seo, Jeong-Sun, Young Seok Ju, Won-Chul Lee, Jong-Yeon Shin, June Koo Lee, Thomas Bleazard, Junho Lee, et al. “The Transcriptional Landscape and Mutational Profile of Lung Adenocarcinoma.” *Genome Research* 22, no. 11 (November 1, 2012): 2109–19. https://doi.org/10.1101/gr.145144.112.

4. Wu, Xiaonan, Jun Zhao, Ling Yang, Xin Nie, Zheng Wang, Ping Zhang, Chao Li, et al. “Next-Generation Sequencing Reveals Age-Dependent Genetic Underpinnings in Lung Adenocarcinoma.” *Journal of Cancer* 13, no. 5 (March 6, 2022): 1565–72. https://doi.org/10.7150/jca.65370.