**Supplement to****Genomic Profiling of Bronchoalveolar Lavage Fluid in Lung Cancer**

**Authors:** Viswam S. Nair,1,2,3†\* Angela Bik-Yu Hui,4† Jacob J. Chabon,4\*\* Mohammad S. Esfahani,4 Henning Stehr,5 Barzin Y. Nabet,4◊ Li Zhou,4§ Aadel A. Chaudhuri,6¶ Jalen Benson,7 Kelsey Ayers,7 Harmeet Bedi,8 Meghan Ramsey,8 Ryan Van Wert,8 Sanja Antic, 9 Natalie Lui,7 Leah Backhus,7 Mark Berry,7 Arthur W. Sung,8 Pierre P. Massion,9∆ Joseph B. Shrager,7 Ash A. Alizadeh,4,10,11\* Maximilian Diehn4,6,11\*

**Affiliations:**

1 Clinical Research Division, Fred Hutchinson Cancer Research Center

2 Division of Pulmonary, Critical Care & Sleep Medicine, University of Washington School of Medicine

3 Department of Radiology, Stanford University School of Medicine

4 Stanford Cancer Institute, Stanford University School of Medicine

5 Department of Pathology, Stanford University School of Medicine

6 Department of Radiation Oncology, Stanford University School of Medicine

7 Division of Thoracic Surgery, Stanford University School of Medicine

8 Division of Pulmonary, Allergy & Critical Care Medicine, Stanford University School of Medicine

9 Division of Allergy, Pulmonary & Critical Care Medicine, Vanderbilt University School of Medicine

10 Division of Oncology, Department of Medicine, Stanford University School of Medicine

11 Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine

† These authors contributed equally to the manuscript

\* Co-corresponding authors

\*\* Current affiliation: Foresight Diagnostics, Aurora, CO

◊ Current affiliation: Department of Oncology Biomarker Development, Genentech, South San Francisco, CA

¶ Current affiliation: Department of Radiation Oncology, Washington University School of Medicine, St. Louis, MO

§ Current affiliation: Grail Inc, Menlo Park, CA

∆ Deceased

**Supplementary Methods**

*Stanford*. All patients were consented for enrollment prior to sample collection and with the Institutional Review Board (IRB) approval at Stanford Health Care. Demographic data were collected from interview and the medical record. Smoking status was defined as current, ever, or never. Tumor profiling of biopsies and surgically resected specimens was performed with two versions of the Stanford Actionable Mutation Panel (STAMP). STAMP is a clinically validated targeted next-generation sequencing assay used for clinical tumor genotyping at Stanford. The first version was in clinical use from 2015 to 2016 and targeted 302 kb of genomic space in 198 genes. The second version was used subsequently that targets 232 kb in 150 genes. STAMP tests for single nucleotide variants, small insertions and deletions, selected copy number alterations and selected gene fusions. STAMP testing was performed in a CLIA certified lab.

*Vanderbilt*. Patients were enrolled as part of VICC THO 0398 “Molecular Predictors of Lung Cancer Behavior” and VICC THO 1078 “Nashville Early Diagnosis of Lung Cancer Project.” Demographic data and samples (see below for collection method) were collected by the Primary Investigator (PMM) after institutional IRB approval.

**Sample collection**

*Stanford*. Prior to bronchoscopy, phlebotomy was performed to collect blood for cfDNA analysis in K2EDTA tubes (Becton Dickinson, Franklin Lakes, NJ). Blood was processed within an hour to plasma and the leukocyte fraction using a refrigerated centrifuge at 4oC at 1200-1500 rcf for 10 minutes. The plasma was pooled and re-spun at 1200-1500 rcf for 10 minutes then stored in 0.5-1 ml aliquots at -80oC. Leukocytes were separated from RBCs after plasma pooling by gentle pipetting and were stored in at -80oC for DNA extraction to identify germline variants.

Patients had BAL performed during bronchoscopic procedures or prior to resection in the operating theatre during airway inspection. During bronchoscopy, the proceduralist (VN, AS, RVW, HB, MR, NL, LB, MB, JBS) was instructed to lavage in the airway leading to the tumor and prior to biopsy when possible. Lavage was performed using 45 to 60 mL of room temperature normal saline in 1-2 fractions after wedging the bronchoscope. The aspirated return using manual suction through a syringe was then placed into 1:100 v/v 0.5 M EDTA for cell stabilization (for all but the first 8 samples enrolled) and sent for processing within 1 hour of collection. BAL fluid was spun at 1200-1500 rcf for 10 minutes at 4oC and the cell free supernatant decanted off from the cell pellet. Both fractions were stored at -80oC for further processing to DNA.

*Vanderbilt.* Each patient had 150 ml of normal saline injected into the lingula of the left lung or the middle lobe of the right lung and the aspirate was spun down and the supernatant stored at -80oC. Blood was drawn at the same time of the procedure into K2EDTA or Streck™ blood tubes and the plasma and PMBC fraction stored at -80oC. PBMCs were processed to genomic DNA at Vanderbilt using Puregene chemistry on an Autopure robot and shipped to Stanford along with plasma aliquots (~4 ml per patient) and BAL fluid (~15 ml per patient).

**DNA extraction**

Stored plasma and BAL fluid aliquots were thawed at room temperature and centrifuged at 13,000 rpm for 15 minutes to remove residual cellular components. Cell free DNA (cfDNA) was extracted using the QiaAmp Circulating Nucleic Acid Kit (Qiagen) according to manufacturer’s instructions. Genomic DNA from whole blood was isolated with DNeasy Blood & Tissue Kit (Qiagen). Lavage cell pellets were processed in the same manner as whole blood. Following isolation, DNA was quantified using the Qubit dsDNA High Sensitivity Kit (Life Technologies). Genomic DNA and lavage fluid / cell pellet DNA was fragmented to smaller sizes using Covaris S2 instrument as described previously (12).

**Library preparation**

Sequencing library preparation was performed using the KAPA LTP Library Prep Kit according to the manufacturer’s protocol with some modifications (Kapa Biosystems), as previously described (17, 19). Briefly, genomic DNA was first subjected to End Repair & A-Tailing, followed by ligation with custom made adapters. This was followed by a hybridization-based enrichment of specific sequences using a custom designed pool of biotinylated DNA oligonucleotides (Roche NimbleGen).

**Statistical analyses**

All data were categorized as continuous or categorical and the distribution of the data assessed in order to apply the correct statistical test for analysis of differences. For independent groups (i.e. case-control or male vs. female), a Student’s t-test or Wilcoxon log-rank test was used to compare continuous variables in two groups. Chi-squared or Fisher exact test’s (for n ≤ 5) were used to compare categorical variables between independent groups. For paired groups (i.e. biofluid comparisons in the cancer cohort), we utilized a Wilcoxon matched-pair signed-rank test for continuous variables and a McNemar’s test for dichotomized data. All p-values reported are two-sided and a p value < 0.05 was considered statistically significant. Analyses were performed using SPSS v26.0 (RRID:SCR\_002865) and Statistical Analysis System SAS v9.4 (RRID:SCR\_008567).

**Supplementary Figure Legends**

**Supplemental** **Figure 1. Study design**

Clinicians use a bronchoscope to access the main airways of the lung and then instill normal saline to wash the area around the tumor. This fluid can be analyzed by cytology to determine whether there is cancer present. Here, we present molecular techniques to profile bronchoalveolar lavage (BAL) fluid in lung cancer using hybrid-capture, targeted deep sequencing (CAPP-Seq).

1. Up to 60 ml of BAL fluid was obtained in the subsegment where the lesion of interest was located in two divided doses. We attempted to collect lavage prior to a tumor biopsy whenever the clinical circumstances allowed. For this study, 9/35 cancer patients (26%) had lavage collected after a biopsy. Blood was obtained prior to biopsy at the same time as the procedure in all cases. BAL fluid was centrifuged and the cell pellet separated from the supernatant to analyze cfDNA and cellular DNA independently. Blood was spun down and the plasma analyzed for cfDNA as previously reported. Plasma depleted whole blood (PDWB) was used for germline analysis and to account for clonal hematopoiesis. STAMP = Solid Tumor Actionable Mutation Panel. CAPP-Seq = CAncer Personalized Profiling by deep Sequencing.
2. Mutations detected by CAPP-Seq in the BAL cell free DNA (cfDNA) and plasma cfDNA were compared to those detected by STAMP in tumor DNA after filtering out germline polymorphisms. In total, 34 tumor samples, 38 plasma samples and 35 lavage cfDNA samples from 38 unique patients were available for analysis.

**Supplemental** **Figure 2. Median sequencing depths for cfDNA samples**

Violin plot comparing BAL cfDNA (blue) and plasma cfDNA (red). Dots represent individual samples and the median depth is shown by a horizontal bar for each biofluid. The width of the plot corresponds to the local density of sample numbers. P value is displayed above the plot.

**Supplemental Figure 3. Tumor informed vs. naïve mutation calls in BAL fluid and plasma.**

1. Overlapping tumor derived mutations for 27 patients using a tumor naïve calling approach. Based on data from Supplemental Table 7.
2. Detection of tumor mutations by tumor informed calling and tumor naïve calling in BAL cfDNA (blue) and plasma cfDNA (red).
3. Detection of cancer driver genes by tumor informed calling and tumor naïve calling in BAL cfDNA (blue) and plasma cfDNA (red). P values are on the right of the figure. NS = non-significant.

**Supplemental** **Figure 4. BAL classifier scores stratified by patient characteristics**

1. Feature importance of gene feature model (x-axis) by ‘mean decrease accuracy’ (y-axis). The most informative features are displayed from right to left. Interquartile ranges are displayed as error bars for each column. See Supplemental Table 9 for description of features in the plot.
2. Risk score (y-axis) is plotted against relevant clinical variables (x-axis). Patients are identified by dots, the median score is denoted by a horizontal line for each group and p-values between groups are denoted on top of the figure. Age was dichotomized at 65 years old, smoking was defined as ever or never, and stage as I/II vs III/IV for this analysis.

**Supplementary Tables**

**Supplemental** **Table 1.** Demographic and biofluid characteristics for all patients enrolled

**Supplemental** **Table 2.** Comparison of BAL cfDNA and BAL cellular DNA tumor variant calls

**Supplemental** **Table 3.** Sequencing characteristics of biofluids

**Supplemental** **Table 4.** All tumor informed mutation SNVs and indels detected in BAL cfDNA and plasma

**Supplemental** **Table 5.** Summary of tumor informed SNVs and indels detected in BAL cfDNA and plasma

**Supplemental** **Table 6.** Tumor naïve mutation SNVs detected in BAL cfDNA and plasma

**Supplemental** **Table 7.** Summary of tumor naïve mutation SNVs detected in BAL cfDNA and plasma

**Supplemental** **Table 8.** Summary of SNVs detected in BAL cfDNA and plasma for case-control cohort

**Supplemental** **Table 9**. BAL genome classifier variables, risk score and cytology