AWARD NUMBER: **W81XWH-18-1-0339**

TITLE: Developing High-Accuracy Sequencing Methods for Use in Early Cancer Detection, Disease Stratification, and Chemotherapy Resistance With Cell-Free DNA

PRINCIPAL INVESTIGATOR: **Scott R Kennedy, PhD**

CONTRACTING ORGANIZATION: **University of Washington, Seattle, WA**

REPORT DATE: **November 2021**

TYPE OF REPORT: **Final**

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
November 2021	Final	01Aug2018-31Jul2021

4. TITLE AND SUBTITLE

Developing High-Accuracy Sequencing Methods for Use in Early Cancer Detection, Disease Stratification, and Chemotherapy Resistance With Cell-Free DNA

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-18-1-0339

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Scott R Kennedy

E-Mail: scottrk@uw.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Washington
Box 359472, 4333 Brooklyn Ave
NE Seattle WA, 98195

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Recent efforts to understand the mutational landscape of tumors has resulted in a detailed cataloguing of diagnostic, prognostic, and clinically actionable mutations. Previous studies have identified a number of "driver" mutations, thought to be responsible for tumor formation, present in a significant proportion of non-small cell lung cancer patients. Detection of these mutations can help in early cancer detection, guide treatment options, or alert to the emergence of chemotherapy resistance, all of which could be harnessed to significantly improve survival. The advent of next-generation sequencing technology(NGS) has opened up the possibility of clinically exploiting ctDNA. Unfortunately, ctDNA from cancer comprises only a small fraction of all the overall amount of cell-free DNA in the blood stream. This issue, in conjunction with the high error rates of NGS technology, has proven to be a major impediment in developing minimally invasive tests to look for rare tumor specific mutations in the sea of normal cell-free DNA. In this grant, we developed a targeted gene panel for use with ultra-high accuracy Duplex Sequencing to detect therapy resistance in non-small cell lung cancer (NSCLC). We enrolled 40 NSCLC patients with a known clinically actional tumor mutation (BRAF, KRAS, EGFR, ALK, ERBB2, or ALK) and collected cell-free DNA (cfDNA) every 3 months for at least 1 year. We demonstrate the ability of Duplex Sequencing to sequence cfDNA with high efficiency (>30%) and high accuracy to detect a tumor therapy resistance mutations.

15. SUBJECT TERMS
Duplex Sequencing, non-small cell lung cancer, somatic mutations, cell-free DNA

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	15	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

I. INTRODUCTION

Efforts to understand the mutational landscape of tumors has resulted in a detailed cataloguing of diagnostic, prognostic, and clinically actionable mutations. Studies have identified a number of "driver" mutations, thought to be responsible for tumor formation, present in a significant proportion of non-small cell lung cancer patients. Detection of these mutations can help in early cancer detection, guide treatment options, or alert to the emergence of chemotherapy resistance, all of which could be harnessed to significantly improve survival. As with most other cancers in the chest cavity, access to tumor tissue by biopsy or surgical resection is often extremely limited or unobtainable and, additionally, not necessarily representative of the entire tumor. For this reason, DNA shed by the tumor into the bloodstream, often referred to as circulating tumor DNA (ctDNA), holds the promise of yielding detailed information about a tumor using a simple, minimally-invasive, blood test. ctDNA is increasingly used to stratify and guide therapy for non-small cell lung cancer (NSCLC) patients. Unfortunately, ctDNA from cancer comprises only a small fraction of all the overall amount of cell-free DNA (cfDNA) in the blood stream. This issue, in conjunction with the relatively high error rates of modern sequencing platforms, has proven to be a major impediment in developing minimally invasive tests to look for rare tumor specific mutations in the sea of normal cell-free DNA (cfDNA). To overcome this issue, we previously developed Duplex Sequencing, which is capable of detecting these low frequency mutations. The goal of this grant was to develop an alternative approach to Duplex Sequencing that we termed Linked-Strand Anchored Multiplex PCR (LS-AMP), a simple PCR-based target enrichment method that maintains the accuracy of Duplex Sequencing while offering reduced costs with increased efficiency and scalability on low amounts of DNA frequently encountered in cfDNA applications. The project aimed to develop a gene panel for high accuracy detection of the most clinically useful mutations found in NSCLC and then validate its ability to detect and track these mutations during chemotherapy treatment.

II. KEYWORDS

Duplex Sequencing, non-small cell lung cancer, somatic mutations, cell-free DNA.

III. ACCOMPLISHMENTS

A. What are the major goals of the project?

The project encompasses three major goals.

i. Develop NSCLC gene panel.

Subtask 1: Optimize LS-AMP protocol, including primer design, multiplex PCR conditions, and reagents.

Subtask 2: Validate with *in vitro* mixing experiments.

ii. Characterize biological background mutations in cfDNA from healthy controls

Subtask 1: Use LS-AMP to sequence cfDNA from healthy controls using gene panel developed in Specific Aim 1. cfDNA will be purified from commercially available plasma

iii. Validation of LS-AMP on cfDNA from NSCLC patients

Subtask 1: IRB approval from Dept. of Defense

Subtask 2: Collection of cfDNA from NSCLC patients every 3 months. Samples will be collected at Seattle Cancer Care Alliance.

Subtask 3: Sequencing of cfDNA using LS-AMP gene panel developed in Specific Aim 1.

Subtask 4: Analysis of data to determine sensitivity, specificity, and lead time for tumor recurrence and disease relapse.

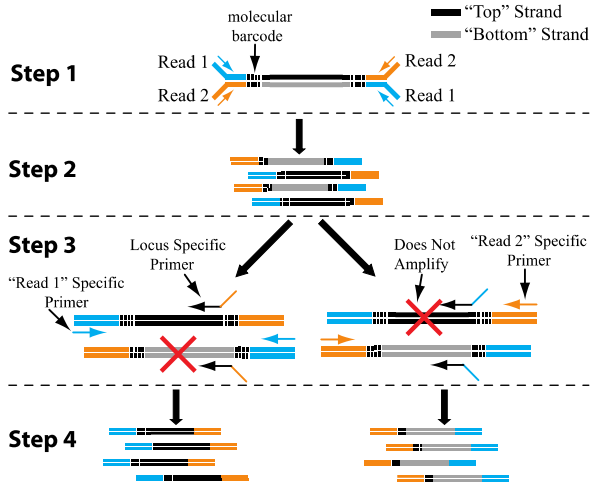

Figure 1. Schematic of SPLiT-Seq. Step 3 shows that all molecules contain PCR primer specific to one of the adapter sequences, which will allow all molecules to linearly amplify.

B. What was accomplished under these goals?

1. Goal 1: Develop gene NSCLC panel

Subtask 1: Optimize LS-AMP protocol, including primer design, multiplex PCR conditions, and reagents.

During the project's period of performance, we made significant progress in developing a panel of PCR primers that could be used to detect NSCLC-related mutations in cfDNA. As part of this work, we identified several technical hurtles. First, we highlighted an issue arising from linear amplification of non-target DNA, which encompasses the vast majority of the DNA within a sample, is occurring by priming of the adapter specific primer (Figure 1). Because *every* molecule an adapter sequenced attached to it, all sequences will linearly amplify, whereas only the locus specific primer would allow for exponential amplification. However, because the off-target sequence is so much more prevalent, even linear amplification will lead to high amounts of non-target DNA molecules with both Illumina flow-cell sequences. We tested a number of approaches to remove this adapter sequence, including methyl-dependent restriction, deoxyuracil or ribobase incorporated adapter, and CRISPR/Cas9 restriction (Figure 2 & 3).

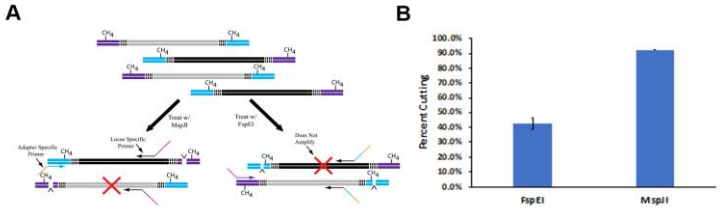

Figure 2. Schematic of enzymatic removal of undesired adapter sequence. (A) The schematic is the equivalent frame of Step 3 in Figure 1. In the presented case, methylated primers, with methylation sites being different between the two adapter-specific PCR primers are introduced during Step 1-2 in Figure 1. These different methylated sites are cut by different restriction endonucleases. The result is that the only fully functional molecules that are able to be sequenced are produced by exponential amplification of the target loci. Linearly amplified molecules will not contain both flow cell primer due to their enzymatic removal. A similar strategy is used with dU and ribonucleic acid containing primers. **(B)** Methylated adapter sequence was introduced *via* PCR and the reaction biproducts removed by SPRI bead cleanup. The PCR reaction was then split and a small 2µL aliquot was removed as a non-cut control to determine cutting efficiency. The indicated enzyme was then added and the reaction incubated for 4 hours. A 2µL aliquot was taken at the 4 hour timepoint and the amount of amplifiable product was determined by qPCR using primers specific for the P5 (FspEI) or P7 (MspJI) adapter sequence and the sequence of the lambda phage used in the test fragment (N=3).

Methyl-dependent restriction

We first tested the use of removing the adapter sequence by methyl-dependent restriction. We tested MspJI and FspEI methyl-dependent restriction enzymes. Our data (Figure 2) indicate that MspJI resulted in >90% removal of the desired adapter sequence. However, FspEI performed with significantly lower efficiency

(<50%), indicating that this enzyme did not perform well with the particular recognition site present in the Duplex Sequencing adapter. This approach was not viable.

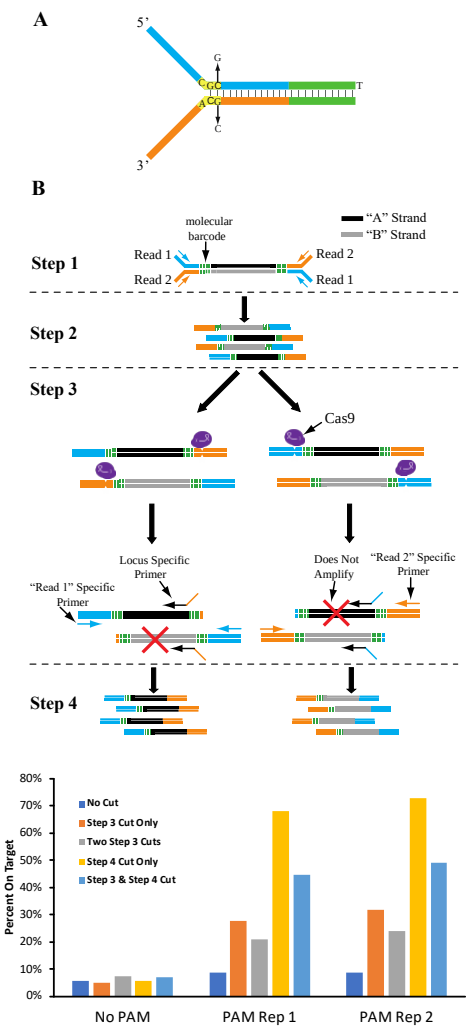

Figure 3. Schematic of LS-AMP with CRISPR restriction. (A) Schematic of the modified Duplex-Seq adapter containing the single base change to for a Cas9 PAM site (*yellow*). **(B)** Modified LS-AMP protocol that incorporates the Cas9 restriction of the adapter sequence. We performed our LS-AMP protocol for a single genomic locus (TH01) (Figure 4B) with different Cas9 cutting strategies. PCR enrichment was done with 20-cycles to ensure a reasonable dynamic range in assaying on-target percentage. The steps in Figure 4B when the Cas9 restriction reactions occurred are indicated in the figure legend. The samples were then sequenced on an Illumina MiSeq platform and the percent of on target reads determined post-alignment. **(C)** We performed our LS-AMP protocol for a single genomic locus (TH01) (Figure 4B) with different Cas9 cutting strategies. PCR enrichment (Fibure 4B, Step 3) was done with 20-cycles to ensure a reasonable dynamic range in assaying on-target percentage.

Deoxyuracil or ribobase

In a similar strategy to the one using methylated primers, we next tested if the enzymatic removal of deoxyruracil or ribobase with enzyme cocktail of uracil RNaseH2, deglycosylase (UDG), AP endonuclease (APE), and S1 nuclease is routinely used to fragment double-stranded DNA. This approach works by introducing an abasic site by UDG or RNaseH2, which is then converted to a single-strand nick by APE. S1 nuclease then recognizes the nick and cleaves the phophodiester bond of the DNA strand opposite the nick, resulting in a double-stranded break. We tested if this enzyme cocktail approach could efficiently remove the adapter sequence of interest. To test this, we changed the 3'-most non-terminal dC in our adapter PCR primer to a dU. We observed substantial sample degradation from S1 digestion of both sides of the DNA molecule, which means that both adapter sides are being degraded. This approach was not viable.

CRISPR/Cas9 restriction

Finally, we tested the programmable Cas9 endonuclease, which we have previously used for *in vitro* targeted genome fragmentation (Figure 3). This approach required the introduction of a single-base change to the adapter (Figure 3A) that introduces a new 5'-NGG PAM site that is present on both sides of the adapter and, conveniently, removes the entire non-desired adapter sequence, something that our other approaches were not able to do. Using qPCR, we quantified the percentage of molecules cut on the P7 (Figure 3B, left branch) or P5 (Figure 3B, right branch) sides. We tested different Cas9 restriction conditions: 1) No cut control; 2) Single cut (at Figure 3B, Step 3); 3) Two sequential cuts (Figure 3B, Step 3); 4) Single cut immediately prior to sequencing (Figure 4B, Step 4); 5) Two non-sequential cuts (One at Step 3 and one immediately prior to sequencing). As reported in Figure 6, removing the unwanted adapter sequencing significantly improved the percentage of on-target molecules at 20 exponential cycles, with two non-sequential cut approach resulting in ~70% of molecules mapping to our desired locus. Importantly, our non-cut and non-PAM site controls did not show any improved enrichment.

PCR Optimization

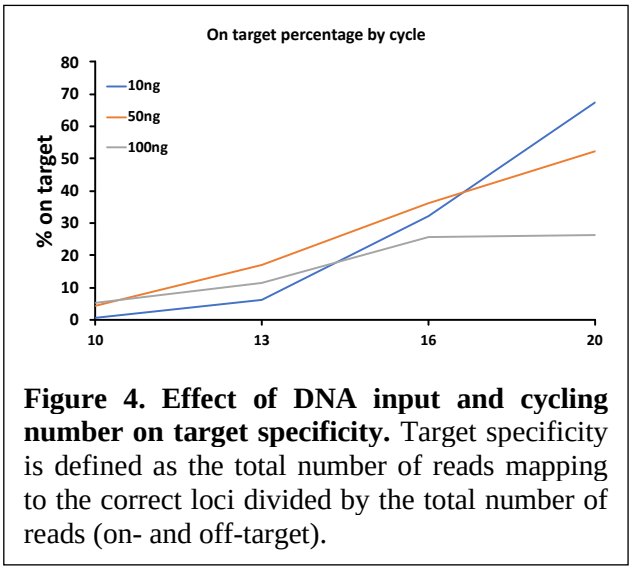

Figure 4. Effect of DNA input and cycling number on target specificity. Target specificity is defined as the total number of reads mapping to the correct loci divided by the total number of reads (on- and off-target).

We optimized the PCR cycling conditions to see if we could improve this enrichment to a level that is comparable to target hybridization capture (~90+%) across a range on input cfDNA amounts. We determined that increasing amounts of input DNA (10ng, 50ng, 100ng) resulted in poorer performance with the majority of reads being from DNA molecules that failed to cut with CRISPR/Cas9. Our data indicated that more cycling increased target enrichment, but became less effective with increasing cfDNA input levels due to the PCR reaction becomes saturated at C(t)=~20, preventing us from using more PCR cycles to further sample enrichment (Figure 4). Therefore, we concluded that LS-AMP is only a viable approach for very low (<10ng) cfDNA input applications. However, the utility of cfDNA to detect rare tumor mutations is blunted at low inputs, regardless of the method efficiency, because it is unlikely to sample these mutations (i.e. one can't detect a mutation that isn't present in your tube).

During this phase of our development, commercially developed Duplex-Sequencing kits, of which LS-AMP is a derivative technology, became available. We decided as a backup strategy to evaluate the performance of this kit on a cfDNA sample. We designed a panel of biotinylated probes against the exons of NSCLC genes known to play an important role in chemotherapy resistance. The list of genes and targeted exons are listed in Table 1.

Table 1

Gene	Targeted Exons
ALK	21,22,23,24,25
BRAF	13,15
EGFR	22,23,24,25
ERBB2(HER2)	24,25
KIT	17
KRAS	2,3
MET	19,22
NTRK1	16,17
NTRK2	20
NTRK3	18,19
PIK3CA	2,5,8,9,10,20,21
RET	11,12,13,14,15,16,18
ROS1	36,37,38,41
STK11	All

Our initial sample, which was from a NSCLC patient that withdrew from our study, had ~50ng of purified cfDNA. We made the sequencing library following the manufacturer's protocol and sequenced on an Illumina NovaSeq platform. The sample exhibited extremely encouraging results. 97.1% of reads were on target and a peak depth of 5711X and a median depth of 3700X (Figure 5A). An important metric is the conversion efficiency which is the percentage of genomes that form a final duplex consensus read compared to the number of input genomes. 50ng of input DNA is equivalent to ~15,000 genome equivalents. Therefore, our peak conversion efficiency is 5711/15000=0.38 or 38%, which is significantly higher than published efficiencies for Duplex Sequencing. This efficiency is very close to the efficiency goal we were hoping to get with our LS-AMP approach. We attribute the efficiency boost to a combination of protocol optimizations on the part of the commercial developer and the naturally fragmented nature of cfDNA that obviates the need for harsh sheering conditions required by genomic DNA. Most importantly, this patient was known to have a ERBB2/HER2 insertion in exon 20 and we were able to detect this variant at <1% VAF without any additional clones being seen in any other gene (Figure 5B).

Conventional Duplex Sequencing makes use of targeted hybridization capture for enrichment. This approach has the advantage of adding additional probes when new information

becomes available without requiring a re-optimization of PCR conditions that would needed for LS-AMP. Based on all these considerations and limitations involving LS-AMP, and the loss of time due to personnel issues and COVID-19, we moved forward with using a conventional Duplex-Sequencing approach for the NSCLC samples that we have been collecting as part of this project (See Goal 3).

Subtask 2: Validate with in vitro mixing experiments.

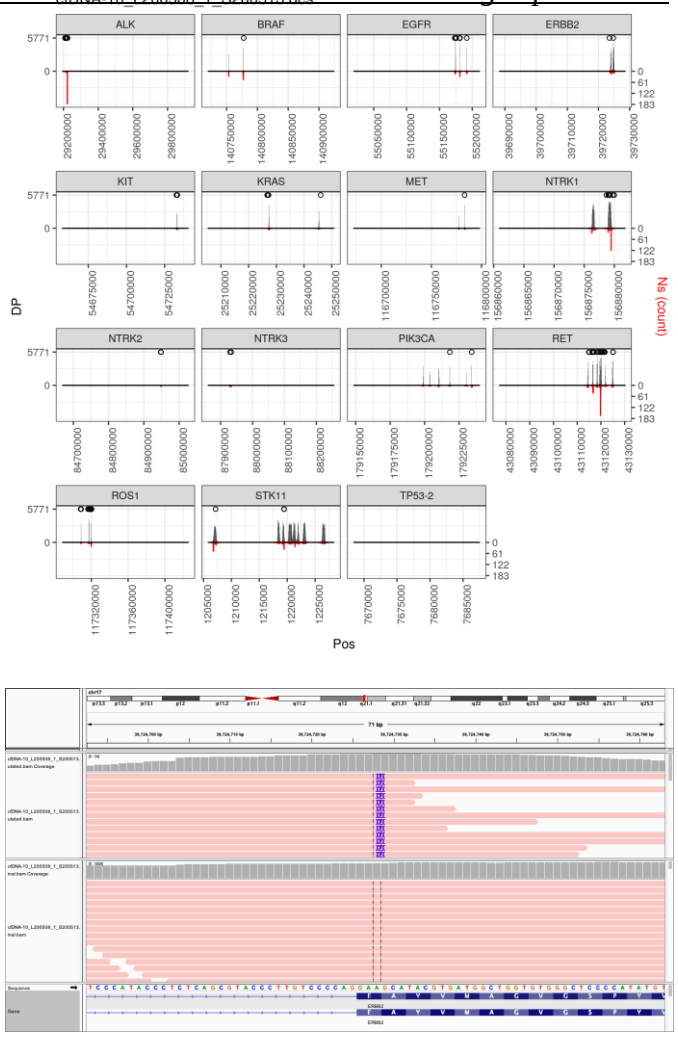

Figure 5. Depth of targeted genes in Table 1. (A) Each gene in reported in a separate panel. The depth of each targeted region (i.e. exon) is reported in the y-axis ad the genomic coordinates are reported in the x-axis. **(B)** Detection of known ERBB2/HER2 exon 20 insertion by conventional Duplex Sequencing panel.

Conventional Duplex Sequencing has been validated in numerous publications and is known to be accurate down to <1 mutation per $1x10^7$ sequenced bases. Because this technique is already so well validated, we dispensed with repeating these mixing experiments that others have previously performed.

2. Goal 2: Characterize biological background mutations in cfDNA from healthy controls

Subtask 1: Use LS-AMP to sequence cfDNA from healthy controls using gene panel developed in Specific Aim 1. cfDNA will be purified from commercially available plasma.

Nothing to report. See Section V. B. for details.

3. Goal 3: Validation of LS-AMP on cfDNA from NSCLC patients

Subtask 1: IRB approval from Dept. of Defense

HRPO approved our human subject protocol on 1/30/19. This subtask was completed without issue.

Subtask 2: Collection of cfDNA from NSCLC patients every 3 months. Samples will be collected at Seattle Cancer Care Alliance.

We have consented 40 individuals with NSCLC, the maximum allowed for by our IRB application. Patient demographics are detailed in Figure 6. We collected multiple longitudinal blood draws from most patients (Table 2). cfDNA and genomic DNA was purified from the samples by the UW Laboratory Medicine Genetics and Solid Tumor Laboratory. We were able to obtain four longitudinal blood draws from 25 of the 40 enrolled patients. The reason 15 patients did not complete the study was due to four patients withdrawing consent early and eights passing away prior to donating the minimum number of blood draws. Several of the deceased

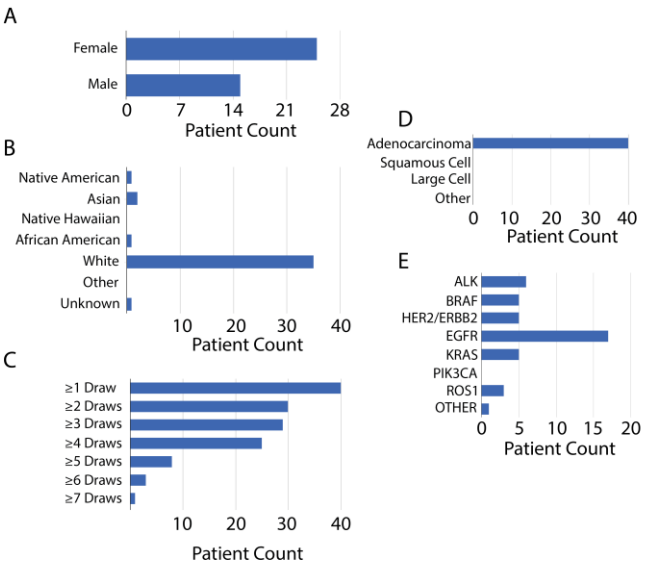

Figure 6. Patient cohort demographics and statistics. (A) Patient sex. **(B)** Patient race. **(C)** Cumulative minimum number of blood draws. **(D)** NSCLC subtype. **(E)** Driver mutation.

patients had multiple blood draws which allowed us to obtain useful data from their samples even without obtaining the minimum goal of 4 time points.

Subtask 3: Sequencing of cfDNA using LS-AMP gene panel developed in Specific Aim 1.

We successfully sequenced the cfDNA from 122 blood draw samples and gDNA from the first blood draw from all 40 patients.

Subtask 4: Analysis of data to determine sensitivity, specificity, and lead time for tumor recurrence and disease relapse.

To date. all sequencing data has been subjected Duplex Sequencing error correction using v2.1.2 of our Duplex-Seq pipeline. All variants have been annotated using Enemble Variant Effect Predictor, release 100 against the human HG38 reference genome. We are continuing to analyze the data and are working with the University of Washington's Institute of Translational Health Sciences to perform statistical analysis to correlate variant detection and relapse prediction. We expect data analysis to be completed in early 2022 and submitted for publication.

Table 2

Patient #	Driver Gene	Blood draw #1	Blood draw #2	Blood draw #3	Blood draw #4	Blood draw #5	Blood draw #6	Blood draw #7	Off Study date/reason
Pt 01	ALK	5/14/19	8/8/19	11/12/19	2/10/20	5/4/20			Complete
Pt 02	ALK	10/15/18	1/28/19	5/13/19	9/16/19	12/16/19	3/16/20	6/15/20	Complete
Pt 03	EGFR	12/12/18	4/29/19	7/16/19	10/2/19	1/31/20			Complete
Pt 04	EGFR	6/4/19	9/6/19	12/17/19	2/27/20				Complete
Pt 05	ROS1	11/19/18	5/2/19	7/29/19	12/16/19	3/9/20	6/15/20		Complete
Pt 06	ALK	12/10/18							Withdrew
Pt 07	ALK	1/16/19	5/6/19	10/28/19	1/31/20	4/13/20			Complete
Pt 08	EGFR	12/11/18	5/8/19	8/27/19	11/18/19	1/22/20	4/15/20		Complete
Pt 09	BRAF	12/7/18	4/26/19	8/2/19					Passed away
Pt 10	ERBB2	12/13/18							Withdrew
Pt 11	EGFR	12/13/18	4/22/19	7/22/19					Withdrew
Pt 12	ROS1	12/5/18	6/12/19	9/25/19	2/5/20	5/20/20			Complete
Pt 13	ERBB2	1/10/19							Passed away
Pt 14	EGFR	5/21/19	8/29/19	1/28/20	4/30/20				Complete
Pt 15	ROS1	6/5/19	9/11/19	1/9/20	5/13/20				Complete
Pt 16	KRAS	6/20/19							Passed away
Pt 17	BRAF	7/10/19	9/30/19	12/12/19	2/28/20				Complete
Pt 18	EGFR	7/15/19	10/9/19	1/6/20	3/26/20				Complete
Pt 19	KRAS	8/9/19	11/4/19	7/20/20	10/8/20				Complete
Pt 20	EGFR	9/10/19							Passed away
Pt 21	ERBB2	8/8/19							Passed away
Pt 22	BRAF	8/7/19	10/29/19	1/8/20	4/10/20				Complete
Pt 23	EGFR	9/16/19							Passed away
Pt 24	EGFR	10/3/19	1/6/20						Passed away
Pt 25	EGFR	9/26/19							Passed away
Pt 26	ERBB2	11/13/19	2/5/20	5/20/20					Withdrew
Pt 27	EGFR	11/25/19	3/11/20	6/10/20	9/23/20				Complete
Pt 28	ALK	1/17/20	6/8/20	10/26/20	5/26/21				Complete
Pt 29	KRAS	10/18/19	1/17/20	4/10/20	8/24/20				Complete
Pt 30	EFGR	10/22/19	1/22/20	4/22/20	7/22/20				Complete
Pt 31	KRAS	10/10/19	1/3/20						Passed away
Pt 32	EFGR	1/29/19	4/29/20	7/29/20	10/28/20				Complete
Pt 33	ALK	10/24/19	1/21/20	4/2/20	7/2/20				Complete
Pt 34	KRAS	10/28/19	2/10/20	6/1/20	10/5/20				Complete
Pt 35	EGFR	11/18/19	2/10/20	5/7/20	12/28/20				Complete
Pt 36	EGFR	11/13/19	2/12/20	6/3/20					Passed away
Pt 37	BRAF	11/8/19	7/6/20	10/5/20	5/7/21				Complete
Pt 38	ERBB2	12/5/19	5/28/20	8/24/20	2/9/21				Complete
Pt 39	EGFR	12/19/19							Passed away
Pt 40	EGFR	1/2/20	4/10/20	7/24/20	10/16/20				Complete

C. What opportunities for training and professional development has the project provided?
Nothing to Report

D. How were the results disseminated to communities of interest?
Nothing to Report

E. What do you plan to do during the next reporting period to accomplish the goals?
Not Applicable

IV. IMPACT

A. What was the impact on the development of the principal discipline(s) of the project?
Nothing to Report

B. What was the impact on other disciplines?
Nothing to Report

C. What was the impact on technology transfer?
Nothing to Report

D. What was the impact on society beyond science and technology?
Nothing to Report

V. CHANGES/PROBLEMS

A. Changes in approach and reasons for change.
We moved to sequencing of the NSCLC patient cfDNA using a commercially available Duplex Sequencing kit. We decided to do so after careful consideration of the facts that 1) we lost significant time due to visa issues of Dr. Monica Sanchez-Contreras and COVID-19, 2) efficiency levels of the commercial kit that met our efficiency goals for LS-AMP, and 3) lower than expected on-target enrichment of LS-AMP, especially at cfDNA input amounts >10ng.

B. Actual or anticipated problems or delays and actions or plans to resolve them.
We encountered 4 main problems during the course of this project.

1) We encountered significant visa issues for Dr. Sanchez-Contreras. Dr. Sanchez-Contreras was unable to renew her J-1 visa and was forced to stop working 12/2/18. To solve this issue, she applied for a H-1b visa sponsored by the University of Washington. However, while this application was pending, she was unable to perform any of the work. This issue negatively impacted our ability to perform the proposed work. This issue was resolved in April 2019 and work resumed.

2) We encountered issues obtaining access to cfDNA from heathy control at a variety of ages. Our initial plan was to purchase cfDNA from a vendor, but the costs associated with this source was substantial.

3) Our PCR enrichment approach, while functional, was not meeting our desired enrichment goal of >90%.

4) The global COVID-19 pandemic substantially impacted this project in terms of benchwork. Our lab presence was dramatically reduced at the end of February 2020 just as we starting to process the collected samples for sequencing and analysis. We resumed our efforts in early July 2020 and made progress in processing samples. Work slowed again due to the emergence of the COVID delta variant. Our collection of blood samples from human subjects was only minimally affected. To lessen the impact of these issues, we moved to using a recently available commercial product for Duplex Sequencing. This product performs well in our initial tests and we processed all cfDNA and linked whole-blood samples without a single failure at the sample prep stage.

C. Changes that had significant impact on expenditures.
Nothing to Report

D. Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents.

Nothing to Report

E. Significant changes in use or care of human subjects.
Nothing to Report

F. Significant changes in use or care of vertebrate animals
Nothing to Report

G. Significant changes in use of biohazards and/or select agents.
Nothing to Report

VI. PRODUCTS

A. Publications, Conference Papers, and Presentations

To date, three publications have been supported by this proposal.

1) Kennedy SR, Zhang Y, Risques RA. (2019) Cancer-associated mutations but no cancer: Insights into the early steps of carcinogenesis and implications for early cancer detection. Trends Cancer 5(9), 531-540.

2) Salk JJ, Kennedy SR. (2019) Next-generation genotoxicology: Using modern sequencing technologies to study genetic toxicology and somatic mutagenesis. Environ Mol Mut 61(1), 135-151.

3) Sanchez-Contreras M, Sweetwyne MT, Kohrn BF, Tsantilas KA, Hipp MJ, Schmidt ES, Fredrickson J, Whitson, JA, Campbell MD, Hurley JB, Rabinovitch PS, Marcinek DJ, Kennedy SR. (2021) A replication-linked mutational gradient drives somatic mutation accumulation and influences germline polymorphisms and genome composition in mitochondrial DNA. Nucleic Acids Res 49(19), 11103-11118.

B. Website(s) or Internet Sites(s)

We have maintained a software development website that allow access for our open-source Duplex Sequencing pipeline: https://github.com/Kennedy-Lab-UW/Duplex-Seq-Pipeline

C. Technologies or Techniques

Nothing to Report

D. Inventions, Patent Applications, and/or License

The LS-AMP intellectual property was licensed from the University of Washington by TwinStrand Biosciences Inc for commercial development.

E. Other Products

Nothing to Report

VII. PARTICIPANTS & COLLABORATING ORGANIZATIONS

A. What individuals have worked on this project?

Name:	Monica Sanchez-Contreras
Project Role:	Post-doctoral Fellow
Research Identifier:	ORCID 0000-0002-3092-2781
Contribution to Project:	Dr. Sanchez-Contreras performed the majority of experiments conducted so far. She optimized assay development and performed sequencing experiments.

Funding Support:	None

Name:	Scott R Kennedy
Project Role:	Principal Investigator
Research Identifier:	ORCID 0000-0002-4444-1145
Contribution to Project:	Dr. Kennedy oversaw all experiments, helped analyze data, and administrated the grant.
Funding Support:	National Institutes of Health R01 X3 National Institute of Justice Safeway Cancer Consortium Early Career Award

Name:	Keith D Eaton
Project Role:	Co-Investigator
Research Identifier:	0000-0001-5025-2001
Contribution to Project:	Dr. Eaton oversaw NSCLC patient recruitment and sample collection, as well as helped analyze data.
Funding Support:	None

Name:	Philimon Sims
Project Role:	Patient Coordinator
Research Identifier:	None
Contribution to Project:	Mr. Sims identified, tracked, consented NSCLC patients, as well as ensured blood sample collection and delivery for processing.
Funding Support:	None

B. Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Dr. Kennedy was awarded two NIH R21 grants and overall effort was reduced to 2.48 calendar months.

C. What other organizations were involved at partners?

Fred Hutchinson Cancer Research Center (FHCRC) and Seattle Cancer Care Alliance (SCCA).

FHCRC oversaw the IRB approval and SCCA is where the NSCLC patients were consented and collections took place.

VIII. SPECIAL REPORTING REQUIREMENTS

Not Applicable

Award Log Number: LC170277 **Award Title** Developing High-Accuracy Sequencing Methods for Use in Early Cancer Detection, Disease Stratification, and Chemotherapy Resistance with Cell-Free DNA
PI: Scott Kennedy, University of Washington, WA **Budget:** Total Award Cost
Topic Area: Lung Cancer Research Program, Investigator-Initiated Translational Research Award
Mechanism: W81XWH-17-LCRP-IITRA

Research Area(s): 0700 Award Status: 8/1/18 – 7/31/21

Study Goals:

1. Develop gene NSCLC panel
2. Characterize biological background mutations in cfDNA from healthy controls
3. Validation of LS-AMP on cfDNA from NSCLC patients
4. Analysis of data to determine sensitivity, specificity, and lead time for tumor recurrence and disease relapse

Specific Aims:

1. Specific Aim 1: Fully develop SPLiT-Seq to make it amenable for ctDNA detection in NSCLC.
2. Demonstrate the utility of SPLiT-Seq for NSCLC detection with ctDNA.
   2a. Characterize biological background mutations in cfDNA from healthy controls.
   2b. Demonstrate the utility of SPLiT-Seq for identification of resistance mutations and prediction of recurrence using ctDNA in NSCLC.

Key Accomplishments and Outcomes:

Publications:

1. Kennedy SR, Zhang Y, Risques RA. (2019) Cancer-associated mutations but no cancer: Insights into the early steps of carcinogenesis and implications for early cancer detection. Trends Cancer 5(9), 531-540.
2. Salk JJ, Kennedy SR. (2019) Next-generation genotoxicology: Using modern sequencing technologies to study genetic toxicology and somatic mutagenesis. Environ Mol Mut 61(1), 135-151.
3. Sanchez-Contreras M, Sweetwyne MT, Kohrn BF, Tsantilas KA, Hipp MJ, Schmidt ES, Fredrickson J, Whitson, JA, Campbell MD, Hurley JB, Rabinovitch PS, Marcinek DJ, Kennedy SR. (2021) A replication-linked mutational gradient drives

somatic mutation accumulation and influences germline polymorphisms and genome composition in mitochondrial DNA. Nucleic Acids Res 49(19), 11103-11118.

Patents: The LS-AMP intellectual property was licensed from the University of Washington by TwinStrand Biosciences Inc for commercial development.

Funding Obtained: None