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**

REPORT DATE:
**AUGUST 2020**

TYPE OF REPORT:
**Annual**

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

## REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| AUGUST 2020 | Annual | 8/1/19-7/31/20 |

**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 Materiel 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. 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. 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. To overcome this issue, we previously developed Duplex Sequencing, which is capable of detecting these low frequency mutations. However, this method requires more DNA than is generally found as cell-free DNA in the blood. As part of previous funding from the Department of Defense, we recently developed a method we call 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 with cell-free DNA applications. At the completion of this project, we will have developed a gene panel for high accuracy detection of the most clinically useful mutations found in non-small cell lung cancer.**

**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 | 13 | 19b. TELEPHONE NUMBER *(include area code)* |

**Standard Form 298 (Rev. 8-98)**
**Prescribed by ANSI Std. Z39.18**

## Table of Contents

## I. INTRODUCTION

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. 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. 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. To overcome this issue, we previously developed Duplex Sequencing, which is capable of detecting these low frequency mutations. However, this method requires more DNA than is generally found as cell-free DNA in the blood. As part of previous funding from the Department of Defense, we recently developed a method we call 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 with cell-free DNA applications. At the completion of this project, we will have developed a gene panel for high accuracy detection of the most clinically useful mutations found in non-small cell lung cancer.

**KEYWORDS**

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

## II. ACCOMPLISHMENTS

### A. What are the major goals of the project?

The project encompasses three major goals.

**1. 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.

**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.

**3. 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.

## 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.

In our previous report, we made significant progress toward developing a panel of PCR primers that could be used to detect NSCLC related mutations in cfDNA. 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. 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. To further complicate the issue, these adapter sequences must be present in order to allow for the production of PCR copies of each strand prior to splitting the sample. This step is essential for how our method works and cannot be modified. We found that treating with CRISPR/Cas9 against the non-target adapter resulted in a significant improvement in on-target efficiency (~70%) with 10ng of input DNA. Therefore, we wanted to optimize 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. Unfortunately, we found 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 indicate that more cycling increases target enrichment, but becomes less effective with increasing cfDNA input levels. Further complicating this issue is that the PCR reaction becomes saturated at C(t)=~20, preventing us from using more PCR cycles to further sample enrichment. Therefore, we conclude 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.

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 (Fig. 1). 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 (Kennedy; Nachmanson). 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 (Fig. 2).

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 (detailed in the last report) and COVID-19,we decided to move 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.

Conventional Duplex Sequencing has been validated in numerous publications and is known to be accurate down to <1 mutation per $1x10_6$ sequenced bases. Because this technique is already so well validated, we have 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 IV. 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 is complete.

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

To date we have consented 40 individuals with NSCLC, the maximum allowed for by our IRB application. We are currently have multiple longitudinal blood draws from most patients. cfDNA and genomic DNA has been purified from the samples by the UW Laboratory Medicine Genetics and Solid Tumor Laboratory. We are aiming to have a minimum of 4 blood draws from all individuals still enrolled in this study. To date, only 4 of the 40 patients have withdraw and 8 of the patients passed away prior to donating the minimum number of blood draws, leaving 28 that have or are expected to have the minimum number of blood draws. Importantly, several of the deceased patients have multiple blood draws which will allow us to obtain useful data from their samples even without obtaining the minimum goal of 4 time points. A summary of the patient collection is provided below.

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

At the time of this report, we have sequenced the cfDNA from 96 blood draw samples.

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

The initial batch of 96 samples were in the process of being analyzed for variants using our data analysis pipeline at the time this report was being written.

**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?**
We expect to finish collecting at least 4 blood draws from all patients still enrolled in the study by 12/2020 with enough time and funds to collect 5 for the majority of currently enrolled donors.

## III. 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

## IV. 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 3 main problems. First, was 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. Second, our PCR enrichment approach, while functional, was not meeting our desired enrichment goal of >90%. Thirdly, 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 have quickly made progress in processing samples. Thankfully, 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 have processed 96 total 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.**
Not Applicable

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

## V. PRODUCTS

**A. Publications, Conference Papers, and Presentations**
Nothing to Report

**B. Website(s) or Internet Site(s)**
The grant has supported the development of an updated Duplex Sequencing pipeline that is able to better analyze large batches of samples. The software can be found at https://github.com/Kennedy-Lab-UW/Duplex-Seq-Pipeline

**C. Technologies or Techniques**
Nothing to Report

**D. Inventions, Patent Applications, and/or License**
Nothing to Report

**E. Other Products**
Nothing to Report

## VI. 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 |
| **Nearest Person Month Worked:** | 11 calendar months |
| **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 |
| **Nearest Person Month Worked:** | 4 calendar month |
| **Contribution to Project:** | Dr. Kennedy oversaw all experiments, helped analyze data, and administrated the grant. |
| **Funding Support:** | National Institutes of Health R01 X3<br>National Institute of Justice<br>Safeway Cancer Consortium Early Career Award |

**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 a Safeway Cancer Consortium Early Career Award
Dr. Kennedy was brought on as a Co-Investigator for a NIH R01 and a NIH P30 grant.
Total effort was reduced from 4.8 to 3.76 calendar months.

**C. What other organizations were involved as partners?**

Nothing to Report

## VII. SPECIAL REPORTING REQUIREMENTS

Not Applicable

**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 |

**Table 2**

| Patient # | 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 | 5/14/19 | 8/8/19 | 11/12/19 | 2/10/20 | 5/4/20 | | | Complete |
| Pt 02 | 10/15/18 | 1/28/19 | 5/13/19 | 9/16/19 | 12/16/19 | 3/16/20 | 6/15/20 | Complete |
| Pt 03 | 12/12/18 | 4/29/19 | 7/16/19 | 10/2/19 | 1/31/20 | 6/24/20 | | Complete |
| Pt 04 | 6/4/19 | 9/6/19 | 12/17/19 | 2/27/20 | | | | |
| Pt 05 | 11/19/18 | 5/2/19 | 7/29/19 | 12/16/19 | 3/9/20 | 6/15/20 | | Complete |
| Pt 06 | 12/10/18 | | | | | | | Withdrew |
| Pt 07 | 1/16/19 | 5/6/19 | 10/28/19 | 1/31/20 | 4/13/20 | | | Complete |
| Pt 08 | 12/11/18 | 5/8/19 | 8/27/19 | 11/18/19 | 1/22/20 | 4/15/20 | | Complete |
| Pt 09 | 12/7/18 | 4/26/19 | 8/2/19 | | | | | Passed away |
| Pt 10 | 12/13/18 | | | | | | | Withdrew |
| Pt 11 | 12/13/18 | 4/22/19 | 7/22/19 | 10/25/19 | | | | Withdrew |
| Pt 12 | 12/5/18 | 6/12/19 | 9/25/19 | 2/5/20 | 5/20/20 | | | Complete |
| Pt 13 | 1/10/19 | | | | | | | Passed away |
| Pt 14 | 5/21/19 | 8/29/19 | 1/28/20 | 4/30/20 | | | | Complete |
| Pt 15 | 6/5/19 | 9/11/19 | 1/9/20 | 5/13/20 | | | | Complete |
| Pt 16 | 6/20/19 | | | | | | | Passed away |
| Pt 17 | 7/10/19 | 9/30/19 | 12/12/19 | 2/28/20 | 5/4/20 | | | Complete |
| Pt 18 | 7/15/19 | 10/9/19 | 1/6/20 | 3/26/20 | 7/2/20 | | | Complete |
| Pt 19 | 8/9/19 | 11/4/19 | 7/20/20 | | | | | |
| Pt 20 | 9/10/19 | | | | | | | Passed away |
| Pt 21 | 8/8/19 | | | | | | | Passed away |
| Pt 22 | 8/7/19 | 10/29/19 | 1/8/20 | 4/10/20 | | | | Complete |
| Pt 23 | 9/16/19 | | | | | | | Passed away |
| Pt 24 | 10/3/19 | 1/6/20 | | | | | | Passed away |
| Pt 25 | 9/26/19 | | | | | | | Passed away |
| Pt 26 | 11/13/19 | 2/5/20 | 5/20/20 | | | | | Withdrew |
| Pt 27 | 11/25/19 | 3/11/20 | 6/10/20 | | | | | |
| Pt 28 | 1/17/20 | 6/8/20 | | | | | | |
| Pt 29 | 10/18/19 | 1/17/20 | 4/10/20 | | | | | |
| Pt 30 | 10/22/19 | 1/22/20 | 4/22/20 | 7/22/20 | | | | |
| Pt 31 | 10/10/19 | 1/3/20 | 4/16/20 | | | | | Passed away |
| Pt 32 | 1/29/19 | 4/29/20 | 7/29/20 | | | | | |
| Pt 33 | 10/24/19 | 1/21/20 | 7/2/20 | | | | | |
| Pt 34 | 10/28/19 | 2/10/20 | 6/1/20 | | | | | |
| Pt 35 | 11/18/19 | 2/10/20 | 5/7/20 | | | | | |
| Pt 36 | 11/13/19 | 2/12/20 | 6/3/20 | | | | | |
| Pt 37 | 11/8/19 | 7/6/20 | | | | | | |
| Pt 38 | 12/5/19 | 5/28/20 | | | | | | |
| Pt 39 | 12/19/19 | 5/4/20 | | | | | | Passed away |
| Pt 40 | 1/2/20 | 4/10/20 | 7/24/20 | | | | | |

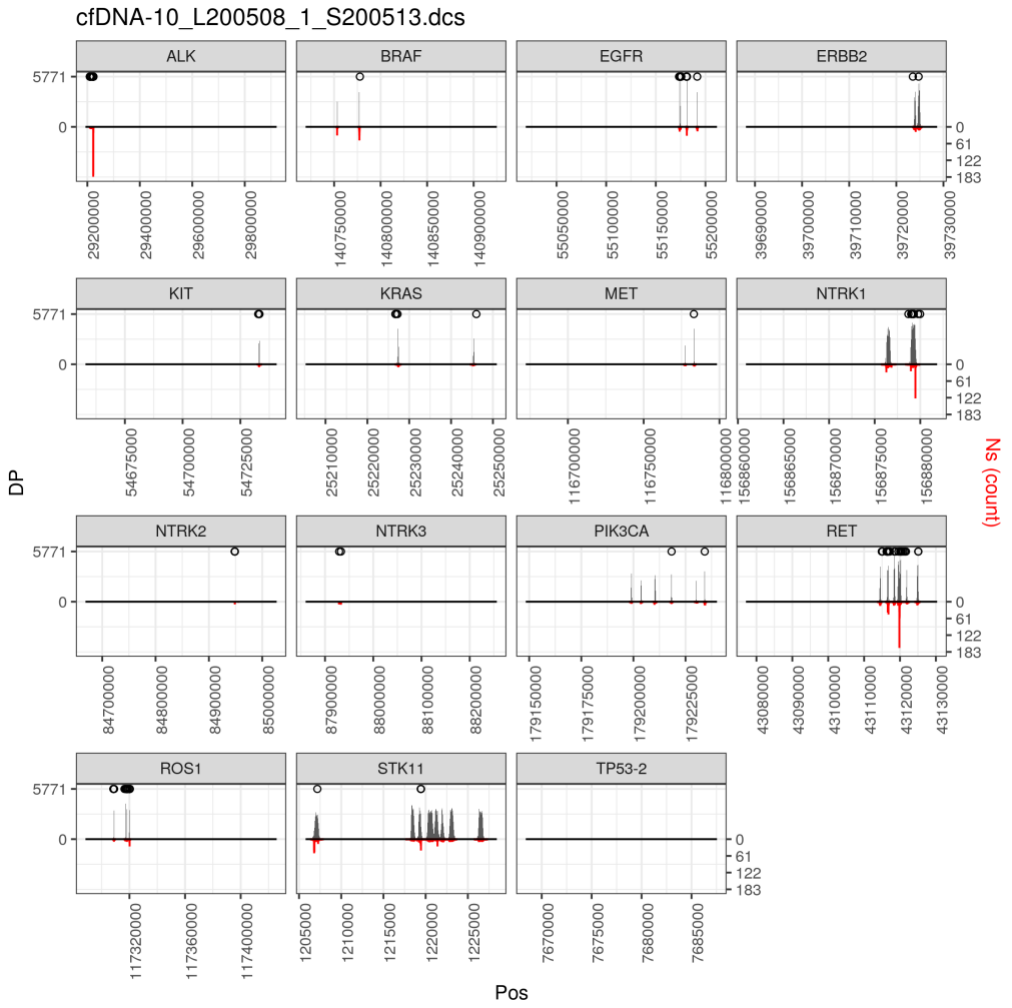


**Fig. 1. Depth of targeted genes in Table 1.** Each gene in reported in a separate panel. The depth of each targeted region (i.e. exon) is reported in the y-axis ad the genomic coordinates are reported in the x-axis.

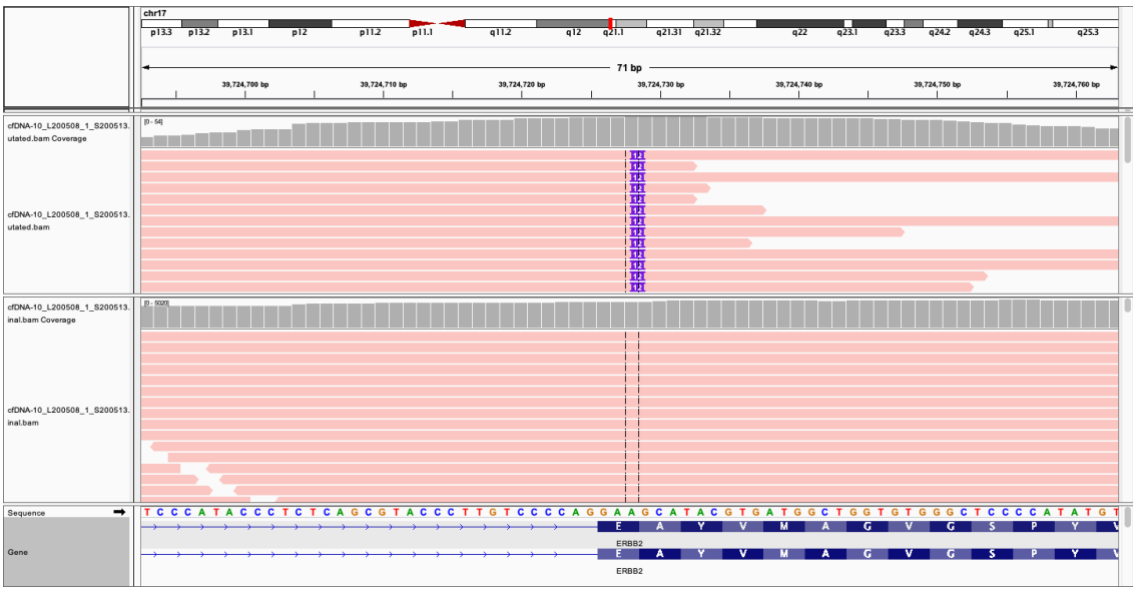


**Fig. 2. Detection of known ERBB2/HER2 exon 20 insertion.**