AWARD NUMBER: **W81XWH-18-1-0357**

TITLE: **Ultrasensitive Detection of Subclinical Lung Cancer by Statistical Analysis of Plasma cfDNA-Derived Whole-Genome Sequencing Data**

PRINCIPAL INVESTIGATOR: **Scott L. Carter**

CONTRACTING ORGANIZATION: Dana-Farber Cancer Institute
450 Brookline Avenue, CLS 11
Boston, MA 02215-5450

REPORT DATE: **SEPTEMBER 2019**

TYPE OF REPORT: **Annual progress report**

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

DISTRIBUTION STATEMENT: Approved for Public Release; Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
SEPTEMBER 2019	ANNUAL	15AUG2018 - 14AUG2019

4. TITLE AND SUBTITLE

Ultrasensitive Detection of Subclinical Lung Cancer by Statistical Analysis of Plasma cfDNA-Derived Whole-Genome Sequencing Data

5a. CONTRACT NUMBER
W81XWH-18-1-0357

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

Dr. Scott L. Carter

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Dana-Farber Cancer Institute
450 Brookline Avenue, CLS 11
Boston, MA 02215-5450

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

We hypothesized that an ultrasensitive test for cancer could be accomplished by analysis of whole-genome sequencing data generated from plasma-derived cell-free DNA (cfDNA). An ultrasensitive test for cancer monitoring is based on paired 'fingerprint' analysis of a cancer sample from the patient. This 'fingerprint' allows detection of cryptic subclinical cancer populations that would not be detectable by other means. W also develop statistical methods for unpaired detection of cancer, suitable for early-detection screening.

15. SUBJECT TERMS

cfDNA, ctDNA, genome sequencing, early cancer detection, cancer monitoring, somatic genetics, cancer DNA fingerprinting

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRDC
Unclassified	Unclassified	Unclassified	Unclassified	30	**19b. TELEPHONE NUMBER** *(include area code)*

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

TABLE OF CONTENTS

	Page
1. Introduction	4
2. Keywords	4
3. Accomplishments	4-19
4. Impact	19-20
5. Changes/Problems	20-22
6. Products	22-24
7. Participants & Other Collaborating Organizations	25-30
8. Special Reporting Requirements	n/a
9. Appendices	n/a

1. INTRODUCTION:

We hypothesized that an ultrasensitive test for cancer could be accomplished by analysis of whole-genome sequencing data generated from plasma-derived cell-free DNA (cfDNA). An ultrasensitive test for cancer monitoring is based on paired 'fingerprint' analysis of a cancer sample from the patient. This 'fingerprint' allows detection of cryptic subclinical cancer populations that would not be detectable by other means. We also develop statistical methods for unpaired detection of cancer, suitable for early-detection screening.

2. KEYWORDS:

cfDNA, ctDNA, genome sequencing, early cancer detection, cancer monitoring, somatic genetics, cancer DNA fingerprinting

3. ACCOMPLISHMENTS:

The PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction.

What were the major goals of the project?

List the major goals of the project as stated in the approved SOW. If the application listed milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

Aim 1: Develop a flexible and robust statistical analysis toolkit to detect minute somatic clones consistent with subclinical cancer in cfDNA samples using moderate depth whole-genome sequencing. Milestone: Completion of statistical methodology for WGS detection of low level tumor DNA within plasma – 12 months. **Progress:** 80% complete – methods for copy-number, allelic ratios, and mutational processes are complete. Methods for statistical identification of sequencing errors are under very active investigation.

Aim 2: Study the sensitivity of whole-genome sequencing of cfDNA using in vitro and in vivo NSCLC specimens, benchmarked against our validated ddPCR assay.
Milestone: Demonstration of plasma WGS sensitivity exceeding sensitivity of ddPCR (18 months).
Progress: 50% complete. We have achieved detection of cancer in samples that were negative for a known ddPCR tracking allele. TBD is validation of sensitivity be serial dilution of control samples.

What was accomplished under these goals?

For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results or key outcomes, including major findings, developments, or conclusions (both positive and negative); and/or 4) other achievements. Include a discussion of stated goals not met. Description shall include pertinent data and graphs in sufficient detail to explain any significant results achieved. A succinct description of the methodology used shall be provided. As the project progresses to completion, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

We have made adequate progress on this award and on track to complete the proposed science successfully. With regard to the tasks outlined in the statement of work (SOW), we have succeeded in developing a flexible and robust statistical analysis toolkit to detect minute somatic clones consistent with subclinical cancer in cfDNA samples using moderate depth whole-genome sequencing **(Aim 1).**

We developed optimal statistical methods for cancer detection using a reference tumor-genome **(Major Task 1A).** To accomplish this, we developed tools for automated inference of tumor phylogeny from using one or more DNA samples (**Figure 1**). Additionally, we developed method to use patient tumor phylogeny to enable optimal cfDNA-based cancer detection. We have applied these methods to a large number of cfDNA samples from clinical samples obtained at Dana-Farber Cancer Institute. Currently, we are performing *in silico* validation of our detection methods by computationally simulating mixture of cfDNA-derived sequencing templates at fractions ranging from 1e-8 – 1e-2. We are evaluating the utility of copy-number, allelic imbalance, and somatic mutation for reference-enabled tumor detection across a panel of patient genomes.

We developed optimal statistical methods for cancer detection that do not require a reference tumor genome (Major Task 1B). Although we have completed the statistical work proposed in this subaim, we and others have concluded that un-paired detection of cancer in cfDNA based on moderate depth whole genome sequencing as employed here is not likely to be a competitive approach for early cancer detection. However, we believe that the statistical tools we developed for this subaim will be highly useful in other contexts.

We developed methods based on tumor aneuploidy, allelic imbalance, and tumor mutational signatures. We investigated deep-learning techniques to filter out false-positive DNA variants, but ultimately concluded that an alternate approach based on weighted logistic regression was superior. This approach has allowed us to develop a novel method for Illumina base-quality recalibration that is superior to the standard base-quality recalibration (BQSR) procedure **(Figure 2a)** and allows for more accurate calling of variants supported by only a single unique sequencing read **(Figure 2b)**. Because this method can be applied in contexts outside of cancer detection, we are writing a separate report on it that will include software.

As we proposed in **Aim 2,** we studied the sensitivity of whole-genome sequencing of cfDNA using *in vivo* NSCLC specimens, benchmarked against our validated ddPCR assay. With regard to the proposed Major Task 2A: Study preparation and activation, IRB activation of protocol #14-147 for prospective plasma collection and HRPO approval of amendment to protocol #14-147 were both successful. A clinical research coordinator (CRC) – Julianna Supplee - was trained on process for ordering, collecting, and labeling plasma specimens and the sample handling work was completed successfully.

With regard to Major Task 2B: Collection of patient plasma specimens, we were successful in obtaining specimens from paired high-low plasma from 15 NSCLC cases, including 7 for which whole-blood was also obtained for gDNA sequencing. Plasma-derived cfDNA from 11 healthy donors was also sequenced for negative controls. A list of samples collected and sequenced for this award is included as **Table 1**.

With regard to Major Task 2C: Cell line analysis – several of the subtasks in this aim are still outstanding, and some are being reconsidered in terms of optimal experimental design. The remaining in this Major task has been disrupted by Covid and we have devised a computational substitute for these experiments that we will submit as a formal amendment proposal to close out this grant.

With regard to Major Task 2D: Analysis of patient cfDNA, we have generated serial ddPCR data to assess tumor concentration in response to therapy. We performed WGS of cfDNA at time of complete plasma response to detect residual tumor DNA. In most cases, no cancer was detected by WGS, however in one case (301, EGFR-mutant lung adenocarcinoma), the residual cancer was robustly detected, despite the failure of all other means to detect cancer in that sample. This case serves as a demonstration of plasma WGS sensitivity exceeding sensitivity of ddPCR.

With regard to Major Task 2E: Blinded plasma analysis, ddPCR and WGS assessment of post-surgical plasma to identify residual tumor DNA – a subset of the cases we have sequenced and analyzed fit this criterion; analysis of WGS-based cancer detection with regard to clinical outcome is currently in progress.

Figure 1: Phylogenetic matrix factorization of tumor copy-number profile for MRD detection. (top-left) observed segmental copy-ratio profile for 3 samples obtained from patient #1: top-to-bottom high-shed, low-shed, whole blood. (top 2nd from left). The observed multi-sample copy-ratio matrix is discretized into unique states using a Bayesian clustering method, in this case grouping 47 genomic segments into 20 clusters, with minimal loss of accuracy (top 3rd from left). (Bottom-left) Reconstruction of the copy-ratio matrix using a phylogenetic factorization into 3 clones (bottom 2nd from left), the 1st of which is the fixed germline, and the 2nd two are somatic cancer clones. The phylogeny matrix here indicates that clone 3 is descended from clone 2, in a linear phylogeny. The detection of clones 2 and 3 (cancer) in the low-shed sample, but not in the whole blood, serves a s proof of principal for phylogenetic detection of MRD.

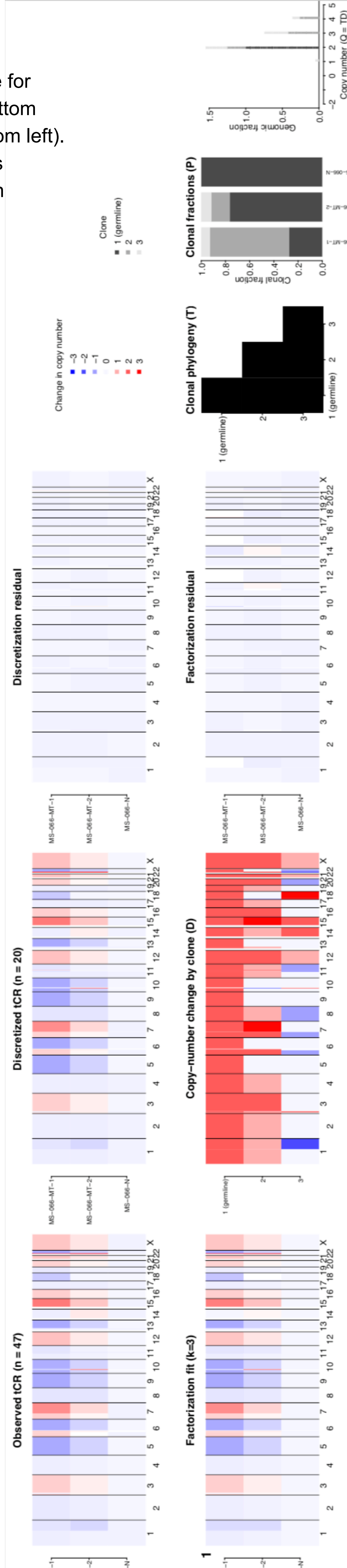

Figure 2. a): Calibration plot of standard BQSR quality scores and "Better base-quality" (BQ; our work). Empirical and observed quality scores are plotted in standard PHRED units. Empirical base qualities were obtained by utilizing discordant pairs of reads cover the same DNA fragment.

Calibration

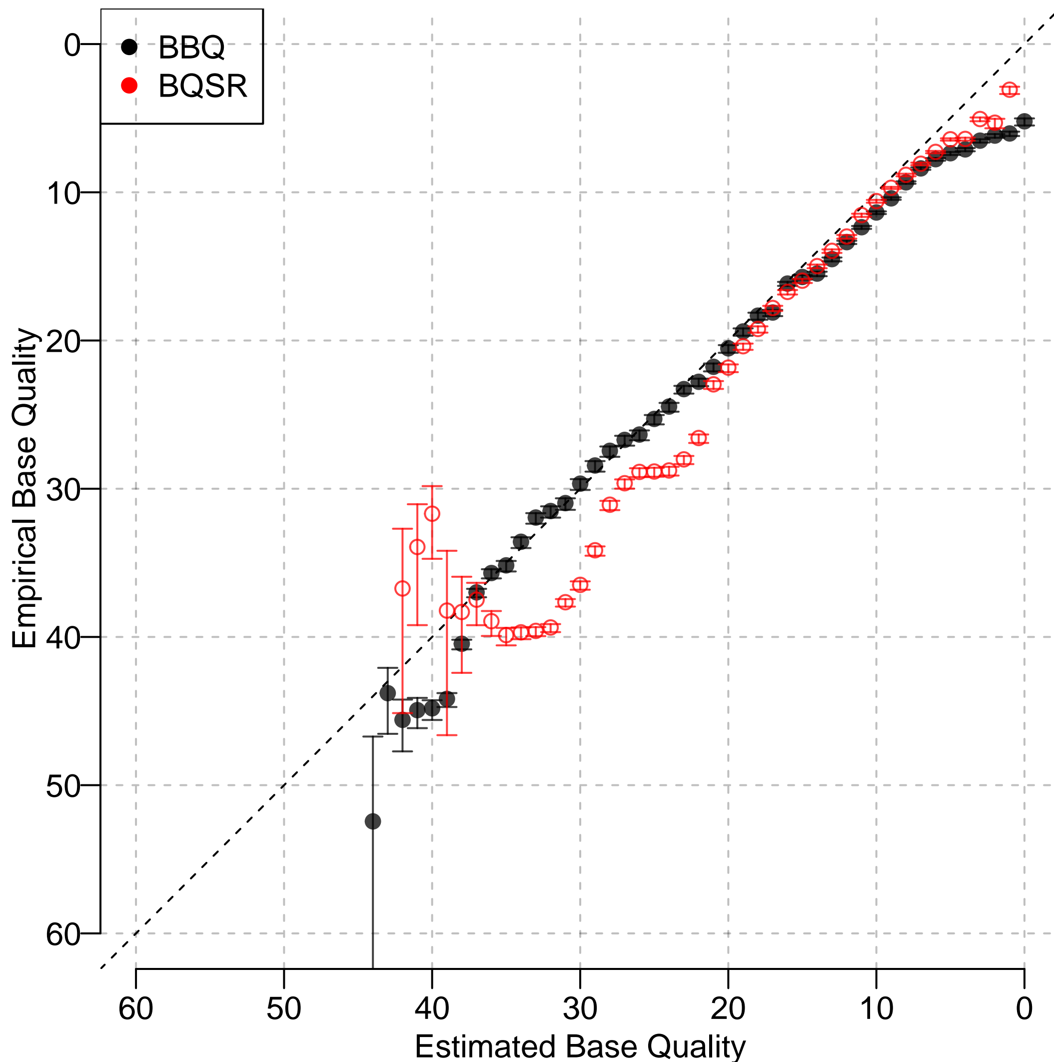

Figure 2b.) Histogram of observed BQSR and BBQ values in reads held out from training the BBQ model.

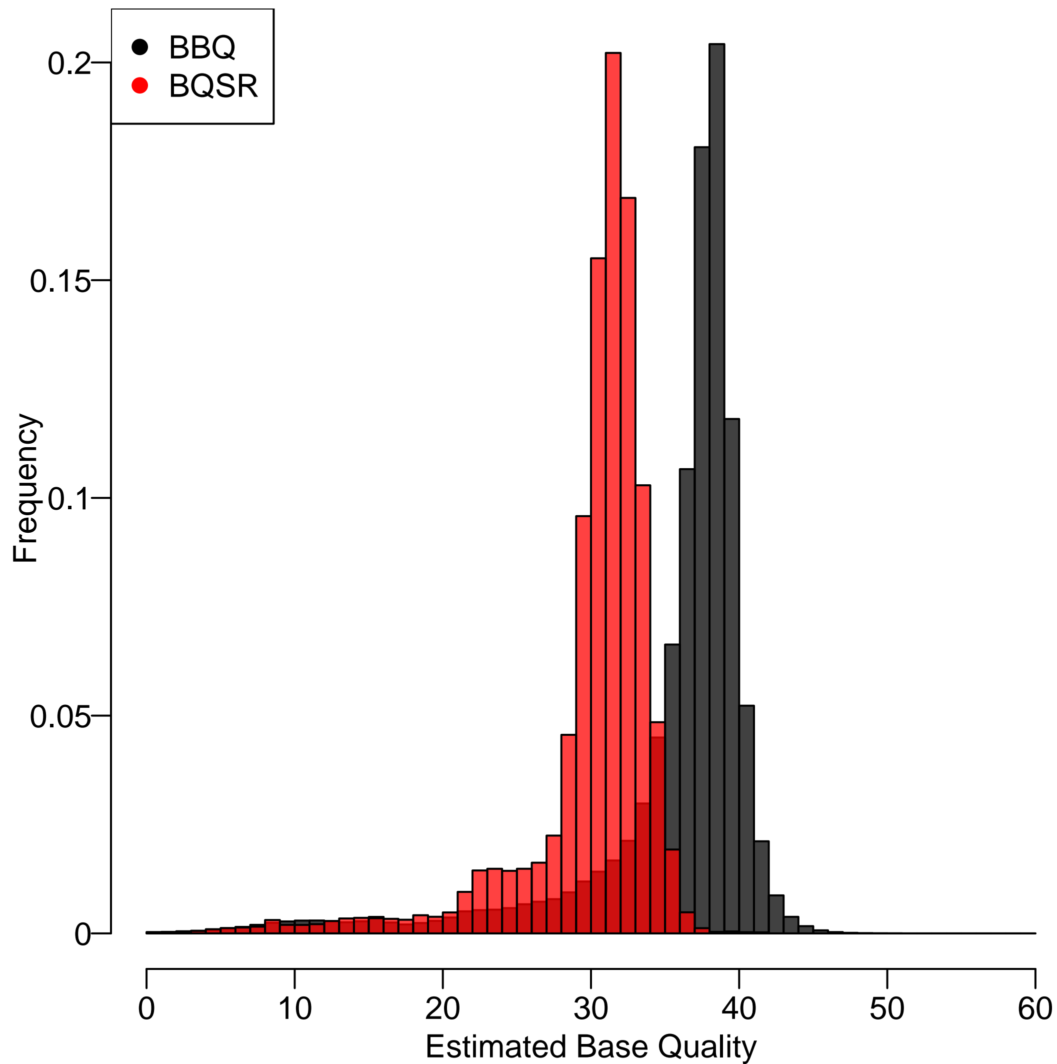

Table 1. Table of samples with completed whole genome sequencing.

individual_id	sample_id	sample_type	SEX_GENOTYPE	display_name	Primary tumor	sequencing tech	DNA type	DNA source	PERCENT_DUPLICATION	concentration_cfDNA(ng/uL)	total_cfDNA_yield(ng)	input_to_LC(ng)	cfDNA_yield(ng/mL)	cfDNA_remaining(ng)
10417_11	FC19309873	Tumor	XX	10417_11__FC19309873	breast	MiSeq	cfDNA	plasma	0.109163	0.228	13.662	3.036	11.4	1.594
10417_11	FC19309874	Tumor	XX	10417_11__FC19309874	breast	MiSeq	cfDNA	plasma	0.135157	0.113	6.768	1.934	5.6	0.79
10417_11	FC19309891	Tumor	XX	10417_11__FC19309891	breast	MiSeq	cfDNA	CSF	0.151905	0.151	9.072	2.11	7.6	1.058
10417_8	FC19665425	Tumor	XX	10417_8__FC19665425	breast	MiSeq	cfDNA	plasma	0.161655	0.69	41.382	9.853	20	19.312
10417_8	FC19269592	Tumor	XX	10417_8__FC19269592	breast	MiSeq	cfDNA	plasma	0.144049	0.19	11.376	2.844	9.48	1.327
10417_8	FC19270024	Tumor	XX	10417_8__FC19270024	breast	MiSeq	cfDNA	plasma	0.143721	0.26	15.57	6.488	12.98	1.816
10417_D5	FC19309880	Tumor	XX	10417_D5__FC19309880	breast	MiSeq	cfDNA	plasma	0.122179	0.418	25.092	4.182	20	3.847
10417_D5	FC19309881	Tumor	XX	10417_D5__FC19309881	breast	MiSeq	cfDNA	CSF	0.139216	0.085	5.076	0.846	4.2	0.592
10417_D5	FC19309882	Tumor	XX	10417_D5__FC19309882	breast	MiSeq	cfDNA	plasma	0.137784	0.353	21.168	3.528	17.6	2.47
10417_D8	FC19309890	Tumor	XX	10417_D8__FC19309890	breast	MiSeq	cfDNA	plasma	0.125044	0.425	25.506	7.287	20	4.251
10417_D8	FC19309892	Tumor	XX	10417_D8__FC19309892	breast	MiSeq	cfDNA	CSF	0.194691	0.074	4.446	1.389	3.7	0.519
10417_D8	FC19309893	Tumor	XX	10417_D8__FC19309893	breast	MiSeq	cfDNA	plasma	0.138546	0.711	42.66	10.157	20	20.548
10417_D2	FC19665684	Tumor	XX	10417_D2__FC19665684	breast	MiSeq	cfDNA	plasma	0.148119	1.098	65.898	NA	20	42.603
10417_D2	FC19309904	Tumor	XX	10417_D2__FC19309904	breast	MiSeq	cfDNA	plasma	0.136445	1.21	72.63	13.704	20	48.998
10417_D2	FC19309905	Tumor	XX	10417_D2__FC19309905	breast	MiSeq	cfDNA	CSF	0.298156	0.06	3.618	0.603	3.1	0.422
98	FC16207246	Tumor	XY	98__FC16207246	melanoma	MiSeq	cfDNA	plasma	0.162376	0.058	3.51	5.014	2.93	NA
98	FC16207248	Tumor	XY	98__FC16207248	melanoma	MiSeq	cfDNA	plasma	0.404583	0.168	10.062	12.578	5	NA
98	SM-61EX7	Tumor	XY	98__SM-61EX7	melanoma	OTHER (GP)	gDNA	frozen tissue	NA	NA	NA	NA	NA	NA

```
98      SM-I7OT4        Normal  XY      98WB_1-22-15    melanoma        OTHER (GP)      gDNA    germline
NA      NA      NA      NA      NA      NA
T33     FC16207280      Tumor   XX      T33__FC16207280 melanoma        MiSeq   cfDNA   plasma  0.328244
0.104   6.228   6.228   5       NA
T33     FC16207281      Tumor   XX      T33__FC16207281 melanoma        MiSeq   cfDNA   plasma  0.292957
0.046   2.754   2.754   2.3     NA
T33     FC16207282      Tumor   XX      T33__FC16207282 melanoma        MiSeq   cfDNA   plasma  0.385847
0.053   3.168   3.52    2.64    NA
T33     SM-9ZE13        Tumor   XX      T33__SM-9ZE13   melanoma        OTHER (GP)      gDNA    frozen
tissue  NA      NA      NA      NA      NA      NA
T33     SM-I7OT3        Normal  XX      T33WB_2-19-15   melanoma        OTHER (GP)      gDNA    germline
NA      NA      NA      NA      NA      NA
HD2     HD2     Normal  XY      HD2__HD2        healthy donor   MiSeq   cfDNA   plasma  NA      1.049
62.964  15.741  20      NA
HD45    HD45    Normal  XY      HD45__HD45      healthy donor   MiSeq   cfDNA   plasma  NA      1.438
86.256  1.725   20      NA
HD46    HD46    Normal  XY      HD46__HD46      healthy donor   MiSeq   cfDNA   plasma  NA      3.135
188.118 3.919   20      NA
02180_41        02180_41_1      Tumor   XX      02180_41__02180_41_1    lung adenocarcinoma     HISEQ_X_10
cfDNA   plasma  0.141183        NA      NA      NA      NA      NA
02180_41        02180_41_2      Tumor   XX      02180_41__02180_41_2    lung adenocarcinoma     HISEQ_X_10
cfDNA   plasma  0.189876        NA      NA      NA      NA      NA
02180_098       02180_098_1     Tumor   XY      02180_098__02180_098_1  lung adenocarcinoma     HISEQ_X_10
cfDNA   plasma  0.209396        NA      NA      NA      NA      NA
02180_098       02180_098_2     Tumor   XY      02180_098__02180_098_2  lung adenocarcinoma     HISEQ_X_10
cfDNA   plasma  0.150692        NA      NA      NA      NA      NA
02180_387       02180_387_1     Tumor   XY      02180_387__02180_387_1  lung adenocarcinoma     HISEQ_X_10
cfDNA   plasma  0.196101        NA      NA      NA      NA      NA
02180_387       02180_387_2     Tumor   XY      02180_387__02180_387_2  lung adenocarcinoma     HISEQ_X_10
cfDNA   plasma  0.170497        NA      NA      NA      NA      NA
02180_509       02180_509_1     Tumor   XX      02180_509__02180_509_1  lung adenocarcinoma     HISEQ_X_10
cfDNA   plasma  0.154958        NA      NA      NA      NA      NA
02180_509       02180_509_2     Tumor   XX      02180_509__02180_509_2  lung adenocarcinoma     HISEQ_X_10
cfDNA   plasma  0.166623        NA      NA      NA      NA      NA
BosBio_Streck_117       BosBio_Streck_117_1     Normal  XY      BosBio_Streck_117__BosBio_Streck_117_1
healthy donor   HISEQ_X_10      cfDNA   plasma  0.150684        NA      NA      NA      NA      NA
BosBio_Streck_130       BosBio_Streck_130_1     Normal  XY      BosBio_Streck_130__BosBio_Streck_130_1
healthy donor   HISEQ_X_10      cfDNA   plasma  0.097476        NA      NA      NA      NA      NA
BosBio_Streck_132       BosBio_Streck_132_1     Normal  XY      BosBio_Streck_132__BosBio_Streck_132_1
```

healthy donor HISEQ_X_10 cfDNA plasma 0.09873 NA NA NA NA NA
BosBio_Streck_38 BosBio_Streck_38_1 Normal XX BosBio_Streck_38__BosBio_Streck_38_1
healthy donor HISEQ_X_10 cfDNA plasma 0.147979 NA NA NA NA NA
BosBio_Streck_4 BosBio_Streck_4_1 Normal XY BosBio_Streck_4__BosBio_Streck_4_1 healthy
donor HISEQ_X_10 cfDNA plasma 0.137309 NA NA NA NA NA
BosBio_Streck_70 BosBio_Streck_70_1 Normal XX BosBio_Streck_70__BosBio_Streck_70_1
healthy donor HISEQ_X_10 cfDNA plasma 0.089495 NA NA NA NA NA
BosBio_Streck_89 BosBio_Streck_89_1 Normal XX BosBio_Streck_89__BosBio_Streck_89_1
healthy donor HISEQ_X_10 cfDNA plasma 0.100134 NA NA NA NA NA
BosBio_Streck_90 BosBio_Streck_90_1 Normal XY BosBio_Streck_90__BosBio_Streck_90_1
healthy donor HISEQ_X_10 cfDNA plasma 0.108573 NA NA NA NA NA
10417_1 FC19269626 Tumor XX 10417_1_FC19269626 breast MiSeq cfDNA CSF 0.182583
0.06 3.582 0.968 2.99 0.418
10417_1 FC19307423 Tumor XX 10417_1_FC19307423 breast MiSeq cfDNA CSF 0.298697
0.074 4.446 1.588 3.71 0.519
10417_1 FC19311536 Tumor XX 10417_1_FC19311536 breast MiSeq cfDNA plasma 0.103525
0.522 31.302 7.114 19.98 9.756
10417_1 FC19667111 Tumor XX 10417_1_FC19667111 breast MiSeq cfDNA plasma 0.154801
0.278 16.704 4.176 13.92 1.949
195 SM-HUS9A Tumor XY 195_a_II_1 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA
88-09 SM-HUS9B Tumor XY 88-09_a_C5D1 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA
310 SM-HUS97 Tumor XX 310_a_T2W2 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA
716 SM-HUS98 Tumor XX 716_a_C4D1 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA
331 SM-HUS99 Tumor XY 331_a_C1W76 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA
399 SM-HPYHQ Tumor XY 399_a_s1 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
405 SM-HPYHR Tumor XY 405_a_s2 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
406 SM-HPYHS Tumor XY 406_a_s3 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
410 SM-HPYHT Tumor XY 410_a_s4 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
441 SM-HPYHU Tumor XX 441_a_s5 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA

465 SM-HPYHV Tumor XY 465_a_s6 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
498 SM-HPYHW Tumor XY 498_a_s7 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
502 SM-HPYHX Tumor XY 502_a_s8 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
512 SM-HPYHY Tumor XX 512_a_s9 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
556 SM-HPYHZ Tumor XX 556_a_s10 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
592 SM-HPYI1 Tumor XY 592_a_s11 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
632 SM-HPYI2 Tumor XY 632_a_s12 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
748 SM-HPYI3 Tumor XY 748_a_s13 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
310 SM-IEMLD Normal XX 310_a_BC_1 lung adenocarcinoma HISEQ_X_10 gDNA
germline NA NA NA NA NA NA
399 SM-IEMLA Normal XY 399_a_g1 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
405 SM-IEMLC Normal XY 405_a_g2 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
406 SM-IEML6 Normal XY 406_a_g3 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
410 SM-IEML4 Normal XY 410_a_g4 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
441 SM-IEML8 Normal XX 441_a_g5 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
465 SM-IEMKZ Normal XY 465_a_g6 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
498 SM-IEML1 Normal XY 498_a_g7 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
502 SM-IEMLB Normal XY 502_a_g8 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
512 SM-IEML9 Normal XX 512_a_g9 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
556 SM-IEML3 Normal XX 556_a_g10 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
592 SM-IEML2 Normal XY 592_a_g11 melanoma HISEQ_X_10 gDNA germline

NA NA NA NA NA NA
632 SM-IEML7 Normal XY 632_a_g12 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
748 SM-IEML5 Normal XY 748_a_g13 melanoma HISEQ_X_10 gDNA germline
NA NA NA NA NA NA
399 SM-HQUXM Tumor XY 399_b_s14 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
405 SM-HQUXN Tumor XY 405_b_s15 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
406 SM-HQUXO Tumor XY 406_b_s16 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
410 SM-HQUXP Tumor XY 410_b_s17 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
441 SM-HQUXQ Tumor XX 441_b_s18 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
465 SM-HQUXR Tumor XY 465_b_s19 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
498 SM-HQUXS Tumor XY 498_b_s20 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
502 SM-HQUXT Tumor XY 502_b_s21 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
512 SM-HQUXU Tumor XX 512_b_s22 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
556 SM-HQUXV Tumor XX 556_b_s23 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
592 SM-HQUXW Tumor XY 592_b_s24 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
632 SM-HQUXX Tumor XY 632_b_s25 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
748 SM-HQUXY Tumor XY 748_b_s26 melanoma Nova UMI cfDNA plasma NA
NA NA NA NA NA
88-09 SM-HKBMI Tumor XY 88-09_b_C17D1 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA
195 SM-HKBMH Tumor XY 195_b_IA_1 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA
310 SM-HKBME Tumor XX 310_b_T2W19 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA
331 SM-HKBMG Tumor XY 331_b_T1W97 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA

716 SM-HKBMF Tumor XX 716_b_C1D1 lung adenocarcinoma Nova UMI cfDNA
plasma NA NA NA NA NA NA
592 SM-IOLDO Tumor XY 592_S1 melanoma HISEQ_X_10 gDNA frozen tissue NA
NA NA NA NA NA
592 SM-IOLDP Tumor XY 592_S2 melanoma HISEQ_X_10 gDNA frozen tissue NA
NA NA NA NA NA
406 SM-IOLDQ Tumor XY 406_S melanoma HISEQ_X_10 gDNA frozen tissue NA
NA NA NA NA NA
406 SM-IOLDR Tumor XY 406_F melanoma HISEQ_X_10 gDNA FFPE tissue NA
NA NA NA NA NA
592 SM-IOLDS Tumor XY 592_F1 melanoma HISEQ_X_10 gDNA FFPE tissue NA
NA NA NA NA NA
592 SM-IOLDU Tumor XY 592_F2 melanoma HISEQ_X_10 gDNA FFPE tissue NA
NA NA NA NA NA
E7800005 SM-J83MQ Tumor XX H1 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
14-147_251 SM-J83MR Tumor XX H2 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
14-147_327 SM-J83MS Tumor XX H3 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
16-093_9 SM-J83MT Tumor XX H4 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
14-147_413 SM-J83MU Tumor XX H5 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
14-147_132 SM-J83MV Tumor XX H6 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
E7800005 SM-J4ACJ Tumor XX L1 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
14-147_251 SM-J4ACK Tumor XX L2 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
14-147_327 SM-J4ACL Tumor XX L3 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
16-093_9 SM-J4ACM Tumor XX L4 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
14-147_413 SM-J4ACN Tumor XX L5 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
14-147_132 SM-J4ACO Tumor XX L6 lung adenocarcinoma HISEQ_X_10 cfDNA
plasma
E7800005 SM-IA4JZ Normal XX G1 lung adenocarcinoma HISEQ_X_10 gDNA

germline
14-147_251 SM-IA4K1 Normal XX G2 lung adenocarcinoma HISEQ_X_10 gDNA
germline
14-147_327 SM-IA4K2 Normal XX G3 lung adenocarcinoma HISEQ_X_10 gDNA
germline
16-093_9 SM-IA4K3 Normal XX G4 lung adenocarcinoma HISEQ_X_10 gDNA
germline
14-147_413 SM-IA4K4 Normal XX G5 lung adenocarcinoma HISEQ_X_10 gDNA
germline
14-147_132 SM-IA4K5 Normal XX G6 lung adenocarcinoma HISEQ_X_10 gDNA
germline
10417_29 SM-GHGQP Tumor XX CSF-029-04-csf breast cfDNA CSF

10417_29 SM-GHGQE Tumor XX CSF-029-05-csf breast cfDNA CSF

10417_43 SM-GHGQO Tumor XX CSF-043-01-csf breast cfDNA CSF

10417_43 SM-GG35M Tumor XX CSF-043-02-b breast gDNA plasma

10417_43 SM-GHGPV Tumor XX CSF-043-02-csf breast cfDNA CSF

10417_43 SM-GG34L Tumor XX CSF-043-03-b breast gDNA plasma

10417_43 SM-GHGQ6 Tumor XX CSF-043-03-csf breast cfDNA CSF

10417_43 SM-GHGQN Tumor XX CSF-043-04-csf breast cfDNA CSF

10417_50 SM-GG356 Tumor XX CSF-050-01 breast cfDNA CSF

399 SM-JP769 Tumor XY Pt399tumDNA melanoma HISEQ_X_10 gDNA frozen
tissue
405 SM-JP76D Tumor XY Pt405tumDNA melanoma HISEQ_X_10 gDNA frozen
tissue
441 SM-JP76C Tumor XX Pt441tumDNA melanoma HISEQ_X_10 gDNA frozen
tissue
465 SM-JP76F Tumor XY Pt465tumDNA melanoma HISEQ_X_10 gDNA frozen
tissue
498 SM-JP766 Tumor XY Pt498tumDNA melanoma HISEQ_X_10 gDNA frozen
tissue

502	SM-JP76E	Tumor	XY	Pt502tumDNA	melanoma	HISEQ_X_10	gDNA	frozen tissue
512	SM-JP76A	Tumor	XX	Pt512tumDNA	melanoma	HISEQ_X_10	gDNA	frozen tissue
556	SM-JP767	Tumor	XX	Pt556tumDNA	melanoma	HISEQ_X_10	gDNA	frozen tissue
632	SM-JP76B	Tumor	XY	Pt632tumDNA	melanoma	HISEQ_X_10	gDNA	frozen tissue
748	SM-JP768	Tumor	XY	Pt748tumDNA	melanoma	HISEQ_X_10	gDNA	frozen tissue

What opportunities for training and professional development has the project provided?
If the project was not intended to provide training and professional development opportunities or there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. "Training" activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency. Training activities may include, for example, courses or one-on-one work with a mentor. "Professional development" activities result in increased knowledge or skill in one's area of expertise and may include workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.

Two trainees have contributed significantly to the development of analysis methods for this project. One of them developed a 'deep learning' framework involving convolutional neural networks for the classification of sequencing reads as errors. That person has now started a PhD in genomics as U. Wisconsin. The second trainee picked up on this work and extended it, refining the deep learning approach and also developing a regression approach that gave more interpretability of the model predictions. That trainee has applied for PhD training and an NSF scholarship based on this work.

How were the results disseminated to communities of interest?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the results were disseminated to communities of interest. Include any outreach activities that were undertaken to reach members of communities who are not usually aware of these project activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

Results have been disseminated to local physician-scientist colleagues. Several clinical trials are in the works based on the preliminary data from this project. This includes application to areas outside of lung cancer. We anticipate submission of a manuscript summarizing our results in the next few months.

Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

We are on track to meet our goals for methods development and generation of patient data.. The main piece of work to be accomplished is the construction of the DNA-mixing experiment. We are designing this experiment at the moment, weighing such considerations such as whether to use genomic or cell-free DNA. Additional considerations include tumor mutational burden, which is a key factor determining what sensitivity is possible.

4. **IMPACT:** D*escribe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:*

What was the impact on the development of the principal discipline(s) of the project?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research in the principal disciplinary field(s) of the project. Summarize using language that an intelligent lay audience can understand (Scientific American style).

We have demonstrated detection of cancer DNA in samples that are negative by all other means, including digital-droplet PCR (ddPCR). We are on track to evaluate this as a predictive biomarker of cancer recurrence after therapy with curative intent, with patients selected for short and long-term disease-free survival. Successful completion of this work would provide the needed citation for IRB approval of several novel precision oncology clinical trials, specifically: i) Determination of a 'safe' level of residual disease for cessation of a toxic therapy e.g. immune-checkpoint blockade; ii) Identification of a patient subset at increased risk of progression based on cfDNA WGS and evaluate stratified trial arms accordingly.

What was the impact on other disciplines?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

Although we wrote this grant for lung cancer, the methods are general. We have generated data from breast cancer and melanoma patients and are evaluation application to those areas simultaneously. Additionally, the genomic data analysis methods we have developed will be broadly applicable, including to fields outside cancer.

What was the impact on technology transfer?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe ways in which the project made an impact, or is likely to make an impact, on commercial technology or public use, including:

- *transfer of results to entities in government or industry;*
- *instances where the research has led to the initiation of a start-up company; or*
- *adoption of new practices.*

We have filed provisional patent applications covering aspects of this work that we feel have commercial application. We are aware of potential for a start-up venture founded on making this technology widely available nationwide.

What was the impact on society beyond science and technology?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world on areas such as:

- *improving public knowledge, attitudes, skills, and abilities;*
- *changing behavior, practices, decision making, policies (including regulatory policies), or social actions; or*
- *improving social, economic, civic, or environmental conditions.*

This research is still very preliminary and major milestones need to be accomplished, mainly demonstrating that early detection of minute cancer DNA in patient plasma is a clinically useful biomarker for reducing cancer mortality and morbidity. We strongly believe in the fundamental scientific principles behind this research and are determined to see our approach reach its full potential.

5. **CHANGES/PROBLEMS:** *The PD/PI is reminded that the recipient organization is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. If not previously reported in writing, provide the following additional information or state, "Nothing to Report," if applicable:*

Nothing to Report

Actual or anticipated problems or delays and actions or plans to resolve them

Describe problems or delays encountered during the reporting period and actions or plans to resolve them.

Nothing to Report

Changes that had a significant impact on expenditures

Describe changes during the reporting period that may have had a significant impact on expenditures, for example, delays in hiring staff or favorable developments that enable meeting objectives at less cost than anticipated.

The cost of whole-genome sequencing has dropped to $750 / sample, from $1,380 in the original budget. We will use the extra funds to sequence more patient samples and synthetic controls.

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

Describe significant deviations, unexpected outcomes, or changes in approved protocols for the use or care of human subjects, vertebrate animals, biohazards, and/or select agents during the reporting period. If required, were these changes approved by the applicable institution committee (or equivalent) and reported to the agency? Also specify the applicable Institutional Review Board/Institutional Animal Care and Use Committee approval dates.

Significant changes in use or care of human subjects

Nothing to Report

Nothing to Report

Significant changes in use of biohazards and/or select agents

Nothing to Report

6. **PRODUCTS:** *List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state "Nothing to Report."*

- **Publications, conference papers, and presentations**
  Report only the major publication(s) resulting from the work under this award.

  Journal publications. *List peer-reviewed articles or papers appearing in scientific, technical, or professional journals. Identify for each publication: Author(s); title; journal; volume: year; page numbers; status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

  Nothing to Report

  Books or other non-periodical, one-time publications. *Report any book, monograph, dissertation, abstract, or the like published as or in a separate publication, rather than a periodical or series. Include any significant publication in the proceedings of a one-time conference or in the report of a one-time study, commission, or the like. Identify for each one-time publication: author(s); title; editor; title of collection, if applicable; bibliographic information; year; type of publication (e.g., book, thesis or dissertation); status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

Nothing to Report

Other publications, conference papers and presentations. *Identify any other publications, conference papers and/or presentations not reported above. Specify the status of the publication as noted above. List presentations made during the last year (international, national, local societies, military meetings, etc.). Use an asterisk (*) if presentation produced a manuscript.*

Nothing to Report

- **Website(s) or other Internet site(s)**
  List the URL for any Internet site(s) that disseminates the results of the research activities. A short description of each site should be provided. It is not necessary to include the publications already specified above in this section.

Nothing to Report

- **Technologies or techniques**
  Identify technologies or techniques that resulted from the research activities. Describe the technologies or techniques were shared.

Nothing to Report

- **Inventions, patent applications, and/or licenses**
  Identify inventions, patent applications with date, and/or licenses that have resulted from the research. Submission of this information as part of an interim research performance progress report is not a substitute for any other invention reporting required under the terms and conditions of an award.

Nothing to Report

- **Other Products**
  Identify any other reportable outcomes that were developed under this project. Reportable outcomes are defined as a research result that is or relates to a product, scientific advance, or research tool that makes a meaningful contribution toward the understanding, prevention, diagnosis, prognosis, treatment and /or rehabilitation of a disease, injury or condition, or to improve the quality of life. Examples include:
  - *data or databases;*
  - *physical collections;*
  - *audio or video products;*
  - *software;*
  - *models;*
  - *educational aids or curricula;*
  - *instruments or equipment;*
  - *research material (e.g., Germplasm; cell lines, DNA probes, animal models);*
  - *clinical interventions;*
  - *new business creation; and*
  - *other.*

We have developed multiple software packages for statistical analysis of genome-sequence data and enabling ultra-sensitive detection of cancer using off-the-shelf WGS technology.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Provide the following information for: (1) PDs/PIs; and (2) each person who has worked at least one person month per year on the project during the reporting period, regardless of the source of compensation (a person month equals approximately 160 hours of effort). If information is unchanged from a previous submission, provide the name only and indicate "no change".

Name: Scott Carter
Project Role: PD/PI
Nearest person month worked: 1.8CM
Contribution to project: Dr. Carter supervised the analysis and interpretation of the genomic data generated in the project. *** No Change from original submission***

Name: Geoffrey Oxnard
Project Role: Co-PI
Nearest person month worked: 1CM
Contribution to project: Dr. Oxnard supervised the clinical translation of the statistical techniques for application to the study of clinical specimens. ***No Change from original submission***

Name: Cloud Paweletz
Project Role: Co_PI
Nearest person month worked: 1CM
Contribution to project: Dr. Paweletz supervised the analysis and reporting of the digital PCR assays at the TRL, and will aid in the comparison of plasma genotyping results across orthogonal assays. ***No Change from original submission***

Name: Julianna Supplee
Project Role: Research Technician
Nearest person month worked: 5CM
Contribution to project: Julianna was responsible for the construction of cell line dilutions, extraction of DNA from plasma, and implementation of digital PCR assays

Name: Andrew Portell
Project Role: Research Associate
Nearest person month worked: 1CM
Contribution to project: Responsible for the construction of cell line dilutions, extraction of DNA from plasma, and implementation of digital PCR assays.

Name: Kevin Vasquez
Project Role: Research Associate
Nearest person month worked: 1CM
Contribution to project: Responsible for the construction of cell line dilutions, extraction of DNA from plasma, and implementation of digital PCR assays.

Name: Jeff Miller
Project Role: Co-Investigator
Nearest person month worked: 2CM
Contribution to project: Developed statistical methodology for copy-number analysis that is optimal for our cfDNA dataset.

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

If the active support has changed for the PD/PI(s) or senior/key personnel, then describe what the change has been. Changes may occur, for example, if a previously active grant has closed and/or if a previously pending grant is now active. Annotate this information so it is clear what has changed from the previous submission. Submission of other support information is not necessary for pending changes or for changes in the level of effort for active support reported previously. The awarding agency may require prior written approval if a change in active other support significantly impacts the effort on the project that is the subject of the project report.

See attached updated other support pages for Scott Carter and Jeff Miller.

What other organizations were involved as partners?
If there is nothing significant to report during this reporting period, state "Nothing to Report."

Describe partner organizations – academic institutions, other nonprofits, industrial or commercial firms, state or local governments, schools or school systems, or other organizations (foreign or domestic) – that were involved with the project. Partner organizations may have provided financial

or in-kind support, supplied facilities or equipment, collaborated in the research, exchanged personnel, or otherwise contributed.

Provide the following information for each partnership:
<u>Organization Name:</u>
<u>Location of Organization: (if foreign location list country)</u>
<u>Partner's contribution to the project</u> (identify one or more)

- *Financial support;*
- *In-kind support (e.g., partner makes software, computers, equipment, etc., available to project staff);*
- *Facilities (e.g., project staff use the partner's facilities for project activities);*
- *Collaboration (e.g., partner's staff work with project staff on the project);*
- *Personnel exchanges (e.g., project staff and/or partner's staff use each other's facilities, work at each other's site); and*
- *Other.*

The Broad Institute of MIT-Harvard
415 Main St., Cambridge MA, USA
Contributed whole-genome sequencing and alignment of raw read data.
(No change from original application)

Massachusetts General Hospital
Boston, MA, USA
Provided patient samples for genome sequencing
(No change from original application)

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS:

9. QUAD CHARTS:

10. APPENDICES:

OTHER SUPPORT: (As of Report Date: November 2019)

CARTER, SCOTT

ACTIVE:

R01CA227156 (Brastianos/Carter) 05/01/18 - 04/30/23 1.8 CM
NIH/NCI
Identification of genomic drivers of brain metastases in lung adenocarcinoma
Specific Aim 1: Identify candidate brain-metastasis drivers through whole exome sequencing of patient-matched BM-LUAD and normal tissues.
Specific Aim 2: Assess whether drivers of BM-LUAD can be detected in clinically sampled primary tumors, and thus potentially serve as predictive biomarkers.
Specific Aim 3: Evaluate the role of candidate BM-LUAD drivers in functional assays of metastasis.
Role: Principal Investigator (Multi PI)

W81XWH-18-1-0357 (Carter) 08/15/18-09/14/20 1.8 CM
Department of Defense
LCRP: Ultra-sensitive detection of Subclinical Lung Cancer by Statistical Analysis of Plasma-cfDNA-Derived Whole-Genome Sequencing Data
We hypothesize that a sequencing approach focused on moderate-depth broad sequencing (whole genome) rather than deep targeted sequencing, combined with our novel advanced statistical methodologies, will be able to achieve exquisitely sensitive detection of tumor DNA in plasma.
Specific Aim 1: Develop a flexible and robust statistical analysis toolkit to detect minute somatic clones consistent with subclinical cancer in cfDNA samples using moderate depth whole-genome sequencing.
Specific Aim 2: Study the sensitivity of whole-genome sequencing of cfDNA using in vitro and in vivo NSCLC specimens, benchmarked against our validated ddPCR assay.

R21CA220253 (Brastianos, Carter, Choueiri) 09/01/2018 - 08/31/2020 0.6 CM
NIH/NCI (total costs for year to DFCI: Carter & Choueiri)
Identification of genomic drivers of brain metastases in renal cell carcinoma
Our proposed genomic analysis will focus on characterizing the molecular alterations in brain metastases from RCC. We will analyze brain metastases, extracranial metastases, cerebrospinal fluid, and plasma samples before and after targeted therapy or immune checkpoint blockade therapy to identify genetic alterations driving their clinical behavior, including resistance to therapy. Identification of these mutations will aid in the design of more effective targeted treatments to treat brain metastases.
Role: Principal Investigator (Multi PI)

PENDING
*Grant for JIT request
R01CA240299 (Gerstner/Brastianos) 09/01/19-08/31/24 1.2CM
NIH/NCI
Using MRI and circulating tumor DNA to improve the interpretation of response to immunotherapy and targeted therapy in CNS metastases
Supervise the analysis and interpretation of the genomic data generated in the project.
Role: Other Significant Contributor (Key Personnel)

R01CA240299 (Miller) 07/01/20-06/30/25 1.2CM
NIH/NCI (Total Costs to DFCI: Carter & Parmigiani)
Statistical methods for cancer genomics and cell-free DNA analysis
Develop robust methods for latent factorization models of mutation count data. The mutational processes that lead to cancer exhibit characteristic genome-wide signatures in the frequency with which various categories of mutations occur. Develop methods for SCNA inference using structured models for sequential data. Cancer genomes exhibit segmental duplications and deletions relative to normal diploid genomes. Develop software,

provide data, and disseminate results to facilitate reproducibility. Our goal is to develop methods and tools that will become the standard for high-performance, statistically sound cancer genomics and cfDNA analysis.
Role: Role: Other Significant Contributor (Key Personnel)

OVERLAP:
None

FOREIGN COLLABORATIONS:
None

Other Support
Jeffrey Miller
Harvard T.H. Chan School of Public Health

ACTIVE

(THIS AWARD)
W81XWH1810357 (Carter) 08/15/2018 – 08/14/2020 1.8 CM
DOD / DFCI
Ultra-Sensitive Detection of Subclinical Lung Cancer by Statistical Analysis of Plasma-cfDNA-Derived Whole-Genome Sequencing Data
Goals: (1) Develop statistical methods and computational tools for the detection of minute somatic clones from cfDNA whole-genome sequencing; (2) Evaluate the performance of these methods and tools on in silico and in vitro data.
Role: Consortium PI

CHANGES IN OTHER SUPPORT

(NEW)
R01GM083084 (Irizarry) 03/01/2019 – 06/30/2020 1.2 CM
NIH / DFCI
Preprocessing and Analysis Tools for High-Throughput Technologies
Develop statistical methods to perform preprocessing and analysis of high-throughput next generation sequencing data, to overcome biases and remove unwanted variability.
Role: Consortium PI

(NEW)
No Award Number (Miller) 10/01/2019 – 09/30/2021 3.6 CM
MGH
Biostatistical analysis of X-Linked Dystonia-Parkinsonism
The objectives of the project are to provide biostatistical analysis in support of the ongoing research program at the Collaborative Center for X-linked Dystonia Parkinsonism (CCXDP) to achieve (1) Exploratory analysis and modeling to understand patterns of XDP progression, (2) Cluster analysis to distinguish XDP subtypes and classify individuals by subtype, (3) Machine learning for accurate early XDP diagnosis and prognosis, and (4) Optimized experimental design for future XDP data collection efforts. Role: PI

(INACTIVE)
No Award Number (Lange) 06/01/15 – 05/31/18 2.40 CM Alzheimer's Disease Research Foundation direct/year
Analytical and statistical tools for sequence analysis for Alzheimer's disease
The goal of this study is to develop statistical analysis tools for the Alzheimer's disease genome project. The methodology is focused on family-based study designs and complex, multivariate phenotypes.
Role: Co-Investigator

(INACTIVE)
McLennan Dean's Challenge (Miller, Mair) 6/15/2017 – 6/14/2018 0.6 CM
HSPH direct/year
Studying mechanisms of aging via closed-loop analysis and experimentation
The objective of this project is to develop statistical methods for predicting response to healthy aging interventions and to define biomarkers of physiological age, using *C. elegans* as a model organism. Role: Co-PI

OVERLAP: None