AWARD NUMBER: W81XWH-18-1-0204

TITLE: Population-Based Identification of Prostate Cancer

PRINCIPAL INVESTIGATOR: L.A. Cannon-Albright, PhD

CONTRACTING ORGANIZATION: University of Utah School of Medicine

REPORT DATE: August 2021

TYPE OF REPORT: ANNUAL

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
AUGUST 2021	ANNUAL	8/1/2020 - 7/31/2021

4. TITLE AND SUBTITLE

Population-Based Identification of Prostate Cancer

5a. CONTRACT NUMBER
W81XWH-18-1-0204

5b. GRANT NUMBER
PC170413

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

LA Cannon-Albright

E-Mail:lisa.albright@utah.edu

5d. PROJECT NUMBER
PC170413

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

UNIVERSITY OF UTAH,
THEGARY GLEDHILL201
PRESIDENTS CIR
SALT LAKE CITY UT 84112-9049

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: **Background:** Prostate cancer (PrCa) is the most common cancer diagnosed in the US and one of the most familial. There is evidence for an inherited contribution to PrCa. Analysis of PrCa pedigrees led to discovery of genes that explain a small number of pedigrees (*ELAC2, RNASEL*, and *HOXB13*); and more than 100 common genetic variants have been reported to confer modest risk to PrCa. However, taken together, these recognized genetic risk factors explain few pedigrees. The likely genetic heterogeneity of PrCa and lack of success in gene identification suggests a different approach is needed to identify responsible predisposition genes. Analysis of related cases in extended high-risk cancer pedigrees is a powerful approach for identification of cancer predisposition genes. In Utah a resource in combining the genealogy of the pioneer founders and their descendants with Utah cancer data allows identification of extended high-risk prostate cancer pedigrees. Analysis of the most clinically significant PrCa cases (those who die from their disease- lethal PrCa or LPrCa) in these pedigrees further enhances the power of this approach. **Objective/Hypothesis: We previously used the same high-risk pedigree approach proposed here to identify multiple cancer predisposition genes in the Utah population (BRCA1- Miki, 1994; BRCA2- Wooster, 1994; CDKN2A- Kamb, 1994). We propose that whole genome sequencing (WGS) of pairs of related LPrCa cases from high-risk PrCa pedigrees and identification of the rare variants shared in these pairs will similarly lead to identification of rare variants in genes that underlie predisposition to PrCa. Extended high-risk pedigrees are likely to evidence a strong role for genetic factors, and the LPrCa case pairs to be sequenced are selected for clinical significance (death from PrCa) to ensure that they exhibit limited genetic heterogeneity. This study relies on unique Utah resources and a powerful high-risk pedigree approach; it is entirely complementary to previous efforts to identify prostate cancer predisposition genes in pedigrees, and it is not possible elsewhere.**

15.SUBJECT TERMS

NONE LISTED

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	9	

TABLE OF CONTENTS

Page

1. INTRODUCTION

Prostate cancer (**PrCa**) is the most common cancer diagnosed in the US and one of the most familial. There is evidence for an inherited contribution to PrCa. Men with a first-degree relative with PrCa have a 2-3 fold risk, and those with ≥1 first-degree relative ≤ 50 years have ~4-fold risk (Albright et al., 2015). These risks are even higher for family history of death from PrCa (Albright et al., 2016). Analysis of PrCa pedigrees had led to discovery of a few genes that explain a small number of pedigrees (*ELAC2*, *RNASEL*, and *HOXB13*); and more than 100 common genetic variants have been reported to confer modest risk to PrCa. However, taken together, these recognized genetic risk factors explain few pedigrees. The likely genetic heterogeneity of PrCa and lack of success in gene identification suggests a different approach is needed to identify responsible predisposition genes.

Although it is a commonly held belief that complex diseases are likely due to many genes with small effects, and the common approach to identification of such predisposition genes is a genome-wide association study (GWAS) with spectacularly large sample sizes, it is clear that there exist extended high-risk prostate cancer pedigrees that are extremely likely to be due to rare variants with at least moderate effect size; such rare variants are not likely to be identified in a GWAS. Analysis of multi-generation families with an excess of disease has been shown to be a powerful approach to identification of predisposition genes. Our previous studies of Utah high-risk cancer pedigrees have identified predisposition genes for breast cancer (*BRCA1*-Miki 1994; *BRCA2*-Wooster 1994) and melanoma (*CDKN2A*-Kamb 1994); similar pedigree studies in other populations have more recently identified predisposition genes for other cancers (*HOXB13*-Ewing 2012; *ATM*-Roberts 2012).

We currently have several funded cancer predisposition gene identification grants using an innovative high-risk pedigree study design based on whole exome sequencing of affected cousin pairs (CA164138 - colon cancer; CA 195614 – melanoma; CA 205796 - small intestine carcinoid cancer). **Here we propose a similar approach using the more clinically significant phenotype of lethal PrCa and analyzing unique and informative extended high-risk Utah PrCa pedigrees to identify PrCa predisposition genes.**

2. KEYWORDS

Prostate cancer
Sequencing
Lethal prostate cancer
High-risk pedigree

3. ACCOMPLISHMENTS

• What were the major goals and objectives of the project?

We previously used the same high-risk pedigree approach proposed here to identify cancer predisposition genes in the Utah population (*BRCA1*- Miki, 1994*; BRCA2*- Wooster, 1994*; CDKN2A*- Kamb, 1994). We are currently funded by NCI to use this high-risk pedigree pair approach for colon cancer, melanoma, and small intestine carcinoid cancer in Utah pedigrees. We propose that WGS of pairs of related LPrCa cases from high-risk PrCa pedigrees will lead to identification of rare variants in genes that underlie predisposition to PrCa. Extended high-risk pedigrees are likely to evidence a strong role for genetic factors and the LPrCa case pairs to be sequenced are selected for clinical significance (death from PrCa) to ensure that they exhibit enhanced genetic contribution and limited genetic heterogeneity. This study relies on unique Utah resources and a powerful high-risk pedigree approach; it is complementary to previous efforts to identify prostate cancer predisposition genes (in pedigrees and GWAS), and it is not possible elsewhere.

SPECIFIC AIMS

Aim 1. Perform WGS on 50 related LPrCa case pairs (cousins) from the most informative set of 50 Utah high-risk PrCa pedigrees. Perform bioinformatics analysis to identify the top 4,000 rare candidate variants shared by a pair of LPrCa cases in a region shared Identical by Descent (IBD).

Subtask 1: Identify all sampled LPrCa and metastatic cancer cases with genealogy data; assemble all clusters including >=2 sampled cases; identify the subset of pedigrees that are high-risk PrCa; select a cousin pair from each of the most informative 50 high-risk pedigrees
Subtask 2: Perform QC on the selected samples; perform whole genome sequencing on the 100 LPrCa cases in the 50 pairs
Subtask 3: Perform bioinformatics analysis of WGS data and PC analysis of NDAR controls using the Genetic Epidemiology pipeline to identify variants of interest; filter for rare variant allele frequency and sharing in at least 1 cousin pair to identify the 4,000 candidate variants
Subtask 4: Share candidate variants with ICPCG
Local IRB Approval

Task 1. We updated data to identify all sampled LPrCa and metastatic cancer cases with genealogy data, assembled all high-risk pedigrees with at least 2 sampled cases, and selected a cousin pair from each.

Task 2. We performed QC on the selected samples, and found replacement samples/pedigrees where necessary. The first set of 50 samples was shipped to MedGenome for whole genome sequencing and the data were received 7/30/19. The second set of 50 samples has also been sent and data received in November 2019.

Task 3. We applied, and were approved, for space on the protected servers at the Center for High Performance Computing where all analysis was accomplished. Analysis is complete and ~4,000 candidate coding and non-coding variants were identified

Aim 2. Assay 4,000 rare shared candidate variants identified in Aim 1 in 1000 additional sampled Utah PrCa cases in the 50 high-risk pedigrees to establish evidence of segregation. Candidate *genes* with significant evidence for segregation will proceed to full candidate gene analysis in Aim 3.

Subtask 1: Design and order the Illumina Infinium iSelectHD assay for up to 5,000 candidate variants and order (minimum order = n=1,052 samples)
Subtask 2: Assemble, perform QC, and plate the ~1,000 additional sampled LPrCa and PrCa cases in the selected 50 high-risk pedigrees to deliver to the core
Subtask 3: Assay the 5,000 candidate variants; identify the carriers of the candidate variants
Subtask 4: Perform RVsharing analysis for each rare variant observed in the 1,000 cases
Subtask 5: Identify the ~400 candidate genes with significant evidence for segregation of at least 1 candidate variant in at least 1 high-risk pedigree

Task 1. Assay designed and orderd
Task 2. 1,195 prostate cancer samples plated for assay
Task 3. All ~4,000 candidate variants assayed in 1,195 cases
Task 4. Thousands of variants were observed in pedigrees. For this reason we removed variants occurring too frequently among the 1,195 cases, and we removed pedigrees that did not have a significant excess of prostate cancer.
Task 5. We have identified 571 candidate variants.

Aim 3. Test the ~400 candidate genes identified in Aim 3 with a targeted sequencing panel in 1,000 Caucasian and AA LPrCa and PrCa cases and compare to 1,000 cancer-free controls, to establish association with PrCa risk.
This Aim was revised to : Validate the ~500 candidate predisposition variants that gave evidence for segregation in Aim 2 with risk association in 2 large resources (UKBiobank with > 8,000 cases and >400,000 controls and PRACTICAL prostate cancer consortium with 79,194 cases and >61,000 controls.
And new tasks:
1: Get UKBiobank approval and import data for cases and controls
2: Utilize public shared PRACTICAL imputed genotypes for validation of candidates
3: perform case/ control association with PrCa for ~500 best candidate variants identified.
4: consolidate risk association validation results to identify significant predisposition variants
5. search for significant variants in existing sequence data for Utah high risk cancer pedigrees in Dr. Cannon-Albright's biorepository to extend high-risk pedigrees

Subtask 1: Design and order the Roche SeqCap EZ Choice targeted capture kit for the 400 candidate genes
Subtask 2: Assemble, perform QC, and plate the 174 AA cases and 926 Utah PrCa cases and deliver to Core
Subtask 3: Assay the 400 candidate genes; identify all rare variants in any candidate gene in cases
Subtask 4: Analyze the WGS data for 1,000 Utah controls and the ~3,000 NDAR matched Controls and perform case/control association analysis for validation testing of all variants in

all candidate genes.

Subtask 5: Share the candidate genes with ICPCG
Subtask 6: Publish validated candidates

Task 1. Received UKBiobank approval and data
Task 3 and 4. Performed UKBiobank case/control association and identified 875 candidate variants with data in UKBiobank and found 12 with significant association with prostate cancer.
Task 5. Analysis underway to identify additional carriers in existing sequence data for other studies.
Task 6. Manuscript is in preparation presenting the 571 candidate variants and some excellent example high-risk pedigrees.

• What opportunities for training and professional development did the project provide?
No formal opportunities yet.

• How were the results disseminated to communities of interest?
No results yet, but manuscript in preparation

• What do you plan to do during the next reporting period to accomplish the goals and objectives?

Analysis is underway and we expect to proceed to submit a manuscript with identification of candidate variants in the next few months.

4. IMPACT

Nothing to report yet.

What was the impact on the development of the principal discipline(s) of the project?

No impact yet.

What was the impact on other disciplines?

The methods and tools developed have been extended to other phenotypes.

What was the impact on technology transfer?

Nothing to report.

What was the impact on society beyond science and technology?

Nothing to report yet.

5. CHANGES/PROBLEMS

Nothing to report.

Changes in approach and reasons for change

Given our finding that validation of risk association is a more powerful next step for identification of predisposition variants than investigation of all possible variants in candidate genes that may not all be associated with risk, we amended Aim 3. We did not

sequence the 400 candidate genes with segregation evidence, but rather sought validation of association with risk. This required development of new software but resulted in a strong set of candidate variants to be published and followed in future projects.

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

Changes that had a significant impact on expenditures

None

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

None

6. PRODUCTS

Nothing to report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	**Lisa A. Albright**
Project Role:	PI/PD
Researcher Identifier:	0000-0003-2602-3668
Nearest person month worked:	1.2
Contribution to Project:	Dr. Cannon-Albright has directed the course of the research, selection of pedigrees/cases, and methods of analysis, as well as supervised the day to day activities of the research team.
Funding Support:	N/A

Name:	**Craig C. Teerlink**
Project Role:	Co-Investigator
Researcher Identifier:	0000-0002-1992-2326
Nearest person month worked:	1.8
Contribution to Project:	Dr. Teerlink has been responsible for oversight of data quality control and genetic analyses (including bioinformatics analysis of sequence data) and method development/testing
Funding Support:	

Name:	**Steven Backus**
Project Role:	Computer Professional
Researcher Identifier:	N/A
Nearest person month worked:	0.8

Contribution to Project:	Mr. Backus has performed data management, data extraction, security, and data storage.
Funding Support:	N/A

Name:	**James Farnham**
Project Role:	Applied Biostatistician
Researcher Identifier:	0000-0002-8213-949X
Nearest person month worked:	0.8
Contribution to Project:	Mr. Farnham has been responsible for the generation and quality control of all data files for pedigree/case selection and tracking.
Funding Support:	N/A

Name:	**Kim Nguyen**
Project Role:	Laboratory Specialist
Researcher Identifier:	N/A
Nearest person month worked:	2.4
Contribution to Project:	Mr. Nguyen has been responsible for testing the concentration and quality of DNA, preparing stored samples, performing quality controls for all samples, and the appropriate storage and inventory and shipping of all biospecimen samples.
Funding Support:	N/A

Name:	**Jeff Stevens**
Project Role:	Laboratory Specialist
Researcher Identifier:	N/A
Nearest person month worked:	3
Contribution to Project:	Mr. Stevens is responsible for bioinformatic analysis of the sequence data.
Funding Support:	N/A

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
Some grants have ended and others have been started; overall FTE on this project was not impacted. It is now in an unfunded continuation so that we can finish analysis and our manuscript-no support has been requested for this.

What other organizations were involved as partners?

Drs. Huff and his team at MD Anderson are our collaborators on multiple funded cancer predisposition gene identification projects (breast cancer, colon cancer, melanoma, pancreas). They have exceptional DNA resources for thousands of PrCa cases with which significant findings can be pursued. They are collaborating on this project using their own MD Anderson funding, but are way behind and have not yet examined our candidates in their samples.

Dr. Diptasri Mandel at University of Louisiana is collaborating with 54 familial african american prostate cancer cases in 13 pedigrees for validation but has not been able to secure a lab to do the validation in her samples.

ICPCG - International Consortium for Prostate Cancer Genetics is no longer a funded entity but we communicated with participants who will assist with confirmation of candidates.

8. SPECIAL REPORTING REQUIREMENTS

none

9. APPENDICES

None

REFERENCES

Albright F, **Stephenson RA, Agarwal N**, & **Cannon-Albright LA.** (2016, Aug 16). Relative risks for lethal prostate cancer based on complete family history of prostate cancer death. *The Prostate*, doi: 10.1002/pros.23247. [Epub ahead of print]

Nelson Q, Agarwal N, Stephenson R, & **Cannon-Albright LA**. (2013). A population-based analysis of clustering identifies a strong genetic contribution to lethal prostate cancer. *Front Genet*, *4*, 152.