AWARD NUMBER: W81XWH-22-1-0247

TITLE: Integrated Meta-Analysis of Prostate Cancer Genomes

PRINCIPAL INVESTIGATOR: Paul C. Boutros

CONTRACTING ORGANIZATION: Regents of the University of California, Los Angeles, CA

REPORT DATE: May 2023

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
May 2023	Annual	15Apr2022-14Apr2023

4. TITLE AND SUBTITLE

Integrated Meta-Analysis of Prostate Cancer Genomes

5a. CONTRACT NUMBER
W81XWH-22-1-0247

5b. GRANT NUMBER
GRANT13434443

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Paul Boutros

E-Mail: pboutros@mednet.ucla.edu

5d. PROJECT NUMBER
PC210269

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

The Regents of the University
of California, Los Angeles

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
None

14. ABSTRACT

Hypothesis. We hypothesize that an integrated database of prostate cancer genomic profiles will improve our understanding of the molecular and evolutionary determinants of lethal disease.
Study Design: Aim #1. There are several dozen prostate cancer genome sequencing studies, comprising almost 10,000 patient specimens. These vary in the DNA sequencing technology and in the portion of the genome analyzed. Most have targeted sequencing but by volume most data is from whole-genome sequencing. We will systematically re-analyze all data through consistent state-of-the-art computational pipelines and aggregate the results into a single community-available database for easy retrieval and analysis for downstream studies.
Study Design: Aim #2. We have DNA whole-genome sequencing data from patients across all prostate cancer disease states. We will apply cutting-edge subclonal reconstruction techniques to identify the specific evolutionary features of each state and use evolutionary timing analysis to pinpoint when specific mutations and mutational processes occur during tumor evolution.
Study Design: Aim #3. Using our aggregated meta-database, we will rigorously identify somatic mutational driver events that occur more frequently than expected by chance. We will search for these at the level of copy number variants, coding and non-coding point variants, genomic rearrangements, and mitochondrial variants. We will then create multivariate models that relate these drivers to clinical features to characterize their timing, and to transcriptional data to understand their phenotypic consequences.

15. SUBJECT TERMS
None listed.

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	9	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. INTRODUCTION:

Prostate cancer is the most common non-skin cancer in men in the United States. It afflicts over 10% of men during their lives. While many prostate cancers are non-lethal, others are extremely aggressive and kill men rapidly. And even the non-lethal cancers can have large effects on a man's quality of life and have large financial costs. To try to reduce prostate cancer death and maximize patient length and quality of life, clinicians and patients make many decisions together. These include whether any treatment is necessary, whether surgery or radiotherapy is helpful, what frequency of monitoring is most appropriate, when and how drugs should be deployed, how to handle castrate-resistant disease and many others. At its heart, the management of prostate cancer is about these joint decisions that men make, supported by their care-givers and clinicians. Today, almost all of these life-changing decisions are made without using any information from the prostate cancer genome. While DNA mutations underlie the growth of prostate cancer, this information has not generally been helpful in improving care for prostate cancer patients. DNA sequencing of prostate cancer has had much less impact on patient care. There are a few important exceptions, like the wide-spread testing for *BRCA2* germline mutations in men with aggressive or metastatic disease. Nevertheless, overall DNA sequencing of prostate cancers has not had the transformative impact it has for other cancer types. **Our objective is to centralize all prostate cancer DNA sequencing data, analyze it with cutting-edge methods and make that available to prostate cancer researchers around the world.** Beyond creating a resource, we will use the re-analyzed prostate DNA sequencing data to answer two key questions. First, how do prostate cancers evolve? We will identify which mutations happen early (and thus are good drug targets) and which ones happen later, just before metastasis or resistance to therapy (and thus might be good biomarkers). Second, what is the full list of all genes that are mutated repeatedly in prostate cancer? This list will give us a set of candidate drug targets and will allow future studies to focus on only those regions of the genome mutated in prostate cancer. This will accelerate research by reducing wasted time and money expended in studying genomic areas not mutated in this cancer type.

2. KEYWORDS:

Prostate
Cancer
Whole Genome Sequencing
Bioinformatics
DNA Mutations
Cancer Evolution
Germline
Big Data
Computational Algorithms
Databases
Drug Targets
RNA Sequencing
Cloud Computing

3. ACCOMPLISHMENTS:

a. **What were the major goals of the project?**

i. Specific Aim 1: Creating and ongoing consistent meta-database of prostate cancer genomes

1. Major Task 1: Obtain IRB and HRPO approval prior to initiating human data usage - COMPLETED

2. Major Task 2: Systematically identify and incorporate all existing prostate cancer DNA-sequencing studies – IN PROGRESS

ii. Specific Aim 2: Reconstruct the evolution of prostate cancer across states

1. Major Task 1: Perform detailed subclonal reconstruction of every tumor with WGA data in our meta-dataset and identify specific features of tumor evolution that vary across difference disease states

iii. Specific Aim 3: Comprehensively identify prostate cancer driver genes

1. Major Task 1: Apply cutting-edge driver-discovery techniques to large prostate cancer datasets

b. **What was accomplished under these goals?**

i. **Overview.** Our Major Activities during this reporting period were three-fold. Our first Major Activity was to standardize and refine all computational pipelines in preparation for high-throughput data processing. Our second Major Activity was to acquire access and download additional datasets. Our third Major Activity was to begin large-scale systematic analysis of samples.

ii. **Major Activity #1: Computational Pipelines.** Our specific objectives here were to standardize the analysis and testing/quality-control of our computational pipelines prior to launching compute- and financially-expensive standardized data-processing. We created a software infrastructure for testing NextFlow-based pipelines called NFTest which is pending open-source release and submission of a manuscript for publication within the next ~two months. We also created a software infrastructure for automating pipeline input/output quality-control called PipeVal, which is also pending open-source release and submission of a manuscript for publication. Finally, we significantly enhanced our pipelines for DNA alignment, somatic copy number analysis and somatic structural variant detection to account for recent algorithmic innovations in the field.

iii. **Major Activity #2: Dataset Access.** We continue to seek and gain access to all published prostate cancer genomic datasets. In some cases, this has required iterative cycles with the data-owners, for example as they request specific wording around IRBs and/or other terms. No significant barriers have occurred, but three datasets have been delayed. First, the GAP6 dataset generation has been slowed by sample-processing delays in that consortium, and so is not yet available. Second, the African prostate cancer dataset is still in process of inter-institutional agreement. Third, the Hispanic prostate cancer dataset generation was delayed by departure of the study pathologist, and so is not yet available. These are all relatively minor and typical delays and are expected to be resolved in the next reporting period.

iv. **Major Activity #3: Systematic Analysis.** We have begun large-scale computational processing. Our data-analysis pipeline starts with fully raw data, and then processes it through to annotation and subclonal reconstruction, at which point manually intervention is required for quality-control and to ensure no parameter-refitting is required. Of course, any issues identified in manual review are then automated for future pipeline iterations. We have begun analysis with WGS datasets because they both take longer to process and provide superior opportunities to ensure robustness of the computational infrastructure at the outset.

c. **What opportunities for training and professional development has the project provided?**

i. While the project was not specifically intended to provide training, both our bioinformatician (Takafumi Yamaguchi) and biostatistician (Stefan Eng) working on this project have been involved in mentorship of PhD students and fellow staff members throughout the course of their work on this project. Additionally, all our staff consistently present their work during our weekly lab meetings, local seminars (including from our Prostate SPORE and Cancer Center) and several other venues.

d. **How were the results disseminated to communities of interest?**

i. Nothing to Report

e. **What do you plan to do during the next reporting period to accomplish the goals?**

i. **Major Activity #1: Computational Pipelines.** We anticipate this activity largely completing during the next reporting cycle, with publication and open-sourcing of novel tools, along with ongoing minor bug-fixes and software quality improvements.

ii. **Major Activity #2: Data Access.** We anticipate that by the end of the next reporting period access to all datasets will be obtained, although there may be risk that the Hispanic prostate cancer dataset will be delayed. That is an independent project, not funded here, but delays in data availability are not blockers to any downstream activity and data can be incorporated into the analysis at a late stage with minimal additional time commitment.

iii. **Major Activity #3: Systematic Analysis.** We plan to complete systematic analysis of all WGS and WXS data over the next cycle.

4. IMPACT:

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

i. **Short-term Impact.** Our work will have three major short-term impacts. First, we will create a harmonized database of prostate tumour DNA sequencing data and open-source community-standard pipelines for analysis of new data. We anticipate these will be extremely widely-used in new grants and proposals to standardize and improve analyses that typically are performed with only TCGA (n=500) or SU2C (n=100) datasets. It will also facilitate analysis of new data in a rapid, systematic way by the community. Second, our compendium of somatic mutation drivers in prostate cancer will determine if there are any as-yet-undetected actionable rare variants and provide a full list of the pathways dysregulated in prostate cancer. Third, our subclonal reconstruction of prostate tumors across the disease spectrum will identify which mutations occur at critical clinical junctures, such as immediately preceding metastasis. This will inform ongoing and new studies of molecular mechanisms underlying clinical features and sharpen biomarker studies by allowing focus on the most informative mutations.

ii. **PCRP Overarching Challenges.** This study will have impact across all stages of prostate cancer, from diagnosis through to metastatic disease. By incorporating genetically-inferred ancestry it will help understand the associations between germline features and somatic mutational profiles. It will define the evolutionary and mutational transitions that distinguish prostate cancers with a low risk of death from those with a high-risk, and between those that have experienced the evolutionary selective pressure of treatment and those that have not.

b. **What was the impact on other disciplines?**

i. **Long-term Impact: Biology.** Prostate cancer occurs through the sequential accumulation of mutations in the tumour genome. This project will provide a fulsome mapping of which mutations occur, when they occur and how they relate to one another. It will create an infrastructure to continually update that mapping with new tumour DNA sequences generated either during routine clinical care or as part of translational studies. In many ways that community resource and infrastructure will impact on our understanding of prostate cancer biology. Consider three examples. First, as new therapeutics are deployed, we will be able to systematically and rapidly compare the genomic profiles of tumors that have evaded them to all other prostate cancers. This will accelerate the identification of genomic determinants of resistance to therapy. A second example: as more prostate tumors from men of diverse ancestries are sequenced, we can begin to rigorously map the relationships between germline variation and somatic variation, better understanding the relative roles of disparities in healthcare and genetic

effects. A final example is in our understanding of the sub-histologies of prostate cancer. While prostate tumors are graded on the ISUP 1-5 scale, within each group there are sub-histologies within each pattern. For example, intraductal carcinoma as a sub-histology of Gleason pattern 4 disease and comedonecrosis and pseudorosetting of Gleason pattern 5 disease. No resource today exists to analyze the somatic features of each of these, particularly during disease evolution. Our aggregated database will provide the first step towards doing so, and as it becomes a routine part of community practice to deposit new sequences in it analyses of sub-histologies will become well-powered and may reveal unanticipated new therapeutic vulnerabilities.

ii. **Long-Term Impact: Biomarkers.** A predictive biomarker provides evidence on whether or not a therapy or intervention will provide benefit to a specific patient. For each genomic feature used in a biomarker, understanding to what it extent it varies clonally across a tumour is key. Some biomarkers occur early, and others are late-occurring features portending imminent aggressive clinical changes. This study characterizes each driver mutation in terms of its clonal features across each disease state. It has been shown in kidney and lung cancer by the TraceRx studies and in prostate cancer by our preliminary data that incorporating evolutionary information into biomarkers can dramatically improve their accuracy. Evolutionarily-aware biomarker development will be a key long-term impact, improving accuracy of patient stratification in all disease stages.

c. **What was the impact on technology transfer?**

i. Nothing to report

d. **What was the impact on society beyond science and technology?**

i. Nothing to report

5. CHANGES/PROBLEMS:

a. **Changes in approach and reasons for change**

i. Nothing to report

b. **Actual or anticipated problems or delays and actions or plans to resolve them**

i. Nothing to report

c. **Changes that had a significant impact on expenditures**

i. Nothing to report

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

i. Nothing to report

e. **Significant changes in use or care of human subjects**

i. Nothing to report

f. **Significant changes in use or care of vertebrate animals.**

i. Not applicable

g. **Significant changes in use of biohazards and/or select agents**

i. Not applicable

6. PRODUCTS:

- **Publications, conference papers, and presentations**
  - Nothing to report
- **Website(s) or other Internet site(s)**
  - Nothing to report
- **Technologies or techniques**
  - Nothing to report
- **Inventions, patent applications, and/or licenses**
  - Nothing to report
- **Other Products**
  - Nothing to report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

- **What individuals have worked on the project?**

Name:	Paul Boutros
Project Role:	PI
Researcher Identifier (e.g. ORCID ID):	PBOUTROS
Nearest person month worked:	1
Contribution to Project:	Dr. Boutros serves as the PI of this project.
Funding Support:	Dr. Boutros receives support from this project.

Name:	Takafumi Yamaguchi
Project Role:	Programmer Analyst III (Bioinformatician)
Researcher Identifier (e.g. ORCID ID):	TYAMAGUCHI
Nearest person month worked:	4
Contribution to Project:	Mr. Yamaguchi serves as a bioinformatician for this project
Funding Support:	Takafumi receives additional funding from UCLA internal grants and departmental resources.

Name:	Stefan Eng
Project Role:	Programmer Analyst III (Biostatistician)
Researcher Identifier (e.g. ORCID ID):	STEFANENG
Nearest person month worked:	4
Contribution to Project:	Stefan Eng serves as the bioinformatician for this project
Funding Support:	Stefan receives additional funding support from another DoD grant and NIH U54 under Dr. Boutros.

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**
    - Since receipt of this award, Dr. Boutros has received several other awards including: DoD W81XWH2210631, DoD W81XWH2210751, DoD W81XWH2210569, NIH U54HG012517, NIH U2CCA271894, and NIH R01CA272678. An additional grant, that was listed as pending is now funded, and is still being processed for IRB approval: NIH R01 CA270108 – "Germline Determinants of Prostate Cancer Evolution." None of these have overlap with the present grant.
- **What other organizations were involved as partners?**
    - Nothing to Report

8. SPECIAL REPORTING REQUIREMENTS

- Nothing to report

9. APPENDICES

- Nothing to report