Let's Take It Personal: Maximizing patient care in late-stage prostate cancer through AI-based approaches in precision medicine

Ralf P. Dagdag[1,2*], James C. Costello[1,2‡], Christopher R. Gignoux[1,3‡]
*Corresponding Author: ralf.dagdag@cuanschutz.edu

Department of Biomedical Informatics
SCHOOL OF MEDICINE
UNIVERSITY OF COLORADO **ANSCHUTZ MEDICAL CAMPUS**

PHARMACOLOGY AND MOLECULAR MEDICINE
PhD Program at The University of Colorado Anschutz Medical Campus

[1]Department of Biomedical Informatics; 1890 N. Revere Court, Mailstop F600, Aurora CO 80045
[2]Department of Pharmacology; 12800 E 19th Avenue, Aurora CO 80045
[3]Colorado Center for Personalized Medicine; 1890 N. Revere Court, Mailstop F563, Aurora CO 80045
‡These authors jointly and equally contributed this study

Study Overview:

Prostate cancer is the most commonly-diagnosed cancer in men and is the second leading cause of cancer-related death. Current research efforts focus on understanding patient's risk of developing prostate cancer, their risk of their cancer spreading in the body, and improving treatments for patients with early-stage prostate cancer.

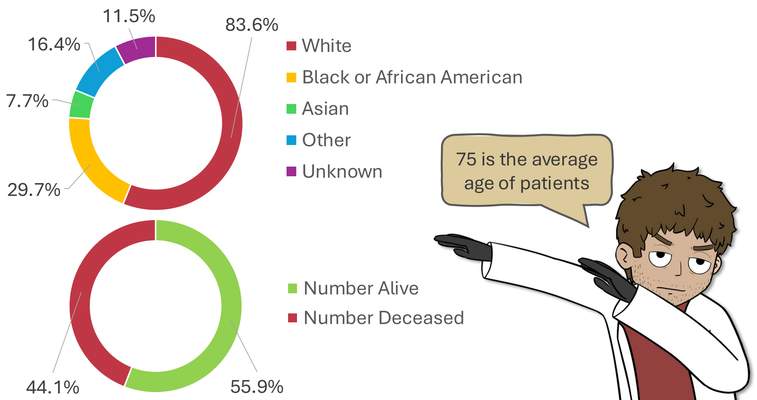
While this is helpful for patients whose cancer is caught early, information for treating patients with later diagnoses is not nearly as strong. This represents in a **critical need** in the field – a better understanding of late-stage prostate cancer at as it relates to drug response will improve patient outcomes.

To address this need, my project tests the **hypothesis that a patient's response to drugs used for treating late-stage prostate cancer can be better understood by looking at their health history, as well as their DNA.** Understanding this relationship can improve and maximize patient care, providing them better armor during their fight against cancer.

Who are the participants?

This research project uses patient health records from four different research studies:

Databank	Description	Total Participant Count	Number of Patients
TriNetX	The largest global source of patient data for use in biomedical research	140,777,937	29,950
flatiron	A cancer-centric study containing data from healthcare providers in the US, Europe, and Asia	461,298	26,154
All of Us Research Program	An initiative by the NIH to create a dataset that is as diverse as the population of the United States	849,100	814
UK biobank	A study that follows half a million participants from the United Kingdom over the course of 30 years	502,146	2,197
Totals:		142,129,183	59,115



The overall approach used in this study consists of the three main objectives:

Preprocessing: ***Define the patient population and form patient groups based on response rates***

Specific Aim 01: ***Determine what aspects of healthcare records are most predictive of response to treatment***

Specific Aim 02: ***Detect aspects of patient DNA that can be useful for precision medicine***

What is the main concept?

Patient data can be messy, so it's important to **preprocess** the data to ensure the rest of the process goes smoothly.

Think of this like baking: trying to grab every ingredient as you go can make things hard, especially if your kitchen is disorganized…

But if you preprocess your ingredients and get them ready to go before you start, the cooking process will be much easier!

Artificial intelligence (AI) uses data from pre-existing information to predict a certain outcome.

Pretend you just had THE best brownie ever – you have a full pantry to recreate it, but don't know which ingredients are important.

AI can "learn" from every brownie recipe you have to predict which ingredients are important for a "good" vs "bad" brownie.

One aspect of **precision medicine** studies how a patient's DNA can influence how well a treatment works for them.

Let's say that the best brownie in the world has a secret ingredient – everyone loves them, but some people have an allergic reaction.

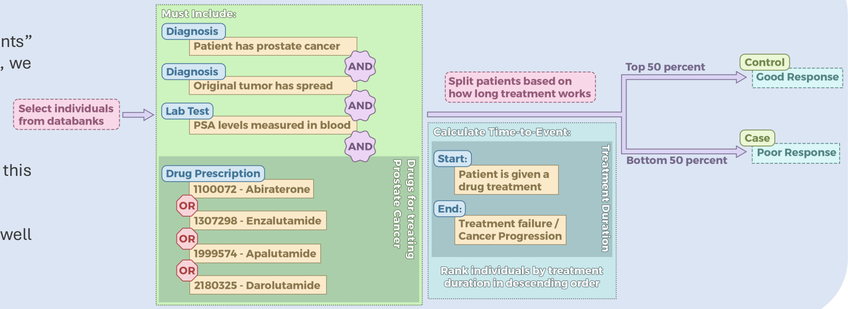
Precision medicine can tell us *why* some people are allergic, which is important in allowing everyone to enjoy the World's Best Brownie™

How is the work done?

Goal:

To ensure we have all the "ingredients" needed for our project set in place, we must:

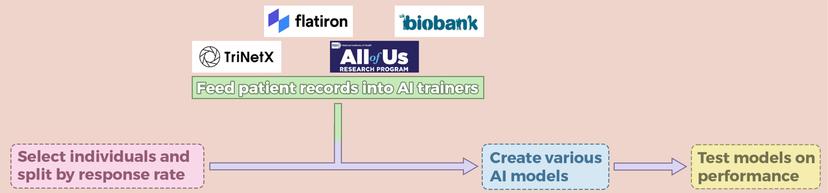
- Define the patient population
- Select patients who meet this definition to form cohorts
- Split the cohorts based on how well they respond to therapy



Goal:

To predict the "recipe" for a good or poor responder and the "ingredients" needed for each, we want to:

- Build AI models to predict whether a patient will have a good or poor response
- Test how well the models perform
- Identify which data are important for deciding response rate



Goal:

To determine how DNA can explain why some patients respond poorly to a treatment "brownie", we will:

- Analyze the DNA of patients who had a poor response to treatment
- Find differences in patient DNA that are consistent with poor response
- Predict how these differences affect and influence the body

Acknowledgments:

This research would not be possible without the partnership and contributions of the participants within UC Health and the Colorado Center for Personalized Medicine, the TriNetX Research Network, the NIH All of Us Research Program, UK Biobank, and Flatiron Health. Additionally, I would like to thank the following individuals for their contributions: Laura S. Graham, MD for providing guidance on framing the clinical scope and relevancy of this project; Michael Orman, Lily Elizabeth Feldman, and Cailin Deiter for their guidance and feedback regarding cancer biology; Lucas Gillenwater and Sutanu Nandi, PhD for their guidance on computational methods; Maizy Brasher for her guidance on UK Biobank and HIPAA and GDPR compliance; Sean Davis, MD, PhD for his guidance on Flatiron; Luke M. Evans, PhD and Chris Arehart for their guidance on drug repurposing.

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award number R01CA276826. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the NIH.
Additionally, this research has been conducted using the UK Biobank Resource under Application Number 100578.