Investigating the genomic profile of inherited prostate cancer using whole exome sequencing data.

Alexandre Sá Ferreira
PG55691
Universidade do Minho

2025

Abstract

A short (ca. 100 word) summary of the dataset being described: what the data covers, how it was collected, how it is stored, and a short description of its reuse potential.

Keywords: hereditary prostate cancer; single nucleotide variations; germline variants, max 6

1 Introduction

Repository location https://github.com/Alexsf35/Projeto_Bioinf.git

Context During their lifetime, men have a 40.1% chance of being diagnosed with any invasive cancer (slightly higher than women, 38.7%). Among these diagnoses, approximately 14.2% are prostate cancers (PrCa). As of 2022, prostate cancer has been determined to be the second most prevalent cancer worldwide in terms of incidence and the fifth most prevalent in terms of mortality among the male population. It has been identified as the most common cancer among men in 118 different countries, including Portugal.
A notable factor in PrCa is its hereditary component. The proportion of PCa attributable to hereditary factors has been estimated to be 5 to 15%, with other estimates stating that up to 20% of men diagnosed with PrCa have a family history of the disease within their paternal or fraternal lineage. In the case of any affected family member, the relative risk is doubled, with an additional increase observed in relation to the number of relatives and their age (under 60 years of age) at the time of diagnosis. In addition, ethnicity has been shown to exert a significant influence on the outcomes of prostate cancer cases, with African ancestry being recognized as a well-established risk factor.
The etiology of PrCa remains to be fully explained; however, mounting evidence suggests that the aforementioned factors are significant indicators of genetic susceptibility to PrCa. This susceptibility is linked to the inheritance of a combination of rare germline variants (defined as having a minor allele frequency [MAF] $< 1\%$) in moderate to high-penetrance genes and common genetic alterations in low-risk genes.
For this reason, PrCa is characterized as a heterogeneous cancer, hence a polygenic/genome-wide approach is recommended. In this context, single nucleotide variations (SNVs) that lead to truncating, missense, or nonsense changes are of particular interest. While few such mutations may have low impact, the sum total of these variations may be problematic.
Whilst the majority of studies on the genetic predisposition to prostate cancer have focused on common variants (MAF ¿ 5%) and very rare, high-risk variants (MAF ¡ 0.1%), the present project aims to bridge this gap by investigating the mutational landscape of rare (0.1% ¡ MAF ¡ 1%) and low-frequency (1% ¡ MAF ¡ 5%) single nucleotide variants. The present study will analyse whole exome sequencing (WES) data from 96 prostate cancer patients across 45 families, with the aim of identifying genes and pathways that exhibit a higher burden of these potentially pathogenic mutations.

2 Epidemiology

PrCa is a major health issue world wide with an average age of diagnosis being 66 years old. The age-standardized incidence rate (ASR) in 2020 was of 31 per 100,000 males, with a lifetime risk of 3.9%. Still, these incidences vary tremendously by region.

For instance, incidences of as high as 83 per 100,000 are seen in Northern Europe, followed by Western Europe (78), the Caribbean (76), and Australia/New Zealand. Conversely, incidences as low as 6.3 per 100,000 have been observed in countries such as South-Central Asia, 14 for South-Eastern Asia, and 17 for Northern Africa. These variations are the reflection of the differences in screening efforts, access to healthcare, and public awareness about the disease. As for mortality, in 2020 the global ASR was approximately 7.7 per 100,000 males. The less developed socioeconomic areas have much higher mortality rates, with the Caribbean at 28 per 100,000, Middle Africa at 25, and Southern Africa at 22. In contrast, Asian regions have significantly lower mortality rate, South-Central Asia at 3.1, Eastern Asia at 4.6, and South-Eastern Asia at 5.4 per 100,000.

In the future, projections are not too bright. It is estimated that new PrCa cases will nearly double, from 1.4 million in 2020 to approximately 2.9 million in 2040. Similarly, PrCa deaths are also estimated to rise by 85%, from 375,000 to nearly 700,000 in the same period. This expected rise is partly due to world population aging and the varied effectiveness of screening and treatment initiatives around the world.

PrCa incidence has been rising since the 1980s, largely as a result of widespread of prostate-specific antigen (PSA) testing and once again an increase in disease awareness. Which in turn results in early diagnosis and treatment, having reflected a decline in mortality rates, that have decreased by 52% from their peak in 1993, with recent patterns showing a moderate decline through 2020.

Lifestyle factors are also a significant in PrCa diagnosis. A meta-analysis in 2016 found that higher physical activity is correlated with a 38% reduction in PrCa-specific mortality, which suggests that an active lifestyle may be important in managing disease progression. Other factors such as dietary and smoking habits, also have demonstrated to have affects in both incidence and mortality.

3 Genetic Etiology

To be diagnosed with hereditary prostate cancer (HPC) families must meet the Johns Hopkins criteria, by either having three or more first-degree relatives diagnosed with PrCa; cases spanning three successive generations; or having at least two relatives diagnosed with early-onset PrCa (before the age of 56 years).

Notably, any affected family member doubles PrCa risk and is increased by 2.5-fold if one first-degree relative is less than 60 years old at diagnosis (1.6 times if the first-degree relative is older than 60 years). If 2 or more relatives are diagnosed before 60 years old, the relative risk can escalate up to 5.7. These elevated risks may be attributed to the existence of rare variants that are present in less than 1% of the population, which get pass down hereditarily, and affect moderate- to high-penetrance genes. These genetic factors are frequently identified in families that also exhibit other hereditary cancer syndromes, which are among the most important risk factors compared to age, race, ethnicity and environmental factors for the development of PCa and this risk is estimated at 40% to 50 %. They are:

HBOC, that is most commonly associated with inherited pathogenic mutations in the BRCA1 and BRCA2 genes, though other genes (such as PALB2, ATM, and CHEK2) can also play a role. PALB2, for instance, links BRCA1 and BRCA2 forming the "BRCA complex" that repairs DNA double-strand breaks via HR. Furthermore, it has been established that they also exert control over centrosome dynamics, chromosome segregation, and cytokinesis, thereby contributing to the temporally and spatially stabilised genome within the cell cycle. The other genes (ATM and CHEK2) are also related to the correction of damaged genome; *CHEK2* encodes a checkpoint kinase that interacts with cell cycle regulators and DNA repair proteins, like the aforementioned ATM serine threonine kinase that recognises double stranded DNA breaks and initiates multiple aspects of the damage response cascade.

LS, that is caused by mutations in DNA mismatch repair (MMR) genes, most commonly MLH1, MSH2, MSH6, and PMS2 (and sometimes the epithelial cell adhesion molecule [EPCAM] through epigenetic silencing of MSH2). Individuals with LS have a higher lifetime risk of developing colorectal and endometrial cancer and a range of other cancers such as ovarian, gastric and urinary tract.

Distinct from the broader hereditary cancer syndromes, HOXB13 is the only gene being specifically and consistently associated with an increased risk of HPC, especially the G84E mutation. Even though the mutations can be found in both affected and healthy man, the carrier rate was found to be considerably elevated among affected males (194 out of 382; 51% in some studies) in comparison to their counterparts within these families. The protein expressed by HOXB13 plays a crucial role in the embryonic development of the prostate gland and continues to be expressed in normal prostate tissue throughout adulthood. Moreover, it has been demonstrated to interact with the androgen receptor, thereby impacting the proliferation and differentiation of both normal and cancerous prostate cells, which underlines its significance as a prostate cancer susceptibility gene.

4 Methods