Genetic landscape of homologous recombination repair genes in early-onset/familial prostate cancer patients

Manuel Teixeira (manuel.teixeira@ipoporto.min-saude.pt)  
Portuguese Oncology Institute of Porto (IPO Porto)  
https://orcid.org/0000-0002-4896-5982

Paula Paulo  
IPO Porto Research Center (CI-IPOP)

Marta Cardoso  
IPO Porto Research Center (CI-IPOP)

Andreia Brandão  
IPO Porto Research Center (CI-IPOP)

Pedro Pinto  
IPO Porto Research Center (CI-IPOP)

Ariane Falconi  
IPO Porto Research Center (CI-IPOP)

Manuela Pinheiro

Nuno Cerveira  
IPO Porto Research Center (CI-IPOP)

Rui Santos

Catarina Santos

Carla Pinto

Ana Peixoto

Sofia Maia

Article

Keywords: Hereditary Prostate Cancer, Next-generation sequencing, Homologous recombination repair genes, germline variants, exonic rearrangements

Posted Date: January 12th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2403119/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Prostate cancer (PrCa) is among the three top most frequent and deadlier cancers worldwide. The discovery of PARP inhibitors for the treatment of tumors having deleterious variants in homologous recombination repair (HRR) genes has placed PrCa in the roadmap of precision medicine. Still, the overall contribution of HRR genes for the 10-20% of the carcinomas arising in men with early-onset/familial PrCa has not been fully clarified. We used Targeted Next Generation Sequencing (T-NGS) covering eight HRR genes (ATM, BRCA1, BRCA2, BRIP1, CHEK2, NBN, PALB2 and RAD51C) and an analysis pipeline querying both small and large genomic variations, to clarify both their global and relative contribution for hereditary PrCa predisposition in a series of 462 early-onset/familial PrCa cases. Deleterious variants were found in 3.9% of the patients, with CHEK2 and ATM being the most frequently mutated genes (38.9% and 22.2% of the carriers, respectively), followed by PALB2 and NBN (11.1% of the carriers, each), and then by BRCA2, RAD51C, and BRIP1 (5.6% of the carriers each). Using the same NGS data, exonic rearrangements were found in two patients, one pathogenic in BRCA2 and one of unknown significance in BRCA1. Additionally, 5.4% of the patients were carriers of variants of unknown significance (VUS). These results support the utility of T-NGS to clarify the genetic heterogeneity that underlies PrCa predisposition, allowing to detect both small and large genomic variations, and unveil CHEK2 and ATM as the major HRR genes associated with early-onset and familial PrCa, respectively.

Introduction

Prostate cancer (PrCa) has long been one of the most frequent and deadlier cancers in men worldwide, with the 2018 estimates pointing to 1.3 million diagnoses and 359 thousand deaths (1). The combination of efforts from several research groups with the most advanced technologies in genome screening have started to unveil the genetic component underlying predisposition to the 10–20% of the PrCa cases occurring at early age and/or in families with aggregation of the disease (2, 3). Lessons from classical hereditary cancer syndromes, such as Hereditary Breast and Ovarian Cancer (HBOC) and Lynch Syndrome (LS), have been fundamental in the identification of germline determinants of PrCa risk. In fact, apart from HOXB13, with the specific G84E mutation defined as a moderately penetrant PrCa risk variant (4), the other well-established genes predisposing to PrCa are those associated with HBOC or LS, namely the homologous recombination repair (HRR) genes BRCA1 and BRCA2 (5–7) and the mismatch repair (MMR) genes MSH2 and MSH6 (8, 9). Screening of variants in the breast cancer predisposing genes ATM, CHEK2, and PALB2 in PrCa patient cohorts have established them as additional moderate penetrance PrCa predisposing genes (10, 11). The involvement in PrCa development of other moderate-risk breast/ovarian cancer predisposing genes, such as BRIP1, NBN and RAD51C, has also been described, although reports are scarcer (12–14).

These observations, along with their collective role in DNA repair and the advent of next generation sequencing (NGS), soon led to the integration of HRR genes in multigene panel testing, fastening our understanding of their involvement in genetic PrCa predisposition and their clinical utility for the genetically informed management of the patients (14–16). In fact, multiple NGS-based screening studies are in agreement for an association between the carrier status for a germline mutation in a HRR gene and higher Gleason score, advanced staging, metastization, and worst outcome (14, 17–19). Additionally, the identification of carriers of HRR deleterious variants may be useful to delineate therapeutic options, considering the positive response to PARP inhibitors observed in patients with metastatic, castration-resistant prostate carcinomas with compromised HRR (20, 21).
The evolution of NGS data analysis methods has also empowered the detection of genomic Structural Variations (SVs). Albeit rare, germline Copy Number Variations (CNVs) leading to loss/gain of exonic regions have been described to explain a fraction of the inherited risk to cancer development, namely in the HBOC and LS predisposing genes (22–24). Apart from the CHEK2 large deletion identified in the Polish population (25) and the Portuguese founder variant c.156_157insAlu in BRCA2 (26), the few examples describing SVs associated with PrCa development derive from genome-wide association studies (GWAS), with the coding genes involved being mostly unknown (27, 28).

An earlier former study of our group has revealed that the most frequent BRCA1 and BRCA2 variants found in Portuguese HBOC families could only explain a fraction of the missing PrCa heritability, suggesting that other variants in the BRCA genes and/or other genes may play a bigger role in PrCa predisposition (26). Using Targeted Next Generation Sequencing (T-NGS) in a pilot study including 121 patients with strong criteria for hereditary disease, not only we unveiled new candidate genes associated with PrCa development, but also confirmed the low frequency of deleterious variants in the BRCA carriers (29). Likewise, MSH2, one of the first genes associated with PrCa predisposition, was not found frequently found mutated in both studies (26, 29).

In this work, we aimed to characterize 462 PrCa patients fulfilling criteria for early-onset and/or familial/hereditary PrCa for the large spectrum of germline variants (both SNVs/INDELs and SVs) present in five well established PrCa predisposing and HRR-related genes (ATM, BRCA1, BRCA2, CHEK2 and PALB2) and in three additional HRR genes proposed to be involved in PrCa predisposition (BRIP1, NBN and RAD51C), in order to clarify the overall and relative contribution of these HRR genes for the genetic predisposition to PrCa, and to validate an analysis pipeline of T-NGS data for the molecular diagnosis of hereditary disease.

Materials And Methods

Biological samples

A total of 462 patients, previously described (26), fulfilling criteria for early-onset and/or familial/hereditary PrCa were enrolled in this study, specifically, PrCa diagnosis before the age of 56 – early-onset criterion – and/or PrCa diagnosis at any age with family history of the disease and at least one family member (the proband or a relative) with PrCa diagnosed before the age of 66 – family history criterion. Among them, 240 patients fulfill the early-onset criterion, 311 patients fulfill the family history criterion, and 89 fulfill both criteria. Clinicopathological characteristics are summarized in Supplementary Table 1.

A series of 701 control samples of healthy individuals from the Northern Portuguese population, also previously described (29), was used (Supplementary Information).

Targeted next-generation sequencing

A custom primer panel was designed using SureDesign (v5.0.1; Agilent Technologies, Santa Clara, CA, USA). The panel covers, among other genes, the coding and splicing regions of eight HRR genes, namely, ATM (NM_00051.3), CHEK2 (NM_007194.3), BRCA1 (NM_007294.3), BRCA2 (NM_000059.3), BRIP1 (NM_032043.2), NBN (NM_002485.5), PALB2 (NM_024675.4), and RAD51C (NM_058216.2), which are the focus of the present study. For library preparation, ~ 50ng of DNA were enriched for the custom primer regions using the
SureSelect\textsuperscript{QXT} protocol (Agilent Technologies), following manufacturer’s recommendations. Enriched libraries were pooled and sequenced by Health\[in\]Code (La Coruña, GA, Spain) in an Illumina HiSeq platform (San Diego, Californía, EUA).

**Variant calling and classification**

For sequence alignment to the reference genome (GRCh37/hg19) and variant calling, the NextGENe software (v2.4.2.2; Softgenetics, State College, PA, USA) was used. Analysis and filtering of single nucleotide variants (SNVs) and INDELs, as well as the software used in analysis of structural variants (SVs) are detailed in Supplementary Information.

**Validation of NGS findings**

Validation of P/LP SNVs and INDELs was performed by Sanger sequencing (primer sequences in Supplementary Table 2, if not previously described (26, 29)). For validation of SVs, we used both Multiplex ligation-dependent probe amplification (MLPA; MRC Holland, Amsterdam, the Netherlands) and array-based Comparative Genomic Hybridization (aCGH) with the SNP microarray CytoScan XON (Thermo Fisher Scientific, Inc.; Waltham, MA, USA). Additional details are described in Supplementary Information.

**Transcript analysis**

To explore the transcriptional consequences of the SV identified in \textit{BRCA1} in patient HPC460, the pattern of \textit{BRCA1} transcripts was analyzed in RNA extracted from patient’s peripheral blood lymphocytes (PBLs) by qRT-PCR, followed by Sanger sequencing (Supplementary Information).

To predict translation initiation sites (TIS) of aberrant transcripts, two online tools, namely, NetStart, (v1.0, https://services.healthtech.dtu.dk/service.php?NetStart-1.0 (30)), and TIS Miner (http://dnafsminer.bic.nus.edu.sg/Tis.html (31)) were used, whenever appropriate.

**Genotyping in Portuguese control subjects**

Variants found to be recurrent were questioned for their prevalence in 701 control samples using KASP (Kompetitive Allele Specific PCR) technology genotyping (LGC, Teddington, UK), if not previously evaluated (29). KASP assay primers (Supplementary Table 2) were designed using the Primer-BLAST design tool from NCBI and acquired from Metabion (Martinsried, Germany). Data were analyzed in the LightCycler 480 Software 1.5.0. (Roche Diagnostics, Basel, Switzerland). The numbers of carriers and non-carriers were compared between patients and control samples using the Fisher’s exact test (two-tailed) in GraphaPad (https://www.graphpad.com/quickcalcs/contingency1/). Differences were considered statistically significant with a $P$ value < 0.05.

**Statistical associations with clinicopathological factors**

To explore statistical associations between the germline mutational status for the eight genes and clinicopathological data of the prostate carcinomas, information regarding PSA at diagnosis, Gleason Score, and TNM staging was gathered. Additionally, patients were assigned into risk groups, according to the NCCN guidelines (v2.2020), and interrogated for associations with germline mutational status. Early-onset disease and family history criteria were also queried. The IBM Statistical Package for the Social Sciences (v.25) was
used to infer statistical significance. An association was considered statistically significant if reaching a $P$ value < 0.05 by two-tailed Fisher's exact test (GraphaPad).

Results

Frequency of carriers of germline deleterious SNVs/INDELs in the eight HRR genes

Using the analysis pipeline for SNVs/INDELs, we found 13 carriers of pathogenic/likely pathogenic (P/LP) variants in well-established PrCa-predisposing genes: seven in CHEK2, four in ATM, and two in PALB2 (Table 1). There were no carriers of P/LP SNVs/INDEL variants in either BRCA1 or BRCA2.
Table 1
Germline P/LP variants found in the eight HRR genes under study.

<table>
<thead>
<tr>
<th>Gene</th>
<th>cDNA change</th>
<th>Protein change</th>
<th>dbSNP ID</th>
<th>ClinVar/ACMG-AMP§</th>
<th>Patient ID</th>
<th>Patient criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Well-established PrCa predisposing HRR genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATM</td>
<td>c.652C&gt;T</td>
<td>p.(Gln218Ter)</td>
<td>rs1555066551</td>
<td>P/LP</td>
<td>*HPC177a</td>
<td>FH</td>
</tr>
<tr>
<td>ATM</td>
<td>c.1236-2A&gt;G</td>
<td>p.(?)</td>
<td>rs80159221</td>
<td>P/LP</td>
<td>HPC408</td>
<td>FH</td>
</tr>
<tr>
<td>ATM</td>
<td>c.8264_8268del</td>
<td>p.(Tyr2755CysfsTer12)</td>
<td>rs730881294</td>
<td>P§</td>
<td>HPC234</td>
<td>FH</td>
</tr>
<tr>
<td>ATM</td>
<td>c.9079dup</td>
<td>p.(Ser3027LysfsTer36)</td>
<td>rs587780645</td>
<td>P/LP</td>
<td>HPC196</td>
<td>FH</td>
</tr>
<tr>
<td>BRCA2</td>
<td>c.156_157insAlu</td>
<td>p.(?)</td>
<td>N.A.</td>
<td>P</td>
<td>*HPC398b</td>
<td>FH</td>
</tr>
<tr>
<td>CHEK2</td>
<td>c.593-1G&gt;T</td>
<td>p.(?)</td>
<td>rs786203229</td>
<td>LP</td>
<td>*HPC353</td>
<td>Age, FH</td>
</tr>
<tr>
<td></td>
<td>c.349A&gt;G</td>
<td>p.(Arg117Gly)</td>
<td>rs28909982</td>
<td>LP#</td>
<td>*HPC50c</td>
<td>Age, FH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*HPC188ac</td>
<td>Age, FH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*HPC274c</td>
<td>Age, FH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*HPC289ac</td>
<td>Age, FH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*HPC290c</td>
<td>Age, FH</td>
</tr>
<tr>
<td>PALB2</td>
<td>c.1438A&gt;T</td>
<td>p.(Lys480Ter)</td>
<td>rs1057520653</td>
<td>P</td>
<td>HPC36</td>
<td>Age</td>
</tr>
<tr>
<td>PALB2</td>
<td>c.2257C&gt;T</td>
<td>p.(Arg753Ter)</td>
<td>rs180177110</td>
<td>p#</td>
<td>HPC223</td>
<td>FH</td>
</tr>
<tr>
<td><strong>Suggested PrCa predisposing HRR genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRIP1</td>
<td>c.2947dup</td>
<td>p.(Ile983AsnfsTer19)</td>
<td>rs774684620</td>
<td>P/LP</td>
<td>HPC427</td>
<td>Age</td>
</tr>
<tr>
<td>NBN</td>
<td>c.156_157del</td>
<td>p.(Ser53CysfsTer9)</td>
<td>rs767454740</td>
<td>P/LP</td>
<td>HPC205</td>
<td>FH</td>
</tr>
<tr>
<td>NBN</td>
<td>c.1914+2_1914+3del</td>
<td>p.(?)</td>
<td>rs1586043528</td>
<td>LP§</td>
<td>HPC95</td>
<td>Age</td>
</tr>
<tr>
<td>RAD51C</td>
<td>c.890_899del</td>
<td>p.(Leu297HisfsTer2)</td>
<td>rs1555602141</td>
<td>P</td>
<td>*HPC186a</td>
<td>Age+ FH</td>
</tr>
</tbody>
</table>

§ American College of Medical Genetics and Genomics and the Association for Molecular Pathology; P, pathogenic; LP, likely pathogenic; N.D., not described; # automatically classified by InverVar; * previously described: a (29), b (26) and c (32); & classification in the context of recessive disease; N.A., not applicable; FH, family history; Age, early-onset diagnosis (<56 years).

The ATM variant found in patient HPC177, as well as the family tree, was previously described (29). The additional carriers of deleterious variants in ATM include two frameshift variants, c.8264_8268del and...
c.9079dup, in patients HPC234 and HPC196, respectively, and a splicing variant (c.1236-2A > G) in patient HPC408. Segregation with the disease was observed in the affected brother of patient HPC234, who agreed to participate in the patients’ recruitment phase (2014–2015; Fig. 1A), now deceased. Although both carriers of the ATM variants c.9079dup and c.1236-2A > G have relatives diagnosed with PrCa (Figs. 1B and 1C) segregation analysis was not possible to be performed.

The five patients carrying the CHEK2 missense variant c.349A > G and the CHEK2 splicing variant found in patient HPC395, have been previously described (29, 32). A second carrier of the same splicing variant (c.593-1G > T) was identified in an early-onset PrCa case (HPC353) without family history of cancer (Fig. 1D).

The two nonsense variants in PALB2 were identified in patients with distinct familial cancer histories, with the variant c.1438A > T being identified in an early-onset PrCa patient (HPC36) with two first degree relatives diagnosed with different carcinomas (Fig. 1E), and the variant c.2257C > T occurring in a patient (HPC223) with history of PrCa in the brother, lung cancer in both parents and two uncharacterized cancers, in brain and leg, in two paternal uncles (Fig. 1F). Segregation of the variant in the affected brother was not possible to be performed.

Among the three HRR genes that have been suggested to predispose to PrCa development, four patients carrying P/LP variants were identified: two in NBN, one in BRIP1, and one in RAD51C, the last having been identified in a former study (29).

The BRIP1 frameshift variant c.2947dup was identified in a patient (HPC427) diagnosed with PrCa at the age of 48 years, and a cancer in the tongue 8 years after PrCa diagnosis. In the family history, two paternal aunts were diagnosed with breast cancer (Fig. 2A). The NBN splicing variant c.1914 + 2_1914 + 3del was also identified in a patient with early-age PrCa diagnosis (HPC95), with a breast carcinoma diagnosed in his maternal grandmother at the age of 48 years (Fig. 2B). The NBN frameshift variant c.156_157del occurs in a patient (HPC205) diagnosed with PrCa at the age of 57 years, with a maternal uncle diagnosed with PrCa at 77 years old (Fig. 2C).

Altogether, deleterious SNVs/INDELs in the five well-established PrCa predisposing genes ATM, BRCA1, BRCA2, CHEK2, and PALB2 represent 2.8% of the 462 patients under study, 2.5% of the 240 patients with early-onset disease (< 56 years) and 2.9% of the 311 patients with at least one relative diagnosed with PrCa. The three genes BRIP1, NBN and RAD51C may explain PrCa development in additional ~ 0.9% of the cases, representing ~ 1.2% of the prostate carcinomas diagnosed at early age and 0.6% of those occurring in patients with family history of PrCa (Fig. 3).

**Frequency of germline structural variants in HRR genes**

Using the TL modules of SureCall and DRAGEN, a single output occurring in the BRCA2 gene in sample HPC398 was obtained with SureCall only. The aligned reads revealed the Portuguese founder variant c.156_157insAlu (Fig. 4A), previously identified (26, 33), and included in Table 1. The three different CNV software outputted two variants with high quality metrics (Supplementary Table 3): a duplication of the exon 2 of BRCA1 in patient HPC498 and a duplication of the exon 5 of RAD51C in patient HPC163 (Fig. 4B).
Using MLPA and aCGH, we validated the duplication in BRCA1 in patient HPC498, encompassing at least 2,718bp of the genomic region containing part of BRCA1 and NBR2 genes (chr17: 43,123,235 – 43,125,953) (Figs. 4C and 4D). Transcriptomic analysis in PBLs from patient HPC498 revealed an undescribed aberrant transcript retaining 299bp of the intron 1–2 of BRCA1 NM_007294.4 (r.-19_-20ins-20 + 236_-20 + 534) (Figs. 4E and 4F). Available online tools to predict the translation initiation site (TIS) of the aberrant BRCA1 transcript highlight the probability of new TISs in the 299bp retained, however, any of the predicted alternative ATGs would lead to the occurrence of a stop codon before the wild-type TIS (Supplementary Table 4), and, thus, the translation of a wild-type BRCA1 protein may not be affected. Segregation analysis in the affected maternal uncle was not possible to be performed (Supplementary Fig. 1), so the classification of this CNV was retained as a variant of unknown significance (VUS), as reported by others (34).

Thus, altogether, structural variants in HRR genes were found in 0.4% of the early-onset/familial PrCa cases, representing 0.3% of the P/LP variants identified in patients with family history of PrCa (Fig. 3).

**Frequency of VUS identified by the SNVs/INDELs NGS pipeline**

Using the filtering criteria described above, 25 patients (5.4%) were found to be carriers of VUS in five of the eight genes under study (Supplementary Table 5). Three of the four recurrent variants, specifically, the CHEK2 variant c.695G > T and the ATM variants c.1595G > A and c.8560C > T, were previously described in a subset of 121 PrCa patients of the same patient series with strong criteria for hereditary predisposition, and discarded as common variants in the Portuguese population (29).

When comparing the overall frequency of each VUS in the whole series of 462 early-onset/familial PrCa with that found in 701 non-cancer Portuguese controls, none is significantly more frequent in Portuguese PrCa patients (P > 0.05; Supplementary Table 6), but both the CHEK2 variant c.695G > T and the ATM variant c.7375C > G reach borderline statistical significance (P = 0.061 and P = 0.065, respectively). Comparing with the frequency described in non-Finish Europeans (NFEs) in gnomAD database (https://gnomad.broadinstitute.org), all recurrent ATM and CHEK2 variants are significantly more frequent in our early-onset/familial PrCa patients (P < 0.001; Supplementary Table 6).

The ATM variant c.7375C > G was found to segregate with the disease in one of the affected brothers of patient HPC460, assigned in the projects’ recruitment phase (Supplementary Fig. 2).

**Associations between the carrier status and clinicopathological data**

When comparing the 18 carriers of P/LP variants (Table 1) with the 420 patients without any HRR variant in the eight genes under study, no statistical significant associations were observed between the carrier status and the clinicopathological characteristics of the prostate cancer patients, namely considering age at diagnosis (< 56y vs. ≥56y), PSA at diagnosis (< 10 vs. ≥10), tumor staging (≤ T2 vs. ≥T3), Gleason grading (GS≤(3 + 4) vs. GS≥(4 + 3)), cancer stage (stage I/II vs. stage III/IV), or NCCN risk group (low/intermediate vs. high) (Supplementary Table 7).

**Discussion**
To our knowledge, this is the first report of full screening of germline variants in a panel of HRR genes focusing in a large series of early-onset/familial PrCa cases. Although the BRCA genes are the best well-established risk genes for PrCa development, and the most frequently mutated genes in the germline of both metastatic patients (18) and patients unselected for family history (35), studies using multigene panel testing to compare the frequencies of deleterious variants in BRCA genes and in other cancer predisposing HRR genes in PrCa patient cohorts with criteria for hereditary disease are scarce.

Thus, we aimed to clarify the overall and relative frequencies of deleterious germline variants in eight HRR genes in our entire series of 462 early-onset/familial PrCa cases. For this purpose, we used a custom T-NGS panel covering the five well-established PrCa predisposing genes, namely, BRCA1, BRCA2, ATM, CHEK2, and PALB2, and three additional genes suggested as PrCa-predisposing in the literature, namely, BRIP1, NBN, and RAD51C. We complemented the analysis pipeline for small variations (SNVs/INDELs) (29) with four different NGS data analysis software designed to detect large exonic structural variants (SVs), either copy number variants (CNV) and/or translocations (TLs).

Following our analysis pipeline for SNVs/INDELs, we identified deleterious variants in 17 patients, representing ~ 3.7% of the 462 patients of our series of early-onset/familial PrCa. By including CNV analyses, we were able to identify the BRCA2 c.156_157insAlu variant, and, additionally, to identify one carrier of a SV in BRCA1. In fact, the c.156_157insAlu variant is the BRCA2 pathogenic variant most frequently found in Portuguese families and, so far, its detection was restricted to a variant-specific detection method (26), being a challenge to NGS approaches. Thus, its identification by an NGS bioinformatic pipeline not only increases the sensitivity of the screening test, but also strengthens the potential of the T-NGS data. A limitation of the SureCall software is that it is only compatible with Agilent’s gene panels design, so other bioinformatic tools must be validated for use with other sequencing strategies. Another possible drawback of the Agilent technology may be the output of false-positive CNVs. In fact, samples from the patients with suspected CNVs (HPC498 and HPC163) were latter submitted to the panel TruSight Cancer (v2, Illumina) and only the BRCA1 duplication was outputted using the CNV tool from NextGENe. Still, in light of the identification of the Alu insertion, Agilent’s technology and associated SureCall TL analysis module may compensate the output of false-positive CNVs, which can ultimately be validated by complementary technologies.

Although aCGH has validated a true-positive CNV in BRCA1 in patient HPC498, it did not allow to define the boundaries of the genomic duplication. In the literature, a large deletion encompassing this region is reported in several HBOC families, resulting in a non-functional BRCA1 transcript (36), but a duplication encompassing the same exons has only been reported in two cases of breast cancer from different HBOC families (34, 37), none of them having identified aberrant transcripts associated with the genomic change. We have identified an aberrant transcript retaining 299bp of the genomic region between exons 1 and 2 of the BRCA1, which is compatible with the validated genomic duplication. Nevertheless, TIS predictions do not support changes at the protein level, thus, in lack of additional evidence, the BRCA1 duplication remains classified as a VUS.

In our patient cohort, CHEK2 is, overall, the most frequently mutated gene, however, ATM ranks first in patients with familial history of PrCa, representing 1.3% of the cases. In patients with early-onset diagnosis, CHEK2 is, again, the most frequently mutated gene, representing 2.1% of the cases. The high frequency of CHEK2 carriers may be attributed to the founder effect of the c.349A > G variant in our population (32). Contrarily, the Portuguese founder variant in BRCA2 (c.156_157insAlu) does not significantly contribute to early-onset/familial
PrCa in our patient cohort. Considering that this is the most frequent P/LP BRCA2 variant in our population, it is possible that patient selection criteria underlie this “bias”, either because carriers of the BRCA2 founder variant do not fit the patient selection criteria, or because BRCA2 carriers are detected by genetic screening in positive HBOC families, and, thus, not considered index patients. Nevertheless, as we found no other P/LP variant in the BRCA genes, and comparing with other studies (14, 38), the overall frequency of BRCA carriers in our patient cohort (0.2%) is the lowest described, so far, increasing just to 0.3% if considering only patients with family history of PrCa.

In 2019, a systematic review including 80 studies of different PrCa subgroups concluded that, contrarily to BRCA1, BRCA2 and ATM, germline deleterious variants in CHEK2, PALB2, NBN, and RAD51C, were more frequent in unselected PrCa cases than in patients with metastatic disease (39). Unfortunately, gene-level statistics for the subgroup of familial PrCa cases analyzed were not computed. In our study, P/LP variants in PALB2 and NBN may explain PrCa development in ~ 0.4% each of the whole series of early-onset/familial PrCa cases, not specifically enriched in either the “early-onset” or the “family history” group. In familial PrCa cases from Poland, a NBN founder variant is found in ~ 2.5% of the cases, being associated with a 4.3-fold increased PrCa risk (13).

We have not found an association between the carrier status for a P/LP variant and disease aggressiveness. However, considering that our cohort is neither enriched in patients with aggressive disease nor in carriers of P/LP variants in BRCA2, associated with aggressive disease in multiple studies (14, 17, 38, 40), we cannot discard the possible effect of a “cumulative bias” driving this lack of association, eventually conditioned by a population-specific effect. P/LP variants in ATM, NBN and PALB2 have also been associated with aggressive PrCa (40, 41), however the low frequency of carriers in our cohort does not allow to reach statistical power.

Our study shows that deleterious variants in HRR genes may explain PrCa development in ~ 3.9% of the PrCa cases arising at early-age and/or with familial aggregation of the disease. CHEK2 and ATM P/LP variants are the most frequent (38.9% and 22.2% of the carriers, respectively), followed by PALB2 and NBN (11.1% of the carriers, each), and then by BRCA2, RAD51C, and BRIP1 (5.6% of the carriers each). By allowing full screen of many genes at the same time and to identify both small and large genomic variations, NGS has a major role in the identification of the genetic components underlying cancer development. Still, further studies are needed to clarify the (prostate) cancer risk of carrying a deleterious variant in several of these genes and their implications in patient management.

Declarations

DATA AVAILABILITY STATEMENT

Data generated during this study can be found within the published article and its supplementary files.

AUTHOR CONTRIBUTION

Conception and design: Manuel Teixeira. Sample collection: Sofia Maia, Rui Silva. Data acquisition: Paula Paulo, Marta Cardoso, Andreia Brandão, Pedro Pinto, Ariane Falconi, Manuela Pinheiro. Analysis and interpretation of data: Paula Paulo, Andreia Brandão, Ariane Falconi, Manuela Pinheiro, Nuno Cerveira, Catarina Santos, Carla Pinto, Ana Peixoto. Statistical analysis: Paula Paulo. Drafting of the manuscript: Paula
Paulo. **Critical revision of the manuscript:** Manuel Teixeira, Andreia Brandão, Catarina Santos, Carla Pinto, Ana Peixoto. **Funding acquisition:** Manuel Teixeira.

**FUNDING**

This work was funded by Fundação para a Ciência e a Tecnologia (FCT; POCI-01-0145-FEDER-016889 and NORTE-01-0145-FEDER-072678) and by IPO-Porto Research Center (CHIPOP-24-2015). The following authors have/had their PhD fellowships or research contracts funded by FCT: AB (POCI-01-0145-FEDER-016889, POCI-01-0145-FEDER-028245, POCI-01-0145-FEDER-006868 and 2021.03835.CEECIND), PPa (CEECINST/00091/2018), PPI (POCI-01-0145-FEDER-006868), MC (SFRH/BD/116557/2016), MP (SFRH/BPD/113014/2015 and NORTE-01-0145-FEDER-072678) and SM (SFRH/BD/71397/2010). RS was funded by Portuguese League Against Cancer (Liga Portuguesa Contra o Cancro – Núcleo Regional do Norte). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**ETHICAL APPROVAL**

Patient samples were collected under the ethical agreement CES 38-010 revised by the Research Ethics Committee (IPO-Porto Comissão de Ética para a Saúde - CES). Informed consent was obtained from all subjects involved in the study.

**COMPETING INTERESTS**

The authors declare that they have no conflict of interest.

**References**


**Figures**

**Figure 1**

A. HPC234: ATM c.8264_8268del; p.(Tyr2755CysfsTer12)

B. HPC196: ATM c.9079dup; p.(Ser3027LysfsTer36)

C. HPC408: ATM c.1236-2A>G; p.(?)

D. HPC353: CHEK2 c.593-1G>T; p.(?)

E. HPC36: PALB2 c.1438A>T; p.(Lys480Ter)

F. HPC223: PALB2 c.2257C>T; p.(Arg753Ter)
Family trees of HPC cases carrying P/LP SNVs/INDELs in well-established PrCa predisposing HRR genes. A, B and C) Patients carrying P/LP variants in ATM. D) Patient carrying the recurrent splicing variant in CHEK2. E and F) Patients carrying P/LP variants in PALB2. Squares represent the males, circles the females, and diamonds unknown gender. Numbers inside the symbols denote the number of individuals with that gender, if known. Deceased individuals are represented by a diagonal line through a symbol and the affected ones are highlighted by colored symbols. The index case is indicated by an upper left arrow and the cancer type and age at diagnosis are indicated whenever known. Identified carriers of the variant are marker by a plus (+) symbol. GaCa- Gastric carcinoma; HeCa- Hepatic carcinoma; LuCa- Lung carcinoma; CoCa- Colon carcinoma; PrCa- Prostate carcinoma; NHL- Non-Hodgkin Lymphoma.

Figure 2

Family trees of HPC cases carrying P/LP SNVs/INDELs in HRR genes suggested PrCa predisposing genes. A) Patient carrying the BRIP1 frameshift variant. B and C) Patients carrying the NBNsplicing and frameshift variants, respectively. Squares represent the males, circles the females, and diamonds unknown gender. Numbers inside the symbols denote the number of individuals with that gender, if known. Deceased individuals are represented by a diagonal line through a symbol and the affected ones are highlighted by colored symbols. The index case is indicated by an upper left arrow and the cancer type and age at diagnosis are indicated whenever known. Identified carriers of the variant are marker by a plus (+) symbol. AMI- Acute Myocardial Infarction; BrCa- Breast carcinoma; LuCa- Lung carcinoma; OvCa- Ovary carcinoma; PrCa- Prostate carcinoma.
Figure 3

Distribution of carriers of P/LP variants by gene considering cases with early-onset or familial PrCa. Blue colors represent well-established and yellow colors not well-established PrCa predisposing genes.

Figure 4

Exonic rearrangements identified by T-NGS. A) Sequence alignment obtained with SureCall TL analysis for sample HPC398, showing insertion of the Alu sequence in exon 3 of BRCA2 (c.156_157insAlu), flanked by a short sequence duplication (TSD)(33). B) Dot-plot of the ratio of the read-depth coverage of the test sample
(HPC498 in the *upper panel* and HPC163 in the *lower panel*) against the median read-depth coverage of a set of ten, randomly selected, samples from the same NGS run. *Green* and *red* horizontal lines denote the empirical threshold to assign gains and losses, respectively. **C** and **D** Validation of the *BRCA1* duplication in patient HPC498 was obtained by MLPA (**C**) and aCGH (**D**). A new transcript retaining 299bp of the alternative exon 1B was identified in patient HPC498 by One-Step RT-PCR in RNA extracted from PBLs. Agarose gel electrophoresis of the RT-PCR amplicons and Sanger sequencing electropherograms are shown in **E** and **F**, respectively. The identified aberrant *BRCA1* transcript, matching Ensembl transcript ENST00000357654.9 (BRCA1-203), contains, additionally, three SNVs (black arrowed), two described in dbSNP with MAF<0.01% [c.-20+415A>G, rs1420583297 (**a**)]; and c.-20+346>A>G, rs1489102499 (**c**)], and one [c.-20+349>A>G (**b**)] not described in gnomAD database. GaCa- Gastric carcinoma; HeCa- Hepatic carcinoma; LuCa- Lung carcinoma; CoCa- Colon carcinoma; PrCa- Prostate carcinoma; NHL- Non-Hodgkin Lymphoma.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformationNoFieldCodes.docx](#)