Title

Association between Single Nucleotide Polymorphisms in DNA Double-Strand Break Repair Genes and Prostate Cancer Aggressiveness in the Spanish Population

Authors: Luis Alberto Henríquez-Hernández[a,b,c] (Ph.D), Almudena Valenciano[b] (B.D.), Palmira Foro-Arnalot[d] (Ph.D., M.D.), María Jesús Álvarez-Cubero[e,f] (Ph.D.), José Manuel Cozar[g] (Ph.D., M.D.), José Francisco Suárez-Novo[h] (M.D.), Manel Castells-Esteve[h] (M.D.), Pablo Fernández-Gonzalo[i] (M.D.), Belén De-Paula-Carranza[i] (M.D.), Montse Ferrer[j] (Ph.D., M.D.), Ferrán Guedea[k] (Ph.D., M.D.), Gemma Sancho-Pardo[l] (Ph.D., M.D.), Jordi Craven-Bartle[l] (Ph.D., M.D.), María José Ortiz-Gordillo[m] (Ph.D., M.D.), Patricia Cabrera-Roldán[m] (Ph.D., M.D.), Juan Ignacio Rodríguez-Melcón[a,b] (M.D.), Estefanía Herrera-Ramos[n,o] (B.D.), Carlos Rodríguez-Gallego[n] (Ph.D.), and Pedro C. Lara[a,b,c] (Ph.D., M.D.)

Affiliations

[a]Radiation Oncology Department, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas (Spain)

[b]Instituto Canario de Investigación del Cáncer, Las Palmas (Spain)

[c]Clinical Sciences Department, Universidad de Las Palmas de Gran Canaria, Las Palmas (Spain)

[d]Institud d'Oncologia Radioteràpica, Hospital de la Esperanza. Parc de Salut Mar, Barcelona (Spain)

[e]Laboratory of Genetic Identification, Legal Medicine and Toxicology Department, Facultad de Medicina, Universidad de Granada, Granada (Spain)

[f]GENYO, Pfizer-University of Granada-Andalusian Government Centre for Genomics and Oncological Research, Granada (Spain)

[g]Department of Urology, Hospital Universitario Virgen de las Nieves, Granada (Spain)

[h]Department of Urology, Hospital Universitari de Bellvitge, L'Hospitalet de Llobregat (Spain)

[i]Radiation Oncology Department, Onkologikoa, Guipuzcoa (Spain)

[j]Health Services Research Group, Institut de Recerca Hospital del Mar (IMIM), Barcelona (Spain)

[k]Department of Radiation Oncology, Institut Català d'Oncologia (ICO), L'Hospitalet de Llobregat (Spain)

[l]Radiation Oncology Department, Hospital de la Santa Creu i Sant Pau, Barcelona (Spain)

[m]Radiation Oncology Department, Hospital Universitario Virgen del Rocío, Sevilla (Spain)

[n]Department of Immunology, Hospital Universitario de Gran Canaria Dr. Negrín, Las Palmas (Spain)

[o]Department of Medical and Surgical Sciences, Universidad de Las Palmas de Gran Canaria, Las Palmas (Spain)

Corresponding author: Luis Alberto Henríquez-Hernández. Address: C/Barranco de La Ballena s/n, CP 35010, Las Palmas de Gran Canaria, Las Palmas (Spain). Phone: 0034928450927, fax: 0034928449805, e-mail: luis.henriquez@ulpgc.es

Running title: SNPs in NHEJ and prostate cancer aggressiveness

Abstract

Background. Novel predictors of prognosis and treatment response for prostate cancer are required to better individualize treatment. Single nucleotide polymorphisms (SNPs) in 4 genes direct (*XRCC5* and *XRCC6*) or indirect (*PARP1* and *MVP*) involved in non-homologous end-joining were examined in 494 Spanish prostate cancer patients.

Methods. A total of 22 SNPs were genotyped in a Biotrove OpenArray® NT Cycler. Clinical tumor stage, diagnostic prostate specific antigen (PSA) serum levels, and Gleason score at diagnosis were obtained for all participants. Genotypic and allelic frequencies were determined using the web-based environment SNPator.

Results. (*XRCC6*) rs2267437 appeared as a risk factor for developing more aggressive prostate cancer tumors. Those patients carrying the GG genotype were at higher risk of develop bigger tumors (Odds ratio (OR) = 2.04, (95% Confidence Interval (CI) 1.26 – 3.29), p = 0.004), present higher diagnostic PSA levels (OR = 2.12, (95%CI 1.19 – 3.78), p = 0.011), higher Gleason score (OR = 1.65, (95%CI 1.01 – 2.68), p = 0.044), and D'Amico higher risk tumors (OR = 2.38, (95%CI 1.24 – 4.58), p = 0.009) than those patients carrying the CC/CG genotypes. Those patients carrying the (*MVP*) rs3815824 TT genotype were at higher risk of present higher diagnostic PSA levels (OR = 4.74, (95%CI 1.40 – 16.07), p = 0.013) than those patients carrying the CC genotype. When both SNPs were analyzed in combination, those patients carrying the risk genotypes were at higher risk of develop D'Amico higher risk tumors (OR = 3.33, (95%CI 1.56 – 7.17), p = 0.002).

Conclusions. For the first time, genetic variants at *XRCC6* and *MVP* genes are associated with risk of more aggressive disease, and would be taken into account when assessing the malignancy of prostate cancer.

Keywords: SNP; NHEJ; *XRCC6*; *MVP*; prostate cancer; OpenArray

1. INTRODUCTION

Prostate cancer (PCa) is a complex disease conditioned by hormonal and genetics factors. Tumor staging, tumor grading in terms of Gleason score and diagnostic prostate specific antigen (PSA) serum levels are used in the daily clinical practice to classify patients into different prognostic risk groups that condition treatment decisions. However, these clinical prognostic factors only explain a proportion of the patient to patient variation observed in the clinical outcome [1], and it seems that better predictors of individualized prognosis and treatment response are urgently needed.

DNA double-strand break (DSB) is the most detrimental form of DNA damage, leading cells to chromosomal breakage and rearrangements [2]. DSBs can have exogenous (i.e. exposure to ionizing radiation (IR)) and endogenous (i.e. due to reactive oxygen species) origin, and can also be generated during somatic recombination, DNA replication and when DNA single-strand breaks are encountered [2]. Unrepaired or misrepaired DSBs would cause cell death, genomic instability and/or oncogenic transformation [3]. In eukaryotic cells, DSBs are repaired by two different pathways: homologous recombination (HR) and non-homologous end-joining (NHEJ) [4]. Recently, single nucleotide polymorphisms (SNPs) – defined as inherited mutations that are present in more than 1% of the population – in HR genes have been proposed as novel factors associated with PCa aggressiveness [5]. Due to the complexity of mammalian genomes, searching of homologous sequences for HR is not efficient. Thus, NHEJ plays a predominant role in DSB repair [6]. This fact makes sense when is taken into account that most of the mammalian genome consists of noncoding sequences [7]. Thus, common variants in genes involved in DNA DSB repair are good candidates for low-penetrance cancer biomarkers [8].

The NHEJ repair mechanism starts with the recruitment of Ku70/Ku80 proteins to the DNA ends. Ku70 is encoded by the x-ray repair complementing defective repair in Chinese hamster cells 6 (*XRCC6*) gene while Ku80 is encoded by the x-ray repair complementing defective repair in Chinese hamster cells 5 (*XRCC5*) gene. Although other proteins are involved in end

ligation (i.e. DNA protein kinase (DNA-PK), XRCC4 or ligase IV (LIG4)), the well function of the Ku70/Ku80 complex is critical when NHEJ starts [9]. To our knowledge, studies concerning polymorphic variants of NHEJ genes have not been done in PCa.

Vaults are ribonucleoprotein particles with a hollow barrel-like structure [10]. It is composed of three proteins: the major vault protein (MVP), the vault poly(adenosine diphosphate-ribose) polymerase also known as VPARP, and telomerase-associated protein-1 (TEP1), together with small untranslated RNA (vRNA). Classically, MVP has been associated with the development of drug resistance [11]. However, in recent years there has been increasing interest in MVP, especially since *MVP* transcription and protein levels are increased in response to DNA-damaging agents including IR [12]. In addition to its implication in the regulation of several cellular processes including transport mechanisms, signal transduction and immune responses [13], the interaction with partners such as PTEN or BCL-2 suggests a role in the cell fate [14,15]. Furthermore, MVP has been recently linked to DSB repair machinery due to an association with Ku70/Ku80 expression [16]. Thus, SNPs in MVP would have a role in PCa behavior, due to the role in the resistance to treatment, the regulation of apoptosis or the association with DNA damage repair mechanisms [13].

Poly[ADP-ribose] synthase 1 is a protein encoded by the gene *PARP1* which is associated with differentiation, proliferation, and tumor transformation [17]. *PARP1* seems to have a role in DNA DSB repair through an interaction with *BRCA1/2* and other genes [18]. Moreover, *PARP1* and *Ku* compete for repair of DNA double strand breaks by different NHEJ pathways [19]. In the last years, PARP1 inhibitors have shown promising activity in ovarian and breast cancers patients with *BRCA1/2* mutation [20], and also in PCa patients [21]. The fact that *PARP1* is a crossroad between HR and NHEJ and it is also of interest as a novel therapeutic target makes this gene a good candidate associated with tumor aggressiveness. Nowadays, the role of single nucleotide variations in *PARP1* has not been studied in the context of prostate cancer.

The aim of this study was to test the hypothesis that genetic variants in genes involved directly (*XRCC6* and *XRCC5*) or indirectly (*MVP* and *PARP1*) in DNA DSB repair confer increased risk of more aggressive prostate cancer. To our knowledge, this is the first time that SNPs located in these genes are studied in prostate cancer patients.

2. MATERIALS AND METHODS

2.1. Patients

A total of 601 patients with non-metastatic localized prostate cancer from 4 different regions of Spain (15.14% from Andalusia, 8.48% from Basque Country, 39.60% from Canary Islands and 36.77% from Catalonia) were included in the study between April 2003 and December 2012, in the context of a multicenter study aimed to evaluate quality of life and clinical outcome of PCa patients [22]. All patients provided written informed consent before sample collection. The study was approved by the Research and Ethics Committee of each institution participant in the study: Hospital Universitario de Gran Canaria Dr. Negrín (Las Palmas de Gran Canaria), Hospital de la Esperanza. Parc de Salut Mar (Barcelona), Hospital Universitario Virgen de las Nieves (Granada), Hospital Universitari de Bellvite (L'Hospitalet de Llobregat), Onkologikoa (Guipuzcoa), Institut Català d'Oncologia (L'Hospitalet de Llobregat), Hospital de la Santa Creu i Sant Pau (Barcelona) and Hospital Universitario Virgen del Rocío (Sevilla). Details of the series are described previously [5,23,24].

Clinical tumor size (cT), diagnostic PSA serum levels, and Gleason score [25] were recruited for all PCa patients. cT was assessed by digital rectal examination (DRE) followed by transrectal ultrasonography (TRUS) and magnetic resonance imaging (MRI); PSA serum levels were assessed by chemiluminescence in an Architect i2000 analyzer (Abbott Laboratories, IL, USA); Gleason score was determined in the biopsy specimen by a specialized genitourinary pathologist from each Institution. Subjects were categorized into three risk-based recurrence groups according to D'Amico classification [26]: low, intermediate, and high risk. After collecting demographic and clinical data, a blood sample was taken after the signature of informed

consent. All samples were sent by courier to the Hospital Universitario de Gran Canaria Dr. Negrín for DNA isolation and genotype analyses.

2.2. DNA isolation and quantification

DNA was isolated from 300 µl of whole-blood in an iPrep™Purification Instrument using the iPrep™ PureLink™ gDNA Blood Kit (Invitrogen, by Life Technologies, Carlsbad, CA), and its integrity was determined by NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE).

2.3. Selection criteria for SNPs

SNPs were selected using data of individuals with European ancestry (CEU) from the HapMap Project (available at: www.hapmap.org). Pairwise linkage disequilibrium (LD) tagging was achieved with Haploview v4.2 software [27]. This strategy is based in the haplotype blocks formed by clustered SNPs [28]. By selecting a modest number of SNPs within a block, it is possible to recover up to 95% of the heterozygosity present at a locus. The R^2 given by Haploview was >0.95 for all genes.

Thus, a total of 22 SNPs in 4 different genes involved in DNA DSB repair were studied (S1 Table): *XRCC6*, rs5751131, rs2267437, rs11912946, rs7291732, rs881092; *XRCC5*, rs16855458, rs9288516, rs1051677, rs1051685; *PARP1,* rs8679, rs3219123, rs1805410, rs3219062, rs1805414, rs1805404, rs3219027; and *MVP*, rs9923649, rs12149514, rs4788186, rs3815824, rs35916172, rs3764944.

2.4. Genotyping

The SNP genotyping was performed in a Biotrove OpenArray® NT Cycler (Applied Biosystems, Foster City, CA) [29]. DNA samples loaded in the OpenArray (OA) had a A260/A280 and A260/230 ratios of 1.7-1.9, and were adjusted to 50ng/µl. A total of 300 ng of genomic DNA was used. A final amount of 150 ng was incorporated into the array with the autoloader, and was genotyped according to the manufacturer's recommendations. A non-

template control (NTC) consisting of DNase-free double-distilled water was introduced within each assay. When the DNA and master mix were transferred, the loaded OA plate was filled with an immersion fluid and sealed with glue. The multiplex TaqMan assay reactions were carried out in a Dual Flat Block (384-well) GeneAmp PCR System 9700 (Applied Biosystems) with the following PCR cycle: an initial step at 93°C for 10 minutes followed by 50 cycles of 45 seconds at 95°C, 13 seconds at 94°C and 2:14 minutes at 53°C; followed by a final step during 2 minutes at 25ºC and holding at 4ºC.

The fluorescence was read using the OpenArray® SNP Genotyping Analysis software version 1.0.5. (Applied Biosystems). The genotyping analysis was made with the TaqMan Genotyper software version 1.0.1. (available at: ttp://www.invitrogen.com) using autocalling as the call method. The quality value of the data points was determined by a threshold above 0.95. Genotype analysis was performed with the same batch of chips and by the same investigator, as previously reported [5,23,24].

About 10% of the samples were randomly selected for confirmation and the results were all consistent.

2.5. Statistical Analysis

Genotype and allelic frequencies were determined using the web-based environment SNPator (SNP Analysis To Results, from the Spain's National Genotyping Centre and the National Institute for Bioinformatics) [30]. Relative excess of heterozygosity was determined to check compatibility of genotype frequencies with Hardy-Weinberg equilibrium (HWE). Thus, p-values from the standard exact HWE lack of fit test were calculated using SNPator. Comparisons of genotypic and allelic frequencies were also done in SNPator. All additional statistical analyses were performed using PASW Statistics 15 (IBM Corporation, Armonk, NY, USA) as follows: descriptive analyses were conducted for clinical variables, categorical variables were represented as absolute numbers and percentages, comparisons between groups

were performed using chi-squared test (categorical variables), the associations between clinical variables and genetic variables were analyzed using a linear additive model employing logistic regressions. Different genetic models were taken into account.

3. RESULTS

3.1. Intra-ethnic variations of SNPs

All the genotyped samples met the quality criteria stated above. A total of 601 PCa patients were genotyped for 22 SNPs. Of the 13,222 possible determinations, 12,608 were successfully genotyped (95.0 %).

We have previously shown that differences in the distribution of genotypes within different populations of the same ethnicity are an important confounding factor in genetics epidemiology [23,24]. Thus, we first analyzed the genotype distribution among subjects from different Spanish regions, highlighting that all PCa patients were from Caucasian origin. The genotypic and allelic frequencies are shown in S2 Table. We observed that 9 of the 22 analyzed SNPs were differentially distributed: (*XRCC6*) rs5751131, rs2267437, rs7291732, rs881092, (*XRCC5*) rs1051677, (*PARP1*) rs3219123, rs1805410, rs1805414, and (*MVP*) rs4788186. As expected, Andalusian population showed a genotypic and allelic distribution clearly different to those reported among subjects from Basque Country, Canary Islands and Catalonia. Thus, we excluded Andalusian patients from subsequent analyses to homogenize the sample and minimize bias. We have previously followed this strategy of analysis [5].

3.2. Clinical variables associated with individual SNPs

A total of 494 PCa were included in the subsequent association studies. Distribution of clinical variables is detailed in Table 1. The majority of PCa patients were cT1a – cT2a (54.7%), PSA < 10 ng/mL (61.9%), and Gleason score < 7 (45.7%). Subsequently, a total of 120 patients (61.5%) were classified as low-intermediate risk tumors according to D’Amico classification.

Among the 22 analyzed SNPs, (*XRCC6*) rs2267437, (*XRCC5*) rs1051685, (*PARP1*) rs3219123, and (*MVP*) rs3815824, were significantly different distributed among PCa patients according to the clinical variables (Table 2). Genotype distribution was consistent with Hardy-Weinberg equilibrium (data not shown).

We observed that rs2267437 was significantly associated with the cT ($\chi 2$ test, p = 0.004). Thus, among the 249 PCa patients with cT1a – cT2a tumors, 37 carried the GG genotype (14.8%), while among the 183 PCa patients with cT2b – cT4 tumors, 48 carried the GG genotype (26.2%). Similar results were observed regarding to Gleason score and D'Amico risk groups ($\chi 2$ test, p = 0.037 and p = 0.015, respectively) (Table 2). In both cases, GG carriers showed more aggressive tumor characteristics; thus, GG genotype was detected in 12 out of 110 low risk tumors and in 73 out of 323 intermediate – high risk tumors (10.9% vs. 22.6%, respectively). Univariate analyses showed that GG genotype for (*XRCC6*) rs2267437 appeared as a risk factor for developing more aggressive PCa tumors (Table 3). The fact that recessive and homozygote models showed significant results reinforce the role of GG genotype as a factor of worst tumor characteristics. In that sense, those PCa patients carrying the GG genotype were at higher risk of develop bigger tumors (Odds ratio (OR) = 2.04, (95% Confidence Interval (CI) 1.26 – 3.29), p = 0.004), present higher diagnostic PSA levels (OR = 2.12, (95%CI 1.19 – 3.78), p = 0.011), higher Gleason score (OR = 1.65, (95%CI 1.01 – 2.68), p = 0.044), and D'Amico higher risk tumors (OR = 2.38, (95%CI 1.24 – 4.58), p = 0.009) than those patients carrying the CC/CG genotypes (Table 3).

We observed that levels of diagnostic PSA was associated with three SNPs located in three different genes (Table 2). Among 404 PCa patients showing diagnostic PSA levels < 20 ng/mL, 78 carried the (*XRCC5*) rs1051685 G allele (19.3%), while 8 out of 79 subjects with initial PSA levels ≥ 20 ng/mL carried the G allele (10.1%) ($\chi 2$ test, p = 0.007). This result suggests that G allele appeared as a protective factor, a fact reinforced in univariate analysis: those PCa patients carrying the AG genotype were at lower risk of present higher diagnostic PSA levels (OR =

0.388, (95%CI 0.16 – 0.93), p = 0.033) than those patients carrying the AA genotype (Heterozygote model, data not shown). On the contrary, those PCa patients carrying the (*MVP*) rs3815824 TT genotype were at higher risk of present higher diagnostic PSA levels (OR = 4.74, (95%CI 1.40 – 16.07), p = 0.013) than those patients carrying the CC genotype (Homozygous model, data not shown). Finally, although (*PARP1*) rs3219123 was significantly associated with diagnostic PSA levels ($\chi 2$ test, p = 0.041; Table 2), this SNP did not appear as a risk factor for that clinical variable in univariate analysis (Data not shown).

3.3. Combined analyses of multiple SNPs

Since it is plausible that the genes considered in the present study contribute to prostate cancer behavior – due to a direct or indirect interaction with NHEJ pathway – we estimated the combined effect of the significantly associated genes (Fig. 1). We considered (*XRCC6*) rs2267437 as the basics for these analyses, because its high impact in tumor characteristics (as previously shown in Tables 2 and 3). While the combination with (*XRCC5*) was not significant (OR = 2.11, (95%CI 0.91 – 4.94)), the combined role of (*XRCC6*) rs2267437 with (*PARP1*) rs3219123 or (*MVP*) rs3815824 were risk factor associated with D'Amico risk group. Thus, PCa patients carrying genotypes different to those considered as reference (CC + CC and CC + GG for rs2267437 + rs3815824 and rs2267437 + rs3219123, respectively) were at higher risk of develop intermediate – high D'Amico risk groups (Fig. 1). Interestingly, the combination of these 3 SNPs did not showed a significant increment in the risk, suggesting that (*XRCC6*) rs2267437 and (*MVP*) rs3815824 have a predominant role in PCa tumor behavior (OR = 3.33, (95%CI 1.56 – 7.17), p = 0.002).

4. DISCUSSION

Although polymorphic alleles of NHEJ genes would predispose carriers to a high risk of developing more aggressive tumors, to date, no direct clinical evidence has been obtained in patients with prostate cancer. We have previously published that common genetic variations in genes involved in HR (such as *XRCC1, ERCC1* or *ERCC2*, among others) are associated with

clinical variables of poor prognosis for prostate cancer [5]. In the present study, we focused in the study of classical genes involved in NHEJ (*XRCC6* and *XRCC5*) as well as other genes associated with DSB DNA repair (*PARP1* and *MVP*). To our knowledge, this is the first comprehensive study that analyzes the possible role of these 4 genes and its influence in prostate cancer behavior. Although it is a candidate gene study, we used an SNP tagging approach which reduces the bias inherit to the selection of well published SNPs and allows the inclusion of novel genetic variants with possible new biological roles.

A total of 601 PCa patients from 4 different Spanish regions were included in the study. Intra-ethnic variations of genotype and allelic frequencies are considered an important confounding factor frequently obviated. Confounding would arise if the population contained several ethnic groups, if allele frequencies at the locus of interest differed between groups, and if disease frequency also differed between groups for reasons quite unrelated to the locus of interest [31]. The phenomenon was previously observed in our series [23,24], a fact explained by differences in the natural history of the populations [32,33]. As expected, the Andalusian population showed significant differences in genotype and allelic distribution, and subsequently, this subset of PCa patients was excluded from the association analyses to avoid the intra-ethnic bias.

Ku70/80 is a protein complex encoded by *XRCC6* and *XRCC5* genes which is involved in starting the NHEJ process. Specifically, ku70/80 is responsible for the recognition of the DNA ends at the site where de DSB lesion takes place [9,34,35]. Since ku70/80 is a complex formed by two different monomers, the functionality of the complex could be compromised if there are structural changes in the proteins, and it is well known that genetic polymorphisms can compromise not only the binding of a protein complex, but also the strength of the bounds between proteins and DNA [34]. *XRCC6* rs2267437 appeared as a risk factor associated with clinical variables of poor prognosis for prostate cancer in terms of tumor size, diagnostic PSA level, Gleason score and even D'Amico risk groups. The rs2267437 polymorphism has been investigated as risk factor for cancer risk with inconsistent results. However, a recent systematic

review and meta-analysis concluded that rs2267437 SNP is associated with a significant increase in risk of breast cancer, renal cell carcinoma and hepatocellular carcinoma [36]. Nonetheless, this study did not include any prostate cancer patient and no associations with clinical variables were analyzed. The role of rs2267437 – specifically the GG genotype – in cancer seems to be established in other meta-analysis [37]. Our results reinforce these previous results and postulate rs2267437 SNP as a novel genetic factor for prostate cancer behavior.

The role of *XRCC5* rs1051685 has been previously evaluated in myeloma patients. Those patients carrying the GG genotype were at lower risk of develop myeloma [38]. This result was proven among Spanish multiple myeloma patients [39] and agree with those reported in the present study in prostate cancer patients. Although is expectable to find a link between SNPs in *XRCC5* and *XRCC6,* when rs2267437 and rs1051685 were analyzed together we did not observe such association, possible due to the small amount of patients carrying the “protective” genotype (n = 27). The minor allele frequency for rs1051685 was around 10% and this result agrees with those reported specifically in the European population (data available at http://www.ncbi.nlm.nih.gov/projects/SNP/). With the data available, the potential role of rs1051685 polymorphism needs to be explored deeply.

We did not find previously published studies aimed to explore the association between genetic variants in *PARP1* and prostate cancer. Other SNPs in *PARP1* seem to be associated with glioma [40] and gastrointestinal cancers [41], but nothing is reported in relation to rs3219123 or prostate cancer. In our series, we observed an association between this SNP and diagnostic PSA levels, but univariate analysis was not statistical significant. Even when rs3219123 was analyzed in combination with rs2267437, the results were less significant than those obtained with rs2267437 alone, suggesting a poor role of (*PARP1*) rs3219123 in the biology of prostate cancer.

There is only one previous publication studying the role of genetic variants in *MVP*. It is an article aimed to explore the role of SNPs in multidrug resistant genes in patients with mesial temporal lobe epilepsy with hippocampal sclerosis [42]. The authors did not observe any role of different SNPs in *MVP* – including rs3815824 – in this type of patients. In that sense, *MVP* is an unknown gene which interest has increased since its association with Ku70/Ku80 [16] and its role in radiation response [43-45] in the clinical setting. We reported here for the first time that prostate cancer patients carrying the TT genotype were at higher risk of develop worst prostate tumors in terms of diagnostic PSA level (a classical prognosis factor for this type of cancer). Moreover, when combined with (*XRCC6*) rs2267437, these SNPs together increased the risk of a higher D'Amico score at diagnosis. The mechanisms behind this association are unknown, but support the existing hypothesis about a potential link between *MVP* and NHEJ [16].

The functional consequences of the SNPs reported in the present study are difficult to address since rs2267437 and rs3815824 are intron variations which are not traduced into protein. However, introns are removed and exons are joined during a complex post-transcriptional modification of RNA named splicing, and the consequences of SNPs located in intron regions on cis regulatory elements which drive the splicing process are unknown but would influence the spliceosome [46]. Moreover, SNPs located in non-coding regions have been associated with human diseases and in any case, are needed to a better understanding of their impact on gene function and health of an individual [47]. Finally, SNPs located in intron regions are considered in *in silico* tools such as "F-SNP: a collection of functional SNPs, specifically prioritized for disease association studies" (available at: http://compbio.cs.queensu.ca/F-SNP/). This fact suggests the biological relevance of this type of genetic variations in human diseases [48,49]. In that sense, functional studies are required to disclose the biological consequences of rs2267437 and rs3815824 in prostate cancer.

The present study provides a number of advantages that contribute to their credibility. First, the high confidence of the genotyping system and the clinical endpoints considered (of total routine

in daily clinical practice) minimize the possibility of inferring results due to the subjectivity of the observations. Second, the present study has been performed in a series of Spanish prostate cancer patients which was homogeneous regarding to intra-ethnic variability, a factor that has been shown to be important in association studies with polymorphisms. And third, all the determinations (13,222 in total) were performed with the same methodology (OpenArray, Applied Biosystems), with the same batch of chips and by the same investigator, thus minimizing biases from technical origin. In the other hand, the present study has some weaknesses that need to be highlighted: i) although 601 PCa patients seem sufficient to obtain statistically reliable results, it is possible that some results may be of stochastic nature (type I error), especially for those SNPs with lower MAF and taken into account that the association studies were performed in a series of 494 PCa patients; ii) other factors associated with prostate cancer (i.e. age, familiar aggregation, toxic habits, prostate volume, testosterone level, other DNA repair genes or some kind of diets) have not been taken into account in the present study; iii) in the present study we did not account for multiple testing (i.e. Bonferroni correction), thus limiting the strength of some statistical analyses; and iv) functional studies to disclose the biological consequences of the SNPs are required to understand the real role of these genetic variations in prostate cancer.

Taken together, our results suggest that genetic variants in DNA DSB repair genes and other genes close related to the NHEJ, could be of relevance in prostate cancer. An inability of cells to respond properly to, or to repair, DNA damage leads to genetic instability, which in turn may enhance the rate of cancer development or condition the aggressiveness of the tumor. Anyhow, to better understand the biological consequences of these genetic variations, other genes involved in DNA repair (such as *p53, ATM, BRCA1* or *BRCA2*) have to be taken into account, especially when there are accumulating evidence connecting deficiencies in cellular responses to DNA DSBs with tumorigenesis [2].

5. CONCLUSIONS

In summary, our study shows for the first time that (*XRCC6*) rs2267437 and (*MVP*) rs3815824 polymorphisms are associated with clinical variables of more aggressive prostate cancer. The present results have to be considered as hypothesis generating, and additional studies with a larger separate cohort of men should be done to validate the results as well as their combined effects. Moreover, functional studies are required to understand the role of both SNP in prostate cancer.

Acknowledgements

We thank the technical support from the Immunology Department (Hospital Universitario de Gran Canaria Dr. Negrín) staff: Nereida González-Quevedo and Yanira Florido-Ortega. This work was subsidized by grants from the Instituto de Salud Carlos III (Ministerio de Economía y Competitividad from Spain), ID: PI12/01867 and PI13/00412; and Instituto Canario de Investigación del Cáncer (ICIC-GR-F-14/11). Almudena Valenciano has a grant from the Instituto Canario de Investigación del Cáncer (ICIC). The founders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Fraser M, Berlin A, Bristow RG, van der Kwast T. Genomic, pathological, and clinical heterogeneity as drivers of personalized medicine in prostate cancer. *Urol Oncol* 2015; **33**(2)**:** 85-94.

2. Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* 2001; **27**(3)**:** 247-254.

3. Sharpless NE, Ferguson DO, O'Hagan RC, Castrillon DH, Lee C, Farazi PA *et al*. Impaired nonhomologous end-joining provokes soft tissue sarcomas harboring chromosomal translocations, amplifications, and deletions. *Mol Cell* 2001; **8**(6)**:** 1187-1196.

4. Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. *Cell Res* 2008; **18**(1)**:** 134-147.

5. Henriquez-Hernandez LA, Valenciano A, Foro-Arnalot P, Alvarez-Cubero MJ, Cozar JM, Suarez-Novo JF *et al*. Single nucleotide polymorphisms in DNA repair genes as risk factors associated to prostate cancer progression. *BMC Med Genet* 2014; **15:** 143.

6. West SC. Molecular views of recombination proteins and their control. *Nat Rev Mol Cell Biol* 2003; **4**(6)**:** 435-445.

7. Strachan T, Read AP. *Human molecular genetics*. BIOS Scientific Publishers ; Wiley-Liss: Oxford New York, 1996, xiv, 597 p.pp.

8. Tseng RC, Hsieh FJ, Shih CM, Hsu HS, Chen CY, Wang YC. Lung cancer susceptibility and prognosis associated with polymorphisms in the nonhomologous end-joining pathway genes: a multiple genotype-phenotype study. *Cancer* 2009; **115**(13)**:** 2939-2948.

9. Bassing CH, Alt FW. The cellular response to general and programmed DNA double strand breaks. *DNA Repair (Amst)* 2004; **3**(8-9)**:** 781-796.

10. Kedersha NL, Heuser JE, Chugani DC, Rome LH. Vaults. III. Vault ribonucleoprotein particles open into flower-like structures with octagonal symmetry. *J Cell Biol* 1991; **112**(2)**:** 225-235.

11. Scheffer GL, Wijngaard PL, Flens MJ, Izquierdo MA, Slovak ML, Pinedo HM *et al*. The drug resistance-related protein LRP is the human major vault protein. *Nat Med* 1995; **1**(6)**:** 578-582.

12. Shimamoto Y, Sumizawa T, Haraguchi M, Gotanda T, Jueng HC, Furukawa T *et al*. Direct activation of the human major vault protein gene by DNA-damaging agents. *Oncol Rep* 2006; **15**(3)**:** 645-652.

13. Lara PC, Pruschy M, Zimmermann M, Henriquez-Hernandez LA. MVP and vaults: a role in the radiation response. *Radiat Oncol* 2011; **6:** 148.

14. Chung JH, Ginn-Pease ME, Eng C. Phosphatase and tensin homologue deleted on chromosome 10 (PTEN) has nuclear localization signal-like sequences for nuclear import mediated by major vault protein. *Cancer Res* 2005; **65**(10)**:** 4108-4116.

15. Ryu SJ, An HJ, Oh YS, Choi HR, Ha MK, Park SC. On the role of major vault protein in the resistance of senescent human diploid fibroblasts to apoptosis. *Cell Death Differ* 2008; **15**(11)**:** 1673-1680.

16. Lloret M, Lara PC, Bordon E, Fontes F, Rey A, Pinar B *et al*. Major vault protein may affect nonhomologous end-joining repair and apoptosis through Ku70/80 and bax downregulation in cervical carcinoma tumors. *Int J Radiat Oncol Biol Phys* 2009; **73**(4)**:** 976-979.

17. Thomas C, Tulin AV. Poly-ADP-ribose polymerase: machinery for nuclear processes. *Mol Aspects Med* 2013; **34**(6)**:** 1124-1137.

18. Schultz N, Lopez E, Saleh-Gohari N, Helleday T. Poly(ADP-ribose) polymerase (PARP-1) has a controlling role in homologous recombination. *Nucleic Acids Res* 2003; **31**(17)**:** 4959-4964.

19. Wang M, Wu W, Rosidi B, Zhang L, Wang H, Iliakis G. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res* 2006; **34**(21)**:** 6170-6182.

20. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M *et al*. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* 2009; **361**(2)**:** 123-134.

21. Zhang J. Poly (ADP-ribose) polymerase inhibitor: an evolving paradigm in the treatment of prostate cancer. *Asian J Androl* 2014; **16**(3)**:** 401-406.

22. Ferrer M, Suarez JF, Guedea F, Fernandez P, Macias V, Marino A *et al*. Health-related quality of life 2 years after treatment with radical prostatectomy, prostate brachytherapy, or external beam radiotherapy in patients with clinically localized prostate cancer. *Int J Radiat Oncol Biol Phys* 2008; **72**(2)**:** 421-432.

23. Henriquez-Hernandez LA, Valenciano A, Foro-Arnalot P, Alvarez-Cubero MJ, Cozar JM, Suarez-Novo JF *et al*. Intraethnic variation in steroid-5-alpha-reductase polymorphismsin prostate cancer patients: a potential factor implicated in 5-alpha-reductase inhibitor treatment. *J Genet* 2015; **94**(2)**:** 335-341.

24. Henriquez-Hernandez LA, Valenciano A, Foro-Arnalot P, Alvarez-Cubero MJ, Cozar JM, Suarez-Novo JF *et al*. Polymorphisms in DNA-repair genes in a cohort of prostate cancer patients from different areas in Spain: heterogeneity between populations as a confounding factor in association studies. *PLoS One* 2013; **8**(7)**:** e69735.

25. Epstein JI, Allsbrook WC, Jr., Amin MB, Egevad LL. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol* 2005; **29**(9)**:** 1228-1242.

26. D'Amico AV, Whittington R, Malkowicz SB, Wu YH, Chen MH, Hurwitz M *et al*. Utilizing predictions of early prostate-specific antigen failure to optimize patient selection for adjuvant systemic therapy trials. *J Clin Oncol* 2000; **18**(18)**:** 3240-3246.

27. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**(2)**:** 263-265.

28. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B *et al*. The structure of haplotype blocks in the human genome. *Science* 2002; **296**(5576)**:** 2225-2229.

29. Henriquez-Hernandez LA, Valenciano A, Herrera-Ramos E, Lloret M, Riveros-Perez A, Lara PC. High-throughput genotyping system as a robust and useful tool in oncology: experience from a single institution. *Biologicals* 2013; **41**(6)**:** 424-429.

30. Morcillo-Suarez C, Alegre J, Sangros R, Gazave E, de Cid R, Milne R *et al*. SNP analysis to results (SNPator): a web-based environment oriented to statistical genomics analyses upon SNP data. *Bioinformatics* 2008; **24**(14)**:** 1643-1644.

31. Clayton D. Population Association. *Handbook of statistical genetics* 2001**:** 519-540.

32. Rando JC, Cabrera VM, Larruga JM, Hernandez M, Gonzalez AM, Pinto F *et al*. Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* 1999; **63**(Pt 5)**:** 413-428.

33. Calafell F, Bertranpetit J. Principal component analysis of gene frequencies and the origin of Basques. *Am J Phys Anthropol* 1994; **93**(2)**:** 201-215.

34. Gomes BC, Silva SN, Azevedo AP, Manita I, Gil OM, Ferreira TC *et al*. The role of common variants of non-homologous end-joining repair genes XRCC4, LIG4 and Ku80 in thyroid cancer risk. *Oncol Rep* 2010; **24**(4)**:** 1079-1085.

35. Rivera-Calzada A, Spagnolo L, Pearl LH, Llorca O. Structural model of full-length human Ku70-Ku80 heterodimer and its recognition of DNA and DNA-PKcs. *EMBO Rep* 2007; **8**(1)**:** 56-62.

36. Jia J, Ren J, Yan D, Xiao L, Sun R. Association Between the XRCC6 Polymorphisms and Cancer Risks: A Systematic Review and Meta-analysis. *Medicine (Baltimore)* 2015; **94**(1)**:** e283.

37. Xu H, Zou P, Chen P, Zhao L, Zhao P, Lu A. Association between the XRCC6 Promoter rs2267437 polymorphism and cancer risk: evidence based on the current literature. *Genet Test Mol Biomarkers* 2013; **17**(8)**:** 607-614.

38. Hayden PJ, Tewari P, Morris DW, Staines A, Crowley D, Nieters A *et al*. Variation in DNA repair genes XRCC3, XRCC4, XRCC5 and susceptibility to myeloma. *Hum Mol Genet* 2007; **16**(24)**:** 3117-3127.

39. Cibeira MT, de Larrea CF, Navarro A, Diaz T, Fuster D, Tovar N *et al*. Impact on response and survival of DNA repair single nucleotide polymorphisms in relapsed or

refractory multiple myeloma patients treated with thalidomide. *Leuk Res* 2011; **35**(9)**:** 1178-1183.

40. Adel Fahmideh M, Schwartzbaum J, Frumento P, Feychting M. Association between DNA repair gene polymorphisms and risk of glioma: a systematic review and meta-analysis. *Neuro Oncol* 2014; **16**(6)**:** 807-814.

41. Hu Y, Zhou M, Li K, Zhang K, Kong X, Zheng Y *et al*. Two DNA repair gene polymorphisms on the risk of gastrointestinal cancers: a meta-analysis. *Tumour Biol* 2014; **35**(3)**:** 1715-1725.

42. Balan S, Radhab SK, Sathyan S, Vijai J, Banerjee M, Radhakrishnan K. Major vault protein (MVP) gene polymorphisms and drug resistance in mesial temporal lobe epilepsy with hippocampal sclerosis. *Gene* 2013; **526**(2)**:** 449-453.

43. Henriquez-Hernandez LA, Moreno M, Rey A, Lloret M, Lara PC. MVP expression in the prediction of clinical outcome of locally advanced oral squamous cell carcinoma patients treated with radiotherapy. *Radiat Oncol* 2012; **7:** 147.

44. Henriquez-Hernandez LA, Lloret M, Pinar B, Bordon E, Rey A, Lubrano A *et al*. BCL-2, in combination with MVP and IGF-1R expression, improves prediction of clinical outcome in complete response cervical carcinoma patients treated by radiochemotherapy. *Gynecol Oncol* 2011; **122**(3)**:** 585-589.

45. Lloret M, Lara PC, Bordon E, Rey A, Falcon O, Apolinario RM *et al*. MVP expression is related to IGF1-R in cervical carcinoma patients treated by radiochemotherapy. *Gynecol Oncol* 2008; **110**(3)**:** 304-307.

46. Jensen CJ, Oldfield BJ, Rubio JP. Splicing, cis genetic variation and disease. *Biochem Soc Trans* 2009; **37**(Pt 6)**:** 1311-1315.

47. Shastry BS. SNPs: impact on gene function and phenotype. *Methods Mol Biol* 2009; **578:** 3-22.

48. Lee PH, Shatkay H. Ranking single nucleotide polymorphisms by potential deleterious effects. *AMIA Annu Symp Proc* 2008**:** 667-671.

49. Lee PH, Shatkay H. F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res* 2008; **36**(Database issue)**:** D820-824.

FIGURE LEGENDS

Figure 1. Forrest plot of odds ratios (OR) with 95% confidence interval (CI) for polymorphisms rs2267437 (*XRCC6*), rs3815824 (*MVP*), rs1051685 (*XRCC5*), and rs3219123 (*PARP1*) analyzed in combination and D'Amico risk recurrence group. Each diamond represents the OR and the horizontal line indicates the 95%CI. For the binary logistic regression, patients were dichotomized in two groups as follows: low vs. intermediate – high D'Amico groups.

Supplementary material

S1 Table. Description of SNPs included in the study and analyzed by OpenArray.

S2 Table. Genotype and allelic frequencies of gene polymorphisms among populations. Statistical differences among genotypes are shown.

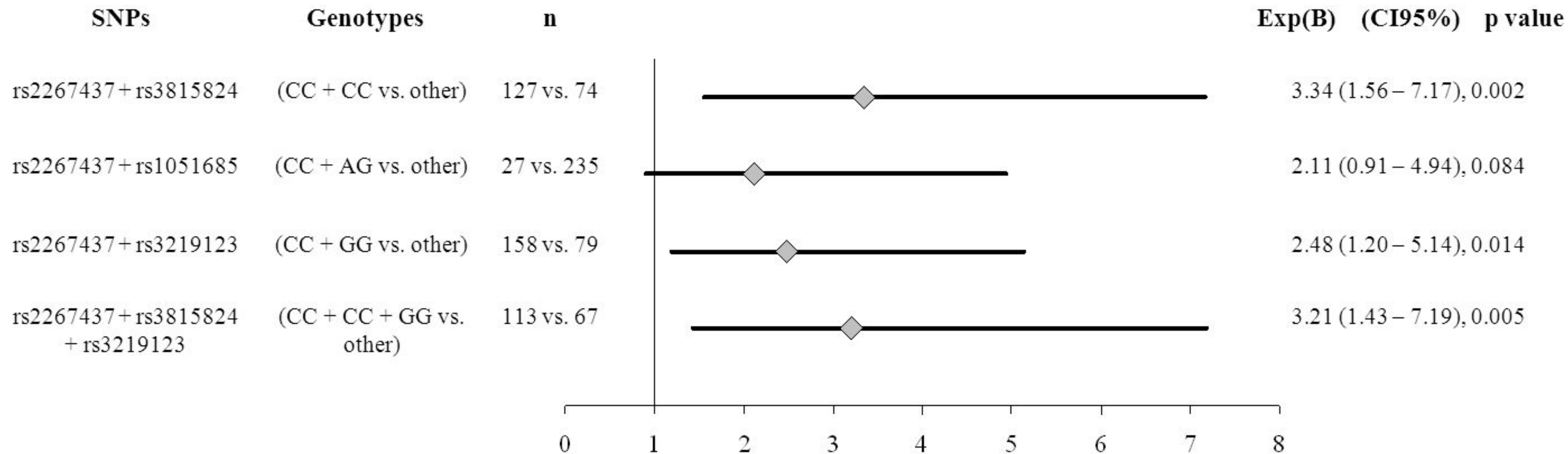
SNPs
Genotypes
n
Exp(B) (CI95%) p value
rs2267437 + rs3815824
(CC + CC vs. other)
127 vs. 74
3.34 (1.56 – 7.17), 0.002
rs2267437 + rs1051685
(CC + AG vs. other)
27 vs. 235
2.11 (0.91 – 4.94), 0.084
rs2267437 + rs3219123
(CC + GG vs. other)
158 vs. 79
2.48 (1.20 – 5.14), 0.014
rs2267437 + rs3815824 + rs3219123
(CC + CC + GG vs. other)
113 vs. 67
3.21 (1.43 – 7.19), 0.005
0
1
2
3
4
5
6
7
8

Table 1. Description of clinical variables (n = 494)

Variable	n	(%)
Clinical tumor size (cT)		
cT1a – cT2a	270	(54.7)
cT2b – cT2c	141	(25.8)
cT3 – cT4	66	(13.4)
NA	17	(3.4)
Initial PSA (ng/mL)		
< 10	306	(61.9)
10 – 19.99	103	(20.9)
> 20	79	(16.0)
NA	6	(1.2)
Gleason score		
< 7	226	(45.7)
7	195	(39.5)
> 7	71	(14.4)
NA	2	(0.4)
D'Amico risk group		
Low	120	(24.3)
Intermediate	184	(37.2)
High	173	(35.0)
NA	17	(3.4)

Abbreviations: PSA, prostate specific antigen; NA, not available.

Table 2. Significant associations between clinical variables and SNPs. The distribution of patients is expressed in absolute numbers and percentages (in brackets)

SNP	Genotypes	Clinical tumor size (cT)			P#
		cT1a – cT2a	cT2b – cT2c	cT3 – cT4	
rs2267437 (*XRCC6*)	CC	105 (24.3)	41 (9.5)	29 (6.7)	0.004
	CG	107 (24.8)	45 (10.4)	20 (4.6)	
	GG	37 (8.6)	38 (8.8)	10 (2.3)	
		Diagnostic PSA (ng/mL)			
		<10	10 – 19.99	≥20	
rs1051685 (*XRCC5*)	AA	253 (52.4)	73 (15.1)	71 (14.7)	0.007
	AG	46 (9.5)	25 (5.2)	6 (1.2)	
	GG	3 (0.6)	4 (0.8)	2 (0.4)	
rs3219123 (*PARP1*)	AA	1 (0.2)	0 (0.0)	1 (0.2)	0.041
	AG	29 (6.0)	18 (3.7)	4 (0.8)	
	GG	274 (56.6)	84 (17.4)	73 (15.1)	
rs3815824 (*MVP*)	TT	4 (0.8)	2 (0.4)	5 (1.0)	0.025
	TC	61 (12.7)	29 (6.0)	18 (3.8)	
	CC	238 (49.6)	69 (14.4)	54 (11.3)	
		Gleason score			
		<7	7	>7	
rs2267437 (*XRCC6*)	CC	91 (20.4)	58 (13.0)	31 (7.0)	0.037
	CG	82 (18.4)	78 (17.5)	19 (4.3)	
	GG	31 (7.0)	41 (9.2)	14 (3.1)	
		D'Amico risk group			
		Low	Intermediate	High	
rs2267437 (*XRCC6*)	CC	52 (12.0)	58 (13.4)	64 (14.8)	0.015
	CG	46 (10.6)	76 (17.6)	52 (12.0)	
	GG	12 (2.8)	34 (7.9)	39 (9.2)	

Chi square test.

Table 3. Univariate analysis for polymorphism rs2267437 (*XRCC6*) and clinical variables

		Recessive model		**Dominant model**		**Homozygote model**		**Heterozygote model**	
		GG vs. CG + CC		CG + GG vs. CC		CC vs. GG		CG vs. CC	
Clinical variable	**n**	OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P
cT, 1a-2a/2b-4	270/207	2.04 (1.26 – 3.29)	0.004	1.18 (0.80 – 1.74)	0.413	1.95 (1.15 – 3.29)	0.013	0.91 (0.59 – 1.40)	0.673
PSA, <20/≥20	409/79	2.12 (1.17 – 3.78)	0.011	1.30 (0.75 – 2.21)	0.355	2.07 (1.08 – 3.98)	0.029	0.96 (0.52 – 1.77)	0.892
Gleason, <7/ ≥ 7	226/266	1.65 (1.01 – 2.68)	0.044	1.37 (0.94 – 2.01)	0.100	1.81 (1.07 – 3.08)	0.027	1.21 (0.80 – 1.83)	0.368
Risk group, Low/Int-High	120/357	2.38 (1.24 – 4.58)	0.009	1.48 (0.95 – 2.29)	0.080	2.59 (1.30 – 5.18)	0.007	1.19 (0.74 – 1.89)	0.475

Abbreviations: OR, odds ratio; CI, confidence interval; Int, intermediate.
Statistical test: binary logistic regression (Reference category: CC).