Gene variants in *MSH3* and *BLM* may influence myelodysplastic syndrome susceptibility and prognosis, respectively, in a Portuguese population group

Mafalda Melo[1], Ana Cristina Gonçalves[2,3,4], Luisa Mota-Vieira[5,6,7,*], Ana Bela Sarmento-Ribeiro[2,3,4,8,*]

[1]Faculty of Medicine, University of Coimbra (FMUC), 3000-548 Coimbra, Portugal

[2]Laboratory of Oncobiology and Hematology (LOH) and University Clinic of Hematology, FMUC, 3000-548 Coimbra, Portugal

[3]Center of Investigation in Environment, Genetics and Oncobiology (CIMAGO), FMUC, 3000-548 Coimbra, Portugal

[4]Center for Neuroscience and Cell Biology and Institute for Biomedical Imaging and Life Sciences (CNC.IBILI), 3004-504 Coimbra, Portugal

[5]Molecular Genetics and Pathology Unit, Hospital of Divino Espírito Santo of Ponta Delgada, EPER, 9500-370 Ponta Delgada, São Miguel Island, Azores, Portugal

[6]Azores Genetics Research Group, Instituto Gulbenkian de Ciência, 2780-156 Oeiras, Portugal

[7]BioISI – Byosystems & Integrative Sciences Institute, University of Lisboa, Faculty of Sciences, 1749-016 Lisbon, Portugal

[8]Clinical Hematology Department, Centro Hospitalar Universitário de Coimbra, EPE (CHUC, EPE), 3000-548 Coimbra, Portugal

*Both authors contributed equally to this work.

TABLE OF CONTENTS

TABLE INDEX

FIGURE INDEX

ABBREVIATIONS

5q-	5q- syndrome
AML	acute myeloid leukemia
BER	base excision repair
BLM	Bloom syndrome, RecQ helicase-like
CI	confidence interval
CMML	chronic myelomonocytic leukemia
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
ERCC2	excision repair cross-complementation group 2
GP	genetic profile
HR	hazard ratio
HRR	homologous recombination repair
HSC	hematopoietic stem cell
HWE	Hardy-Weinberg equilibrium
int-1	intermediate-1
int-2	intermediate-2
IPSS	International Prognostic Score System
LIG1	ligase I, DNA, ATP-dependent
MAF	minor allele frequency
MDS	myelodysplastic syndrome
MDS-U	unclassifiable myelodysplastic syndrome
MMR	mismatch repair
MSH3	mutS homolog 3
NHEJ	non-homologous end-joining
OGG1	8-oxoguanine DNA glycosylase
OR	odds ratio
OS	overall survival
PCR	polymerase chain reaction
RAD52	RAD52 homolog, DNA repair protein
RAEB-1	refractory anemia with excess blasts type 1
RAEB-2	refractory anemia with excess blasts type 2
RARS	refractory anemia with ring sideroblasts
RCMD	refractory cytopenia with multilineage dysplasia
RCUD	refractory cytopenia with unilineage dysplasia
RMI1	RecQ mediated genome instability 1
SNP	single nucleotide polymorphism
TOP3A	topoisomerase (DNA) III alpha
WHO	World Health Organization
XRCC3	X-ray repair complementing defective repair in Chinese hamster cells 3
XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5

ABSTRACT

Background: Myelodysplastic syndrome (MDS) is a heterogeneous group of hematopoietic stem cell disorders, characterized by peripheral cytopenias, ineffective hematopoiesis and frequent transformation into acute myeloid leukemia (AML). Several mechanisms are involved in disease development and progression as a consequence of stepwise accumulation of DNA mutations, which infers a defect in DNA repair mechanisms. Variants in genes involved in these mechanisms have been identified for their potential role in cancer susceptibility. However, in MDS, the relevance of these variants remains to be fully established and correlated with prognosis.

Methods: We performed a hospital-based case control-study to investigate the association of DNA repair genes with MDS susceptibility and prognosis in a group of Portuguese patients. To that end, we genotyped by TaqMan® real-time PCR 10 SNPs (one *per* gene: *XRCC5*, *RMI1*, *RAD52*, *XRCC3*, *BLM*, *TOP3A*, *OGG1*, *LIG1*, *ERCC2*, and *MSH3*) in 60 MDS patients and 120 age-sex matched controls. Frequencies of alleles, genotypes, and genotypic profiles were estimated and compared between patients and controls. The role of these genes in MDS susceptibility was studied by logistic regression analysis. The influence in MDS prognosis was evaluated by estimating, through Kaplan-Meier analysis, the rate of MDS transformation into AML and the overall survival.

Results: There was no significant difference in frequencies of *XRCC5*, *RMI1*, *RAD52*, *XRCC3*, *BLM*, *TOP3A*, *OGG1*, *LIG1* and *ERCC2* variants between patients and controls. In contrary, we found that heterozygous individuals for *MSH3* c.2655+5137C>G had an increased susceptibility to MDS development (OR = 6.882, 95% CI 1.789-26.479, $p < 0.003$), being the increased risk attributed to G allele

(OR = 6.405, 95% CI 1.552-30.469, $p < 0.003$). In addition, homozygous for *BLM* c.-4-889A>C showed higher rate of MDS transformation into AML (HR = 7.646, 95% CI 1.362-24467, $p < 0.023$).

Conclusion: The study suggests that *MSH3* c.2655+5137C>G variant influences MDS susceptibility, and *BLM* c.-4-889A>C variant may be implicated in the propensity to AML transformation observed in MDS patients. Thus, these variants could be used as a risk and prognostic biomarkers, in MDS, if these associations were replicated in a larger case-control study and/or with other populations.

KEYWORDS: Myelodysplastic syndrome, DNA repair, gene variants, susceptibility, prognosis

RESUMO

Introdução: A síndrome mielodisplásica (SMD) engloba um grupo heterogéneo de doenças clonais da célula estaminal hematopoiética, caracterizado por citopenias no sangue periférico, hematopoiese ineficaz e transformação frequente em leucemia mielóide aguda (LMA). Diversos mecanismos estão envolvidos no desenvolvimento e progressão da doença como a acumulação de mutações do DNA consecutivas, o que infere a existência de defeitos nos mecanismos de reparação do DNA. Variantes nos genes envolvidos nestes mecanismos têm sido identificadas pelo seu papel potencial na suscetibilidade para o cancro. Contudo, na SMD, a relevância destas variantes precisa ser mais bem estudada e correlacionada com o prognóstico.

Métodos: Realizou-se um estudo de caso-controlo para investigar a associação de genes de reparação do DNA com a suscetibilidade e o prognóstico da SMD num grupo de doentes portugueses. Para este fim, genotipou-se por PCR em tempo real 10 SNPs (um por gene: *XRCC5*, *RMI1*, *RAD52*, *XRCC3*, *BLM*, *TOP3A*, *OGG1*, *LIG1*, *ERCC2* e *MSH3*) em 60 doentes e 120 controlos emparelhados por idade e sexo. As frequências dos alelos, genótipos e perfis genéticos foram calculadas e comparadas entre doentes e controlos. O papel destes genes na suscetibilidade para MDS foi estudada por análise de regressão logística. A influência no prognóstico foi avaliada estimando-se a taxa de transformação em LMA e a sobrevivência através de curvas de Kaplan-Meier.

Resultados: Não se observou diferença significativa nas frequências das variantes *XRCC5*, *RMI1*, *RAD52*, *XRCC3*, *BLM*, *TOP3A*, *OGG1*, *LIG1* e *ERCC2* entre doentes e controlos. Pelo contrário, indivíduos heterozigóticos para *MSH3* c.2655+5137C>G apresentaram uma maior suscetibilidade para SMD (OR = 6.882, 95% IC 1.789-26.479, $p < 0.003$), sendo esse risco atribuído ao alelo G (OR = 6.405, 95% IC 1.552-30.469,

$p < 0.003$). Além disso, os homozigóticos para *BLM* c.-4-889A>C apresentaram maior taxa de progressão para LMA (HR = 7.646, 95% IC 1.362-24467, $p < 0.023$).

Conclusão: O presente estudo sugere que a variante *MSH3* c.2655+5137C>G influencia a suscetibilidade para SMD, e que a variante *BLM* c.-4-889A>C poderá estar implicada na propensão para a transformação em LMA observada nos doentes com SMD. Assim, estas variantes genéticas poderão vir a ser usadas como marcadores de risco e de prognóstico, em SMD, se as associações forem replicadas em estudos de caso-controlo mais alargados e/ou de preferência com populações de outro fundo genético.

PALAVRAS-CHAVE: Síndrome mielodisplásica, reparação de DNA, variantes genéticas, suscetibilidade, prognóstico

INTRODUCTION

Myelodysplastic syndrome (MDS) is a heterogeneous group of hematopoietic stem cell (HSC) disorders, characterized by ineffective hematopoiesis and frequent disease transformation into acute myeloid leukemia (AML).[1,2] This syndrome is most commonly found in the elderly population, *i.e.*, about 80% cases are diagnosed after 60 years old.[3,4] The 2008 edition of the World Health Organization (WHO) monograph classifies MDS according to (1) presence of citopenias (2) percentage of bone marrow and peripheral blood blasts, (3) type and number of dysplastic cell lineages, (4) presence or absence of ring sideroblasts, and (5) presence or absence of specific chromosomal abnormalities. These criteria allow the discrimination of MDS in the following seven subtypes: isolated -5q syndrome (5q-), refractory anemia with excess blasts type 1 (RAEB-1), refractory anemia with excess blasts type 2 (RAEB-2), refractory anemia with ringed sideroblasts (RARS), refractory cytopenia with multilineage dysplasia (RCMD), refractory cytopenia with unilineage dysplasia (RCUD), and unclassifiable MDS (MDS-U).[5] The International Prognostic Scoring System (IPSS), the most widely used prognostic classification system, stratifies patients into four risk groups – low, intermediate-1 (int-1), intermediate-2 (int-2), and high risk –, which allow the management of treatment MDS patients.

The etiology of MDS is multifactorial and complex. These syndromes can be induced by environmental and occupational toxins, such as benzene and its derivates (*de novo*/ primary MDS), or by large genotoxic insults, such as chemotherapy and radiotherapy treatments for primary cancer (therapy-related/ secondary MDS);[6,7,8] however, in most cases the etiology remains unknown. The stepwise disease progression is a consequence of the accumulation of genomic alterations in HSC that leads to deregulated

differentiation, proliferation, survival and apoptosis, causing clonal expansion of aberrant cells.[9,10,11]

DNA repair mechanisms ensure genomic integrity of HSC, by preventing the occurrence of genetic and epigenetic abnormalities.[12] These mechanisms include the (1) homologous recombination repair (HRR), (2) non-homologous end-joining (NHEJ), (3) mismatch repair (MMR), (4) base excision repair (BER), and (5) nucleotide excision repair (NER). The first three mechanisms – HRR, NHEJ and MMR – deals with mistakes made during DNA replication, while the last two – BER and NER – deals with DNA damage that arise from ionizing radiation or oxidative stress.[13,14] Specifically, HRR and NHEJ act in DNA-strand breaks, the most important damage in HSC;[15,16] MMR removes mispaired nucleotides; BER targets a single damaged DNA base, which is removed by a DNA glycosylase-type enzyme; and, finally, NER corrects bulky helix-distorting lesions. It is also known that inefficacy of these mechanisms in cells surviving toxic stress is crucial for cancer development.[17]

Single nucleotide polymorphisms (SNPs) are the most prevalent class of inherited genetic variation and thought to be responsible for most of individual genetic variability. Functionally, they may influence the genetic susceptibility (or protection) to different diseases, including hematologic malignancies, and the sensitivity (or resistance) to therapeutics. Therefore, the study of specific gene variants involved in biological pathways may contribute (1) to clarify the molecular basis of diseases, (2) to identify predisposed individuals, and (3) to find molecular markers of prognosis and/or therapeutic targets.[18]

The impact of genetic background on MDS is only beginning to be elucidated. In the last few years, DNA repair genes have been surmised as candidate genes to cancer

susceptibility. The present study investigates this hypothesis in a group of MDS Portuguese patients, through a hospital-based case-control design. We also analyzed the association of 10 variants in 10 DNA repair genes with MDS prognosis, namely: the rate of transformation into AML and the overall survival.

MATERIALS AND METHODS

Ethics statement

The present investigation follows the guidelines stablished on the Helsinki Declaration and the Oviedo Convention; the project was approved by the Ethics Committee of Faculty of Medicine of University of Coimbra (Coimbra, Portugal). The study design includes informed consent, confidentiality, anonymity of personal data, and abandonment option in case of expressed will.

Study design and population

We conducted a hospital-based case-control study comparing 60 MDS patients and 120 controls. All enrolled participants (n = 180) were recruited from the Department of Clinical Hematology of Centro Hospitalar e Universitário de Coimbra, EPE (CHUC,EPE) and Hospital Distrital da Figueira da Foz, EPE (HDFF,EPE) from 2010 to 2016. Patients were diagnosed according to the 2008 WHO classification, and stratified into the IPSS risk categories. Overall survival (OS) and AML transformation were selected as study endpoints. The OS was measured from diagnosis date (patients who were still alive at the date last contact were censored). The AML transformation was measured from MDS diagnosis date to time of AML transformation, defined according to the 2008 WHO classification. Control group included non-neoplastic individuals matched for gender and age (± 5 years) recruited from these two Hospitals. Basic demographic characteristics of patients and controls are shown in Table 1.

Table 1. Demographic and clinical characteristics of MDS patients and controls.

Characteristics	Patients	Controls
Demographic features	*n* = 60 (%)	*n* = 120 (%)
Gender		
Male	28 (46.7)	55 (45.8)
Female	32 (53.3)	65 (54.2)
Age (years); Median (range)	74 (53 – 89)	74 (46 – 90)
Clinical features		
MDS type*	*n* = 60 (%)	
5q-	2 (3.3)	
RAEB-1	5 (8.3)	
RAEB-2	1 (1.7)	
RARS	5 (8.3)	
RCMD	31 (51.7)	
RCUD	6 (10.0)	
CMML	10 (16.7)	
IPSS risk groups	*n* = 48 (%)	
Low	24 (50.0)	
Int-1	18 (37.5)	
Int-2	6 (12.5)	

*Subtypes, according to WHO classification; MDS, myelodysplastic syndrome; WHO, World Health Organization; WBC, white blood cells; Hb, hemoglobin; 5q-, 5q syndrome; RAEB-1, refractory anemia with excess blasts type 1; RAEB-2, refractory anemia with excess blasts type 2; RARS, refractory anemia with ring sideroblasts; RCMD, refractory cytopenia with multilineage dysplasia; RCUD, refractory cytopenia with unilineage dysplasia; CMML, chronic myelomonocytic leukemia; IPSS, international prognostic scoring system; Int-1, intermediate-1; Int-2, intermediate-2.

Selection of genes and SNPs

Studied genes were selected due to their well-defined biological function and involvement in DNA repair pathways. SNPs were chosen according to their reported genetic association on MDS susceptibility, as well as their validation in other case-control studies. These information was available in Pubmed[1] and dbSNP[2] databases. In total, we selected 10 SNPs among 10 candidate genes (one SNP *per* gene: *XRCC5*, *RMI1*, *RAD52*, *XRCC3*, *BLM*, *TOP3A*, *OGG1*, *LIG1*, *ERCC2*, and *MSH3*). The characteristics of all SNPs are shown in Table 2.

[1] Pubmed: http://www.ncbi.nlm.nih.gov/pubmed
[2] dbSNP: http://www.ncbi.nlm.nih.gov/SNP/

Table 2. Relevant information of SNPs showing significant associations with myelodysplastic syndromes.

DNA repair pathways	Gene symbol*	dbSNP	Chr. position (GRCh38.p2)	Variant	Molecular consequence†	SNP functional effect			MAF‡		Reported association			
						FS§	Category	Pathogenicity	Global	IBS	Allele/ genotype	OR	(95% CI)	Ref.
Non-homologous end joining	*XRCC5*	rs3835	2:216201914	c.2110-2408G>A	IVS	0.065	Trans reg	Not changed	0.2061 (A)	0.1262 (A)	AG	0.10	(0.03-0.29)	Ribeiro, 2014
Homologous recombination repair	*RMI1*	rs1982151	9:84002350	c.1364A>G	Asn455Ser	0.902	Prot coding Splicing reg Post trans	Possibly damaging Changed Exist	0.3289 (A)	0.2710 (A)	GA+AA	1.90	(1.10-3.30)	Broberg, 2007
	RAD52	rs11226	12:912647	c.744C>T	3'-UTR	0.5	Trans reg	Changed	0.4301 (T)	0.4439 (T)	TC	1.80	(1.31-2.48)	Belickova, 2013
	XRCC3	rs861539	14:103699416	c.722C>T	Thr241Met	0.5	Prot coding Splicing reg Post trans	Benign Changed Exist	0.2169 (T)	0.3972 (T)	CC	0.094	(0.012-0.73)	Aktuglu, 2014
	BLM	rs6496724	15:90746500	c.-4-889A>C	IVS	0.208	Trans reg	Changed	0.3704 (C)	0.2757 (C)	CC	0.34	(0.12-0.95)	Broberg, 2009
	TOP3A	rs12945597	17:18271367	g.18174681G>A	Intergenic	–	–	–	0.2268 (A)	0.3084 (A)	AA	4.90	(1.70-14)	Broberg, 2009
Base excision repair	*OGG1*	rs1052133	3:9757089	c.977C>G	Ser326Cys	0.294	Prot coding Splicing reg Trans reg Post trans	Benign Changed Exist Exist	0.3021 (G)	0.1916 (G)	CG + GG	2.66	(1.17-6.01)	Aktuglu, 2014
	LIG1	rs20580	19:48151296	c.510A>C	Ala170=	0.5	Prot coding Splicing reg	Synonymous Changed	0.4671 (C)	0.4766 (A)	AC	2.03	(1.50-2.75)	Belickova, 2013
Nucleotide excision repair	*ERCC2 (XPD)*	rs13181	19:45351661	c.2251A>C	Lys751Gln	0.749	Prot coding Splicing reg Post trans Conserved	Benign Changed Exist Conserved	0.2366 (C)	0.3131 (C)	CC	4.07	(1.77-9.39)	Aktuglu, 2014
Mismatch repair	*MSH3*	rs3797896	5:80797981	c.2655+5137C>G	IVS	0.242	Trans reg	Changed	0.1575 (G)	0.0514 (G)	GC	0.20	(0.07-0.54)	Belickova, 2013

*According to HUGO Gene Nomenclature Committee (HGNC). †According to Sequence Ontology. §According to F-SNP. The F-SNP functional score (FS) incorporates functional effects of SNPs predicted at splicing, transcriptional, translational, and post-translational level. ‡MAF source: 1000 Genomes. OR, odds ratio; CI, confidence interval; Ref, references. ᵉThe functional effect from protein coding category was predicted using PredictSNP, and from transcriptional and splicing regulation categories were predicted by F-SNP. Trans reg, transcriptional regulation; Prot coding, protein coding; Splicing reg, splicing regulation; Post trans, Post translational modification. *XRCC5*, X-ray repair complementing defective repair in Chinese hamster cells 5; *RMI1*, RecQ mediated genome instability 1; *RAD52*, RAD52 homolog, DNA repair protein; *XRCC3*, X-ray repair complementing defective repair in Chinese hamster cells 3; *BLM*, Bloom syndrome, RecQ helicase-like; *TOP3A*, topoisomerase (DNA) III alpha; *OGG1*, 8-oxoguanine DNA glycosylase; *LIG1*, ligase I, DNA, ATP-dependent; *ERCC2*, excision repair cross-complementation group 2; *MSH3*, mutS homolog 3.

DNA extraction and SNP genotyping

Peripheral blood samples from patients at diagnosis as well from patients were collected by venipuncture into EDTA tubes. Genomic DNA was extracted following Barlett and Whites's protocol (Barlett *et al.*, 2003), and quantified using a NanoDrop ND 1000 spectrophotometer (NanoDrop Technologies, USA).

All SNPs were genotyped by real-time PCR using TaqMan® Pre-Designed SNP Genotyping Assays on an ABI 7500 Fast Real Time PCR System, according to manufacturer's instructions (Applied Biosystems, USA). The thermal cycle conditions were as follows: pre-PCR holding at 60ºC for 1 min, followed by enzyme activation at 95ºC for 10 min, then two-stage polymerase run of 40 cycles at 95ºC for 15s, and 60ºC for 1 min. Two reviewers independently scored all genotypes.

Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics (version 23), GraphPad Prism, Arlequin (version 3.5.2), and Interactive Statistical Pages[3]. Normality and differences of confounding variables (age and gender) between groups (MDS patients *versus* controls) were assessed by the Kolmogorov Smirnov test and nonparametric Mann Whitney U test, respectively. Allele and genotype frequencies were determined by direct counting. Deviation of the genotype proportions from Hardy-Weinberg equilibrium (HWE) was assessed in patients and controls. Genotypic profile frequencies were inferred using the maximum likelihood method (expectation maximization algorithm). Differences in frequencies of alleles, genotypes and genetic profiles between patients and controls were tested using Pearson's chi-square test. To test the hypothesis of association between variants and MDS susceptibility, we used methods based on

[3] Interactive Statistical Pages*:* http://statpages.org/

logistic regression analysis. Odds ratios (OR) and 95% confidence intervals (CI) were calculated for each genotype compared with the homozygous for the major allele, which were set as the reference genotype. Analyses were performed under codominant, dominant, and recessive inheritance models. The influence of variants in overall survival and MDS transformation into AML was analyzed by Kaplan Meier curves, and verified by log-rank tests. The hazard ratio (HR) and its 95% CI were calculated using the Cox proportional hazard model. All statistical analyses were two sided, and a *p*-value < 0.05 was considered statistically significant.

RESULTS

Characterization of MDS patients and controls

The present study included 60 MDS patients, with a median age of 74 years (53 to 89 years), being 32 female (53.3%) and 28 male (46.7%), as well as 80 non-neoplastic controls with a median age of 74 (46 to 90 years), being 65 females (54.2%) and 55 males (45.8%). In order to avoid confounding bias and to confirm adequate matching between MDS and controls, we assessed differences in the demographic features. There were no significant differences concerning age or gender. These results indicated adequate group matching.

Patients were grouped according to the 2008 WHO classification into the following seven subtypes: myelodysplastic syndrome with isolated del(5q) (5q-: n = 2; 3.3%), refractory anemia with excess blasts type I (RAEB-1: n = 5; 8.3%), refractory anemia with excess blasts type 2 (RAEB-2: n = 1; 1.7%), refractory anemia with ringed sideroblasts (RARS: n = 5; 8.3%), refractory cytopenia with multilineage dysplasia (RCMD: n = 31; 51.7%), refractory cytopenia with unilineage dysplasia (RCUD: n = 6; 10.0%), and chronic myelomonocytic leukemia (CMML: n = 10; 16.7%). In addition, patients were stratified according to the IPSS into the following risk groups: low-risk (low: n = 24; 50.0%), intermediate-I risk (int-1: n = 18; 37.5%), and intermediate-2 risk (int-2: n = 6; 12.5%).

Frequencies of alleles, genotypes and genotypic profiles, and their association with MDS susceptibility

In order to evaluate the contribution of DNA repair gene variants to MDS susceptibility, first, we calculated the allele and genotype frequencies of selected SNPs and, second, we estimated the OR by logistic regression.

The allele frequencies of the 10 selected SNPs (one *per* gene) in MDS patients as well in controls are shown in Table 3. All SNPs were in HWE ($p < 0.05$) in patients or controls, with two exceptions: variants in *RAD52* and *XRCC3* genes. For this reason, these two variants were not considered in further analyses. Regarding the association analysis, we found that allele *MSH3* G increases the predisposition to MDS (OR = 6.405, 95% CI 1.552-30.469, $p = 0.003$). None of other studied variants showed significant association.

Table 3. Allele frequencies of selected SNPs in MDS patients and controls, and its association with risk of myelodysplastic syndrome.

Gene: dbSNP	**Minor allele**[†]	**MAF**		**Association analysis**		
		Patients	Controls	Patients *vs* controls		
				OR	(95% CI)	*p*-value
XRCC5: rs3835	A	0.092	0.092	1.000	(0.437-2.257)	1.000
RMI1: rs1982151	A	0.267	0.308	0.816	(0.486-1.367)	0.463
BLM: rs6496724	C	0.246	0.263	0.914	(0.532-1.568)	0.797
TOP3A: rs12945597	A	0.283	0.254	1.160	(0.688-1.952)	0.612
OGG1: rs1052133	G	0.233	0.225	1.048	(0.602-1.819)	0.894
LIG1: rs20580	A	0.412	0.500	0.701	(0.434-1.133)	0.136
ERCC2: rs13181	C	0.333	0.371	0.848	(0.521-1.380)	0.560
MSH3: rs3797896	**G**	0.075	0.013	**6.405**	(**1.552-30.469**)*	**0.003**

[†]Minor allele of controls and 1000 Genomes database (Iberian population in Spain). Bold indicates statistically significant association (*susceptibility). MDS, myelodysplastic syndrome. MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.

The genotype frequencies of selected SNPs were compared between MDS patients and controls. Results are shown in Table 4. We observed that *MSH3* CG heterozygous genotype increases MDS risk (codominant model: OR = 6.882, 95% CI 1.789-26.479, $p = 0.005$). On the other hand, there were no significant differences in genotypes of *XRCC5*, *RMI1*, *BLM*, *TOP3A*, *OGG1*, *LIG1*, and *ERCC2* between patients and controls.

Table 4. Genotype frequencies of selected SNPs in MDS patients and controls, and its association with risk of myelodysplastic syndrome.

Gene: dbSNP	**Genotype frequencies**				**Association analysis**		
	Patients		Controls		Patients *vs* controls		
	n	%	*n*	%	OR	(95% CI)	*p*-value
XRCC5: rs3835							
GG	50	83.3	99	82.5	Ref.		
GA	9	15.0	20	16.7	0.891	(0.378-2.099)	0.792
AA	1	1.7	1	8.0	1.980	(0.121-32.320)	0.632
Dominant model					0.943	(0.413-2.154)	0.889
Recessive model					2.017	(0.124-32.816)	0.622
RMI1: rs1982151							
GG	33	55.0	53	44.2	Ref.		
GA	22	33.7	60	50.0	0.589	(0.306-1.132)	0.112
AA	5	8.3	7	5.8	1.147	(0.336-3.914)	0.826
Dominant model					0.647	(0.347-1.207)	0.171
Recessive model					1.468	(0.446-4.834)	0.528
BLM: rs6496724							
AA	33	55.9	68	56.7	Ref.		
AC	23	39.0	42	35.0	1.156	(0.598-2.233)	0.666
CC	3	5.1	10	8.3	0.618	(0.159-2.398)	0.487
Dominant model					1.051	(0.560-1.971)	0.878
Recessive model					0.584	(0.154-2.208)	0.428
TOP3A: rs12945597							
GG	30	50.0	64	53.3	Ref.		
GA	26	43.3	51	42.5	1.088	(0.573-2.065)	0.797
AA	4	6.7	5	8.3	1.707	(0.427-6.814)	0.449
Dominant model					1.143	(0.615-2.125)	0.673
Recessive model					1.643	(0.425-6.356)	0.472
OGG1: rs1052133							
CC	37	61.7	70	58.3	Ref.		
CG	18	30.0	46	38.3	0.740	(0.377-1.454)	0.383
GG	5	8.3	4	3.3	2.365	(0.599-9.342)	0.219
Dominant model					0.870	(0.461-1.641)	0.668
Recessive model					2.636	(0.681-10.204)	0.160
LIG1: rs20580							
CC	20	35.1	37	31.1	Ref.		
CA	27	47.4	50	42.0	0.999	(0.487-2.048)	0.998
AA	10	17.5	32	26.9	0.578	(0.236-1.414)	0.230
Dominant model					0.835	(0.428-1.628)	0.596
Recessive model					0.578	(0.262-1.279)	0.177
ERCC2: rs13181							
AA	26	43.3	47	39.2	Ref.		
AC	28	46.7	57	47.5	0.888	(0.459-1.716)	0.724
CC	6	10.0	16	13.3	0.723	(0.250-2.089)	0.549
Dominant model					0.854	(0.455-1.601)	0.622
Recessive model					0.770	(0.283-2.099)	0.610
MSH3: rs3797896							
CC	51	85.0	117	97.5	Ref.		
CG	9	15.0	3	2.5	**6.882**	**(1.789-26.479)***	**0.005**
GG	0	0.0	0	0.0	–		–
Dominant model					**6.882**	**(1.789-26.479)***	**0.005**
Recessive model					–		–

The OR (95% CI) and *P*-value were calculated by logistic regression according the following genetic models: codominant model (MM *vs* MM, MM *vs* Mm, and MM *vs* mm, *i.e*. each genotype was compared with major allele homozygous genotype as reference); dominant model (MM *vs* Mm + mm); and recessive model (MM + Mm *vs* mm). Bold indicates statistically significant association (*susceptibility). M, major allele; m, minor allele; MDS, myelodysplastic syndrome; OR, odds ratio; CI, confidence interval; Ref., reference.

In order to assess the frequencies of multilocus genotypes, we performed a genotypic profile (GP) analysis. GPs were inferred using Arlequin software, and grouped in two different pathways: homologous recombination repair (*RMI1* + *BLM* + *TOP3A*) and base excision repair (*OGG1* + *LIG1*). Determination of the relative risk of disease development associated with different genetic profiles was performed, using Fisher's exact test. Results are presented in Table 5. In the homologous recombination repair pathway, we observed a total of 19 GPs, being one unique to the MDS group, and three unique to the control one. No statistically significant associations between GPs and MDS susceptibility were observed.

Table 5. Significant genotypic profiles frequencies of selected SNPs in MDS patients and controls, and its association with risk of myelodysplastic syndrome.

Pathway:	**Profile frequencies**		**Association analysis**		
Genotypic profile	Patients	Controls	Patients *vs* controls		
	%	%	OR	(95% CI)	*p*-value
Homologous recombination repair: *RMI1* + *BLM* + *TOP3A*					
GP1: AA AA GA	3.3	3.3	1.018	(0.125-6.726)	1.000
GP2: AA AA GG	0.0	0.8	–		–
GP3: AA AC GA	1.7	1.7	1.017	(0.036-14.712)	1.000
GP4: AA AC GG	1.7	0.0	–		–
GP5: GA AA AA	3.3	1.8	2.070	(0.202-21.216)	0.599
GP6: GA AA GA	10.2	10.0	1.132	(0.356-3.478)	0.794
GP7: GA AA GG	8.5	14.2	0.561	(0.170-1.737)	0.339
GP8: GA AC GA	6.8	6.7	1.091	(0.263-4.229)	1.000
GP9: GA AC GG	8.5	14.2	0.561	(0.170-1.737)	0.339
GP10: GA CC GA	0.0	0.8	–		–
GP11: GA CC GG	1.7	2.5	0.672	(0.026-7.486)	1.000
GP12: GG AA AA	3.3	0.8	4.175	(0.289-118.945)	0.253
GP13: GG AA GA	11.9	14.9	0.816	(0.286-2.259)	0.817
GP14: GG AA GG	15.3	11.7	1.363	(0.504-3.643)	0.487
GP15: GG AC AA	0.0	1.7	–		–
GP16: GG AC GA	8.5	4.2	2.130	(0.507-8.947)	0.301
GP17: GG AC GG	11.9	6.7	1.885	(0.577-6.117)	0.259
GP18: GG CC GA	1.7	1.7	1.000	(0.035-14.457)	1.000
GP19: GG CC GG	1.7	3.3	0.492	(0.020-4.829)	0.666

Table 5 (cont). Significant genotypic profiles frequencies of selected SNPs in MDS patients and controls, and its association with risk of myelodysplastic syndrome.

Pathway:	**Profile frequencies**		**Association analysis**		
Genotypic profile	Patients	Controls	Patients *vs* controls		
	%	%	OR	(95% CI)	*p*-value
Base excision repair: *OGG1* + *LIG1*					
GP1: CC AA	12.3	15.1	0.786	(0.276-2.163)	0.818
GP2: CC AC	33.3	26.1	1.419	(0.675-2.978)	0.372
GP3: CC CC	17.5	16.8	1.053	(0.420-2.602)	1.000
GP4: CG AA	5.3	11.8	0.417	(0.091-1.644)	0.275
GP5: CG AC	8.8	15.1	0.540	(0.165-1.662)	0.340
GP6: CG CC	14.0	11.8	1.224	(0.436-3.380)	0.808
GP7: GG AC	5.7	0.8	6.556	(0.588-167.524)	0.100
GP8: GG CC	3.5	2.5	1.406	(0.159-10.741)	0.659

MDS, myelodysplastic syndrome; OR, odds ratio; CI, confidence interval; GP, genetic profile. Underlining indicates reported risk genotypes.

Association analysis of genotypes with MDS subtypes and IPSS risk groups

We also investigated if DNA repair gene variants were associated with MDS subtypes and IPSS risk groups, by logistic regression analysis. None of the genotypes showed a statistically significant association neither with MDS subtypes (5q-, RAEB-1, RAEB-2, RARS, RCMD, RCUD, CMML) nor with IPSS risk groups (low, int-1, and int-2; data not shown).

Analysis of prognostic impact

Finally, we analyzed the effect of DNA repair gene variants on disease prognosis by estimating, through Kaplan-Meier analysis, the rate of transformation into AML and the overall survival (OS) in MDS patients. Patients were stratified according to their SNP genotypes.

The results demonstrated that variants in *XRCC5*, *RMI1*, *TOP3A*, *OGG1*, *LIG1*, *ERCC2*, and *MSH3* genes do not exert an interactive effect on the MDS prognosis (data not

shown). In contrast, *BLM* CC genotype proved to increase the AML transformation rate, when compared to *BLM* AA+AC genotypes (HR = 7.646, 95% CI 1.362-24467, $p = 0.023$), as observed in Figure 1.

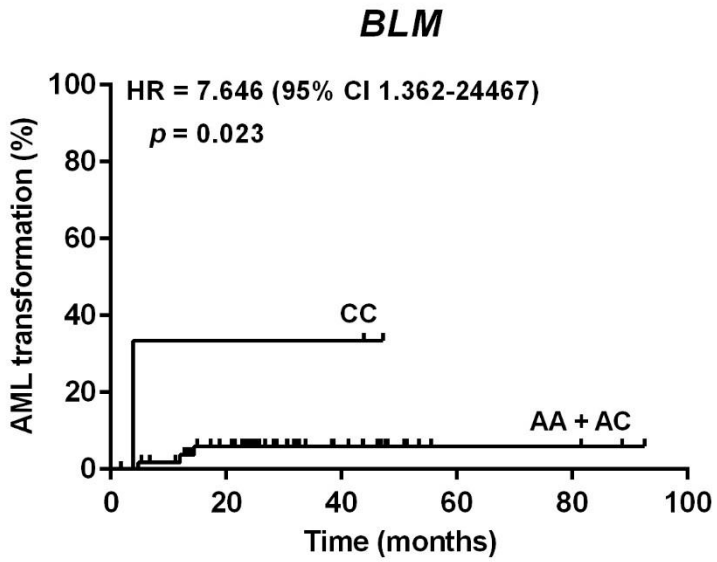

Figure 1. Time to AML transformation in MDS patients, according to *BLM* genotypes. Time to AML transformation was performed by Kaplan Meier analysis, and the hazard ratio (HR) with 95% confidence interval (CI) were calculated using Cox proportional hazard model.

DISCUSSION

Myelodysplastic syndrome (MDS) represent a heterogeneous group of clonal hematopoietic stem cell disorders associated with genetic instability and frequent progression to acute myeloid leukemia (AML).[1,2] DNA repair genes play a pivotal role in maintaining genome stability. Variants within these genes may give rise to different susceptibilities in individuals, leading to the development of diseases, such as MDS.[19] In the present study, we assessed the influence of relevant DNA repair gene variants on the genetic susceptibility to MDS development, by conducting a hospital-based case-control study in a group of Portuguese patients. We also assessed associations with MDS prognosis, namely: the rate of MDS transformation into AML and the overall survival. Two main results were obtained: first, *MSH3* c.2655+5137C>G was associated with MDS susceptibility and, second, *BLM* c.-4-889A>C was associated with the disease prognosis (transformation into AML).

Regarding MDS susceptibility, we found that *MSH3* CG heterozygous genotype increases the risk of MDS, albeit with a large confidence interval (OR = 6.882, 95% CI 1.789-26.479, p = 0.005). The codominant and dominant models showed that the presence of a single allele C increases the risk for MDS. Although we did not observe any GG homozygous genotype, it would be expected that GG homozygous individuals would have an increased risk for MDS. In contrast to our result, Belickova *et al*. found an association between *MSH3* CG and MDS protection (OR = 0.20, 95% CI 0.07-0.54, p = 0.0002) in a Czech population (198 MDS patients).[20] Considering this discrepancy, we propose to perform an association analysis in additional large case-control study. Moreover, a larger sample would decrease the upper confidence interval and increase the probability of finding GG homozygous individuals.

Here, we observed no significant difference in allele and genotype frequencies in the DNA repair gene variants – *XRCC5*, *RMI1*, *BLM*, *TOP3A*, *OGG1*, *LIG1*, and *ERCC2* – between MDS patients and controls (Tables 3 and 4). However, variants in these genes have been associated to MDS susceptibility in other population groups, such as: Czech (*LIG1* AC and *RAD52* TC), Swedish (*RMI1* GA+AA, *BLM* CC, *TOP3A* AA), and Turkish (*OGG1* CG + GG and *ERCC2* CC).[20,22,23,24] In contrary, in Brazilians *XRCC5* AG is associated with a reduced chance of developing MDS.[21] These discrepancies between studies may be explained by differences in ethnic or genetic background of populations, and may support the complexity and heterogeneous pathogenesis of MDS.[21]

Concerning MDS prognosis, the present study correlates a *BLM* variant with AML transformation rate. We observed, that *BLM* CC genotype may increase the AML transformation rate when compared to AA and AC genotypes, albeit with a very large confidence interval (HR = 7.646, 95% CI 1.362-24467, $p = 0.023$). To our knowledge, there was no published data on prognosis relevance of DNA repair genes on MDS patients until recently. Here, we found no significant association regarding *XRCC5*, *RMI1*, *TOP3A*, *OGG1*, *LIG1*, *ERCC2*, and *MSH3* variants and MDS prognosis (namely, AML transformation and overall survival). However, Gonçalves *et al.* recently reported that the *OGG1* GG genotype influenced the AML transformation rate and the survival of MDS patients.[25] This discrepancy may be explained by differences in genotyping techniques and/or by differences in samples, since our sample comprised CMML (16.7%), while Gonçalves *et al.* sample did not include this MDS subtype, Moreover, we reinforce the need to perform an association analysis in additional large case-control study, in order to achieve a narrow confidence interval and prove this association.

MDS emerge from a complex interplay between environmental factors and several DNA changes in many different genes affecting entire biological pathways. We hypothesize that the join action of variants within the same DNA repair pathway may have a more significant role in MDS development. To this end, we performed a multilocus genotype analysis in which we inferred genotypic profiles (GPs) from two different DNA repair pathways: homologous recombination repair (*RMI1* + *BLM* + *TOP3A*) and base excision repair (*OGG1* + *LIG1*). Results were not predictive of the GPs effect in the risk for MDS development, as none profiles showed statistical significance. This can be sustained by the reduced number of patients that was not enough to establish proper analysis, as many genotypic profiles accounted a very small number of individuals.

In the present study, five selected SNPs are located in the protein coding region and lead to a change in the translated aminoacids (missense variants). The other ones are non-coding SNPs located within regulatory regions (introns and UTRs). It is well known that non-coding SNPs can disrupt gene expression, making them of prime importance to be considered for candidate gene association studies.[26,27] Here, we found significant associations of MDS with two non-coding SNPs (the *MSH3* variant with susceptibility and the *BLM* with prognosis), which reinforces the role of non-coding SNPs in cancer development.

The main limitation of our study was the relatively small size of the sample. Nevertheless, allele frequencies observed in cases and controls were in HWE, suggesting that the sample was sufficiently random. Also, the genotypes should be confirmed by DNA sequencing (about 5-10% of individuals). Secondly, we used a hospital-based case-control design, which may cause selection bias. Finally, we did not

evaluate the biological and functional consequences of the studied SNPs. Due to these restrictions, associations should be interpreted with caution.

In conclusion, the results here presented suggest that *MSH3* CG genotype is associated with individual susceptibility to MDS. We also propose that *BLM* CC genotype may be implicated in the increased propensity to AML transformation in MDS patients. These data support that DNA repair genes have an impact not only in the susceptibility for the disease, but also in their prognosis. Nevertheless, these conclusions should be confirmed by additional case-control studies with larger numbers of subjects and/or with other populations.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflict of interest.

ACKNOWLEDGEMENTS

This project was supported by grants from the Direção Regional da Ciência e Tecnologia of Governo dos Açores (Ref. M121/I/ OLD/2002).

We thank Dr. Sara Bulhões and Dr. Tânia Pereirinha for their valuable assistance in the genotyping analysis.

REFERENCES

1. Kyle RA, Rajkumar SV. ASH 50th anniversary review. *Blood*, 111:2962–72 (2008)

2. Tefferi A, Vardiman J. Myelodysplastic syndromes. *N Engl J Med*, 361:1872–85 (2009)

3. Visconte V, Tiu RV, Rogers HJ. Pathogenesis of myelodysplastic syndromes: An overview of molecular and non-molecular aspects of the disease. *Blood Res*, 49:216–27 (2014)

4. Cazzola M, Malcovati L, Invernizzi R. Myelodysplastic/Myeloproliferative neoplasms. *Hematology Am Soc Hematol Educ Program*, 2011:264-72 (2011)

5. Vardiman JW, Thiele J, Arber DA, Brunning RD, et al. The 2008 revision of the WHO classification of myeloid neoplasms and acute leukemia: Rationale and important changes. *Blood*,114:937–52 (2015)

6. Descatha A, Jenabian A, Conso F, Ameille J. Occupational exposures and haematological malignancies: Overview on human recent data. *Cancer Causes Control*, 16:939–53 (2005)

7. Godley LA, Larson RA. Therapy-Related Myeloid Leukemia. *Semin Oncol*, 35:418–29 (2008)

8. Sekeres MA. The epidemiology of myelodysplastic syndromes. *Hematol Oncol Clin North Am*. Elsevier Ltd, 24:287–94 (2010)

9. Jädersten M, Hellström-Lindberg E. Myelodysplastic syndromes: Biology and treatment. *J Intern Med*, 265:307–28 (2009)

10. Raza A, Galili N. The genetic basis of phenotypic heterogeneity in myelodysplastic syndromes. *Nat Rev Cancer*, 12:849–59 (2012)

11. Greenberg PL. Molecular and genetic features of myelodysplastic syndromes. *Int J Lab Hematol*, 34:215–22 (2012)

12. Wyrick JJ, Roberts SA. Genomic approaches to DNA repair and mutagenesis. *DNA Repair (Amst)*, 36:146-55 (2015)

13. Lahtz C, Pfeifer GP. Epigenetic changes of DNA repair genes in cancer. *J Mol Cell Biol*, 3:51–8 (2011)

14. Maugeri-Saccà M, Bartucci M, De Maria R. DNA damage repair pathways in cancer stem cells. *Mol Cancer Ther*, 11:1627–36 (2012)

15. Jeggo PA, Löbrich M. How cancer cells hijack DNA double-strand break repair pathways to gain genomic instability. *Biochem J*, 471:1–11 (2015)

16. Economopoulou P, Pappa V, Kontsioti F, Papageorgiou S, et al. Expression analysis of proteins involved in the non homologous end joining DNA repair mechanism, in the bone marrow of adult de novo myelodysplastic syndromes. *Ann Hematol*, 89:233–9 (2010)

17. Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. *Nature*, 411:366–74 (2001)

18. Altieri F, Grillo C, Maceroni M, Chichiarelli S. DNA damage and repair: From molecular mechanisms to health implications. *Antioxid Redox Signal*. 10:891–937 (2008)

19. Sterpone S, Cozzi R. Influence of XRCC1 Genetic Polymorphisms on Ionizing Radiation-Induced DNA Damage and Repair. *J Nucleic Acids*. (2010)

20. Belickova M, Merkerova MD, Stara E, Vesela J, et al. DNA repair gene variants are associated with an increased risk of myelodysplastic syndromes in a Czech population. *J Hematol Oncol*, 6:9 (2013)

21. Ribeiro-Junior HL, Oliveira RT, Maia AR, Filho LI, et al. Polymorphisms of DNA repair genes are related to the pathogenesis of myelodyplastic syndrome. *Hematol Oncol*, 22:220–8 (2015)

22. Broberg K, Hoglund M, Gustafsson C, Ingvar C, et al. Genetic variant of the human homologous recombination-associated gene RMI1 (S455N) impacts the risk of AML/MDS and malignant melanoma. *Cancer Lett*, 258:38–44 (2007)

23. Broberg K, Huynh E, Schläwicke Engström K, Björk J, et al. Association between polymorphisms in RMI1, TOP3A, and BLM and risk of cancer, a case-control study. *BMC Cancer*, 9:140 (2009)

24. Aktuglu MB, Ayer M, Bireller ES, Rencuzogullari C, et al. Investigation of DNA repair gene variants on myelodysplastic syndromes in a Turkish population. *Med Oncol*, 31:174 (2014)

25. Gonçalves AC, Alves R, Baldeiras I, Cortesão E, et al. Genetic variants involved in oxidative stress, base excision repair, DNA methylation, and folate metabolism pathways influence myeloid neoplasias susceptibility and prognosis. *Mol Carcinog*, [Epub ahead of print] (2016)

26. Patnala R, Clements J, Batra J. Candidate gene association studies: A comprehensive guide to useful in silico tools. *BMC Genet*, 14:39 (2013)

27. Prokunina L, Alarcón-Riquelme ME. Regulatory SNPs in complex diseases: Their identification and functional validation. *Expert Rev Mol Med*, 6:1–15 (2004)