Genome-wide association study identifies the GLDC/IL33 locus associated with survival of osteosarcoma patients

Roelof Koster¹, Orestis A. Panagiotou¹, William A. Wheeler², Eric Karlins¹,¹⁵, Julie M. Gastier-Foster³, Silvia Regina Caminada de Toledo⁴, Antonio S. Petrelli⁴, Adrienne M. Flanagan⁶,⁷, Roberto Tirabosco⁷, Irene L. Andrulis⁸, Jay S. Wunder⁸, Nalan Gokgoz⁸, Ana Patiño-Garcia⁹, Fernando Lecanda⁹, Massimo Serra¹⁰, Claudia Hattinger¹⁰, Piero Picci¹⁰, Katia Scotlandi¹⁰, David M. Thomas¹³, Mandy L. Ballinger¹³, Richard Gorlick⁵, Donald A. Barkauskas¹¹, Logan G. Spector¹⁴, Margaret Tucker¹, Belynda D. Hicks¹,¹⁵, Meredith Yeager¹,¹⁵, Robert N. Hoover¹, Sholom Wacholder¹,¹, Stephen J. Chanock¹, Sharon A. Savage¹, and Lisa Mirabello¹,*

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
²Information Management Services (IMS), Inc. Rockville, MD, USA
³Nationwide Children’s Hospital, and The Ohio State University Department of Pathology and Pediatrics, Columbus, OH, USA
⁴Laboratorio de Genética, Pediatric Oncology Institute, GRAACC/UNIFESP, São Paulo, Brazil
⁵Albert Einstein College of Medicine, The Children’s Hospital at Montefiore, New York, NY, USA
⁶UCL Cancer Institute, Huntley Street, London, UK
⁷Royal National Orthopaedic Hospital NHS Trust, Stanmore, Middlesex, UK
⁸Litwin Centre for Cancer Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada
⁹Department of Pediatrics, University Clinic of Navarra, Universidad de Navarra, Pamplona, Spain
¹⁰Laboratory of Experimental Oncology, Orthopaedic Rizzoli Institute, Bologna, Italy
¹¹Keck School of Medicine, University of Southern California, Los Angeles, California, USA
¹²Stanford University & Lucile Packard Children's Hospital, Palo Alto, CA, USA
¹³The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Darlinghurst, NSW, Australia
¹⁴Department of Pediatrics, University of Minnesota Minneapolis, MN, 55455, USA
¹⁵Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

*Corresponding Author: Lisa Mirabello, Ph.D., Investigator, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, 9609 Medical Center Drive, Room 6E524, Bethesda, MD 20850, Tel: 240-276-7258, mirabellol@mail.nih.gov.
†Deceased.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.
Abstract
Survival rates for osteosarcoma, the most common primary bone cancer, have changed little over the past three decades and are particularly low for patients with metastatic disease. We conducted a multi-institutional genome-wide association study (GWAS) to identify germline genetic variants associated with overall survival in 632 patients with osteosarcoma including 523 patients of European ancestry and 109 from Brazil. We conducted a time-to-event analysis and estimated hazard ratios (HR) and 95% confidence intervals (CI) using Cox proportional hazards models, with and without adjustment for metastatic disease. The results were combined across the European and Brazilian case sets using a random-effects meta-analysis. The strongest association after meta-analysis, was for rs3765555 at 9p24.1, which was inversely associated with overall survival (HR=1.76; 95% CI 1.41-2.18, \( P = 4.84 \times 10^{-7} \)). After imputation across this region, the combined analysis identified two SNPs that reached genome-wide significance. The strongest single association was with rs55933544 (HR=1.9; 95% CI 1.5-2.4; \( P =1.3 \times 10^{-8} \)), which localizes to the \textit{GLDC} gene, adjacent to the \textit{IL33} gene and was consistent across both the European and Brazilian case sets. Using publicly available data, the risk allele was associated with lower expression of \textit{IL33} and low expression of \textit{IL33} was associated with poor survival in an independent set of patients with osteosarcoma. In conclusion, we have identified the \textit{GLDC}\textit{IL33} locus on chromosome 9p24.1 as associated with overall survival in patients with osteosarcoma. Further studies are needed to confirm this association and shed light on the biological underpinnings of this susceptibility locus.

Keywords
Osteosarcoma; overall survival; genome-wide association study; osteosarcoma specific survival

INTRODUCTION
Osteosarcoma is the most common primary bone cancer in children and adolescents.\(^1\)-\(^4\) The introduction of effective neo-adjuvant and adjuvant chemotherapy in the 1980s resulted in improved long-term survival for non-metastatic osteosarcoma patients, increasing survival from 20\%-30\% to more than 70\%.\(^4\)-\(^7\) However, over the past three decades, there has been little improvement in survival rates for patients with metastatic disease at diagnosis\(^8\) with 5-year overall survival rates remaining at 25\%-30\%.\(^9\),\(^10\) Several factors have been suggested to influence survival of patients with osteosarcoma\(^11\), including age at diagnosis, metastatic disease at presentation, tumor histology, blood alkaline phosphatase levels, tumor size and location, and response to chemotherapy (estimated by the percentage of tumor necrosis after chemotherapy).\(^12\)-\(^14\) Recently, we reported that germline genetic variants in the \textit{NFIB} gene locus (9p23-9p22.3) are associated with metastatic disease at osteosarcoma diagnosis, suggesting that genetic susceptibility could contribute to survival.\(^15\)

There is growing interest in whether germline genetic variants could influence outcomes of patients with cancer. A population-based family study showed that cancer-specific survival
can be partly related to inherited factors within families, suggesting there are germline genetic determinants of survival. While earlier candidate gene studies have identified variants associated with prognosis that have not been confirmed, more recent large genome-wide association studies (GWAS) have identified associations between common SNPs and survival in adult cancers of the pancreas, breast, ovary, and in a rare pediatric cancer, neuroblastoma. There have also been exploratory pharmacogenomic studies of pediatric Ewing sarcoma and osteosarcoma that have identified SNPs associated with response to treatment and survival, although most await further validation. We performed a GWAS in order to explore whether germline genetics may contribute to survival in patients with osteosarcoma.

METHODS

Study populations

A summary of the participating studies is provided in Supplemental Table 1. Previously, we reported 689 histologically confirmed osteosarcoma patients of >80% European ancestry based on a STRUCTURE analysis employing principal components analyses of a set of 12,000 unlinked markers (pairwise $r^2 < 0.004$). A subset of 523 European ancestry osteosarcoma patients had available data on mortality (European set) for a survival analysis. After performing the GWAS for survival in this data set, we evaluated the most promising SNPs from the European set ($P < 10^{-4}$) in 109 Brazilian osteosarcoma patients from the Instituto de Oncologia Pediátrica GRAACC/UNIFESP and Universidade Federal de Sao Paulo, Brazil (Brazilian set; Supplemental Table 1). Participating subjects provided informed consent under the auspices of local Institutional Review Boards.

Genotyping and quality control

All subjects were previously genotyped as part of our osteosarcoma susceptibility GWAS (dbGaP Study Accession: phs000734.v1.p1). In brief, genotyping was conducted on the Illumina OmniExpress BeadChip. SNPs included in the analyses were autosomal with a minor allele frequency (MAF) of 5% or more; had a 90% or more completion rate; and no evidence of deviation from Hardy-Weinberg equilibrium ($P > 10^{-7}$). SNPs were also excluded if they had abnormal heterozygosity values. After quality control assessment, 510,856 SNPs were advanced in our survival analysis.

Statistical analyses

We conducted a time-to-event analysis to investigate the effect of genetic variation on overall survival. The outcome variable of interest was time until the event of death. The overall survival time was calculated as the time from the date of osteosarcoma diagnosis until the date of death for those deceased or the last date known to be alive; patients were censored at the last date known to be alive or when lost to follow-up. All events were identified and verified through medical record review and/or death certificates at each participating study center. We modeled each genome-wide association between one or more SNPs and survival using Cox proportional hazards regression and estimated hazard ratios (HR) and 95% confidence intervals (CI) per copy of the minor allele (log-additive genetic model). We tested the hazards proportionality assumption (i.e., the hazard ratio is constant
over time)\textsuperscript{35} of the Cox model using Schoenfeld’s residuals;\textsuperscript{36, 37} we did not detect nominally significant violations of the proportional hazards assumption.

Cox models were adjusted for age at diagnosis, gender, significant principal components, and study/center (except for the Brazilian study, since all samples were from the same hospital). We did not adjust for metastatic disease agnostically at the genome-wide level, because some SNPs may be associated with metastatic disease and survival, as we have shown previously,\textsuperscript{15} and thus adjusting for potential intermediate factors that lie on the causal pathway could introduce bias.\textsuperscript{38–40} However, since metastatic disease is a prognostic factor for overall survival,\textsuperscript{12} we performed sensitivity analyses for the top signals by also adjusting for metastatic disease.\textsuperscript{41} We constructed Kaplan-Meier survival curves\textsuperscript{42} for the top SNPs under dominant models and compared their statistical equivalence for each model with the log-rank test.

SNPs that reached $P < 10^{-4}$ in the European set were followed-up in the Brazilian set. We used a random-effects meta-analysis with inverse-variance weighting to estimate the summary effect across the sets.\textsuperscript{43} We evaluated between-sets heterogeneity using Cochran’s Q chi-squared statistic and quantified heterogeneity with the I$^2$ metric.\textsuperscript{44} SNPs that demonstrated significant between-set heterogeneity ($P < 0.05$) were excluded.

Based on our GWAS results, we conducted region-specific imputation analysis of flanking SNPS, namely, 1 Mb region on either side of the strongest SNP using the IMPUTE2 software and the reference data from the 1000 Genomes project (Phase 3 genotype data).\textsuperscript{45}

To investigate whether signals in the same genomic region represent independent associations, we conducted conditional analyses by adjusting the Cox models for the top SNP in each region as applicable.

Statistical analyses were performed in Stata version 13 and R 3.0.2. All P-values are two-sided.

**eQTL and survival-expression analyses**

We performed expression quantitative trait locus (eQTL) based analyses using publicly available genotype and expression data from 29 osteosarcoma tumors (GSE33383)\textsuperscript{46} and separately in osteosarcoma tumors from 89 patients included in the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) osteosarcoma dataset (http://ocg.cancer.gov/programs/target).\textsuperscript{47} The data used for TARGET are available at dbGaP accession phs000218, accession phs000468, in which we conducted survival-expression analysis.\textsuperscript{47} For the 127 patients from the combined dataset (Kuijjer et al\textsuperscript{46} and Buddingh et al\textsuperscript{48}) survival-expression was analyzed using the R2: Genomics analysis and visualization platform (http://r2.amc.nl/).\textsuperscript{49}

**Bioinformatic analyses**

Linkage disequilibrium (LD) was evaluated with $r^2$ based on the 1000 Genomes Project (Phase 3 genotype data)\textsuperscript{45} using LDlink (http://analysistools.nci.nih.gov/LDlink/).\textsuperscript{23} Chromosome location for human genome assembly hg19 was retrieved from the National
Center for Biotechnology Information’s (NCBI) Gene database (http://www.ncbi.nlm.nih.gov/gene/). Genomic annotation of SNP markers was conducted using the Encyclopedia of DNA Elements (ENCODE)\(^\text{50}\) tool HaploReg\(^\text{51}\) and RegulomeDB\(^\text{52}\) for all cell lines. Surrogate SNPs were identified using the 1000 Genomes data for individuals of European ancestry with an \(r^2 > 0.4\) and within ±500kb. LocusZoom\(^\text{53}\) was used to plot regional associations.

**RESULTS**

**Patient characteristics**

Table 1 shows the characteristics of the patients included in this study. There were 632 total osteosarcoma patients included in the overall survival analyses. In the European set, there were 170 (33%) mortality events, and 37 (34%) in the Brazilian set. Age (\(P < 0.001\)) and presence of metastases at diagnosis (\(P < 0.001\)) were associated with patient survival (Table 1, Supplemental Figure 1).

**SNPs associated with overall survival**

In a case-case analysis, 81 SNPs were associated with overall survival at \(P < 10^{-4}\) in the European set (Supplemental Table 2) and were followed-up in the independent set of 109 osteosarcoma cases from Brazil. The strongest association with overall survival in the European set is at chromosome 5q11.1 tagged by rs1030228 (located 14kb 5' of NDUF54) with a HR for mortality of 1.71 (95% CI 1.39-2.12, \(P = 7.10 \times 10^{-7}\); Supplemental Figure 2, Supplemental Table 2). However, this SNP is not associated with survival in the Brazilian study (\(P=0.80\)). There is a large degree of between-set heterogeneity (\(P_{\text{het}}=0.052, I^2=73.5\%\) for this variant (Supplemental Table 2).

In the combined analysis, the strongest association was SNP rs3765555, which is inversely associated with overall survival (HR=1.76 per copy of the A allele, 95% CI 1.41-2.18, \(P=4.84 \times 10^{-7}\); \(I^2=0\%\); Table 2, Figure 1A). This SNP is located in intron 23 of the glycine dehydrogenase (decarboxylating) gene (GLDC) on chromosome 9p24.1. The MAF for rs3765555 in the European (MAF=0.26) and Brazilian (MAF=0.21) populations are similar and we did not observe significant between-set heterogeneity (\(P_{\text{het}}=0.34\); Table 2, Supplemental Table 2).

In a further exploration of the promising regions, we imputed SNPs across a 1 Mb region centered on rs3765555 to further evaluate this locus. After a random-effects meta-analysis for the imputed SNPs, we identified rs55933544 (Table 2, Figure 1A) as the SNP most strongly associated with overall survival (HR=1.89 per copy of the T allele, 95% CI 1.50-2.37; \(P=4.81 \times 10^{-8}\); \(I^2=0\%\)), which is in strong LD with rs3765555 (\(r^2=0.86\) in Europeans and \(r^2=0.94\) in admixed Americans). The results remained the same after adjustment for metastatic disease at diagnosis (Table 2), suggesting that the 9p24.1 locus marked by rs55933544 affects overall survival independent of metastatic disease at genome-wide significance (HR=1.92 per copy of the T allele, 95% CI 1.53-2.41; \(P=1.34 \times 10^{-8}\); \(I^2=0\%\). rs55933544 (chr9:6534080) was not correlated with SNP rs7034162 at 9p23-9p22.3
(chr9:14190287) that we previously identified as associated with metastatic disease at osteosarcoma diagnosis \( (r^2=0.0008, 1,000 \text{ Genomes Project CEU data}).^{15} \\

A second SNP at 9p24.1, rs74438701 located approximately 25kb downstream of the interleukin 33 \((IL33)\) gene, is also inversely associated with overall survival in the combined analysis \( (HR=2.00 \text{ per copy of the } C \text{ allele}, 95\% \text{ CI } 1.56-2.57, P=4.90 \times 10^{-8}; I^2=0\% ; \text{ Figure 1B}). \) However, the conditional analysis showed that rs74438701 is not an independent signal \( (\text{Figure 1C}).\)

Kaplan-Meier curve analysis indicate a statistically significant difference between survival rates over time \( (\log\text{-rank test } P < 0.001) \) for both the dominant \( (\text{Figure 1D}) \) and a multiplicative model \( (\text{data not shown}) \) for rs55933544. In addition, we confirmed this association in an independent dataset of 89 cases \( (\text{TARGET}^{47}; \log\text{-rank } P=0.013; \text{ Figure 2A}).\)

We examined the set of highly correlated surrogate SNPs \( (n=31) \) across the 9p24.1 region \( (\text{based on an } r^2>0.4, 1,000 \text{ Genomes Project CEU data}) \) to identify putative regulatory elements using the ENCODE data resource\(^50\) tools HaploReg\(^51\) and RegulomeDB\(^52\) \( (\text{Supplemental Table 3}).\) A subset of the surrogate SNPs \( (n=29) \) are located in predicted promoter and/or enhancer histone marks, DNAse sensitivity regions, and/or transcription factor binding sites in a variety of different cell types and may have a functional impact \( (\text{Supplemental Table 3}).\)

**IL33 expression levels associated with survival**

We performed expression quantitative trait locus \( (\text{eQTL}) \) analyses to evaluate whether rs55933544 was associated with expression of GLDC, IL33 or other neighboring protein-encoding genes, using publicly available expression and genotyping data on osteosarcoma tumors. Interestingly, a previous eQTL was observed between rs55933544 and IL33 expression in human skin\(^54\) and human brain tissue\(^55\) \( (\text{Supplemental Table 3}).\) We found that the risk allele \( (T) \) of rs55933544 was significantly associated with a decrease in IL33 expression in both osteosarcoma tumor data sets from TARGET\(^47\) \( (N=83, \ P=0.041) \) and Kuijjer \textit{et al}.\(^46\) \( (N=29, \ P=0.020) \) \( (\text{Figure 2B and Supplemental Figure 3}).\) In addition, lower expression of IL33 in osteosarcoma tissue was independently associated with worse osteosarcoma patient survival in TARGET\(^47\) \( (\log\text{-rank test } P=0.023; \text{ Figure 2C}) \) and Kuijjer \textit{et al}.\(^46\) \( (\text{raw } P=7.9 \times 10^{-3}; \text{ Supplemental Figure 3}).\) There was no association between rs55933544 genotypes and expression of GLDC or other nearby protein-encoding genes \( (TPD52L3, UHRF2 \text{ and KDM4C}; \text{ data not shown}).\)

**DISCUSSION**

We conducted a genome-wide association study for overall survival in osteosarcoma cases using data from a multi-stage, international collaborative effort.\(^33\) One locus, GLDC/IL33 at 9p24.1, was associated with overall survival of patients with osteosarcoma, which suggests that germline genetics can influence osteosarcoma outcomes, independent of metastatic disease. Here we observed moderate to large effect sizes for a SNP associated with overall survival, a finding similar to that observed in our GWAS of metastatic disease at
osteosarcoma diagnosis. The observed effect sizes are also comparable to GWAS of other pediatric and young adulthood cancers, and higher than those observed in adult GWAS of common cancer susceptibility.

The most promising signal for overall survival in patients with osteosarcoma localizes to the 9p24.1 region, downstream and independent of the NFIB gene locus (9p23-9p22.3) previously reported for metastatic disease. The SNP marker, rs55933544, in the GLDC gene region is associated with decreased survival. High expression of GDLC has been associated with poor survival in other cancers, however, we did not detect an eQTL for rs55933544 and GDLC in osteosarcoma cells. Interestingly, rs55933544 alleles have also been associated with expression of the nearby gene, IL33, and we detected an eQTL with IL33 in osteosarcoma cells. In addition, lower expression of IL33 was associated with poor survival in patients with osteosarcoma. IL-33 is an inhibitor of bone reabsorption, blocking osteoclastic activity, which may be important in osteosarcoma. Lower levels of IL-33 have also been associated with worse prognosis or more advanced disease in several other tumor types, consistent with our data.

This exploratory study requires further follow up and has limitations. Treatment likely varied among patients in our studies, but we could not control for this because treatment information was not available. However, it is unlikely that individual treatment modalities varied systematically by germline genotype. We also performed a combined analysis with osteosarcoma patients from Brazil, with a relatively small sample size. LD patterns differ in many parts of the genome between admixed Brazil populations and Europeans, and this could lead to false negative results in our analysis. This could explain the lack of replication in Brazilian patients of some of the top SNPs in the European osteosarcoma patients. However, an important strength, is that this two-stage design also reduces the likelihood of false positive results.

In conclusion, we provide evidence that germline genetic variants are associated with overall survival in osteosarcoma patients. These findings warrant follow-up in additional populations and functional characterization to investigate the biologic mechanisms by which polymorphisms at this locus impact survival.

Acknowledgments

Funding Support

This study was funded by the intramural research program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health.

This work was supported by the Bone Cancer Research Trust UK to A.M.F.

Research is supported by the Chair's Grant U10 CA98543 and Human Specimen Banking Grant U24 CA114766 of the Children's Oncology Group from the National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. Additional support for research is provided by a grant from the DEB Foundation, Inc. to the Children's Oncology Group.
This work was supported by grants to I.L.A. and J.S.W. from the Ontario Research Fund, and Canadian Foundation for Innovation.

This study was also supported by biobank grants from the Regione Emilia-Romagna, by the infrastructure and personnel of the Royal National Orthopaedic Hospital Musculoskeletal Research Programme and Biobank. Support was also provided to A.M.F. (UCL) by the National Institute for Health Research UCLH Biomedical Research Centre, and UCL Experimental Cancer Centre, funding from PI13/01476, FIS, ISCIII and La Fundación Bancaria "La Caixa", Fundación Caja Navarra to AP-G, and AECC project to F.L.

The International Sarcoma Kindred Study was supported by the Rainbows for Kate Foundation, the Liddy Shriver Sarcoma Initiative, the Victorian Cancer Agency, the Australian National Health and Medical Research Council (APP1004017) and Cancer Australia (APP1067094).

**Abbreviations used**

- **GWAS** genome-wide association study
- **HR** hazard ratio
- **CI** confidence intervals
- **MAF** minor allele frequency
- **LD** linkage disequilibrium
- **ENCODE** Encyclopedia of DNA Elements
- **NCBI** National Center for Biotechnology Information
- **eQTL** expression quantitative trait loci
- **SNP** single nucleotide polymorphisms
- **GLDC** glycine dehydrogenase (decarboxylating) gene
- **IL33** Interleukin 33 gene
- **NFIB** Nuclear Factor I B gene
- **TPD52L3** Tumor Protein D52 Like 3 gene;
- **KDM4C** Lysine Demethylase 4C gene

**References**


47. TARGET. Therapeutically Applicable Research to Generate Effective Treatments.


**Novelty and Impact**

To date, few prognostic factors have been identified associated with survival in patients with osteosarcoma. The authors conducted a genome-wide association study (GWAS) of overall survival in two sets of patients with osteosarcoma. They identified a common single nucleotide polymorphism (SNP), rs55933544, located in the *GLDC* gene on chromosome 9, associated with poor survival. The rs55933544 risk allele was associated with lower expression of the nearby gene, *IL33*. These findings, if replicated in additional populations, form the foundation for future studies of the molecular basis of the association of the *GLDC/IL33* (rs55933544) variant with survival in osteosarcoma.
Figure 1.
Regional plots of the combined association results, recombination hotspots and linkage disequilibrium (LD) for the 9p24.1 region that harbors rs55933544 and rs74438701 that are associated with overall survival. Results are shown for unconditional (A, B) and conditional (C) analyses. Also shown is the Kaplan-Meier curve (D) for overall survival for the strongest SNP (rs55933544) under a dominant model in the combined European and Brazilian sets. In panels A-C, Y-axes represent the statistical significance (−log10 transformed $P$ values) of SNP association results from a trend test (left) and the recombination rate (right). SNPs are color-coded based on pairwise linkage disequilibrium ($r^2$) with the most statistically significant SNP. The most statistically significant SNP is labeled and shown in purple. Allelic P values are the P-values from Cox models. Physical locations of the SNPs are based on NCBI human genome build 36, and gene annotation was based on the NCBI RefSeq genes from the UCSC Genome Browser.
Figure 2. 
*IL33* expression levels associated with survival in osteosarcoma patients. (A) Kaplan-Meier curve for overall survival by rs55933544 genotype; (B) eQTL of rs55933544 and IL33; and, (C) patient survival by *IL33* low and high expression levels (independent of genotype), all using TARGET data.
Table 1

Patient characteristics in the European and Brazilian set

<table>
<thead>
<tr>
<th>Endpoint</th>
<th>European</th>
<th>Brazil</th>
<th>Combined Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=523</td>
<td>N=109</td>
<td>N=632</td>
</tr>
<tr>
<td><strong>Overall Survival</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vital status, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dead</td>
<td>170 (33)</td>
<td>37 (34)</td>
<td>207 (33)</td>
</tr>
<tr>
<td>Alive</td>
<td>353 (67)</td>
<td>72 (66)</td>
<td>425 (67)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25</td>
<td>432 (83)</td>
<td>93 (85)</td>
<td>525 (83)</td>
</tr>
<tr>
<td>≥25 to&lt;60</td>
<td>69 (13)</td>
<td>12 (11)</td>
<td>81 (13)</td>
</tr>
<tr>
<td>≥60</td>
<td>21 (4)</td>
<td>0 (0)</td>
<td>21 (3)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0)</td>
<td>4 (4)</td>
<td>5 (1)</td>
</tr>
<tr>
<td>Gender, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>295 (56)</td>
<td>58 (53)</td>
<td>353 (56)</td>
</tr>
<tr>
<td>Females</td>
<td>228 (44)</td>
<td>51 (47)</td>
<td>279 (44)</td>
</tr>
<tr>
<td>Metastasis at diagnosis, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>131 (25)</td>
<td>40 (37)</td>
<td>171 (27)</td>
</tr>
<tr>
<td>No</td>
<td>392 (75)</td>
<td>69 (63)</td>
<td>455 (72)</td>
</tr>
</tbody>
</table>

* All subjects included in the European set were of >80% European ancestry.
† Shows the time at which 75% of patients had not experienced the event of interest (i.e. death or progression).
‡‡ P-values are from log-rank test.

ST: survival time; NA: non-applicable
SNPs associated with overall survival.

Table 2

<table>
<thead>
<tr>
<th>SNP</th>
<th>Method</th>
<th>Gene Locus</th>
<th>**</th>
<th>Position†</th>
<th>Set</th>
<th>MAF</th>
<th>Unadjusted for metastatic disease</th>
<th>Adjusted for metastatic disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3765555-C</td>
<td>Genotyped</td>
<td>GLDC</td>
<td></td>
<td>Chr9: 6535956</td>
<td>European</td>
<td>0.259</td>
<td>1.67 (1.32–2.13)</td>
<td>2.70×10⁻⁵</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Brazil</td>
<td>0.205</td>
<td>2.23 (1.31–3.81)</td>
<td>3.29×10⁻³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>1.76 (1.41–2.18)</td>
<td>4.84×10⁻⁷</td>
<td>1.78 (1.43–2.21)</td>
</tr>
<tr>
<td>rs55933544-C</td>
<td>Imputed</td>
<td>GLDC</td>
<td></td>
<td>Chr9: 6534080</td>
<td>European</td>
<td>0.231</td>
<td>1.85 (1.44–2.37)</td>
<td>1.58×10⁻⁶</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Brazil</td>
<td>0.231</td>
<td>2.11 (1.21–3.69)</td>
<td>8.44×10⁻³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>1.89 (1.50–2.37)</td>
<td>4.81×10⁻⁸</td>
<td>1.92 (1.53–2.41)</td>
</tr>
</tbody>
</table>

* Alleles are shown as major/minor.
** Gene locus information is based on the GENCODE data from HaploReg v2.
† Position is based on hg19.

Hazard ratios are shown per copy of the minor allele in the discovery stage.

MAF: minor allele frequency; HR: hazard ratio; CI: confidence interval