A pleiotropic variant in DNAJB4 is associated with multiple myeloma risk

Marco Dicanio | Matteo Giaccherini | Alyssa Clay-Gilmour
Angelica Macauda | Juan Sainz | Mitchell J. Machiela
Malwina Rybicka-Ramos | Aaron D. Norman | Agata Tyczynska
Stephen J. Chanock | Torben Barington | Shaji K. Kumar
Parveen Bhatti | Wendy Cozen | Elizabeth E. Brown | Anna Suska
Eva K. Hastrup | Robert Z. Orlofski | Marek Dudziński
Ramon Garcia-Sanz | Marcin Kruszewski | Joaquin Martinez-Lopez
Katia Beider | Elżbieta Iskierka-Jazdzewska | Matteo Pelosini
Sonja I. Berndt | Małgorzata Rażyń | Krzysztof Jamroziak
S. Vincent Rajkumar | Artur Jurczyszyn | Annette Juul Vangsted
Pilar Garrido Collado | Ulla Vogel | Jonathan N. Hofmann | Mario Petrini
Aleksandra Butrym | Susan L. Slager | Elad Ziv | Edyta Subocz
Graham G. Giles | Niels Frost Andersen | Grzegorz Mazur
Marzena Watek | Fabienne Lesueur | Michelle A. T. Hildebrandt
Daria Zawirska | Lene Hyldahl Ebbesen | Herlander Marques
Federica Gemignani | Charles Dumont | Judit Várkonyi | Gabriele Buda
Arnon Nagler | Agnieszka Druzd-Sitek | Xifeng Wu | Katalin Kadar
Nicola J. Camp | Norbert Grzasko | Rosalie G. Waller | Celine Vachon
Federico Canzian | Daniele Campa

Correspondence
Daniele Campa, Department of Biology, University of Pisa, 56126 Pisa, Italy.
Email: daniele.campa@unipi.it

Present address
Matteo Pelosini, Ospedale Santa Chiara, Pisa, Italy.

Abstract
Pleiotropy, which consists of a single gene or allelic variant affecting multiple unrelated traits, is common across cancers, with evidence for genome-wide significant loci shared across cancer and noncancer traits. This feature is particularly relevant in multiple myeloma (MM) because several susceptibility loci that have been

Abbreviations: CI: confidence interval; eQTL: expression quantitative trait loci; GTEx: genotype-tissue expression project; GWAS: genome-wide association studies; HSC: hematopoietic stem cell; HWE: Hardy-Weinberg equilibrium; IMMEnSE: International Multiple Myeloma RESEARCH; IMWG: International Myeloma Working Group; InterLymph: International Lymphoma Epidemiology Consortium; LD: linkage disequilibrium; lncRNA: long non-coding RNA; MM: multiple myeloma; OR, odds ratio; PC: principal component; SNP: single nucleotide polymorphism.

For affiliation refer to page 245

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. International Journal of Cancer published by John Wiley & Sons Ltd on behalf of UICC.
identified to date are pleiotropic. Therefore, the aim of this study was to identify novel pleiotropic variants involved in MM risk using 28,684 independent single nucleotide polymorphisms (SNPs) from GWAS Catalog that reached a significant association ($P < 5 \times 10^{-8}$) with their respective trait. The selected SNPs were analyzed in 2434 MM cases and 3446 controls from the International Lymphoma Epidemiology Consortium (InterLymph). The 10 SNPs showing the strongest associations with MM risk in InterLymph were selected for replication in an independent set of 1955 MM cases and 1549 controls from the International Multiple Myeloma rESearch (IMMEnSE) consortium and 418 MM cases and 147,282 controls from the FinnGen project. The combined analysis of the three studies identified an association between DNAJB4-rs34517439-A and an increased risk of developing MM (OR = 1.22, 95%CI 1.13-1.32, $P = 4.81 \times 10^{-7}$). rs34517439-A is associated with a modified expression of the FUBP1 gene, which encodes a multifunctional DNA and RNA-binding protein that it was observed to influence the regulation of various genes involved in cell cycle regulation, among which various oncogenes and oncosuppressors. In conclusion, with a pleiotropic scan approach we identified DNAJB4-rs34517439 as a potentially novel MM risk locus.

KEYWORDS
genetic susceptibility, multiple myeloma, pleiotropy, pleiotropy scan, polymorphisms

What’s new?
Genetic variants can have multiple, seemingly unrelated, effects. Often, these so-called “pleiotropic” variants play a role in cancer. Here, the authors set out to identify new pleiotropic variants involved in multiple myeloma (MM) risk. They analyzed 28,684 independent single nucleotide polymorphisms (SNPs) that had been identified in genome-wide association studies as having an effect on a human trait. This analysis revealed an association between increased MM risk and a variant called DNAJB4-rs34517439-A. That variant has been associated with changes in expression of a DNA- and RNA-binding protein that helps regulate cell cycle genes.

1 | INTRODUCTION

Multiple myeloma (MM) is an incurable hematological disease originating from plasma cells in the bone marrow.1 MM is the third most common hematological tumor with a crude incidence rate of 6.8/100000 new cases per year in Europe2 (https://gco.iarc.fr/today/home).

Several studies have investigated the genetic susceptibility of MM, using genome-wide association studies (GWAS)3-10 or more targeted approaches.8,11-15 Twenty-four risk loci have been found to be associated with MM risk through GWAS; however, these variants explain approximately only 16% of MM heritability.9 Due to the large number of tests conducted in a GWAS, a stringent statistical significance threshold, $P < 5 \times 10^{-8}$, is employed. The associated increase in type II error rates means that many true associations may remain undetected and unreported.

One strategy to avoid this loss of information is to investigate pleiotropic variants. Pleiotropy is a genetic phenomenon consisting of a single gene or allelic variant affecting multiple, often unrelated traits.16 GWASs have identified several loci associated with cancer and noncancer phenotypes. Recently, it was estimated that half of all the SNPs showing an association with a $P_{value}$ threshold of $10^{-8}$ are pleiotropic, and 12.34% show associations with 10 or more phenotypes.17

Pleiotropy is frequent across cancers,18-20 with a portion of genome-wide associated loci shared with other traits and several regions harboring single nucleotide polymorphisms (SNP) associated with multiple cancer types. For example, the TERT and 8q24 loci are associated with risk of multiple cancer types, including MM.12,21-25 This feature is particularly evident in MM since we observed, using GWAS Catalog, that 17 of the 24 associated polymorphisms (~71%) are pleiotropic or in high linkage disequilibrium (LD) ($r^2 > 0.8$) with SNPs associated with other traits (https://www.ebi.ac.uk/gwas/), including telomere length,26 BMI27 and risk of various cancers, such as myeloproliferative neoplasms28 and pancreatic cancer.25 Therefore, the aim of this study was to test the impact of all the SNPs associated at genome-wide significant level with any human trait on risk of MM. This strategy has been employed for several cancer sites,18-20,29-31 but never for MM.
2 | MATERIALS AND METHODS

2.1 | Study populations

The study employed a two-step approach, consisting of a discovery phase in which the International Lymphoma Epidemiology Consortium (InterLymph) GWAS data were analyzed, and a validation phase performed using both summary statistics from the FinnGen project MM GWAS (https://www.finngen.fi/en/access_results) and cases and controls of the International Multiple Myeloma RESEarch (IMMEnSE) consortium.

InterLymph has been described elsewhere. Briefly, InterLymph generated GWAS data, using multiple platforms, for 2434 MM cases and 3446 controls of European ancestry from the United States of America, Canada and Australia genotyped using the Affymetrix 6.0 and Illumina (610 Quad, Human660W-quad Beadchip, Omni5, OmniExpress Beadchip, Oncoarray) platforms. Imputation was performed using the Haplotype Reference Consortium as reference panel, and the Michigan imputation server (https://imputationserver.sph.umich.edu/). After imputation, each site was filtered to include only imputed variants with information score > 0.6 and further quality control checks were implemented (genotype rate > 95%, minor allele frequencies >0.01, and Hardy-Weinberg equilibrium [HWE] P > 10\(^{-8}\) in controls). Finally, the data were pooled and final quality controls were performed on the pooled GWAS set, including checks for mississngness, duplicates, sex mismatch, abnormal heterozygosity, genetic relatedness, population outliers (through principal component [PC] analysis), and genomic inflation. After applying quality control measures to the imputed data, 5,864,648 SNPs remained for analysis.

The FinnGen project is a cohort of 176,899 Finnish individuals genotyped with Illumina and Affymetrix platforms (https://www.finngen.fi). Subjects with sex discrepancies, call rate < 95%, excess heterozygosity (\(\pm 4SD\)) and non-Finnish ancestry were removed. SNPs with call rate < 98%, deviation from HWE (P < 10\(^{-6}\)) and low minor allele count < 3 were removed. Subsequently, imputation was conducted using the 1000 Genomes v1 reference panel with Beagle 4.1 (version 08Jun17.d8b). Imputed variants with information score < 0.7 were removed. A GWAS testing association of 16,962,023 SNPs with call rate < 98%, deviation from HWE (P < 10\(^{-6}\)) and low minor allele count < 3 were removed. Subsequently, imputation was conducted using the Haplotype Reference Consortium as reference panel, and the Michigan imputation server (https://imputationserver.sph.umich.edu/). After imputation, each site was filtered to include only imputed variants with information score > 0.6 and further quality control checks were implemented (genotype rate > 95%, minor allele frequencies >0.01, and Hardy-Weinberg equilibrium [HWE] P > 10\(^{-8}\) in controls). Finally, the data were pooled and final quality controls were performed on the pooled GWAS set, including checks for mississngness, duplicates, sex mismatch, abnormal heterozygosity, genetic relatedness, population outliers (through principal component [PC] analysis), and genomic inflation. After applying quality control measures to the imputed data, 5,864,648 SNPs remained for analysis.

The IMPhEN project is a cohort of 7,212,284 European ancestry Finnish individuals genotyped with Illumina and Affymetrix platforms (https://www.finngen.fi). Subjects with sex discrepancies, call rate < 95%, excess heterozygosity (\(\pm 4SD\)) and non-Finnish ancestry were removed. SNPs with call rate < 98%, deviation from HWE (P < 10\(^{-6}\)) and low minor allele count < 3 were removed. Subsequently, imputation was conducted using the Haplotype Reference Consortium as reference panel, and the Michigan imputation server (https://imputationserver.sph.umich.edu/). After imputation, each site was filtered to include only imputed variants with information score > 0.6 and further quality control checks were implemented (genotype rate > 95%, minor allele frequencies >0.01, and Hardy-Weinberg equilibrium [HWE] P > 10\(^{-8}\) in controls). Finally, the data were pooled and final quality controls were performed on the pooled GWAS set, including checks for mississngness, duplicates, sex mismatch, abnormal heterozygosity, genetic relatedness, population outliers (through principal component [PC] analysis), and genomic inflation. After applying quality control measures to the imputed data, 5,864,648 SNPs remained for analysis.

The FinnGen project is a cohort of 176,899 Finnish individuals genotyped with Illumina and Affymetrix platforms (https://www.finngen.fi). Subjects with sex discrepancies, call rate < 95%, excess heterozygosity (\(\pm 4SD\)) and non-Finnish ancestry were removed. SNPs with call rate < 98%, deviation from HWE (P < 10\(^{-6}\)) and low minor allele count < 3 were removed. Subsequently, imputation was conducted using the Haplotype Reference Consortium as reference panel, and the Michigan imputation server (https://imputationserver.sph.umich.edu/). After imputation, each site was filtered to include only imputed variants with information score > 0.6 and further quality control checks were implemented (genotype rate > 95%, minor allele frequencies >0.01, and Hardy-Weinberg equilibrium [HWE] P > 10\(^{-8}\) in controls). Finally, the data were pooled and final quality controls were performed on the pooled GWAS set, including checks for mississngness, duplicates, sex mismatch, abnormal heterozygosity, genetic relatedness, population outliers (through principal component [PC] analysis), and genomic inflation. After applying quality control measures to the imputed data, 5,864,648 SNPs remained for analysis.

2.2 | SNP identification and selection

We downloaded a list of all the SNPs associated with at least one human phenotype at P < 5 × 10\(^{-8}\) from the GWAS Catalog web site (https://www.ebi.ac.uk/gwas/). The list was downloaded in January 2020 and all SNPs without “rs” identifier (n = 1667) were excluded, leaving 66,296 unique SNPs for analysis.

2.3 | Sample preparation, genotyping and quality control in IMMEnSE

DNA samples from IMMEnSE consortium participants were extracted from whole blood and genotyped using TaqMan (Thermo Fisher Applied Biosystems, Waltham MA, USA) assay technology, according to the manufacturer’s recommendations. Genotyping was conducted in 384 well plates, including n = 203 duplicate samples (6%) for quality control purpose (concordance rate was higher than 98%). The distribution of cases and controls was unknown to the operator performing the genotyping. The fluorescent emission of the genotyping assay was detected by a QuantStudio 5 Real-Time PCR system (Thermofisher) and the genotyping calls were made with QuantStudio software (Thermofisher). The average call rate of the SNPs was 86.30%. Subjects with a call rate < 70% (331 cases and 216 controls) were excluded from further analysis. Pearson’s chi-square test (\(\chi^2\)) was performed to assess if genotype frequencies were in HWE. The analysis, restricted to controls, was performed overall and separately for each country. All SNPs were in HWE except for rs10187103

<table>
<thead>
<tr>
<th>Table 1: Study populations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis</strong></td>
</tr>
<tr>
<td>InterLymph</td>
</tr>
<tr>
<td>2434</td>
</tr>
<tr>
<td>IMMEnSE</td>
</tr>
<tr>
<td>1955</td>
</tr>
<tr>
<td>FinnGen</td>
</tr>
<tr>
<td>418</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td>5880</td>
</tr>
<tr>
<td>3504</td>
</tr>
<tr>
<td>147 700</td>
</tr>
<tr>
<td>157 084</td>
</tr>
<tr>
<td><strong>Median age (25%-75%)</strong></td>
</tr>
<tr>
<td>MM cases</td>
</tr>
<tr>
<td>61 (26-90)</td>
</tr>
<tr>
<td>Controls</td>
</tr>
<tr>
<td>51 (43-61)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>61%</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>39%</td>
</tr>
</tbody>
</table>

Note: Details on age and sex distribution of FinnGen individuals are not available.

\(^{a}\)Median age values of MM cases and controls with 25th and 75th percentile.
### RESULTS

#### Discovery

Among the 46,264 variants selected for testing, 4439 were not genotyped or imputed in the InterLymph MM GWAS and therefore could not be analyzed. Of the 59,852 remaining SNPs, 2979 SNPs were typed or imputed in the InterLymph MM GWAS and therefore could be analyzed.

Several bioinformatic tools were used to evaluate possible functional features of the SNPs associated with MM. The occurrence of regulatory motif alterations was investigated through HaploReg (https://haploreg.broadinstitute.org/haploreg/haploreg.php). The Genotype-Tissue Expression project (GTEx portal) was used to assess if selected SNPs are expression quantitative trait loci (eQTLs), that is, their alleles are associated with differential expression of genes in cis.

#### Annotation of functional effect of the SNPs

Several bioinformatic tools were used to evaluate possible functional features of the SNPs associated with MM. The occurrence of regulatory motif alterations was investigated through HaploReg (https://haploreg.broadinstitute.org/haploreg/haploreg.php). The Genotype-Tissue Expression project (GTEx portal) was used to assess if selected SNPs are expression quantitative trait loci (eQTLs), that is, their alleles are associated with differential expression of genes in cis.

#### 2.4 | Statistical analysis

Unconditional logistic regression analysis was performed using InterLymph data to assess the association between the pleiotropic SNPs and risk of developing MM, reporting odds ratios (ORs) with 95% confidence intervals (CIs). Analyses were adjusted for age at diagnosis/old) were then analyzed in the validation populations. Summary statistics were used for FinnGen, whereas for IMMEnSE, additive and co-dominant unconditional logistic regression analysis was performed, sex study site and the first five PCs. LD pruning (r = 0.8) was applied to eliminate SNPs representing the same locus.

#### TABLE 2

<table>
<thead>
<tr>
<th>SNP</th>
<th>Region (Alleles) M/m</th>
<th>InterLymph OR (95%CI)</th>
<th>P_value</th>
<th>IMMeNSE OR (95%CI)</th>
<th>P_value</th>
<th>FinnGen OR (95%CI)</th>
<th>P_value</th>
<th>Combined analysis OR (95%CI)</th>
<th>P_value</th>
<th>I^2</th>
<th>Q</th>
<th>Q-P_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs34517439</td>
<td>1p31.1 C/A</td>
<td>1.30 (1.15-1.48)</td>
<td>4.70 x 10^-5</td>
<td>1.09 (0.89-1.34)</td>
<td>0.402</td>
<td>1.21 (1.09-1.34)</td>
<td>0.071</td>
<td>1.22 (1.13-1.32)</td>
<td>4.81 x 10^-7</td>
<td>7.30%</td>
<td>2.25</td>
<td>0.325</td>
</tr>
<tr>
<td>rs6674512</td>
<td>1p22.3 G/A</td>
<td>1.40 (1.21-1.62)</td>
<td>8.71 x 10^-6</td>
<td>1.10 (0.88-1.36)</td>
<td>0.402</td>
<td>1.04 (0.91-1.20)</td>
<td>0.765</td>
<td>1.17 (0.96-1.43)</td>
<td>0.115</td>
<td>77.50%</td>
<td>9.74</td>
<td>0.008</td>
</tr>
<tr>
<td>rs10187103</td>
<td>2q24.3 C/T</td>
<td>0.84 (0.77-0.91)</td>
<td>6.06 x 10^-5</td>
<td>0.99 (0.88-1.12)</td>
<td>0.915</td>
<td>0.99 (0.90-1.09)</td>
<td>0.924</td>
<td>0.95 (0.83-1.08)</td>
<td>0.397</td>
<td>81.10%</td>
<td>10.38</td>
<td>0.006</td>
</tr>
<tr>
<td>rs1022206</td>
<td>3q13.3 C/T</td>
<td>1.18 (1.09-1.28)</td>
<td>8.61 x 10^-5</td>
<td>1.11 (1.00-1.24)</td>
<td>0.061</td>
<td>0.99 (0.92-1.07)</td>
<td>0.923</td>
<td>1.09 (0.97-1.22)</td>
<td>0.136</td>
<td>80.00%</td>
<td>10.25</td>
<td>0.006</td>
</tr>
<tr>
<td>rs4143832</td>
<td>5q31.1 G/T</td>
<td>1.24 (1.12-1.37)</td>
<td>4.66 x 10^-6</td>
<td>0.95 (0.83-1.08)</td>
<td>0.439</td>
<td>0.91 (0.84-0.99)</td>
<td>0.275</td>
<td>1.02 (0.84-1.25)</td>
<td>0.819</td>
<td>91.30%</td>
<td>22.96</td>
<td>0.003</td>
</tr>
<tr>
<td>rs537930</td>
<td>5q31.1 G/T</td>
<td>0.83 (0.76-0.91)</td>
<td>5.09 x 10^-5</td>
<td>0.97 (0.86-1.09)</td>
<td>0.571</td>
<td>0.96 (0.88-1.04)</td>
<td>0.956</td>
<td>0.92 (0.83-1.01)</td>
<td>0.088</td>
<td>70.10%</td>
<td>6.51</td>
<td>3.80 x 10^-2</td>
</tr>
<tr>
<td>rs1063348</td>
<td>6p21.32 G/A</td>
<td>1.19 (1.09-1.30)</td>
<td>7.44 x 10^-5</td>
<td>0.95 (0.85-1.07)</td>
<td>0.398</td>
<td>NA</td>
<td>NA</td>
<td>1.07 (0.87-1.32)</td>
<td>0.534</td>
<td>89.6%</td>
<td>14.49</td>
<td>0.001</td>
</tr>
<tr>
<td>rs13252276</td>
<td>8p23.1 C/T</td>
<td>0.77 (0.68-0.87)</td>
<td>3.10 x 10^-5</td>
<td>1.01 (0.91-1.12)</td>
<td>0.870</td>
<td>1.03 (0.96-1.11)</td>
<td>0.711</td>
<td>0.93 (0.79-1.10)</td>
<td>0.415</td>
<td>88.00%</td>
<td>17.50</td>
<td>2.00 x 10^-4</td>
</tr>
<tr>
<td>rs465530</td>
<td>12q14.3 G/T</td>
<td>1.18 (1.09-1.28)</td>
<td>4.94 x 10^-5</td>
<td>1.01 (0.98-1.03)</td>
<td>0.926</td>
<td>0.95 (0.88-1.02)</td>
<td>0.458</td>
<td>1.03 (0.90-1.19)</td>
<td>0.657</td>
<td>87.80%</td>
<td>17.06</td>
<td>2.00 x 10^-4</td>
</tr>
<tr>
<td>rs8132680</td>
<td>2q22.12 T/C</td>
<td>1.21 (1.11-1.33)</td>
<td>2.96 x 10^-5</td>
<td>0.99 (0.88-1.12)</td>
<td>0.924</td>
<td>1.04 (0.96-1.13)</td>
<td>0.646</td>
<td>1.08 (0.96-1.21)</td>
<td>0.197</td>
<td>77.40%</td>
<td>8.06</td>
<td>0.018</td>
</tr>
</tbody>
</table>

aM = the most common allele in controls; m = the less common allele in controls. The more common allele found in the control population has been used as reference and the other allele was defined as the exposure factor.

bMeasures of heterogeneity between the combined analysis components.

cThis polymorphism is not present in FinnGen, and therefore the combined analysis refers only to InterLymph and IMMeNSE populations.
developing MM or in LD ($r^2 > 0.8$) with these known SNPs were further excluded. Residual LD ($r^2 > 0.4$) between the remaining 16 SNPs was evaluated, resulting in the exclusion of an additional six SNPs. The remaining 10 SNPs were selected for evaluation in FinnGen and IMMEnSE. Figure S1 shows a flowchart of the process.

3.2 Replication and combined analysis

We performed the combined analysis using exclusively the allelic model, since this was available for the FinnGen dataset. We found a statistically significant association (using $P < 1.74 \times 10^{-5}$) between the A allele of DNAJB4-rs34517439 and an increased risk of developing MM (OR$_{\text{combined-analysis}} = 1.22$, 95%CI 1.13-1.32, $P = 4.81 \times 10^{-5}$) with very low study heterogeneity ($I^2 = 7.30\%$, $Q = 2.25$) (Table 2 and Figure 1). However, we observed some degree of heterogeneity in the combined analysis for the SNPs that did not show any statistically significant results, that could be due to the different direction of the association (Table 2) or to differences of the allelic frequencies between Finnish and non-Finnish Europeans (Table S2).

Though not statistically significant, we observed an association between LOC105374037-rs1022206 and risk of developing MM in the IMMEnSE population when comparing homozygotes for the rare allele (T/T) to homozygotes for the common allele (C/C) (OR$_{\text{homozygotes}} = 1.29$, 95%CI 1.03-1.62, $P = .028$) and in the recessive model (OR$_{\text{recessive}} = 1.28$, 95%CI 1.04-1.57, $P = .019$) in the IMMEnSE population (Table 3).

3.3 Functional effect of the SNPs

RegulomeDB assigned a rank of 5 to both SNPs, implying a possible influence on a transcription factor binding site. In addition, GTEx and HaploReg indicated that DNAJB4-rs34517439 is an eQTL affecting far upstream element binding protein 1 gene (FUBP1) expression in several human tissues. This included a statistically significant correlation between the A allele of this SNP and a reduced expression of FUBP1 in cultured fibroblasts ($P = 1.10 \times 10^{-4}$). GTEx also indicated an association between the T allele of LOC105374037-rs1022206 and increased expression of the nectin cell adhesion molecule 3 (NECTIN3) gene in cultured fibroblasts ($P = 2.10 \times 10^{-2}$).

4 DISCUSSION

Pleiotropic variants are commonly found associated with complex diseases, like cancer. To investigate the possible association between common SNPs and the risk of developing MM, we analyzed all the genetic variants associated with any human trait in 4807 patients and 152 277 controls. We observed a statistically significant association ($P < 1.74 \times 10^{-5}$) between DNAJB4-rs34517439 and risk of developing MM. This SNP is pleiotropic since it is associated with numerous traits such as cutaneous melanoma, diastolic blood pressure, fat-free mass, hair color, hand grip strength, height, lung cancer, noncognitive aspects of educational attainment, psoriasis, serum alkaline phosphatase levels, smoking initiation and BMI, with a wide range of effect sizes and $P$ values as reported in Table S3. The association between this SNP and BMI is of particular interest since BMI is also associated with MM risk.\(^{39}\) The A allele that was associated with increased MM risk in our study has previously been associated with increased risk of lung cancer.\(^{40}\) This pleiotropic effect, especially in cancers, may be attributed to the LD of rs34517439 with the variants inside or in the proximity of DNAJB4 heat shock protein family (Hsp40) member B4 (DNAJB4), situated on chromosome 1. This protein is a molecular chaperone specifically recognizing wild-type from mutant E-cadherin protein. Since E-cadherin is an important element in the suppression of tumor invasion, the function of DNAJB4 represents a significant pleiotropic mechanism of tumoral invasion inhibition.\(^{41,42}\) DNAJB4 also acts as a co-chaperone, forming a complex with Hsp70 which participates in protein folding.\(^{43}\) The expression and activity of heat shock proteins increases during cellular stress,\(^{44}\) and DNAJB4 has been observed to reduce tumor metastasis and progression in several cancer types, including lung carcinoma and melanoma.\(^{45,46}\) This mechanism could also be relevant to MM. In bioinformatic analysis, GTEx showed an association between DNAJB4-rs34517439-A and a reduced expression of the FUBP1 gene in various tissues, including lung, skin (sun exposed and not sun exposed), adipose tissue, the arteries and the heart. The altered expression of the gene in these tissues mirrors the associations found in the GWASs (eg, altered expression in lung for lung cancer, or skin for melanoma).\(^{47}\) FUBP1 encodes a multifunctional DNA and RNA-binding protein involved in cell cycle regulation and self-renewal of hematopoietic stem cells (HSCs).\(^{47}\) FUBP1 acts both as oncogene and as tumor suppressor, and its activity is tissue-dependent.\(^{48}\) In various tumor types FUBP1 promotes the overexpression of the MYC oncogene,\(^{48}\) whereas FUBP1 silencing does not influence the expression of MYC in normal
# TABLE 3  Associations between MM risk and the top 10 SNPs in IMMEnSE population

<table>
<thead>
<tr>
<th>SNP</th>
<th>Region</th>
<th>( Alleles )</th>
<th>M/m #</th>
<th>Allelic model</th>
<th>Codominant model</th>
<th>Dominant model</th>
<th>Recessive model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OR</td>
<td>95% CI</td>
<td>P-value</td>
<td>OR</td>
</tr>
<tr>
<td>rs34517439</td>
<td>1p31.1</td>
<td>C/A</td>
<td>1.09</td>
<td>0.89-1.34</td>
<td>.402</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>rs6674512</td>
<td>1p22.3</td>
<td>G/A</td>
<td>1.10</td>
<td>0.88-1.36</td>
<td>.402</td>
<td></td>
<td>1.14</td>
</tr>
<tr>
<td>rs10187103</td>
<td>2q24.3</td>
<td>C/T</td>
<td>0.99</td>
<td>0.88-1.12</td>
<td>.915</td>
<td></td>
<td>0.90</td>
</tr>
<tr>
<td>rs102206</td>
<td>3q13.13</td>
<td>C/T</td>
<td>1.11</td>
<td>1.00-1.24</td>
<td>.061</td>
<td></td>
<td>1.01</td>
</tr>
<tr>
<td>rs4143832</td>
<td>5q31.1</td>
<td>G/T</td>
<td>0.95</td>
<td>0.83-1.08</td>
<td>.439</td>
<td></td>
<td>0.94</td>
</tr>
<tr>
<td>rs537930</td>
<td>5q31.1</td>
<td>G/T</td>
<td>0.97</td>
<td>0.86-1.09</td>
<td>.571</td>
<td></td>
<td>0.86</td>
</tr>
<tr>
<td>rs1063348</td>
<td>6p21.32</td>
<td>G/A</td>
<td>0.95</td>
<td>0.85-1.07</td>
<td>.398</td>
<td></td>
<td>0.92</td>
</tr>
<tr>
<td>rs13252276</td>
<td>8p23.1</td>
<td>C/T</td>
<td>1.01</td>
<td>0.91-1.12</td>
<td>.870</td>
<td></td>
<td>1.02</td>
</tr>
<tr>
<td>rs465530</td>
<td>12q14.3</td>
<td>G/T</td>
<td>1.01</td>
<td>0.89-1.13</td>
<td>.926</td>
<td></td>
<td>0.83</td>
</tr>
<tr>
<td>rs8132680</td>
<td>21q22.12</td>
<td>T/C</td>
<td>0.99</td>
<td>0.88-1.12</td>
<td>.924</td>
<td></td>
<td>1.01</td>
</tr>
</tbody>
</table>

Note: In bold it is reported the statistically significant association.

# M = the more common allele in controls; m = the less common allele in controls. The more common allele found in the control population has been used as the reference and the other allele was defined as the exposure factor.

# Also termed quantitative additive (m vs M).

# Mm vs MM (het) and mm vs MM (hom).

# Mm + mm vs MM.

# mm vs Mm + MM.

# All the analyses were adjusted for age at diagnosis/recruitment, sex and country of origin. Results in bold show associations with $P < .05$. 
Moreover, FUBP1 is involved in the upregulation and downregulation of the cell cycle inhibitor p21. The pleiotropic role of rs34517439 could be explained by its involvement in the altered expression of FUBP1 in different tumors, that determines their development, by interacting with specific factors in each tissue. Specifically for hematopoietic lineages, the cooperation of FUBP1 with RUNX1 in promoting the expression of the c-KIT oncogene was reported. In summary, a possible explanation for the DNAJB4-rs34517439 association with increased risk of developing MM could be that the A allele modifies the expression of DNAJB4 leading to an incorrect folding of FUBP1. In turn, this mechanism could be responsible for the disruption of the HSCs homeostasis equilibrium, thus increasing the risk of developing MM. Although we lack an experimental validation of the proposed mechanism in MM cell lines, our hypothesis relies on the experimental and genomic data collected from GTEx, which are broadly used in the scientific community.

In addition, the T allele of LOC105374037-rs1022206 in homozygosity showed an increased risk of developing MM in the IMMEnSE population. LOC105374037 is an uncharacterized long noncoding RNA (lncRNA) for which possible functional roles are unknown. Interestingly, LOC105374037-rs1022206 is an eQTL for NECTIN3 gene expression in cultured fibroblasts. Fibroblasts are the cells with the highest expression of NECTIN3 in humans. Cancer-associated fibroblasts (CAF) are one of the known cellular elements participating in MM tumoral microenvironment in the BM. In fact, they show a bidirectional loop with myeloma cells providing chemotaxis, adhesion; apoptosis resistance and proliferation through cytokines, growth factors, angiogenic factors and cell-cell contact. However, considering the lack of data in the literature on the topic it is not clear how the increased expression of NECTIN3 mediated by LOC105374037-rs1022206 might be involved in MM risk. Furthermore, Nectin-3 is involved in cell survival through PDGF receptor signal and in inhibition of cell movement. The SNP LOC105374037-rs1022206 is in LD with variants associated in with various traits: balding type 1, smoking initiation, self-reported math ability, feeling hurt and neuroticism. Although these traits apparently do not show a shared mechanism or pathway, the LOC105374037 lncRNA could be the common element explaining the pleiotropic effect of the SNP on different cellular functions.

In conclusion, our results suggest the involvement of a pleiotropic region on chromosome 1 in MM development and highlight pleiotropy as an approach to uncover additional risk variants in cancer susceptibility.

**AUTHOR CONTRIBUTIONS**

Daniele Campa conceived and designed the study. Marco Dicanio and Matteo Giaccherini performed the lab work, Marco Dicanio, Matteo Giaccherini and Alyssa Clay-Gilmour analyzed the data. All the authors contributed with the interpretation of the data. Marco Dicanio, Matteo Giaccherini and Daniele Campa wrote the first draft of the manuscript and all authors contributed to the writing and approve of the final version of the manuscript. The work reported in the paper has been performed by the authors, unless clearly specified in the text.

**AFFILIATIONS**

1. Department of Biology, University of Pisa, Pisa, Italy
2. Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Greenville, South Carolina, USA
3. Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany
4. Genomic Oncology Area, GENYO. Center for Genomics and Oncological Research: Pfizer, University of Granada/Andalusian Regional Government, Granada, Spain
5. Department of Hematology, Virgen de las Nieves University Hospital, Granada, Spain
6. Department of Medicine, University of Granada, Granada, Spain
7. Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA
8. Department of Hematology Specialist Hospital No. 1, Bytom, Poland
9. Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Ontario, USA
10. Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Ontario, USA
11. Department of Hematology and Transplantology, Medical University of Gdańsk, Gdańsk, Poland
12. Department of Hematology, Rigshospitalet, Copenhagen, Denmark
13. Division of Hematology, Department of Internal Medicine, Mayo Clinic, Rochester, Ontario, USA
ACKNOWLEDGMENT

Open Access Funding provided by Universita degli Studi di Pisa within the CRUI-CARE Agreement.

FUNDING INFORMATION

This study was partially supported by intramural funds of University of Pisa and DKFZ.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The primary data for this work will be made available to researchers who submit a reasonable request to the corresponding author, conditional to approval by the centers participating in this study. Data will be stripped from all information allowing identification of study participants.

ETHICS STATEMENT

The IMMEnSE study protocol was approved by the Ethics Committee of the Medical Faculty of the University of Heidelberg (reference number: S-004/2020). Following the guidelines of the Declaration of Helsinki, written informed consent was obtained from each participant. For InterLymph, contributing studies were approved by local ethics review committees. Summary statistics were used from the Finngen study, that was approved by the ethical Review Board of the Hospital District of Helsinki and Uusimaa. FinnGen participants provided written, informed consent.


15. Rand KA, Song C, Dean E, et al. A meta-analysis of multiple myeloma...


17. Shikov AE, Skitchenko RK, Predeus AV, Barbitoff YA. Phenome-wide functional dissection of pleiotropic effects highlights key molecular pathways for human complex traits.

18. Pierce BL, Ahsan H. Genome-wide “pleiotropy scan” identifies HNF1A region as a novel pancreatic cancer susceptibility locus.

REFERENCES


SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.