Identification of miRSNPs associated with the risk of multiple myeloma

Angelica Macauda[1,2], Diego Calvetti[1,2], Giuseppe Maccari[3], Kari Hemminki[4], Asta Försti[4], Hartmut Goldschmidt[5], Niels Weinhold[5], Richard Houlston[6], Vibeke Andersen[7,8], Ulla Vogel[9], Gabriele Buda[10], Judit Varkonyi[11] Anna Sureda[12], Joaquin Martinez Lopez[13], Marzena Watek[14], Aleksandra Butrym[15], Maria Eugenia Sarasquete[16], Marek Dudziński[17], Artur Jurczyszyn[18], Agnieszka Druzd-Sitek[19] Marcin Kruszewski[20], Edyta Subocz[21], Mario Petrini[10], Elzbieta Iskierka-Jażdżewska[22], Malgorzata Raźny[23], Gergely Szombath[11], Herlander Marques[24], Daria Zawirska[18], Dominik Chraniuk[25], Janusz Halka[21], Svend Erik Hove Jacobsen[26], Grzegorz Mazur[27], Ramón García Sanz[16], Charles Dumontet[28], Victor Moreno[29], Anna Stępień[30], Katia Beider[31], Matteo Pelosini[32], Rui Manuel Reis[24,33], Malgorzata Krawczyk-Kulis[34], Marcin Rymko[25], Hervé Avet-Loiseau[35], Fabienne Lesueur[36,37,38,39], Norbert Grząśko[40], Olga Ostrovsky[31], Krzysztof Jamroziak[41], Annette J. Vangsted[42], Andrés Jerez[43], Waldemar Tomczak[44], Jan Maciej Zaucha[45], Katalin Kadar[11], Juan Sainz Pérez[46], Arnon Nagler[31], Stefano Landi[1], Federica Gemignani[1*], Federico Canzian[2*]

[1] Department of Biology, University of Pisa, Pisa, Italy

[2] Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany

[3] The Pirbright Institute, Pirbright, Woking, Surrey, GU24 0NF, United Kingdom

[4] Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany

[5] Department of Internal Medicine V, University of Heidelberg, Heidelberg, Germany

[6] Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK

[7] Institute of Regional Health Research, and Institute of Molecular Medicine, Faculty of Health Sciences, University of Southern Denmark, Odense C, Denmark

[8] Research Unit of Molecular Diagnostics and Clinical Research, Laboratory Center, Hospital of Southern Jutland, Aabenraa, Denmark

[9] National Research Centre for the Working Environment, DK-2100 Copenhagen, Denmark

[10] UO Hematology, Department of Internal and Experimental Medicine, University of Pisa, Pisa, Italy

[11] Department of Hematology, Semmelweis University, Budapest, Hungary

[12] Hematology Department, Catalan Institute of Oncology (ICO) and IDIBELL, Barcelona, Spain

[13] Department of Hematology. Hospital Universitario 12 de Octubre, Complutense School of Medicine, CNIO, Madrid, Spain

[14] Holycross Medical Center, Kielce, Poland

[15] Medical University, Wrocław, Poland

[16] Department of Hematology, University Hospital of Salamanca, Salamanca, Spain

[17] Department of Hematology, Specialist District Hospital, Rzeszow, Poland

[18] Department of Hematology, Cracow University Hospital, Cracow, Poland

[19] Department of Lymphoid Malignancies Maria Sklodowska-Curie Memorial Institute and Oncology Centre Warsaw, Warsaw, Poland

[20] Department of Hematology, University Hospital, Bydgoszcz, Poland

[21] Department of Hematology, Military Institute of Medicine, Warsaw, Poland

[22] Department of Hematology, Medical Universityof Lodz, Łódź, Poland

[23] Department of Hematology, Rydygier Hospital, Cracow, Poland

[24] Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal and ICVS/3B's – PT Government Associate Laboratory, Braga/Guimarães, Portugal

[25] Department of Hematology, Copernicus Hospital, Torun, Poland

[26] Clinic of Immunology, Laboratory Center, Hospital of Southern Jutland, Sønderborg, Denmark

[27] Department of Internal Diseases, Hypertension and Occupational Medicine, Medical University,

Wroclaw, Poland

[28] INSERM UMR 1052 / CNRS 5286, Université Claude Bernard Lyon I, Lyon, 69622, France

[29] Cancer Prevention and Control Program, Catalan Institute of Oncology (ICO), IDIBELL, CIBERESP and Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain

[30] Laboratory of Clinical and Transplant Immunology and Genetics, Copernicus Memorial Hospital, Łódź, Poland

[31] Chaim Sheba Medical Center, Tel-Hashomer, Israel

[32] U.O Dipartimento di Ematologia, Azienda USL Toscana Nord Ovest, Livorno

[33] Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, Brazil

[34] Department of Hematology and Bone marrow Transplantation, Silesian Medical University, Katowice, Poland

[35] Laboratoire d'hématologie, Pôle biologie, Institut Universitaire du Cancer de Toulouse – Oncopole 1, avenue Irène Joliot-Curie - 31059 Toulouse, France

[36] Institut Curie, 26 rue d'Ulm, F-75005 Paris, France

[37] PSL Research University, F-75005 Paris, France

[38] Inserm, U900, F-75005, Paris France

[39] Mines Paris Tech, F-77305 cedex Fontainebleau, France

[40] Hematology Department, St. John's Cancer Center, Lublin, Poland

[41] Department of Hematology, Institute of Hematology and Transfusion Medicine, Warsaw, Poland

[42] Department of Haematology, Rigshospitalet, Copenhagen University, Copenhagen, Denmark

[43] Hematology and Medical Oncology Department, University Hospital Morales Meseguer, IMIB, Murcia, Spain

[44] Department of Hematooncology and Bone Marrow Transplantation, Medical University of Lublin, Lublin, Poland

[45] Department of Hematology, Sea Hospital, Gdynia, Poland

[46] Centro Pfizer - Universidad de Granada - Junta de Andalucía de Genómica e Investigación Oncológica (GENYO), Granada, Spain

[*] These authors contributed equally

Running title: Identification of miRSNPs associated with the risk of multiple myeloma

Key words: Multiple Myeloma, miRNA, SNP, genetic susceptibility

Financial support: This work was partially funded by: intramural funds of German Cancer Research Center (DKFZ), Grant ref. HUS412A12-1 from the "Gerencia Regional de Salud de la Junta de Castilla y Léon". This work was supported by grants from the Instituto de Salud Carlos III (Madrid, Spain; PI12/02688). Catalan Government DURSI grant 2014SGR647 and Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– grants PI11-01439 and PIE13/00022.

Corresponding author:

Dr. Federico Canzian

Genomic Epidemiology Group,

German Cancer Research Center (DKFZ),

Im Neuenheimer Feld 280

69120 Heidelberg

Heidelberg, Germany

f.canzian@dkfz.de

Abstract

Multiple myeloma (MM) is a malignancy of plasma cells usually infiltrating the bone marrow, associated with the production of a monoclonal immunoglobulin (M protein) which can be detected in the blood and/or urine. Multiple lines of evidence suggest that genetic factors are involved in MM pathogenesis, and several studies have identified single nucleotide polymorphisms (SNPs) associated with the susceptibility to the disease. SNPs within miRNA-binding sites in target genes (miRSNPs) may alter the strength of miRNA–mRNA interactions, thus deregulating protein expression. MiRSNPs are known to be associated with risk of various types of cancer, but they have never been investigated in MM. We performed an *in silico* genome-wide search for miRSNPs predicted to alter binding of miRNAs to their target sequences. We selected 12 miRSNPs and tested their association with MM risk. Our study population consisted of 1,832 controls and 2,894 MM cases recruited from 7 European countries and Israel in the context of the IMMEnSE (International Multiple Myeloma rESEarch) consortium. In this population two SNPs showed an association with $p<0.05$: rs286595 (located in gene *MRLP22*) and rs14191881 (located in gene *TCF19*). Results from IMMEnSE were meta-analyzed with data from a previously published genome-wide association study (GWAS). The SNPs rs13409 (located in the 3'UTR of the *POU5F1* gene), rs1419881 (*TCF19*), rs1049633, rs1049623 (both in *DDR1*) showed significant associations with MM risk. In conclusion, we sought to identify genetic polymorphisms associated with MM risk starting from genome-wide prediction of miRSNPs. For some mirSNPs we have shown promising associations with MM risk.

What's new? Even though deregulation of miRNA expression has been associated with human cancers little information is available regarding their relation with MM susceptibility. We performed an *in silico* genome-wide search for miRSNPs and selected the most promising ones for an association study. The SNPs with the strongest associations with MM risk are localized in genes which have never been related with MM.

Introduction

Multiple myeloma (MM) is a malignancy of plasma cells, usually infiltrating the bone marrow, and associated with the production of a monoclonal immunoglobulin (M protein) which can be detected in the blood and/or urine. MM affects the places where bone marrow is normally active in an adult. MM is a relatively infrequent cancer among both sexes. On a worldwide scale, it is estimated that about 86,000 incident cases occur annually, accounting for about 0.8% of all new cancer cases. About 63,000 subjects are reported to die from the disease each year, accounting for 0.9% of all cancer deaths and nearly 10% of all haematological neoplastic diseases.

Several factors are known or suspected to cause myeloma or trigger an already abnormal or damaged pre-myeloma cell population in the bone marrow. Exposure to toxic chemicals, ionizing radiations, immunodeficiency, or infection with cancer-causing viruses have all been implicated as causes or triggers of MM[1]. Converging evidence of MM in monozygotic twins and familial aggregation of MM strongly suggest that MM aetiology has a robust genetic component as well[2]. Several risk *loci* have been proposed and a few have been identified through candidate gene and genome-wide association studies (GWAS)[3–10]. Some of these *loci* are involved in complex pathways related to cell cycle, cell proliferation and DNA repair, in which micro-RNAs (miRNAs) have a proven regulatory role[11].

MiRNAs are small non-coding RNA molecules, 20-25 nucleotides long, highly conserved throughout evolution. In mammals, miRNAs are predicted to control the activity of more than 60% of all protein-coding genes and participate in the regulation of almost every cellular process investigated to date[12]. They play a major role in post-transcriptional regulation processes, mainly silencing target mRNAs and thus decreasing their corresponding protein expression[13].

These small RNAs post-transcriptionally repress gene expression by recognizing complementary target sites most often in the 3' untranslated region (UTR) of target messenger RNAs (mRNAs). Several miRNAs were found to be directly involved in human cancers, including lung, breast, brain, liver, colon cancer and leukemia. In addition, some miRNAs may function as oncogenes or tumor

suppressors in tumor development. Furthermore, a widespread down-regulation of miRNAs is commonly observed in human cancers, which promotes cellular transformation and tumorigenesis[14–17]. Recent studies have revealed a role of miRNAs in MM. Deregulated miRNA expression in plasma cells has been associated with tumor progression, molecular subtypes, clinical staging, prognosis, and drug response in MM[18].

Although mutation in miRNA seed sequence seems to be a rare event, sequence variation in miRNA target sites is relatively frequent and may play a role in cancer etiology. *In silico* analyses of expressed sequence tag SNP databases indicate different allele frequencies of miRNA-binding sites in tumors versus normal tissues[19]. Polymorphisms in miRNA-binding sites in target genes may alter the strength of miRNA–mRNA interactions, thus deregulating protein expression. SNPs belonging to this category are called miRSNPs[20]. MiRSNPs have been shown to be associated with the risk of several cancers [21–23] .

We hypothesize that miRSNPs may have a role in the susceptibility of MM. To this aim, we performed an *in silico* genome-wide search of miRSNPs predicted to affect binding of micro-RNAs to their target genes and tested the most promising ones in a case-control association study.

Materials and methods

Study population

The study population consisted of 1,832 controls and 2,894 MM cases recruited from 7 European countries and Israel in the context of the IMMEnSE (International Multiple Myeloma rESEarch) consortium (table 1)[24] .

Diagnosis of patients with symptomatic MM was carried out by haematologists according to the International Myeloma Working Group (IMWG) criteria. Demographic and clinicopathological characteristics including age, gender, country of origin, disease stage (Durie-Salmon and/or International Staging System) and serum creatinine levels were retrospectively gathered from medical records in each participant institution.

Controls were selected among the general population (Italian), blood donors (Danish, Spanish, Polish, Portuguese and French), and hospitalized subjects with different diagnoses excluding cancer (Hungarian, Spanish). Gender and age at recruitment were collected for every subject enrolled. Cases and controls are not matched individually but age distribution and sex ratio were similar between cases and controls (table 1). Control samples have been collected in the same centers of the cases, or at least in the same geographic areas.

In keeping with the guidelines of the Declaration of Helsinki, written informed consent was obtained from each participant and approval for collection and use of the samples was obtained from local Institutional Review Boards.

SNP selection criteria

A search for mirSNPs was carried out on the whole genome. Polymorphisms were selected using bioinformatic tools and consultation of results of a previously published GWAS on MM risk. The selection of miRSNPs was conducted with the following criteria (Figure 1):

- location in 3'UTR of known genes
- high conservation score, $\geq +1$

- MAF (minor allele frequency) ≥ 0.05 in Caucasians
- the strength of binding between each miRNA and its target sequence with the major and the minor allele (measured *in silico*), using mrSNP [25]
- location in genes that are expressed in MM cells, using Oncomine [26]

Putative miRNA-binding sites within the 3'UTR (defined as transcribed sequences from the stop codon to the end of the last exon of each gene) were detected by jSNPSelector version 1.0.1.7 (http://sourceforge.net/projects/jsnpselector/), a custom-made algorithm that interrogates the ENSEMBL database. The conservation score is assigned to each nucleotide in a multiple species alignment to determine how conserved the nucleotide is.

The strength of binding with the major and the minor allele was measured *in silico* using the publicly available mrSNP software (http://mrsnp.osu.edu). mrSNP showed a good performance to predict SNPs experimentally validated to affect miRNA binding, correctly identifying 69% (11/16) of the SNPs disrupting binding[25].

The selection with the mentioned workflow resulted in a list of 394 SNPs with high difference in Gibbs free energy between the two alleles (ΔΔG) (Supplementary table 1). For mirSNPs predicted to bind more than one miRNA, we summed up all the ΔΔGs for each miRNA, obtaining thus a total ΔΔG.

As a final step, we verified if these miRSNPs were associated with MM risk in a previously published GWAS data set[4,6].

Based on the process described above, we started from 2,817 SNPs, and we finally selected 12 SNPs reported in table 2. The final list included 3 SNPs (rs1050239, rs735794 and rs1052536) with highest predicted difference between the two alleles in the strength of binding between miRNAs and their target sequence (total ΔΔG), regardless of association in the previously performed GWAS, with a threshold of total ΔΔG>550 (another SNP, rs13505, had ΔΔG>550 as shown in supplementary table 1, but it could not be designed as TaqMan assay and was therefore excluded). Three more SNPs

(rs1803275, rs12509103, rs3197716) had both a high total ΔΔG (ΔΔG>300) and $p<0.05$ for association with MM risk in the GWAS. The other 6 SNPs were selected according to the GWAS p-value criterion, regardless of total ΔΔG, with an arbitrary threshold of $p<0.005$.

SNP genotyping, quality control and data filtering

Genomic DNA was extracted from peripheral blood or whole blood of MM cases and controls, using QIAampR 96 DNA QIAcubeR HT Kit. A whole genome amplification was carried out on samples with low levels of DNA. All the genotyping assays were carried out in 384-well format, with 10 ng of DNA from each subject. The order of DNAs from cases and controls was randomized on plates in order to ensure that an equal number of cases and controls was analyzed simultaneously. For quality control purpose, duplicates of 10% of the samples were interspersed throughout the plates. Genotyping of the 12 selected miRSNPs was carried out by using the TaqMan Real Time PCR method. The quality of the genotype data was assessed on the basis of sample call rates (we accepted samples with at least 8 out of 12 SNPs successfully genotyped), concordance rates between duplicate DNAs (≥99%) and test for Hardy-Weinberg Equilibrium (HWE) among controls in each population. We applied a Bonferroni correction for multiple testing to calculate the significance threshold for HWE: $0.05/(12 \text{ SNPs} \times 8 \text{ countries}) = 5.2 \times 10^{-4}$.

Statistical analysis

Analysis of association between SNPs and MM risk was performed with multivariate logistic regression models, adjusting for a set of covariates including age (at diagnosis for MM cases, at recruitment for controls), gender and country of origin. The association between miRSNPs and MM risk was calculated by estimating odds ratios (OR) and their 95% confidence intervals (C.I.). For all genotypes, we performed a statistical analysis with the allelic, dominant, codominant and recessive models. Since 3 of our SNPs (rs1049633, rs1049628, rs1049623) were located in the same gene *DDR1* (r^2 ranging from 0.377 to 0.237 in Caucasians) and two more SNPs (rs1419881 and rs13409)

are physically close to each other (r^2=0.887), we calculated also M_{EFF} correction. M_{EFF} is a simple correction for multiple testing of SNPs in linkage disequilibrium (LD) with each other, based on the spectral decomposition of matrices of pairwise LD between SNPs[27]. This method provides a useful alternative to more computationally intensive permutation tests. For performing this correction we used an interface available online (http://gump.qimr.edu.au/general/daleN/SNPSpD). We found a M_{EFF} = 11.50.

A P-value below 0.001, calculated with the formula 0.05/(11.5x4) was considered as threshold of statistical significance, considering the Bonferroni correction for M_{EFF} and the number of inheritance models tested.

Finally, we performed a meta-analysis between the data obtained in IMMEnSE and the data from two GWAS, respectively conducted in German and English populations[4,6] . We considered the different countries participating in IMMEnSE and in the GWAS as separate groups and carried out meta-analyses according to the fixed effects model.

All the statistical analyses were carried out using STATA software, version 11 for Windows.

Results

IMMeNSE. As showed in table 3, we did not find strong associations between SNPs and MM risk. Table 3 reports the results of the analyses performed according to an allelic and codominant inheritance model. Two SNPs showed a p-value < 0.05: rs286595 (located in gene *MRLP22*) and rs14191881 (located in gene *TCF19*). Analyses of recessive and dominant models did not show any association at the level of 0.05 or below (data not shown). For all the analyses none of the SNPs passed the threshold of significance after correction for multiple testing.

Meta-analysis. We performed a meta-analysis between the results from IMMEnSE and the published data of two GWASs conducted in the English and the German populations, respectively, for a total of 1,675 MM cases and 5,903 controls. Complete results are shown in supplementary table 2. We found p-values $<10^{-3}$ for the SNPs rs13409, rs1049623, rs1049633 and rs1419881 according to allelic or co-dominant model and no or minor heterogeneity, suggesting an association with a decreased risk of MM (table 4).

Discussion and conclusion

The first steps of our study were the selection of the SNPs and the genotyping of cases and controls in the context of the IMMEnSE consortium. Although the sample size was adequate, none of the statistical analyses passed the significance threshold after the correction for multiple testing. However SNPs rs286595 (*MRPL22)* and rs14191881 (*TCF19*) showed a p-value < 0.05. The latter came up also in the meta-analysis, as did the SNPs rs1049633, rs1049623 (both located in the 3'UTR of the *DDR1*) and rs13409 (*POU1F5*), which all showed an association with a decreased risk of MM, with no heterogeneity between IMMEnSE and the GWAS except for one case (rs1049633, $p_{heterogeneity}$=0.038).

The protein encoded by the gene *DDR1* belongs to a subfamily of tyrosine kinase receptors that are activated by various types of collagen. Upregulation of DDR1 in multiple human cancers implies that DDR1 is involved in tumor progression [28–30]. Four different miRNAs are predicted to bind *DDR1* at the location of rs1049633: miR-2355-3p, miR-7, miR-3915, miR-4689. When A substitutes G, miR-2355-3p and miR-4689 are predicted to bind more tightly to the *DDR1* 3'UTR, instead, for miR-7 and miR-3915 the A allele is predicted to decrease its binding affinity. MicroRNA-7 (miR-7) acts as a potential tumor suppressor, but the opposite effect also has been reported[31]. MiR-7 regulates diverse fundamental biological processes of cancer cells including initiation, proliferation, migration, invasion, survival and death by targeting a number of oncogenic signaling pathways. Two miRNAs are predicted to bind the *DDR1* gene where the SNP rs1049623 is located: miR-4499 and miR-4513. We found out that carriers of C allele show a decreased risk of MM. The binding force of miRNAs for this SNP is predicted to be stronger when this allele is present, therefore it may be that when a C is present *DDR1* is down-regulated, whereas usually this gene is up-regulated in cancer [32]. Thus, the association we observed is consistent with the known biological function of DDR1.

SNP rs1419881 is located in the gene *TCF19* (transcription factor 19), which encodes a protein containing a PHD-type zinc finger domain and likely functions as a transcription factor. *TCF19* was found to be involved in cell cycle progression and proliferation in the pancreatic β-cell and it plays a

role in the pathogenesis of diabetes[33]. Type 2 diabetes has been frequently associated with MM and it is thought to influence the myelomagenesis through hyperglycaemia and insulin-dependent and -independent mechanisms[34]. Moreover, rs1419881 is in linkage disequilibrium (r^2=0.82) with the SNP rs3130453 which is located in the gene *CCHCR*, also known as *HCR* (Supplementary table 2). The detailed function of the gene is still largely unknown. In particular, rs3130453 creates either a codon for tryptophan (G) or a stop codon (A). The stop codon (allele *Iso3*) results in the shorter isoform whereas the codon for tryptophan enables the usage of an earlier translation start site in exon 1b, thus leading to a protein with 89 additional amino acids in its N-terminal domain[35]. The most significant association in our analyses was found in the codominant model (p = 0.0002), where carriers of AA genotype have a decreased risk of MM.

The gene *POU5F1,* where rs13409 is located, encodes a transcription factor containing a POU homeodomain that plays a key role in embryonic development and stem cell pluripotency. Interestingly, POU5F1 is also named as OCT-4 and it is one of the four main factors involved in the formation of the induced pluripotent stem cells[36]. It has a pivotal role in the maintenance of the differentiation status of, in practice, all types of cells and its regulation could be very important also for maintaining the appropriate differential state in B lymphocytes.

This study has several strengths: large sample size and systematic, genome-wide search for miRSNPs. The genome-wide approach ensures that the whole set of common miRSNPs has been considered for this study. We cannot exclude that rarer miRSNPs have a role in MM risk, but a much larger sample size would be needed to address this issue.

Predictions of miRNA-mRNA binding were carried out only with *in silico* tools and not experimentally confirmed. This could be a weakness of the study. Moreover, the vast majority of study subjects were Caucasians, thus it is not clear if these results can be generalized to other populations as well. Finally, the associations were largely driven by the German component of the GWAS results, and were weaker in IMMEnSE and in the English GWAS. Indeed, when the German population is excluded from the meta-analysis none of the SNPs were significant associated with MM

risk considering a Bonferroni-corrected threshold, although rs286595 was associated with $p<0.05$ (supplementary table 3). Thus, the results of the meta-analysis should be taken with caution. However, we did not observe statistically significant heterogeneity among the IMMEnSE subgroups and the GWAS populations (except for rs1049633, $p_{heterogeneity}=0.038$).

In conclusion, we found promising associations between mirSNPs in *DDR1, TCF19, POU5F1* and MM risk, which should be further replicated in independent studies. If these associations are confirmed, it would be interesting to test the function of the SNPs with *in vitro* and *in vivo* studies.

Conflicts of interest. Vibeke Andersen received compensation as a member of an advisory board (Merck/MSD) and as a consultant (Merck/MSD and Janssen). Hartmut Goldschimidt has potential conflicts of interest as follow: Research support (Celgene, Janssen, Chugai, Novartis, BMS, Millenium, Morphosys), Advisory Boards (Janssen, Celgene, Novartis, Onyx, Amgen Takeda, BMS), Speakers Bureau (Celgene, Janssen, Novartis, Chugai, Onyx, Millenium). Jan Zaucha received a travel grant from Celgene.

References

1. Gold LS, Milliken K, Stewart P, Purdue M, Severson R, Seixas N, Blair A, Davis S, Hartge P, De Roos AJ. Occupation and multiple myeloma: an occupation and industry analysis. *Am J Ind Med* 2010;53:768–79.

2. Kristinsson SY, Björkholm M, Goldin LR, Blimark C, Mellqvist U-H, Wahlin A, Turesson I, Landgren O. Patterns of hematologic malignancies and solid tumors among 37,838 first-degree relatives of 13,896 patients with multiple myeloma in Sweden. *Int J Cancer* 2009;125:2147–50.

3. Martino A, Campa D, Buda G, Sainz J, García-Sanz R, Jamroziak K, Reis RM, Weinhold N, Jurado M, Ríos R, Szemraj-Rogucka Z, Marques H, Szemraj J, Stein A, Kumar R, Orciuolo E, Gemignani F, Landi S, Goldschmidt H, Petrini M, Dumontet C, Canzian F, Rossi AM. Polymorphisms in xenobiotic transporters ABCB1, ABCG2, ABCC2, ABCC1, ABCC3 and multiple myeloma risk: a case-control study in the context of the International Multiple Myeloma rESEarch (IMMEnSE) consortium. *Leukemia* 2012;26:1419–22.

4. Broderick P, Chubb D, Johnson DC, Weinhold N, Försti A, Lloyd A, Olver B, Ma YP, Dobbins SE, Walker BA, Davies FE, Gregory WA, Child JA, Ross FM, Jackson GH, Neben K, Jauch A, Hoffmann P, Mühleisen TW, Nöthen MM, Moebus S, Tomlinson IP, Goldschmidt H, Hemminki K, Morgan GJ, Houlston RS. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet* 2012;44:58–61.

5. Weinhold N, Johnson DC, Chubb D, Chen B, Försti A, Hosking FJ, Broderick P, Ma YP, Dobbins SE, Hose D, Walker BA, Davies FE, Kaiser MF, Li NL, Gregory WA, Jackson GH, Witzens-Harig M, Neben K, Hoffmann P, Nöthen MM, Mühleisen TW, Eisele L, Ross FM, Jauch A, Goldschmidt H, Houlston RS, Morgan GJ, Hemminki K. The CCND1 c.870G>A polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. *Nat Genet* 2013;45:522–5.

6. Chubb D, Weinhold N, Broderick P, Chen B, Johnson DC, Försti A, Vijayakrishnan J, Migliorini G, Dobbins SE, Holroyd A, Hose D, Walker BA, Davies FE, Gregory WA, Jackson GH, Irving JA, Pratt G, Fegan C, Fenton JAL, Neben K, Hoffmann P, Nöthen MM, Mühleisen TW, Eisele L, Ross FM, Straka C, Einsele H, Langer C, Dörner E, Allan JM, Jauch A, Morgan GJ, Hemminki K, Houlston RS, Goldschmidt H. Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nat Genet* 2013;45:1221–5.

7. Campa D, Martino A, Sainz J, Buda G, Jamroziak K, Weinhold N, Vieira Reis RM, García-Sanz R, Jurado M, Ríos R, Szemraj-Rogucka Z, Marques H, Lesueur F, Bugert P, Moreno V, Szemraj J, Orciuolo E, Gemignani F, Rossi AM, Dumontet C, Petrini M, Goldschmidt H, Landi S, Canzian F. Comprehensive investigation of genetic variation in the 8q24 region and multiple myeloma risk in the IMMEnSE consortium. *Br J Haematol* 2012;157:331–8.

8. Campa D, Martino A, Varkonyi J, Lesueur F, Jamroziak K, Landi S, Jurczyszyn A, Marques H, Andersen V, Jurado M, Brenner H, Petrini M, Vogel U, García-Sanz R, Buda G, Gemignani F, Ríos R, Vangsted AJ, Dumontet C, Martínez-López J, Moreno MJ, Stępień A, Wątek M, Moreno V, Dieffenbach AK, Rossi AM, Butterbach K, Jacobsen SEH, Goldschmidt H, Sainz J, Hillengass J, Orciuolo E, Dudziński M, Weinhold N, Reis RM, Canzian F. Risk of multiple myeloma is associated with polymorphisms within telomerase genes and telomere length. *Int J Cancer J Int Cancer* 2015;136:E351–8.

9. Swaminathan B, Thorleifsson G, Jöud M, Ali M, Johnsson E, Ajore R, Sulem P, Halvarsson B-

M, Eyjolfsson G, Haraldsdottir V, Hultman C, Ingelsson E, Kristinsson SY, Kähler AK, Lenhoff S, Masson G, Mellqvist U-H, Månsson R, Nelander S, Olafsson I, Sigurðardottir O, Steingrimsdóttir H, Vangsted A, Vogel U, Waage A, Nahi H, Gudbjartsson DF, Rafnar T, Turesson I, Gullberg U, Stefánsson K, Hansson M, Thorsteinsdóttir U, Nilsson B. Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. *Nat Commun* 2015;6:7213.

10. Mitchell JS, Li N, Weinhold N, Försti A, Ali M, van Duin M, Thorleifsson G, Johnson DC, Chen B, Halvarsson B-M, Gudbjartsson DF, Kuiper R, Stephens OW, Bertsch U, Broderick P, Campo C, Einsele H, Gregory WA, Gullberg U, Henrion M, Hillengass J, Hoffmann P, Jackson GH, Johnsson E, Jöud M, Kristinsson SY, Lenhoff S, Lenive O, Mellqvist U-H, Migliorini G, Nahi H, Nelander S, Nickel J, Nöthen MM, Rafnar T, Ross FM, da Silva Filho MI, Swaminathan B, Thomsen H, Turesson I, Vangsted A, Vogel U, Waage A, Walker BA, Wihlborg A-K, Broyl A, Davies FE, Thorsteinsdottir U, Langer C, Hansson M, Kaiser M, Sonneveld P, Stefansson K, Morgan GJ, Goldschmidt H, Hemminki K, Nilsson B, Houlston RS. Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nat Commun* 2016;7:12050.

11. Morgan GJ, Johnson DC, Weinhold N, Goldschmidt H, Landgren O, Lynch HT, Hemminki K, Houlston RS. Inherited genetic susceptibility to multiple myeloma. *Leukemia* 2014;28:518–24.

12. Lee RC, Ambros V. An extensive class of small RNAs in Caenorhabditis elegans. *Science* 2001;294:862–4.

13. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116:281–97.

14. Melo SA, Esteller M. Dysregulation of microRNAs in cancer: playing with fire. *FEBS Lett* 2011;585:2087–99.

15. Jacobsen A, Silber J, Harinath G, Huse JT, Schultz N, Sander C. Analysis of microRNA-target interactions across diverse cancer types. *Nat Struct Mol Biol* 2013;20:1325–32.

16. Chen X, Lu P, Wu Y, Wang D-D, Zhou S, Yang S-J, Shen H-Y, Zhang X-H, Zhao J-H, Tang J-H. MiRNAs-mediated cisplatin resistance in breast cancer. *Tumour Biol J Int Soc Oncodevelopmental Biol Med* 2016;

17. Tagscherer KE, Fassl A, Sinkovic T, Richter J, Schecher S, Macher-Goeppinger S, Roth W. MicroRNA-210 induces apoptosis in colorectal cancer via induction of reactive oxygen. *Cancer Cell Int* 2016;16:42.

18. Bi C, Chng WJ. MicroRNA: important player in the pathobiology of multiple myeloma. *BioMed Res Int* 2014;2014:521586.

19. Cipollini M, Landi S, Gemignani F. MicroRNA binding site polymorphisms as biomarkers in cancer management and research. *Pharmacogenomics Pers Med* 2014;7:173–91.

20. Dzikiewicz-Krawczyk A. MicroRNA polymorphisms as markers of risk, prognosis and treatment response in hematological malignancies. *Crit Rev Oncol Hematol* 2015;93:1–17.

21. Naccarati A, Rosa F, Vymetalkova V, Barone E, Jiraskova K, Di Gaetano C, Novotny J, Levy M, Vodickova L, Gemignani F, Buchler T, Landi S, Vodicka P, Pardini B. Double-strand break repair and colorectal cancer: gene variants within 3' UTRs and microRNAs binding as modulators of cancer risk and clinical outcome. *Oncotarget* 2015;

22. Cipollini M, Landi S, Gemignani F. MicroRNA binding site polymorphisms as biomarkers in cancer management and research. *Pharmacogenomics Pers Med* 2014;7:173–91.

23. Landi D, Gemignani F, Pardini B, Naccarati A, Garritano S, Vodicka P, Vodickova L, Canzian F, Novotny J, Barale R, Landi S. Identification of candidate genes carrying polymorphisms associated with the risk of colorectal cancer by analyzing the colorectal mutome and microRNAome. *Cancer* 2012;118:4670–80.

24. Martino A, Sainz J, Buda G, Jamroziak K, Reis RM, García-Sanz R, Jurado M, Ríos R, Szemraj-Rogucka Z, Marques H, Lesueur F, Moreno V, Orciuolo E, Gemignani F, Landi S, Rossi AM, Dumontet C, Petrini M, Campa D, Canzian F. Genetics and molecular epidemiology of multiple myeloma: the rationale for the IMMEnSE consortium (review). *Int J Oncol* 2012;40:625–38.

25. Deveci M, Catalyürek UV, Toland AE. mrSNP: software to detect SNP effects on microRNA binding. *BMC Bioinformatics* 2014;15:73.

26. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: A Cancer Microarray Database and Integrated Data-Mining Platform. *Neoplasia N Y N* 2004;6:1–6.

27. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004;74:765–9.

28. Kim H-G, Hwang S-Y, Aaronson SA, Mandinova A, Lee SW. DDR1 receptor tyrosine kinase promotes prosurvival pathway through Notch1 activation. *J Biol Chem* 2011;286:17672–81.

29. Cader FZ, Vockerodt M, Bose S, Nagy E, Brundler M-A, Kearns P, Murray PG. The EBV oncogene LMP1 protects lymphoma cells from cell death through the collagen-mediated activation of DDR1. *Blood* 2013;122:4237–45.

30. Valencia K, Ormazábal C, Zandueta C, Luis-Ravelo D, Antón I, Pajares MJ, Agorreta J, Montuenga LM, Martínez-Canarias S, Leitinger B, Lecanda F. Inhibition of collagen receptor discoidin domain receptor-1 (DDR1) reduces cell survival, homing, and colonization in lung cancer bone metastasis. *Clin Cancer Res Off J Am Assoc Cancer Res* 2012;18:969–80.

31. Yu Z, Ni L, Chen D, Zhang Q, Su Z, Wang Y, Yu W, Wu X, Ye J, Yang S, Lai Y, Li X. Identification of miR-7 as an oncogene in renal cell carcinoma. *J Mol Histol* 2013;44:669–77.

32. Roig B, Virgos C, Franco N, Martorell L, Valero J, Costas J, Carracedo A, Labad A, Vilella E. The discoidin domain receptor 1 as a novel susceptibility gene for schizophrenia. *Mol Psychiatry* 2007;12:833–41.

33. Krautkramer KA, Linnemann AK, Fontaine DA, Whillock AL, Harris TW, Schleis GJ, Truchan NA, Marty-Santos L, Lavine JA, Cleaver O, Kimple ME, Davis DB. Tcf19 is a novel islet factor necessary for proliferation and survival in the INS-1 β-cell line. *Am J Physiol Endocrinol Metab* 2013;305:E600–10.

34. Ríos R, Lupiañez CB, Campa D, Martino A, Martínez-López J, Martínez-Bueno M, Varkonyi J, García-Sanz R, Jamroziak K, Dumontet C, Cayuela AJ, Wętek M, Landi S, Rossi AM, Lesueur F, Reis RM, Moreno V, Marques H, Jurczyszyn A, Andersen V, Vogel U, Buda G, Orciuolo E, Jacobsen SEH, Petrini M, Vangsted AJ, Gemignani F, Canzian F, Jurado M, Sainz J. Type 2 diabetes-related variants influence the risk of developing multiple myeloma: results from the IMMEnSE consortium. *Endocr Relat Cancer* 2015;22:545–59.

35. Tervaniemi MH, Siitonen HA, Söderhäll C, Minhas G, Vuola J, Tiala I, Sormunen R, Samuelsson L, Suomela S, Kere J, Elomaa O. Centrosomal Localization of the Psoriasis Candidate Gene Product, CCHCR1 , Supports a Role in Cytoskeletal Organization. *PLOS ONE* 2012;7:e49920.

36. Takahashi K, Yamanaka S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat Rev Mol Cell Biol* 2016;17:183–93.

Table 1. Description of the study population

	Cases	Controls	Total
Geographic origin			
Denmark	299	489	788
France	360	176	536
Hungary	155	101	256
Israel	93	97	190
Italy	298	228	526
Poland	1,254	227	1.481
Portugal	152	195	347
Spain	283	319	602
Total	2,894	1,832	4,726
Median age (25–75% percentiles)	62 (55–68)	53 (42–65)	
Gender			
Males	1,378	991	2,414
Females	1,302	841	2,159

Table 2. List of the selected SNPs

miRSNP	Gene	Alleles	Chromosome position	Total ΔΔG	p values (GWAS)	Regulomea
rs735794	SH3BP2	G>C	4:2,837,711	604.1	0.61608	2b: TF binding + any motif + DNase Footprint + DNase peak
rs12509103	MFAP3L	A>T	4:169,986,644	315.4	0.016448	3a: TF binding + any motif + DNase peak
rs286595	MRPL22	C>T	5:154,968,992	171.3	0.0034897	3a: TF binding + any motif + DNase peak
rs1049623	DDR1	T>C	6:30,897,052	53.4	0.00014557	5: TF binding or DNase peak
rs1049633	DDR1	G>A	6:30,898,750	132.4	0.0017839	1f: eQTL + TF binding/DNase peak
rs1049628	DDR1	C>T	6:30,899,329	182.2	0.0050759	5: TF binding or DNase peak
rs1419881	TCF19	G>A	6:31,162,816	340.3	0.00094442	1f: eQTL + TF binding/DNase peak
rs13409	POU5F1	G>A	6:31,164,363	164.1	0.0009308	4: TF binding + DNase peak
rs1050239	SMPD1	G>A	11:6,394,233	638.6	0.74808	5. TF binding or DNase peak
rs11628336	DCH24	G>A	14:23,043,713	301.6	0.028482	No Data
rs1803275	IL16	G>A	15:81,306,075	338.2	0.02965	1f: eQTL + TF binding/DNase peak
rs1052536	LIG3	C>T	17:35,004,556	550.1	0.52649	1f: eQTL + TF binding/DNase peak

Table 3. Results of the case/control study in IMMeNSE

			Cases[b]			Controls[b]			Allelic model			Codominant model					
SNP	Gene	Alleles (M/m)[a]	MM	Mm	Mm	MM	Mm	mm	OR[c]	95% CI[c]	p	OR_{het}[c]	95% CI	p	OR_{hom}[c]	95% CI	p
rs735794	*SH3BP2*	G/C	1431	1206	302	848	771	177	0.99	0.88–1.10	0.79	0.88	0.75–1.02	0.09	1.1	0.85–1.40	0.45
rs12509103	*MFAP3L*	A/T	2412	483	6	1484	278	1	1.04	0.86–1.27	0.64	1.04	0.85–1.27	0.69	1.72	0.15–18.8	0.66
rs286595	*MRPL22*	C/T	799	1471	640	538	857	357	1.12	1.01–1.24	**0.03**	1.14	0.95–1.34	0.14	1.25	1.01–1.54	**0.03**
rs1049623	*DDR1*	T/C	417	1333	1185	711	841	228	1.08	0.96–1.19	0.18	0.91	0.73–1.14	0.43	1.08	0.85–1.36	0.51
rs1049633	*DDR1*	G/A	2465	409	3	1498	276	6	0.82	0.67–1.00	0.05	0.85	0.66–1.01	0.07	0.49	0.10–2.36	0.38
rs1049628	*DDR1*	C/T	1781	932	139	1077	594	67	1.01	0.88–1.14	0.93	0.93	0.80–1.09	0.38	1.24	0.87–1.78	0.24
rs1419881	*TCF19*	G/A	1124	1278	439	594	873	288	0.98	0.88–1.09	0.75	0.82	0.70–0.97	**0.02**	1.05	0.84–1.30	0.66
rs13409	*POU5F1*	G/A	1194	1304	418	666	864	265	1.02	0.92–1.13	0.68	0.91	0.77–1.06	0.22	1.12	0.89–1.39	0.31
rs1050239	*SMPD1*	G/A	1872	902	133	1069	593	93	0.92	0.81–1.04	0.22	0.92	0.78–1.07	0.29	0.89	0.62–1.22	0.42
rs11628336	*DCH24*	G/A	1025	1375	533	646	853	293	0.99	0.89–1.10	0.91	0.96	0.82–1.24	0.63	1	0.81–1.24	0.98
rs1803275	*IL16*	G/A	2567	379	21	1528	259	11	0.98	0.81–1.19	0.9	0.95	0.77–1.17	0.66	1.42	0.55–3.64	0.47
rs1052536	*LIG3*	C/T	878	1402	654	590	825	365	1.01	0.90–1.11	0.89	1.05	0.89–1.24	0.57	1	0.81–1.23	0.96

a M = major allele (i.e., more common in controls); m = minor allele (less common in controls).

b Numbers may not add up to 100% due to genotyping failure, DNA depletion or covariate missing values.

c OR: odds ratio; CI: confidence interval; ORhet: odds ratio obtained comparing heterozygotes with homozygotes for the major allele; ORhom: odds ratio obtained comparing homozygotes for the minor allele with homozygotes for the major allele; all analyses were adjusted by age, sex and country of origin. Values in bold are statistically significant (p>0.05).

Table 4. Results of the meta-analysis for the allelic and codominant model

SNP	Gene	Model	OR	95% CI	p	p_heterogeneity
rs13409	POUF51	Allelic	0.92	0.872-0.964	6.50×10^{-4}	0.079
		Codominant	0.85	0.762-0.938	2.00×10^{-3}	0.068
rs1049623	DDR1	Allelic	0.92	0.870-0.965	9.00×10^{-4}	0.183
		Codominant	0.86	0.765-0.956	6.00×10^{-3}	0.504
rs1049633	DDR1	Allelic	0.87	0.804-0.939	3.70×10^{-4}	0.038
		Codominant	0.69	0.582-0.819	2.14×10^{-5}	0.599
rs1419881	TCF19	Allelic	0.91	0.867-0.958	2.80×10^{-4}	0.069
		Codominant	0.83	0.747-0.917	2.80×10^{-4}	0.261

Figure 1. Workflow of the SNP selection.

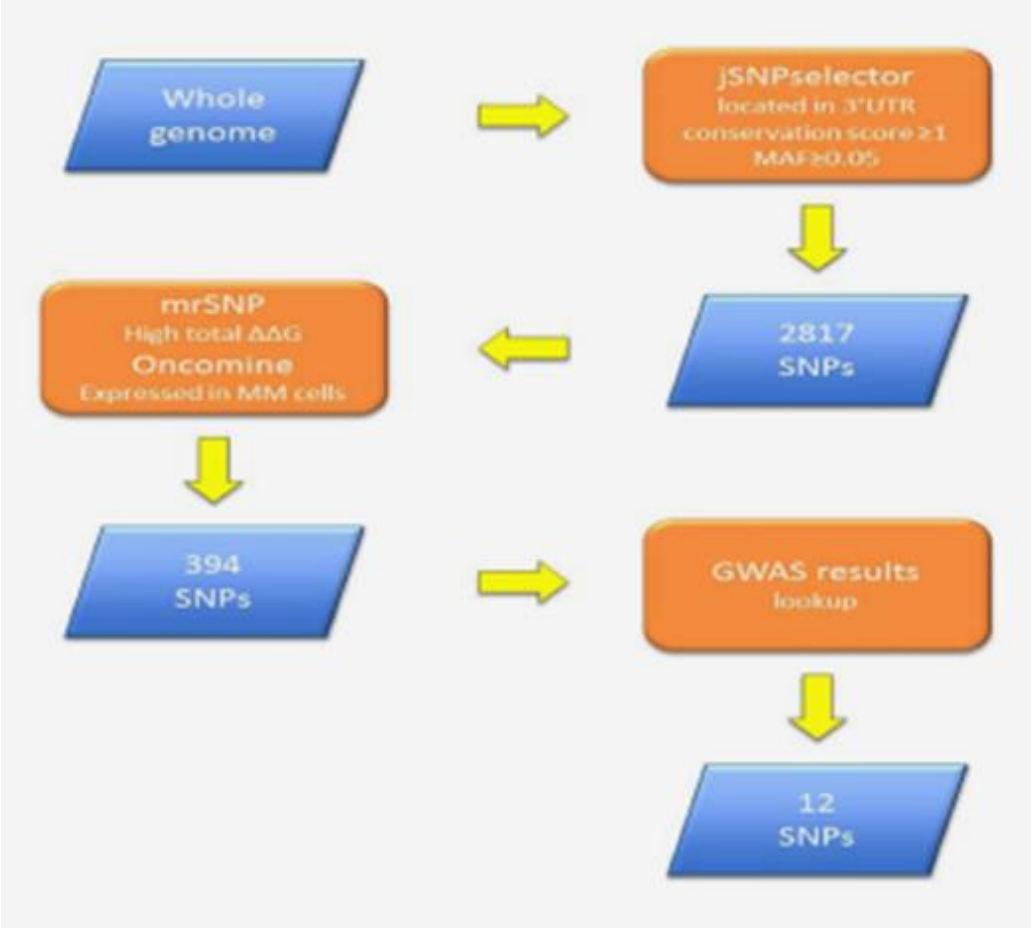

Figure 2. Forest plots of the most significant results.