Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk

Peter Broderick1,15, Daniel Chubb1,15, David C Johnson2,15, Niels Weinhold3,15, Asta Försti4, Amy Lloyd1, Bianca Oliver1, Yussanne P Ma1, Sara E Dobbins1, Brian A Walker2, Faith E Davies2, Walter A Gregory5, J Anthony Child5, Fiona M Ross6, Graham H Jackson7, Kai Neben3, Anna Jauch8, Per Hoffmann9, Thomas W Mühleisen9, Markus M Nöthen9,10, Susanne Moebus11, Ian P Tomlinson12, Hartmut Goldschmidt3,13, Kari Hemminki14,16, Gareth J Morgan2,16 & Richard S Houlston1,2,16

To identify risk variants for multiple myeloma, we conducted a genome-wide association study of 1,675 individuals with multiple myeloma and 5,903 control subjects. We identified risk loci for multiple myeloma at 3p22.1 (rs1052501 in ULK4; odds ratio (OR) = 1.32; P = 7.47 × 10−9) and 7p15.3 (rs4487645, OR = 1.38; P = 3.33 × 10−15). In addition, we observed a promising association at 2p23.3 (rs6746082, OR = 1.29; P = 1.22 × 10−7). Our study identifies new genomic regions associated with multiple myeloma risk that may lead to new etiological insights.

Multiple myeloma is a malignancy of the plasma cells that primarily localize to the bone marrow1. In the United States, ~16,000 individuals are diagnosed each year with this disease and ~11,000 die from it2. Monoclonal gammopathy of undetermined significance (MGUS; a pre-malignant clone of plasma cells producing a monoclonal paraprotein) is present in ~2% of individuals older than 50 years, and the risk of progressing to multiple myeloma increases by 1% each year3. The increased risk of multiple myeloma in relatives of individuals with MGUS is consistent with MGUS being a marker of genetic susceptibility4. To date, no lifestyle or environmental exposure factors have been consistently linked to an increased risk of multiple myeloma or MGUS. On the basis of the hypothesis that part of the two- to fourfold elevated risk of multiple myeloma in relatives of individuals with multiple myeloma5 is a consequence of the co-inheritance of multiple low-risk variants, we conducted two genome-wide association studies (GWAS) of multiple myeloma.

The two GWAS were conducted in the UK and Germany. Genotyping of both series of individuals with multiple myeloma (cases) was conducted using Illumina OmniExpress BeadChips. Cases for the UK GWAS comprised 1,371 individuals recruited through the UK Medical Research Council (MRC) Myeloma-IX trial6. Genotype frequencies were compared with publicly accessible genotype data generated by the UK Wellcome Trust Case Control Consortium 2 (WTCCC2) study of 2,699 individuals from the 1958 British Birth Cohort (known as 58C)7 and 2,501 individuals from the UK Blood Service (UKBS) collections that had been genotyped using Illumina Human 1.2M-Duo Custom_v1 Array BeadChips (see Online Methods). The German GWAS performed genotyping of 384 multiple myeloma cases recruited through the Heidelberg University Clinic. Genotype frequencies were compared with publicly accessible genotype data generated by the Heinz-Nixdorf Recall (HNR) study of 704 individuals8 from the German population who had been genotyped using Illumina Human Omni1-Quad BeadChips (see Online Methods).

Genotype data from the two GWAS were filtered on the basis of pre-specified quality control measures (see Online Methods). Individual SNPs were excluded from further analysis if they showed deviation from Hardy-Weinberg equilibrium with P < 1.0 × 10−6 in controls, an individual SNP genotype yield <95% or a minor allele frequency <1%. After this filtering, 422,839 autosomal SNPs common to both case-control series remained for further study. A total of 80 case samples were removed during quality control processing for several reasons, including if sample genotyping failed, if samples belonged to pairs of unknown duplicates or closely related individuals or if individuals were of different ancestry from the cohort of Utah residents of Northern and Western European descent (CEU) (Supplementary Figs. 1 and 2).

Prior to undertaking the meta-analysis of the two GWAS, we searched for potential errors and biases in the data sets.

1Molecular and Population Genetics, Division of Genetics and Epidemiology, Institute of Cancer Research, Surrey, UK. 2Haemat-Oncology Research Unit, Division of Molecular Pathology, Institute of Cancer Research, Surrey, UK. 3Department of Internal Medicine V, University of Heidelberg, Heidelberg, Germany. 4German Cancer Research Center, Heidelberg, Germany. 5Clinical Trials Research Unit, University of Leeds, Leeds, UK. 6Cytogenetics Group, Wessex Regional Cytogenetic Laboratory, Salisbury, UK. 7Royal Victoria Infirmary, Newcastle upon Tyne, UK. 8Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany. 9Institute of Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany. 10Molecular and Population Genetics, Wellcome Trust Centre for Human Genetics, Oxford, UK. 11National Centre of Tumour Diseases, Heidelberg, Germany. 12Center for Primary Health Care Research, Lund University, Malmo, Sweden. 13These authors contributed equally to this work. 14These authors jointly directed this work. Correspondence should be addressed to R.S.H. (richard.houlston@icr.ac.uk).

Received 8 September; accepted 3 October; published online 27 November 2011; doi:10.1038/ng.993
Quantile-quantile plots of genome-wide \( \chi^2 \) squared values showed that there was minimal inflation of the test statistics, indicating that substantial cryptic population substructure or differential genotype calling between cases and controls was unlikely in either GWAS (genomic control inflation factor \( \lambda = 1.010 \) and 1.005 in the UK and German GWAS, respectively; Supplementary Fig. 3). To fully examine the effects of population structure on our findings, we performed principal-components analysis using the EIGENSTRAT software (\( \lambda_{\text{corrected}} = 1.010 \) and 1.005 in the UK and German GWAS, respectively; Supplementary Fig. 3).

Using data from all cases and controls from both series, we derived the combined OR and confidence interval (CI) for each SNP under a fixed-effects model along with the associated \( P \) value. In the combined analysis, 19 SNPs, which map to three genomic regions, showed evidence for an association with multiple myeloma at \( P_{\text{combined}} < 5.0 \times 10^{-7} \) and with evidence for association at \( P < 0.05 \) in both data sets (Supplementary Table 1). We successfully genotyped the most highly associated SNPs mapping to the three regions in 169 multiple myeloma cases recruited through the UK MRC Myeloma VII trial (see Online Methods). For controls, we used Illumina Hap550K BeadChip genotype data generated for 927 healthy individuals from the United Kingdom as part of a study of colorectal cancer we had previously conducted\(^\text{12}\). In combined analysis, the association of the rs1052501 SNP with multiple myeloma risk was confirmed for the UK GWAS \((P = 8.65 \times 10^{-7})\), and for the German GWAS \((P = 6.04 \times 10^{-7})\). The association of the rs6746082 SNP with multiple myeloma risk was confirmed for the UK GWAS \((P = 6.36 \times 10^{-7})\) and for the German GWAS \((P = 1.10 \times 10^{-8})\). The association of the rs4487645 SNP with multiple myeloma risk was confirmed for the UK GWAS \((P = 8.65 \times 10^{-7})\) and for the German GWAS \((P = 7.47 \times 10^{-9})\). The rs1052501 SNP localizes to exon 17 of the ULK4 gene (encoding ulk4-like kinase 4) (41,900,402 bp) within a 516-kb region of multiple myeloma20.

### Table 1 Summary of results for three SNPs associated with multiple myeloma risk

| SNP             | RAF\(^a\) Case genotypes | RAF          | Control genotypes | OR      | 95% CI        | \( P \) value | \( P_{\text{adj}} \) adjusted |
|-----------------|---------------------------|--------------|-------------------|---------|---------------|--------------|--------------------------|
| rs4487645 (7p15.3) |                            |              |                   |         |               |              |                         |
| UK GWAS         | 0.71                      | 0.69         | 0.65              | 0.65    | 2,216         | 3,333        | 464          | 1.34                  | 1.22–1.47              | 1.07 \( \times 10^{-9} \) | 3.23 \( \times 10^{-9} \) |
| German GWAS     | 0.76                      | 0.26         | 0.67              | 0.67    | 320           | 300          | 83           | 1.55                  | 1.26–1.90              | 2.66 \( \times 10^{-5} \) | 6.40 \( \times 10^{-5} \) |
| UK replication  | 0.73                      | 0.92         | 0.65              | 0.65    | 383           | 421          | 113          | 1.50                  | 1.16–1.95              | –                       |                         |
| Combined        |                           |              |                   |         |               |              |              |                       | 1.38                  | 1.28–1.50              | 3.33 \( \times 10^{-15} \) | 2.62 \( \times 10^{-14} \) |
| rs1052501 (3p22.1) |                            |              |                   |         |               |              |              |                       |                       |                         | \( P_{\text{het}} = 0.36, \lambda^2 = 1\% \) |
| UK GWAS         | 0.20                      | 0.94         | 0.16              | 0.16    | 137           | 1,391        | 3,668        | 1.31                  | 1.18–1.46              | 8.65 \( \times 10^{-7} \) | 6.04 \( \times 10^{-7} \) |
| German GWAS     | 0.19                      | 0.10         | 0.16              | 0.16    | 17            | 185          | 502          | 1.32                  | 1.04–1.68              | 0.021                   | 0.085                   |
| UK replication  | 0.21                      | 0.19         | 0.16              | 0.16    | 25            | 248          | 643          | 1.34                  | 1.00–1.79              | –                       |                         |
| Combined        |                           |              |                   |         |               |              |              |                       | 1.32                  | 1.20–1.45              | 7.47 \( \times 10^{-9} \) | 1.81 \( \times 10^{-8} \) |
| rs6746082 (2p23.3) |                            |              |                   |         |               |              |              |                       |                       |                         | \( P_{\text{het}} = 0.99, \lambda^2 = 0\% \) |
| UK GWAS         | 0.82                      | 0.94         | 0.28              | 0.79    | 3,202         | 1,758        | 238          | 1.26                  | 1.13–1.41              | 4.17 \( \times 10^{-5} \) | 3.79 \( \times 10^{-5} \) |
| German GWAS     | 0.84                      | 0.84         | 0.10              | 0.77    | 415           | 258          | 31           | 1.50                  | 1.18–1.90              | 8.21 \( \times 10^{-4} \) | 3.63 \( \times 10^{-3} \) |
| UK replication  | 0.82                      | 0.11         | 0.77              | 0.79    | 574           | 294          | 49           | 1.22                  | 0.91–1.64              | 0.177                   | –                       |
| Combined        |                           |              |                   |         |               |              |              |                       | 1.29                  | 1.17–1.42              | 1.22 \( \times 10^{-7} \) | 4.02 \( \times 10^{-7} \) |

\( \text{RAF, risk allele frequency; OR, odds ratio; 95\% CI, 95\% confidence interval.} \)

\( ^{a} \text{Eigenstrat-adjusted } P \text{ values.} \)
located within a putative transcription factor–binding site or enhancer element, we conducted a bioinformatic search of the region of association using the TRANSFAC Matrix Database\(^3\) and PReMod\(^2\) software. These analyses did not provide evidence for rs4487645, rs1052501, rs6746082 or any closely correlated SNPs mapping within a known or predicted transcription regulatory region (Supplementary Table 2).

To explore whether the associations of rs4487645, rs1052501 and rs6746082 with multiple myeloma reflect cis–acting regulatory effects on their surrounding genes, DNAH11 or CDC7L, ULK4 or TRAK1 and DTNB, respectively, we examined mRNA expression in the plasma cells of 191 multiple myeloma cases using Human Genome U133 Plus 2.0 arrays (Affymetrix)\(^2\) and 90 EBV-infected lymphoblastoid cell lines using Sentrix Human-6 Expression BeadChips (Illumina)\(^2\). There was no consistent statistically significant relationship between rs4487645 or rs6746082 and expression of their surrounding genes after adjustment for multiple testing (Supplementary Fig. 4). These results do not preclude the possibility of subtle effects of genotype on gene expression with a cumulative long-term impact, as we could only detect >5% difference in expression by genotype with 80% power, and, furthermore, levels of RNA at a single time point may not adequately capture the impact of differential expression on tumor development.

Multiple myeloma occurs predominantly in males. We assessed the relationship between sex, age at diagnosis and the rs4487645, rs1052501 and rs6746082 SNPs by case-only analysis using data from all series (Supplementary Table 3). The association with multiple myeloma was not related to age or sex (Supplementary Table 3). Several subtypes of multiple myeloma have been recognized that have unique clinicopathological phenotypes\(^26\). Hierarchically, multiple myeloma can be divided into hyperdiploid and non-hyperdiploid subtypes\(^27\). The latter is primarily composed of cases harboring IGH translocations (t(11;14)(q13;q32), t(4;14)(p16;q32) and t(14;16)(q32;q23)) and is typified by more aggressive disease. Trisomies and a more indolent form of the disease characterize hyperdiploid multiple myeloma.\(^26\) Case-only analysis provided no evidence for subtype-specific associations at these SNPs after adjustment for multiple testing, consistent with the risk variants having general effects on multiple myeloma (Supplementary Table 3).

The risks for multiple myeloma conferred by rs4487645, rs1052501 and rs6746082 are modest, collectively accounting for only ~4% of the familial risk of multiple myeloma. However, the carrier frequencies of risk alleles are high in the European population; therefore, the loci make a significant contribution to the development of multiple myeloma in terms of the attributable fraction in the population, underlying ~37% of cases. It will be intriguing to explore how our findings translate to non-European populations, which have a lower incidence of MGUS and multiple myeloma.\(^30\) As the frequencies of rs4487645, rs1052501 and rs6746082 genotypes in the Han Chinese in Beijing (CHB), Japanese in Tokyo (JPT) and Yoruba in Ibadan (YRI) populations are substantially different from the CEU population, it is possible that variations within 2p23.3, 3p22.1 and 7p15.3 contribute, in part, to differences in disease incidence.

Our findings provide evidence that common genetic variation influences multiple myeloma risk. These findings, in conjunction with recent observations from GWAS of Hodgkin’s lymphoma\(^31\) and chronic lymphocytic leukemia\(^32\), raise the possibility that genetically determined dysregulation of MYC is a common mechanism underlying the predisposition to hematological malignancies of the B-cell lineage. Given the modest size of our study and the evidence for an over-representation of association signals after excluding SNPs mapping to the LD regions of association at 2p23.3, 3p22.1 and 7p15.3 (Supplementary Fig. 3), it is likely that additional risk variants for multiple myeloma will be identified through further study.

**URLs.** R statistical software, http://www.r-project.org/; Illumina SNP panel, http://www.illumina.com/products/human_omni_express.llm/; dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/; HapMap, http://hapmap.ncbi.nlm.nih.gov/; 1000 Genomes Project, http://www.1000genomes.org/; SNP annotation and proxy search (SNAP), http://www.broadinstitute.org/mpg/snp/snap; IMPUTE, https://mathgen.stats.ox.ac.uk/impute/impute.html; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snpstat/snpstat.html; PReMod, http://genomewebqec.bcgsc.ca/PReMod/welcome/; TRANSFAC Matrix Database, http://www.biobase-international.com/product/transcription-factor-bindingsites/; JASPAR2 database, http://jaspar.cgb.ki.se/; EIGENSTRAT, http://geopath.med.harvard.edu/~reich/Software.htm; WTCCC, www.wtccc.org.uk; meta analysis helper (METAL), www.sph.umich.edu/sgc/abecasis/metal/; Mendelian Inheritance In Man, http://www.ncbi.nlm.nih.gov/omim; MRC Myeloma-IX trial,
http://ctru.leeds.ac.uk/myelomaIX/; MRC Myeloma-VII trial, http://ctru.leeds.ac.uk/Haematological/; SIFT, http://sift.jcvi.org/; PolyPhen, http://genetics.bwh.harvard.edu/pph/; Illuminus software, http://www.sanger.ac.uk/resources/software/illuminus/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

ACKNOWLEDGMENTS

We are grateful to all study participants and investigators at the individual centers for taking part. We are extremely grateful to all investigators who contributed to the generation of this data set. We also thank the staff of the Clinical Trials Research Unit (CTRU), University of Leeds, and the National Cancer Research Institute (NCRI) Haematology Clinical Studies Group. We are grateful to all investigators who contributed to the Colorectal Tumour Gene Identification (CORGI) consortium, from which controls in the replication analysis were drawn. The current study made use of genotyping data on the 1958 Birth Cohort. Genotyping data for control subjects was generated by the Wellcome Trust Sanger Institute. A full list of the investigators who contributed to the generation of the genotyping data from the population was made by the Wellcome Trust Sanger Institute. A full list of the investigators who contributed to the generation of the data is available from the WTCCC website (see URLs). The GWAS made use of genotyping data from the population-based HNR study. The HNR study was supported by the Heinz-Nixdorf Foundation. The genotyping of the HNR subjects on Illumina HumanOmni1-Quad BeadChips was financed by the German Centre for Neurodegenerative Disorders (DZNE), Bonn. Myeloma UK provided the main funding for the study. Additional funding was provided by Cancer Research UK (C1298/AR8362 supported by the Bobby Moore Fund), the Leukaemia Lymphoma Research Fund and the National Health Services (NHS) through the Biological Research Centre of the National Institute for Health Research at the Royal Marsden Hospital NHS Trust. In Germany, funding was provided to Dietmar-Hopp-Stiftung in Walldorf, the University Hospital Heidelberg and a grant from APO-STS (European Union Health-F4-2007-200767). Additionally, the study was funded by the German Ministry of Education and Science and the German Research Council (DFG; Projects SI 236/8-1, SI236/9-1 and ER 155/6-1).

AUTHOR CONTRIBUTIONS

R.S.H. designed the study, R.S.H. and G.J.M. obtained financial support in the UK, and K.H. and H.G. obtained funding in Germany. D.C. performed the main statistical and bioinformatic analyses, and Y.P.M. and S.E.D. performed additional related analyses. P.B. coordinated laboratory studies. A.L. and B.O. performed genotyping of the UK samples. P.H., T.W.M. and M.M.N. performed and coordinated genotyping of the German controls; K.H. and A.F. performed genotyping of the German cases. D.C.J. managed and prepared the Myeloma-VII and Myeloma-IX case study DNA samples. H.G., K.N. and N.W. coordinated and managed the German DNA samples. G.J.M., E.F.D., W.A.G., G.H.J. and J.A.C. ascertainment and collected case study samples from the UK Myeloma-VII and Myeloma-IX studies. S.M. obtained and managed the HNR samples. I.P.T. conducted the meta-analysis and Cost-Effectiveness Analysis (Yorkshire University Press, New York, USA, 1994).

ADDITIONAL INFORMATION

© 2012 Nature America, Inc. All rights reserved.

1. Kyle, R.A. & Rajkumar, S.V. Multiple myeloma. N. Engl. J. Med. 351, 1860–1873 (2004).
2. Palumbo, A. & Anderson, K. Multiple myeloma. N. Engl. J. Med. 364, 1046–1060 (2011).
3. Kyle, R.A. et al. A long-term study of prognosis in monoclonal gammapathy of undetermined significance. N. Engl. J. Med. 346, 564–569 (2002).
4. Landgren, O. et al. Risk of plasma cell and lymphoproliferative disorders among 14621 first-degree relatives of 4458 patients with monoclonal gammapathy of undetermined significance in Sweden. Blood 114, 791–795 (2009).
5. Alteri, A., Chen, B., Bermejo, J.L., Castro, F. & Hemminki, K. Familial risks and temporal incidence trends of multiple myeloma. Eur. J. Cancer 42, 1661–1670 (2006).
6. Morgan, G.J. et al. First-line treatment with zoledronic acid as compared with clodronic acid in multiple myeloma (MRC Myeloma IX): a randomised controlled trial. Lancet 376, 1989–1993 (2010).
7. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int. J. Epidemiol. 35, 34–41 (2006).
8. Schlemmer, A. et al. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. Am. Heart J. 144, 212–218 (2002).
9. Dib, A. et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat. Genet. 37, 1243–1246 (2005).
10. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).

NATURE GENETICS VOLUME 44 | NUMBER 1 | JANUARY 2012

LETTERS

61
ONLINE METHODS

Genome-wide association studies. The UK GWAS included multiple myeloma cases (International Classification of Diseases, 10th Revision (ICD-10) C90.0; 819 male; mean age at diagnosis of 64.1 years, s.d. = 10.3) ascertained through the UK MRC Myeloma IX trial. All cases were UK residents. For controls, we used publicly accessible data generated by the WTCCC, as the 58C (also known as the National Child Development Study) and UKBS. Genotyping of both sets of controls was conducted using Illumina Human 1.2M-Duo Custom v1 Array BeadChips. SNP calling was performed using Illuminus software (see URLs). Full details of genotyping, SNP calling and quality control filtering have been previously reported (WTCCC, see URLs). Concordant with previous findings, comparison of the two control series showed little evidence for systematic bias (inflation factor λ = 1.019; Supplementary Fig. 1)31. The German GWAS was comprised of 384 multiple myeloma cases (ICD-10 C90.0: 229 male; mean age at diagnosis of 54.5 years, s.d. = 8.0) ascertained through the Heidelberg University Centre. For controls, we used publicly accessible genotype data on 704 healthy individuals with no past history of malignancy enrolled in the HNR study35.

Replication. For replication, we genotyped 169 additional cases of multiple myeloma (93 male) ascertained through the UK MRC Myeloma-VII trial. For controls, we made use of previously generated data from a UK GWAS of colorectal cancer (420 males, 507 females, aged from 18 to 69 years)32.

Ethics. Collection of blood samples and clinico-pathological information from subjects was undertaken with informed consent and the approval of the relevant German or UK ethical review board, in accordance with the tenets of the Declaration of Helsinki.

Karotyping. Conventional cytogenetic studies of multiple myeloma cells were conducted using standard karotyping methodologies, and standard criteria for the definition of a clone were applied. Where possible, metaphase fluorescence in situ hybridization (FISH) was used to confirm whether abnormalities seen with immunofluorescence in situ hybridization (iFISH) were present in the same cells as the abnormalities detected by conventional cytogenetics. iFISH and ploidy classification of the UK samples was conducted as previously33, and iFISH and ploidy classification of the German samples was performed as described34,35.

Genotyping. DNA was extracted from EDTA-venous blood samples using Qiagen FlexiGene or QIAamp methodologies and quantified using PicoGreen (Invitrogen). Genotyping of cases in the GWAS was conducted using Illumina OmniExpress BeadChips according to the manufacturer’s protocols. To ensure quality of genotyping, duplicates were included on each sample plate (showing a concordance of >99.99%). DNA samples with GenCall scores <0.25 at any locus were considered to not represent calls. In each sample series, sequencing of an SNP was considered unsuccessful if <95% of DNA samples generated a genotype at the locus. Cluster plots were manually inspected for all SNPs and used these as dissimilarity measures on which to perform principal-component analysis. The first two principal components for each individual were plotted, and any individual not present in the main CEU cluster was excluded from analyses. We removed 35 cases with non-CEU ancestry and one 58C control who had previously been diagnosed with Hodgkin’s lymphoma31. In each sample series, we filtered out SNPs having a minor allele frequency <1% and a call rate <95%. We also excluded SNPs showing departure from Hardy-Weinberg equilibrium at P < 1 × 10^-6 in controls. For replication and validation analysis, call rates were >95% per 384-well plate for each SNP. Cluster plots were visually examined by two researchers.

Analyses were primarily undertaken using R (v2.6), STATA v.10 (State College) and PLINK (v1.06)36 software. The association between each SNP and risk of multiple myeloma was assessed by the Cochran-Armitage trend test. The adequacy of case-control matching and the possibility of differential genotyping of cases and controls were formally evaluated using quantile-quantile plots of test statistics. The inflation factor λ was calculated based on the 90% of SNPs with the least significant association3. We adjusted for possible population substructure using EIGENSTRAT software. Odds ratios and associated 95% confidence intervals were calculated by unconditional logistic regression. Meta-analysis was conducted using standard methods31. We calculated Cochran’s Q statistic to test for heterogeneity31 and the I² statistic to quantify the proportion of the total variation that was caused by heterogeneity37. I² values ≥75% were considered to indicate substantial heterogeneity37,38. To conduct a pooled analysis incorporating EIGENSTRAT-adjusted P values from the GWAS, we used the weighted Z method implemented in the meta analysis helper (METAL)39 program. Associations by sex, age and clinico-pathological phenotypes were examined by logistic regression in case-only analyses.

The familial relative risk of multiple myeloma attributable to any locus is given by the formula40:

\[ \lambda^* = \frac{p(p_2 + q_2)^2 + q(p_1 + q_1)^2}{(p_2 + 2pq_1 + q_2)^2} \]

where p is the population frequency of the minor allele, q = 1 - p, and \( r_1 \) and \( r_2 \) are the relative risks (approximated by odds ratios) for heterozygotes and the rarer homozygote relative to the more common homozygote. From \( \lambda^* \), it is possible to quantify the influence of the locus on the overall familial risk of multiple myeloma in first-degree relatives. Assuming a multiplicative interaction between risk alleles, the proportion of the overall familial risk deletions), and -2% were excluded from further analysis. Data were reviewed by a second individual to determine whether any metrics were missed or if further editing was required. Overall, this process resulted in substantially increased genotyping accuracy.

Validation of Illumina SNP genotypes. To confirm genotyping accuracy for rs4487645, rs1052501 and rs6746082, we confirmed genotypes by ABI 3730xl Sanger sequencing in >184 randomly selected samples from each of the UK case series and the 58C and HNR control series (concordance of >99.5%). Sequences for the PCR primers used are available on request.

Replication genotyping. Replication of rs4487645, rs1052501 and rs6746082 associations were performed by ABI 3730xl Sanger sequencing of all MRC Myeloma-VII trial samples.

Statistical and bioinformatic analysis. We applied predetermined quality control metrics to the GWAS data. We restricted analyses to samples for which >95% of SNPs were successfully genotyped, thus eliminating 28 cases. We computed identity-by-state (IBS) probabilities for all pairs (cases and controls) to search for duplicates and closely related individuals among samples (defined as IBS ≥0.80, thereby excluding first-degree relatives). For all identical pairs, the sample having the higher call rate was retained, eliminating 17 multiple myeloma cases. To identify individuals who might not be of Western European ancestry, we merged our case and control data with HapMap phase 2 samples (60 CEU; 60 Yoruba from Badan (YRI), 90 Japanese from Tokyo (JPT) and 90 Han Chinese from Beijing (CHB)). For each pair of individuals, we calculated genome-wide IBS distances on markers shared by HapMap and our SNP panel and used these as dissimilarity measures on which to perform principal-component analysis. The first two principal components for each individual were plotted, and any individual not present in the main CEU cluster was excluded from analyses. We removed 35 cases with non-CEU ancestry and one 58C control who had previously been diagnosed with Hodgkin’s lymphoma31. In each sample series, we filtered out SNPs having a minor allele frequency <1% and a call rate <95%. We also excluded SNPs showing departure from Hardy-Weinberg equilibrium at P < 1 × 10^-6 in controls. For replication and validation analysis, call rates were >95% per 384-well plate for each SNP. Cluster plots were visually examined by two researchers.

Analyses were primarily undertaken using R (v2.6), STATA v.10 (State College) and PLINK (v1.06)36 software. The association between each SNP and risk of multiple myeloma was assessed by the Cochran-Armitage trend test. The adequacy of case-control matching and the possibility of differential genotyping of cases and controls were formally evaluated using quantile-quantile plots of test statistics. The inflation factor λ was calculated based on the 90% of SNPs with the least significant association. We adjusted for possible population substructure using EIGENSTRAT software. Odds ratios and associated 95% confidence intervals were calculated by unconditional logistic regression. Meta-analysis was conducted using standard methods31. We calculated Cochran’s Q statistic to test for heterogeneity31 and the I² statistic to quantify the proportion of the total variation that was caused by heterogeneity37. I² values ≥75% were considered to indicate substantial heterogeneity37,38. To conduct a pooled analysis incorporating EIGENSTRAT-adjusted P values from the GWAS, we used the weighted Z method implemented in the meta analysis helper (METAL)39 program. Associations by sex, age and clinico-pathological phenotypes were examined by logistic regression in case-only analyses.

The familial relative risk of multiple myeloma attributable to any locus is given by the formula40:

\[ \lambda^* = \frac{p(p_2 + q_2)^2 + q(p_1 + q_1)^2}{(p_2 + 2pq_1 + q_2)^2} \]

where p is the population frequency of the minor allele, q = 1 - p, and \( r_1 \) and \( r_2 \) are the relative risks (approximated by odds ratios) for heterozygotes and the rarer homozygote relative to the more common homozygote. From \( \lambda^* \), it is possible to quantify the influence of the locus on the overall familial risk of multiple myeloma in first-degree relatives. Assuming a multiplicative interaction between risk alleles, the proportion of the overall familial risk...
attributable to the locus is given by \( \log (\lambda^*) / \log (\lambda_0) \), where \( \lambda_0 \), the overall familial risk of multiple myeloma, is assumed on the basis of epidemiological studies to be 2.45 (ref. 5).

The population attributable fraction was estimated from \( 1 - \Pi_i (1 - x_i) - x_i \), where \( x_i = p \times OR_{pa} - 1) / (p \times (OR_{pa} - 1) + 1) \), \( p \) being the frequency of the risk allele in the population and \( OR_{pa} \) being the per-allele OR.

Prediction of the ungenotyped SNPs was carried out using IMPUTEv2 based on the HapMap phase 3 haplotypes release 2 (HapMap Data Release 27, phase 3 Feb 2009 on NCBI B36 assembly, dbSNP26) and data from the 1000 Genomes Project. Imputed data were analyzed using SNPTYPE v2 to account for uncertainties in SNP prediction. Measurements of LD between HapMap SNPs were based on Data Release 27, phase 3 (Feb 2009 on NCBI B36 assembly, dbSNP26), viewed using Haploview software (v4.2) and plotted using SNP annotation and proxy search (SNAP). LD blocks were defined according to HapMap recombination rate (centimorgans per megabase), as determined using the Oxford recombination hotspots41 and the previously set distribution of confidence intervals42. To annotate potential regulatory sequences within disease loci, we implemented in silico searches using TRANSFAC Matrix Database v7.29 and PReMod10 software 21,22. We used the in silico algorithms SIFT and PolyPhen to predict the impact of amino acid substitutions on protein function.

Relationship between SNP genotype and mRNA expression. To explore a relationship between SNP genotype and expression levels of CDCA7L, DNAH11, TRAK1 and DTNB in multiple myeloma, we made use of Affymetrix Human Genome U133 Plus 2.0 array data we previously generated on the plasma cells from 192 multiple myeloma cases from the MRC Myeloma IX trial23. To explore a relationship between SNP genotype and expression levels in lymphocytes, we made use of publicly available expression data generated from the analysis of 90 lymphoblastoid cell lines derived from individuals of European descent transformed by Epstein-Barr virus using Sentrix Human-6 Expression BeadChips (Illumina)24,25. Online recovery of data was performed using WGAViewer Version 1.25 software. Differences in the distribution of mRNA expression levels between SNP genotypes were compared using a Wilcoxon-type trend test43. Estimates of the power of the assays to establish a relationship between genotype and expression were made using STATA software, assuming an allele-based test of the difference in normalized log values of RNA expression (imposing a Bonferroni-corrected \( p \) value of 0.005 to address multiple testing).

33. Chiecchio, L. et al. Deletion of chromosome 13 detected by conventional cytogenetics is a critical prognostic factor in myeloma. Leukemia 20, 1610–1617 (2006).
34. Neben, K. et al. Combining information regarding chromosomal aberrations t(4;14) and del(17p13) with the International Staging System classification allows stratification of myeloma patients undergoing autologous stem cell transplantation. Haematologica 95, 1150–1157 (2010).
35. Wuilleme, S. et al. Ploidy, as detected by fluorescence in situ hybridization, defines different subgroups in multiple myeloma. Leukemia 19, 275–278 (2005).
36. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
37. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558 (2002).
38. Ioannidis, J.P., Ntzani, E.E. & Trikalinos, T.A. ‘Racial’ differences in genetic effects for complex diseases. Nat. Genet. 36, 1312–1318 (2004).
39. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
40. Houlston, R.S. & Ford, D. Genetics of coeliac disease. QJM 89, 737–743 (1996).
41. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. Science 310, 321–324 (2005).
42. Gabriel, S.B. et al. The structure of haplotype blocks in the human genome. Science 296, 2225–2229 (2002).
43. Cuzick, J. A Wilcoxon-type test for trend. Stat. Med. 4, 87–90 (1985).