Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk

Daniel Chubb1,22, Niels Weinhold2,22, Peter Broderick1, Bowang Chen3, David C Johnson4, Asta Försti3,5, Jayaram Vijayakrishnan1, Gabriele Migliorini1, Sara E Dobbins1, Amy Holroyd1, Dirk Hose2,6, Brian A Walker4, Faith E Davies4, Walter A Gregory7, Graham H Jackson8, Julie A Irving9, Guy Pratt10, Chris Fegan11, James A L Fenton12, Kai Neben2, Per Hoffmann13,14, Markus M Nöthen13–15, Thomas W Mühleisen13,14, Lewin Eisele16, Fiona M Ross17, Christian Straka18, Hermann Einsele19, Christian Langer20, Elisabeth Dörner2, James M Allan9, Anna Jauch21, Gareth J Morgan4,23, Kari Hemminki3,5,23, Richard S Houlston1,23 & Hartmut Goldschmidt2,6,23

To identify variants for multiple myeloma risk, we conducted a genome-wide association study with validation in additional series totaling 4,692 individuals with multiple myeloma (cases) and 10,990 controls. We identified four risk loci at 3q26.2 (rs10936599, P = 8.70 × 10−14), 6p21.33 (rs2285803, PSORS1C2, P = 9.67 × 10−11), 17p11.2 (rs4273077, TNFRSF13B, P = 7.67 × 10−9) and 22q13.1 (rs877529, CBX7, P = 7.63 × 10−16). These data provide further evidence for genetic susceptibility to this B-cell hematological malignancy, as well as insight into the biological basis of predisposition.

Multiple myeloma is a malignancy of plasma cells1. There are around 20,000 new cases of multiple myeloma in the United States each year, and just over half that number die of the disease each year2. We previously reported results of a genome-wide association study (GWAS) of multiple myeloma that was based on an analysis of UK and German series in which we identified risk loci at 2p23.3, 3p22.1 and 7p15.3 through fast-track analysis of SNPs with the smallest P values3. We subsequently conducted further follow-up analyses using an expanded German GWAS track analysis of SNPs with the smallest P values3.

After quality control, the previously reported German GWAS data comprised 1,321 cases recruited from the UK Medical Research Council (MRC) Myeloma IX trial6 genotyped using Illumina OmniExpress BeadChips. We compared the genotype frequencies in these cases to publicly accessible genotype data generated by the UK Wellcome Trust Case Control Consortium 2 (WTCCC2) study of 2,698 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 (Online Methods).

Genotype data from the GWAS were filtered on the basis of pre-specified quality-control measures (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, individual SNP genotype yield <95% or minor allele frequency (MAF) < 1%. After filtering, 414,804 autosomal SNPs common to both case-control series were analyzable (Online Methods, and Supplementary Figs. 1 and 2).

Before undertaking a meta-analysis of the two GWAS, we searched for potential errors and biases in the data sets. Quantile-quantile plots of the genome-wide χ2 values showed that there was minimal inflation of the test statistics, rendering substantial cryptic population substructure or differential genotype calling between cases and the German population who had been genotyped using Illumina Human Omni1-Quad BeadChips or Illumina OmniExpress BeadChips (Online Methods).

After quality control, the previously reported UK GWAS3 comprised 1,321 cases recruited from the UK Medical Research Council (MRC) Myeloma IX trial6 genotyped using Illumina OmniExpress BeadChips. We compared the genotype frequencies in these cases to publicly accessible genotype data generated by the UK Wellcome Trust Case Control Consortium 2 (WTCCC2) study of 2,698 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 (Online Methods).

Received 6 March; accepted 26 July; published online 18 August 2013; doi:10.1038/ng.2733
controls unlikely in either GWAS (the genomic control inflation factor $\lambda$ was 1.033 in the UK GWAS and 1.17 in the German GWAS; Supplementary Fig. 3). For completeness, we performed a principal components analysis using Eigenstrat software to determine the effects of population substructure on our findings ($\lambda_{\text{corrected}} = 1.014$ in the UK GWAS and $\lambda_{\text{corrected}} = 1.029$ in the German GWAS; Supplementary Fig. 3).

Using data on all cases and controls from both GWAS, we derived joint odds ratios (ORs) and confidence intervals (CIs) under a fixed-effects model for each SNP and the associated $P$ values. In the combined analysis, we identified nine SNPs showing good evidence of association ($P < 5.0 \times 10^{-6}$) and that have not previously been associated with multiple myeloma risk (Supplementary Table 1). The $P$ value threshold used does not exclude the possibility that other SNPs represent genuine association signals, but it was a pragmatic strategy for prioritizing replication here.

To validate our findings, we conducted a replication study of the nine SNPs by genotyping samples from three additional series: UK replication 1, comprising 812 cases ascertained through the UK MRC Myeloma IX and XI trials and 1,110 controls; UK replication 2, comprising 396 cases collected through UK hematology centers and 992 controls; and the German replication, comprising 1,149 cases collected through the German Myeloma Study Group (DSMM), Heidelberg University Clinic and Ulm University Clinic and 1,582 regional controls (Online Methods). In the combined analysis, four SNPs, rs10936599 (3q26.2, $P = 8.70 \times 10^{-14}$), rs2285803 (6p21.33, $P = 9.67 \times 10^{-11}$), rs4273077 (17p11.2, $P = 7.67 \times 10^{-9}$) and rs877529 (22q13.1, $P = 7.63 \times 10^{-10}$), showed genome-wide significant evidence for an association with multiple myeloma (Table 1, Online Methods and Supplementary Table 2).

rs10936599 at 3q26.2 ($P = 8.70 \times 10^{-14}$; Table 1) is responsible for the H717H polymorphism in MYNN, the myoneurin gene (MIM606042). The rs10936599 G risk allele has also previously been shown to influence colorectal cancer risk. Whereas MYNN encodes a zinc finger protein of unknown function that is expressed particularly in muscle, rs10936599 (169,492,101 bp) maps within a 250-kb region of linkage disequilibrium (LD) that also encompasses TERC, the telomerase RNA component gene (MIM602322). Telomerase reactivation and telomerase-mediated elongation of shorter telomeres is a feature of multiple myeloma. Because carrier status for the rs10936599 G risk allele is associated with significantly longer telomeres ($P = 1.3 \times 10^{-9}$ for rs10936599 in a meta-analysis of GWAS data; Fig. 1), the A allele of rs2293607 was recently shown to be associated with TERC mRNA expression and longer telomeres in vitro, supporting variation in TERC as the basis of the 3q26.2 cancer association.

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

| SNP                  | Risk allele | RAF in cases | Case genotypes | RAF in controls | Control genotypes | OR      | 95% CI          | $P$            | $P_{\text{adjusted}}$ |
|----------------------|-------------|--------------|----------------|-----------------|-------------------|---------|----------------|----------------|---------------------|
| rs10936599 (3q26.2)  | G           | 0.80         | 843 (49)       | 0.75            | 2,960 (1,914)     | 1.31    | 1.18–1.46      | $4.33 \times 10^{-7}$ | $5.18 \times 10^{-7}$ |
|                      |             | 0.79         | 632 (33)       | 0.75            | 1,187 (778)       | 1.25    | 1.10–1.41      | $6.62 \times 10^{-4}$ | $1.48 \times 10^{-3}$ |
|                      |             | 0.80         | 520 (29)       | 0.76            | 628 (415)         | 1.32    | 1.13–1.55      | $4.98 \times 10^{-4}$ | –                   |
|                      |             | 0.79         | 244 (126)      | 0.75            | 559 (372)         | 1.23    | 1.01–1.51      | $4.22 \times 10^{-2}$ | –                   |
|                      |             | 0.78         | 714 (363)      | 0.76            | 898 (585)         | 1.16    | 1.02–1.31      | $2.56 \times 10^{-2}$ | –                   |
| Combined              |             | 1.26         | 1.18–1.33      | $8.70 \times 10^{-14}$ | $1.74 \times 10^{-13}$ |
| rs2285803 (6p21.3)   | A           | 0.32         | 125 (603)      | 0.28            | 444 (2,055)       | 1.21    | 1.10–1.32      | $6.67 \times 10^{-5}$ | $7.64 \times 10^{-5}$ |
|                      |             | 0.36         | 129 (462)      | 0.31            | 226 (833)         | 1.24    | 1.11–1.39      | $1.15 \times 10^{-4}$ | $1.18 \times 10^{-4}$ |
|                      |             | 0.32         | 78 (362)       | 0.28            | 88 (424)          | 1.22    | 1.06–1.41      | $5.11 \times 10^{-3}$ | –                   |
|                      |             | 0.29         | 32 (152)       | 0.26            | 51 (402)          | 1.14    | 0.94–1.38      | $1.82 \times 10^{-1}$ | –                   |
|                      |             | 0.33         | 130 (491)      | 0.30            | 140 (674)         | 1.12    | 1.00–1.26      | $5.86 \times 10^{-2}$ | –                   |
| Combined              |             | 1.19         | 1.13–1.26      | $9.67 \times 10^{-11}$ | $1.18 \times 10^{-10}$ |
| rs4273077 (17p11.2)  | G           | 0.12         | 15 (284)       | 0.10            | 48 (926)          | 1.24    | 1.08–1.42      | $1.88 \times 10^{-3}$ | $2.65 \times 10^{-3}$ |
|                      |             | 0.14         | 25 (239)       | 0.11            | 27 (390)          | 1.40    | 1.20–1.64      | $2.80 \times 10^{-5}$ | $6.17 \times 10^{-4}$ |
|                      |             | 0.12         | 18 (148)       | 0.09            | 12 (179)          | 1.28    | 1.04–1.57      | $1.96 \times 10^{-2}$ | –                   |
|                      |             | 0.11         | 3 (77)         | 0.10            | 8 (178)           | 1.12    | 0.85–1.48      | $4.20 \times 10^{-1}$ | –                   |
|                      |             | 0.12         | 17 (252)       | 0.11            | 21 (298)          | 1.17    | 0.99–1.38      | $6.79 \times 10^{-2}$ | –                   |
| Combined              |             | 1.26         | 1.16–1.36      | $7.67 \times 10^{-9}$ | $1.41 \times 10^{-7}$ |
| rs877529 (22q13.1)   | A           | 0.51         | 346 (654)      | 0.44            | 1,000 (2,560)     | 1.33    | 1.22–1.45      | $1.08 \times 10^{-10}$ | $9.11 \times 10^{-11}$ |
|                      |             | 0.45         | 214 (483)      | 0.43            | 389 (1,026)       | 1.09    | 0.98–1.21      | $1.18 \times 10^{-1}$ | $1.09 \times 10^{-1}$ |
|                      |             | 0.49         | 176 (436)      | 0.44            | 195 (555)         | 1.24    | 1.08–1.41      | $2.01 \times 10^{-3}$ | –                   |
|                      |             | 0.47         | 86 (192)       | 0.42            | 166 (485)         | 1.24    | 1.05–1.47      | $1.21 \times 10^{-2}$ | –                   |
|                      |             | 0.46         | 238 (586)      | 0.41            | 274 (754)         | 1.23    | 1.10–1.37      | $2.69 \times 10^{-4}$ | –                   |
| Combined              |             | 1.23         | 1.17–1.29      | $7.63 \times 10^{-16}$ | $2.29 \times 10^{-16}$ |

*Supplementary Table 1* provides the $P$ values for the nine SNPs, as well as the combined results. A **bold** is used to highlight each new SNP analyzed, as well as the combined results.
Figure 1 Regional plots of association results and recombination rates for the 3q26.2, 6p21.33, 17p11.2 and 22q13.1 susceptibility loci. (a–d) Association results for both genotyped (triangles) and imputed (circles) SNPs in the GWAS samples and recombination rates within the four loci, 3q26.2, 6p21.33, 17p11.2 and 22q13.1. For each plot, $-\log_{10} P$ values (y axis) of the SNPs are shown according to their chromosomal positions (x axis). The top genotyped SNP in each combined analysis is represented by a large triangle and is labeled by its rsID. The color intensity of each symbol reflects the extent of LD with the top genotyped SNP, ranging from white ($r^2 = 0$) through to dark red ($r^2 = 1.0$). Genetic recombination rates (cM/Mb), estimated using HapMap CEU samples, are shown with light blue lines. Physical positions are based on NCBI build 36 of the human genome. Also shown are the relative positions of genes and transcripts mapping to each region of association. Genes have been redrawn to show the relative positions; therefore, maps are not to physical scale. On the bottom are shown the region of interest together with all transcripts and the chromatin-state segmentation track (ChromHMM) for lymphoblastoid cells using data from the HapMap ENCODE Project. CNV, copy-number variation.
rs2285803 ($P = 9.67 \times 10^{-11}$; Table 1 and Fig. 1) localizes in intron 5 of PSORS1C1, the putative psoriasis susceptibility gene (MIM 135255) at 2p21.33 (31,107,245 bp). The 163-kb region of LD also encompasses CCHCR1 (MIM 605310), CDSN (MIM 602593), TCF19, the transcription factor 19 gene (MIM 600912) and POU5F1, the POU domain, class 5, transcription factor 1 gene (MIM 164177). Although there is currently no evidence for a role of POU5F1 in multiple myeloma, the gene encodes OCT3 (also known as OCT4), which regulates pluripotency, lineage commitment and tissue-specific gene expression. Variation at 6p21.33 has previously been shown to be associated with risk for follicular and Hodgkin’s lymphomas. The associations for follicular and Hodgkin’s lymphomas were defined by rs6457327 in the HLA class I region13 and rs10484561 and rs2647012 in the HLA class II region14,15. The HLA association at 6p21.33 is marked by rs6903608 in the HLA class II region16. The multiple myeloma risk associated with each of these SNPs was nonsignificant (Supplementary Table 3). To further investigate the rs2285803 signal for multiple myeloma, we imputed classical HLA alleles from SNP data from both GWAS using HLA*IMPM17,18. The strongest HLA association was provided by HLA-DRB5*01 ($P = 1.42 \times 10^{-5}$; Supplementary Table 4), which was substantially weaker than that provided by rs2285803 ($P = 3.07 \times 10^{-8}$). To evaluate the independence of the associations, we conducted regression jointly on rs2285803 and the imputed HLA alleles. Conditional analysis showed that most, but not all, of the MHC variation defined by SNP genotype could be explained by rs2285803 (Supplementary Table 4).

rs4273077 ($P = 7.67 \times 10^{-9}$; Table 1) maps within intron 2 of TNFRSF13B, the gene for Homo sapiens tumor necrosis factor receptor superfamily member 13B (MIM 604907) at 17p11.2 (16,849,139 bp; Fig. 1). TNFRSF13B (also known as TACI) is a strong candidate for multiple myeloma predisposition a priori. TNFRSF13B is a key regulator of B- and T-cell function, as it is required for the development of transitional (T2) and mature B lymphocytes and the regulation of normal B-cell homeostasis19. Variation at TNFRSF13B influences circulating IgG concentrations20, and Tnfrsf13b−/− mice show an expanded B-cell population with lymphoproliferation and lymphoma risk21. As TNFRSF13B mutation is a risk factor for antibody-deficient (MIM 240500) and selective immunoglobulin (lg) deficiency (MIM 609529) associated with lymphoproliferation, it is probable that loss of TNFRSF13B function impairs isotype switching. Primary multiple myeloma cells with high TNFRSF13B expression (TACIhigh) resemble bone marrow plasma cells, which depend on interaction with the bone marrow environment. In contrast, multiple myeloma cells with low expression of TNFRSF13B (TACIlow) resemble plasmablasts22. TACI-Ig, a soluble receptor that blocks the TNFRSF13B ligands BAF and APRIL, inhibits the growth of TACIhigh but not TACIlow myeloma cells in the severe combined immunodeficient mouse-human (SCID-hu) model23.

rs877529 localizes to intron 2 of CBX7, the chromobox homolog 7 gene (MIM 608457) at 22q13.1 (39,542,292 bp; $P = 7.63 \times 10^{-16}$; Tables 1 and Fig. 1). CBX7 encodes a polycomb group protein; proteins from this group form part of a gene regulatory mechanism that determines cell fate during development and also contribute to the control of normal cell growth and differentiation24. CBX7-mediated repression of transcription acts through CDKN2A (also called INK4A or ARF)25, cooperating with MYC to promote aggressive B-cell lymphomagenesis, with high amounts of CBX7 being a feature of germinal center–derived follicular lymphoma26.

To explore whether any of the associations reflect cis-acting regulatory effects, we studied mRNA expression in CD138–selected plasma cells27 and lymphoblastoid cell line28–30 (Online Methods and Supplementary Table 5). Although we found no association between genotype and the expression of either mRNA transcript, steady-state levels of RNA at a single time point may not adequately capture the impact of differential expression in tumorigenesis. To explore the epigenetic profiles of the association signals, we used chromatin state segmentation in lymphoblastoid cell–line data generated by the ENCODE Project31. rs2293607 maps to a region of active chromatin that is predicted by the ENCODE data to be an active promoter, and rs877529 maps to within a stronger enhancer element in CBX7 (Fig. 1).

Hierarchically, multiple myeloma can be divided broadly into hyperdiploid and nonhyperdiploid subtypes32,33. The latter is composed primarily of patients harboring IgH translocations, principally t(11;14)(q13;q32) and t(4;14)(p16;q32)34,35. A case-only analysis provided no evidence for a subtype-specific association with genotype for rs10936599, rs2285803 or rs4237077, which is consistent with each variant having a generic effect on multiple myeloma risk (Supplementary Table 6). In contrast, rs877529 showed evidence that the association is driven by non-t(11;14) multiple myeloma, which remained significant after correction for multiple testing ($P_{\text{adjusted}} = 8.0 \times 10^{-4}$).

Our findings provide further evidence for an inherited genetic susceptibility to multiple myeloma and insight into the development of this hematological malignancy. We estimate that the seven loci we have identified so far account for ~13% of the familial risk of multiple myeloma. Although the power of our study to detect the major common loci conferring risk ≥1.3 was high, we had low power to detect alleles with smaller effects and/or MAF < 0.1. By implication, variants with such profiles probably represent a much larger class of susceptibility loci for multiple myeloma because of the truly small effect sizes or sub-maximal LD with tagging SNPs. Thus, it is probable that a large number of variants remain to be discovered. This assertion is supported by the continued excess of associations observed over those expected, in addition to the regions studied here. Further efforts to expand the scale of GWAS, in terms of both sample size and SNP coverage, and increase the number of SNPs taken forward to large-scale replication may therefore identify additional risk variants. In addition, as we have recently shown, stratified analysis of multiple myeloma by karyotype may lead to the discovery of additional subtype-specific risk variants4.

**METHODS**

Methods and any associated references are available in the online version of the paper.

---

**URLs.** The R suite, http://www.r-project.org/; detailed information on the tag SNP panel, http://www.illumina.com/; dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/; HapMap, http://www.hapmap.org/; 1000 Genomes, http://www.1000genomes.org; KBioscience is now LCL Genomics: http://www.lgcgenomics.com/; SNAP, http://www.broadinstitute.org/mpg/snap/; IMPUTE, https://mathgen.stats.ox.ac.uk/impute/impute.html; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/Software.htm; Wellcome Trust Case Control Consortium, http://www.wtccc.org.uk; METAL, http://www.sph.umich.edu/csg/abecasis/metal; Mendelian Inheritance in Man, http://www.ncbi.nlm.nih.gov/omim; MRC Myeloma-IX trial, http://public.ukcrn.org.uk; MRC Myeloma-XI trial, http://ctru.leeds.ac.uk/myelomaXI; USCS genome browser, http://genome.ucsc.edu/; HLA*IMPM, https://oxfordhla.well.ox.ac.uk/hla/.
ACKNOWLEDGMENTS

Leukaemia Lymphoma Research and Myeloma UK provided principal funding for this study in the UK. Additional funding was provided by Cancer Research UK (CI298/AB362 supported by the Bobby Moore Fund) and the NHS through the Biological Research Centre of the National Institute for Health Research at the Royal Marsden Hospital NHS Trust. This study made use of genotyping data from the 1958 Birth Cohort. Genotyping data on controls were generated by the Wellcome Trust Sanger Institute. A full list of the investigators who contributed to the generation of the data is available at http://www.wtccc.org.uk. In Germany (Heidelberg), funding was provided to Dietmar-Hopp-Stiftung Walldorf, the University Hospital Heidelberg, Deutsche Krebshilfe and the Systems Medicine funding from the German Ministry of Education and Science. We are grateful to all investigators who contributed to the National Study of Colorectal Cancer Genetics (NSCCG) and the Genetic Lung Cancer Predisposition Study (GELCAPS), from which controls in the replication were drawn. The GWAS was made use of genotyping data from the population-based HNR study. The HNR study is supported by the Heinz Nixdorf Foundation (Germany). Additionally, the study is funded by the German Ministry of Education and Science and the German Research Council (DFG; projects SI 236/8-1, SI236/9-1, ER 155/6-1 and DFG CRU 216). Funding was provided to L.E. by the Medical Faculty of the University Hospital of Essen (IFORES). The genotyping of the Illumina HumanOmni-1 Quad BeadChips of the HNR subjects was financed by DZNE, Bonn. We are extremely grateful to all investigators who contributed to the generation of this data set. The German replication controls were collected by P. Bugert, Institute of Transfusion Medicine and Immunology, Medical Faculty Mannheim, Heidelberg University, German Red Cross Blood Service of Baden-Württemberg-Hessen, Mannheim, Germany.

We are grateful to all the patients and investigators at the individual centers for their participation. We also thank the staff of the Clinical Trials Research Unit University of Leeds and the National Cancer Research Institute Haematology Clinical Studies Group.

AUTHOR CONTRIBUTIONS

R.S.H. and K.H. designed the study. R.S.H. and G.M. obtained financial support in the UK, and K.H. and H.G. obtained support in Germany. R.S.H. drafted the manuscript. D.C., B.C. and N.W. performed the principal statistical and bioinformatic analyses. S.E.D. and G.M. performed additional statistical and bioinformatic analyses. P.R. coordinated the UK laboratory analyses. J.V. and A.H. performed genotyping in the UK. D.C.J. managed and prepared Myeloma IX and Myeloma XI case study DNA samples. J.M.A. conceived of the Newcastle-based myeloma study (NMS). J.M.A. established the study and supervised data collection and sample management of the NMS. J.A.L., G.H.J., G.P., J.A.L.F. and C.F. developed protocols for the recruitment of individuals with myeloma and performed sample collection of cases within the NMS. H.G., D.H., K.N. and C.F. developed protocols for the recruitment of individuals with myeloma and performed the principal statistical and R.S.H. and K.H. designed the study. R.S.H. and G.M. obtained financial support in the UK. Additional funding was provided by Cancer Research UK (CI298/AB362 supported by the Bobby Moore Fund) and the NHS through the Biological Research Centre of the National Institute for Health Research at the Royal Marsden Hospital NHS Trust. This study made use of genotyping data from the 1958 Birth Cohort. Genotyping data on controls were generated by the Wellcome Trust Sanger Institute. A full list of the investigators who contributed to the generation of the data is available at http://www.wtccc.org.uk. In Germany (Heidelberg), funding was provided to Dietmar-Hopp-Stiftung Walldorf, the University Hospital Heidelberg, Deutsche Krebshilfe and the Systems Medicine funding from the German Ministry of Education and Science. We are grateful to all investigators who contributed to the National Study of Colorectal Cancer Genetics (NSCCG) and the Genetic Lung Cancer Predisposition Study (GELCAPS), from which controls in the replication were drawn. The GWAS was made use of genotyping data from the population-based HNR study. The HNR study is supported by the Heinz Nixdorf Foundation (Germany). Additionally, the study is funded by the German Ministry of Education and Science and the German Research Council (DFG; projects SI 236/8-1, SI236/9-1, ER 155/6-1 and DFG CRU 216). Funding was provided to L.E. by the Medical Faculty of the University Hospital of Essen (IFORES). The genotyping of the Illumina HumanOmni-1 Quad BeadChips of the HNR subjects was financed by DZNE, Bonn. We are extremely grateful to all investigators who contributed to the generation of this data set. The German replication controls were collected by P. Bugert, Institute of Transfusion Medicine and Immunology, Medical Faculty Mannheim, Heidelberg University, German Red Cross Blood Service of Baden-Württemberg-Hessen, Mannheim, Germany. We are grateful to all the patients and investigators at the individual centers for their participation. We also thank the staff of the Clinical Trials Research Unit University of Leeds and the National Cancer Research Institute Haematology Clinical Studies Group.

AUTHOR CONTRIBUTIONS

R.S.H. and K.H. designed the study. R.S.H. and G.M. obtained financial support in the UK, and K.H. and H.G. obtained support in Germany. R.S.H. drafted the manuscript. D.C., B.C. and N.W. performed the principal statistical and bioinformatic analyses. S.E.D. and G.M. performed additional statistical and bioinformatic analyses. P.R. coordinated the UK laboratory analyses. J.V. and A.H. performed genotyping in the UK. D.C.J. managed and prepared Myeloma IX and Myeloma XI case study DNA samples. J.M.A. conceived of the Newcastle-based myeloma study (NMS). J.M.A. established the study and supervised data collection and sample management of the NMS. J.A.L., G.H.J., G.P., J.A.L.F. and C.F. developed protocols for the recruitment of individuals with myeloma and performed sample collection of cases within the NMS. H.G., D.H., K.N. and C.F. coordinated and managed the German DNA samples, and K.H. and A.F. coordinated the German genotyping. H.E., C.L. and C.S. ascertainment and collected DSSM and Ulm case study samples, and C.L. prepared DNA samples. E.D. and N.W. performed genotyping of German replication cases and controls. B.A.W. performed UK expression analyses. F.M.R. performed UK and A.J. performed German FISH analyses. G.J.M., F.E.D., W.A.G., G.H.J. and J.A.I. performed ascertainment and collection of case study samples. P.H., T.W.M. and M.M.N. performed and coordinated the GWAS of German cases and controls. L.E. ascertainment and managed the HNR sample. All authors contributed to the final paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
ONLINE METHODS

Ethics. Collection of samples and clinicopathological information from subjects was undertaken with informed consent and relevant ethical review board approval in accordance with the tenets of the Declaration of Helsinki.

GWAS. UK GWAS. Details of this study have been reported previously. Briefly, 1,371 cases (ICD-10 C00.0; 469 male; mean age at diagnosis, 63.9 years; s.d., 9.9 years) were ascertained through the UK MRC Myeloma IX trial. Genotyping of cases was performed using Illumina Human OmniExpress-12 v1.0 arrays according to the manufacturer’s protocols (Illumina, San Diego, USA). For controls, we used publicly accessible data generated by WTCCC2 from the 1958 Birth Cohort (58C; also known as the National Child Development Study) and the UKBS. Genotyping of controls was conducted using Illumina Human 1-2M-Duo Custom v1. Array chips. SNP calling was performed using Illuminus Software. The full details of genotyping, SNP calling and quality control have been reported previously (http://www.wtccc.org.uk).

German GWAS. The German GWAS comprised 384 cases (229 male; mean age at diagnosis, 54.5 years; s.d., 8.0 years), which were the subject of a previous publication, and an additional series of 698 cases (389 male; mean age at diagnosis, 59 years; s.d., 9.3 years) recruited by the German Multiple Myeloma Study Group (GMMG) coordinated by the University Clinic, Heidelberg. All cases were genotyped using Illumina Human OmniExpress-12 v1.0 arrays according to the manufacturer’s protocols (Illumina, San Diego, USA). For controls, we used genotype data on 2,132 healthy individuals enrolled in the HNR controls; of these controls, 704 were genotyped using Illumina HumanOmni1-Quad v1. and 1,428 were genotyped using the OmniExpress-12 v1.0.

Quality control of GWAS data sets. DNA samples with GenCall scores <0.25 at any locus were considered to be ‘no calls’. A SNP was deemed to have failed if <95% of the DNA samples generated a genotype at the locus. Cluster plots were manually inspected for all SNPs considered for replication. The same quality-control metrics on the new German GWAS data were applied as in our previous multiple myeloma study. We restricted our analyses to samples in which >95% of SNPs were successfully genotyped, thus eliminating ten SNPs having MAF <1% and call rate <95% in cases or controls. We removed SNPs not present in the main CEU cluster was excluded from analyses. We removed measures on which to perform principal components analysis. The first two shared between HapMap and our SNP panel and used these as dissimilarity each pair of individuals, we calculated genome-wide IBS distances on markers 60 Nigerian (YRI), 90 Japanese (JPT) and 90 Han Chinese (CHB) samples). For control data with phase II HapMap samples (60 western European (CEU), who might have non–western European ancestry, we merged our case and

Sample preparation. For German cases, DNA was prepared from ethylenediaminetetraacetic acid (EDTA)–venous blood samples, 100% of the original GWAS, 42% of the additional GWAS and 91% of the replication samples; for the remaining cases, the source was the CD138–negative fraction of bone marrow cells, with <5% contamination by tumor cells. For all UK cases, DNA was prepared from EDTA–venous blood samples. Samples were obtained before delivery of chemotherapy in the vast majority of UK cases and in at least 80% of the German cases. All DNAs were extracted using QIagen FlexiGene or QIAmp methodologies and quantified using PicoGreen (Invitrogen). Statistical and bioinformatic analyses. The main analyses were undertaken using R (v2.6), STATA v10.0 (State College, Texas, US) and PLINK (v1.06) software. ORs, 95% CIs and associated P values were calculated by unconditioned logistic regression. The adequacy of the case–control matching and the possibility of differential genotyping of cases and controls were formally evaluated using quintile–quintile plots of test statistics. The inflation factor, λ, was based on the 90% least significant SNPs. We undertook adjustment for possible population substructure using EIGENSTRAT software. The meta-analysis was conducted using standard methods, Cochran’s Q statistic to test for heterogeneity and the F statistic to quantify the proportion of the total variation due to heterogeneity were calculated. P values <0.05% are considered to be characteristic of large heterogeneity. To conduct a pooled analysis incorporating EIGENSTRAT-adjusted P values from the GWAS, we used the weighted Z-method implemented in the program METAL. Because not all the HNR controls were genotyped using the same Illumina array, the robustness of the genome-wide associations was formally assessed by deriving ORs for the different German case–control combinations and incorporation of these data in the meta-analysis. We examined each SNP for dose response by comparing 1-degree-of-freedom (1-d.f.) and 2-d.f. logistic regression models, adjusting for stage using a likelihood ratio test, and examined the combined effects of multiple SNPs by evaluating the effect of adding an interaction term to the model using a likelihood ratio test and adjusting for stage. Associations by tumor karyotype were examined by logistic regression in case–only analyses. The sibling relative risk attributable to a given SNP was calculated using the formula:

where p is the population frequency of the minor allele, q = 1 - p, and r1 and r2 are the relative risks (estimated as ORs) for heterozygotes and rare homozygotes, respectively, relative to common homozygotes. Assuming a multiplicative interaction, the proportion of the familial risk attributable to a SNP was calculated as log λe/ log λe, where λe is the overall familial relative risk estimated from epidemiological studies, which was assumed to be 2.45 (ref. 42).

Prediction of the untyped SNPs was carried out using IMPUTE2 based on the 1000 Genomes phase 1 integrated variant set (b37) from March 2012. Imputed data were analyzed using SNPPTEST v2 to account for uncertainties in SNP prediction, and the meta-analysis was performed using METAV1.4 (ref. 43).
LD metrics were calculated in PLINK using 1000 Genomes data and plotted using SNAP. LD blocks were defined on the basis of the HapMap recombination rate (cM/Mb) defined using the Oxford recombination hotspots and on the basis of the distribution of CIs defined by Gabriel et al. We imputed classical HLA alleles from GWAS SNPs using HLA*IMP17,18.

To explore the epigenetic profiles of the association signals, we used chromatin state segmentation in lymphoblastoid cell–line data generated by the ENCODE Project. The states were inferred from ENCODE Histone Modification data (histone H4 Lys20 methylation (H4K20me1), histone H3 Lys9 acetylation (H3K9ac), H3K4me3, H3K4me2, H3K4me1, H3K36me3, H3K27me3, H3K27ac and CTCF) binarized using a multivariate hidden Markov model.

**Karyotyping and FISH.** Conventional cytogenetic studies of multiple myeloma cells were conducted using standard karyotyping methodologies, and standard criteria for the definition of a clone were applied. FISH and ploidy classification of UK samples were performed as previously described. The XL IGH Break Apart probe (MetaSystems, Altlussheim Germany) was used to detect any IgH translocation in German samples.

**Relationship between SNP genotype and mRNA expression.** To examine for a relationship between SNP genotype and mRNA expression in multiple myeloma, we used Affymetrix Human Genome U133+2.0 array data on the plasma cells from 192 patients with multiple myeloma from the MRC Myeloma IX trial. To assay TERC, which was not captured on the U133+2.0 array, we used Affymetrix GeneChip miRNA 2.0 data. To examine for a relationship between SNP genotype and expression levels in lymphocytes, we used publicly available expression data generated on lymphoblastoid cell lines from HapMap3 (ref. 29), Geneva GenCord individuals and the MuTHER resource using Sentrix Human-6 Expression BeadChips (Illumina, San Diego, USA)46,49.

36. Eisen, T., Matakidou, A. & Houlston, R. Identification of low penetrance alleles for lung cancer: the GEnetic Lung CANcer Predisposition Study (GELCAPS). *BMC Cancer* 8, 244 (2008).
37. Penegar, S. et al. National study of colorectal cancer genetics. *Br. J. Cancer* 97, 1305–1309 (2007).
38. 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).
39. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558 (2002).
40. Ioannidis, J.P., Ntzani, E.E. & Trikalinos, T.A. ‘Racial’ differences in genetic effects for complex diseases. *Nat. Genet.* 36, 1312–1318 (2004).
41. Wilter, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191 (2010).
42. Altieri, 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).
43. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* 39, 906–913 (2007).
44. 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).
45. Gabriel, S.B. et al. The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229 (2002).
46. Chiecchio, L. et al. Deletion of chromosome 13 detected by conventional cytogenetics is a critical prognostic factor in myeloma. *Leukemia* 20, 1610–1617 (2006).
47. 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).
48. Stranger, B.E. et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet.* 1, e78 (2005).
49. Strand, U. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315, 848–853 (2007).