Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma

Molly Went[1*], Amit Sud[1*], Asta Försti[2,3*], Britt-Marie Halvarsson[4*], Niels Weinhold[5,6*], Scott Kimber[7], Mark van Duin[8], Gudmar Thorleifsson[9], Amy Holroyd[1], David C. Johnson[7], Ni Li[1], Giulia Orlando[1], Philip J. Law[1], Mina Ali[4], Bowang Chen[2], Jonathan S. Mitchell[1], Daniel F. Gudbjartsson[9,10], Rowan Kuiper[8], Owen W. Stephens[5], Uta Bertsch[2,11], Peter Broderick[1], Chiara Campo[2], Obul R Bandapalli[2], Hermann Einsele[12], Walter A. Gregory[13], Urban Gullberg[4], Jens Hillengass[6], Per Hoffmann[14,15], Graham H. Jackson[16], Karl-Heinz Jöckel[17], Ellinor Johnsson[4], Sigurður Y. Kristinsson[18], Ulf-Henrik Mellqvist[19], Hareth Nahi[20], Douglas Easton[21,22], Paul Pharoah[21,22], Alison Dunning[21], Julian Peto[23], Federico Canzian[24], Anthony Swerdlow[1,25], Rosalind A. Eeles[1,26], ZSofia Kote-Jarai[1], Kenneth Muir[27,28], Nora Pashayan[21,29], the PRACTICAL consortium[¥], Jolanta Nickel[6], Markus M. Nöthen[14,30], Thorunn Rafnar[9], Fiona M. Ross[31], Miguel Inacio da Silva Filho[2], Hauke Thomsen[2], Ingemar Turesson[32], Annette Vangsted[33], Niels Frost Andersen[34], Anders Waage[35], Brian A. Walker[5], Anna-Karin Wihlborg[4], Annemiek Broyl[8], Faith E. Davies[5], Unnur Thorsteinsdottir[9,36], Christian Langer[37], Markus Hansson[4,32], Hartmut Goldschmidt[6,11†], Martin Kaiser[7], Pieter Sonneveld[8], Kari Stefansson[9,†] Gareth J. Morgan[5†], Kari Hemminki[2,3†‡], Björn Nilsson[4,38†‡], and Richard S. Houlston[1,7†‡]

[1]Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK

[2]German Cancer Research Center, 69120 Heidelberg, Germany

[3]Center for Primary Health Care Research, Lund University, SE-205 02 Malmo, Sweden

[4]Hematology and Transfusion Medicine, Department of Laboratory Medicine, BMC B13, SE-221 84 Lund University, Sweden

[5]Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, USA

[6]Department of Internal Medicine V, University of Heidelberg, 69117 Heidelberg, Germany

[7]Division of Molecular Pathology, The Institute of Cancer Research, Surrey SM2 5NG, UK

[8]Department of Hematology, Erasmus MC Cancer Institute, 3075 EA Rotterdam, The Netherlands

[9]deCODE Genetics, Sturlugata 8, IS-101 Reykjavik, Iceland

[10]School of Engineering and Natural Sciences, University of Iceland, IS-101 Reykjavik, Iceland

[11]National Centre of Tumor Diseases, 69120 Heidelberg, Germany

[12]University Clinic of Würzburg, 97080 Würzburg, Germany

[13]Clinical Trials Research Unit, University of Leeds, Leeds LS2 9PH, UK

[14]Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany

[15]Division of Medical Genetics, Department of Biomedicine, University of Basel, 4003 Basel, Switzerland

[16]Royal Victoria Infirmary, Newcastle upon Tyne, NE1 4LP, UK

[17]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University of Duisburg–Essen, Essen, Germany

[18]Department of Hematology, Landspitali, National University Hospital of Iceland, IS-101 Reykjavik, Iceland

[19]Section of Hematology, Sahlgrenska University Hospital, Gothenburg 413 45, Sweden

[20]Center for Hematology and Regenerative Medicine, pau, SE-171 77 Stockholm, Sweden

[21]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK

[22]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[23]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

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

[25]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK

[26]Royal Marsden NHS Foundation Trust, London, UK.

[27]Institute of Population Health, University of Manchester, Manchester, UK.

[28]Warwick Medical School, University of Warwick, Coventry, UK.

[29]University College London, Department of Applied Health Research, London, UK.

[30]Department of Genomics, Life & Brain Center, University of Bonn, D-53127 Bonn, Germany

[31]Wessex Regional Genetics Laboratory, University of Southampton, Salisbury SP2 8BJ, UK

[32]Hematology Clinic, Skåne University Hospital, SE-221 85 Lund, Sweden

[33]Department of Haematology, University Hospital of Copenhagen at Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark

[34]Department of Haematology, Aarhus University Hospital, Tage-Hansens Gade 2, DK-8000 Aarhus C, Denmark

[35]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Box 8905, N-7491 Trondheim, Norway

[36]Faculty of Medicine, University of Iceland, IS-101 Reykjavik, Iceland

[37]Department of Internal Medicine III, University of Ulm, D-89081 Ulm, Germany

[38]Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA

‡Corresponding authors: Richard S Houlston (richard.houlston@icr.ac.uk), Kari Hemminki (K.Hemminki@dkfz-heidelberg.de), or Björn Nilsson (bjorn.nilsson@med.lu.se)

*These authors contributed equally to this work. †These authors jointly supervised the work.

¥Additional members from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium are provided in the **End Notes**.

ABSTRACT

Genome-wide association studies (GWAS) have transformed our understanding of susceptibility to multiple myeloma (MM), but much of the heritability remains unexplained. We report a new GWAS, a meta-analysis with previous GWAS and a replication series, totalling 9,974 MM cases and 247,556 controls of European ancestry. Collectively, these data provide evidence for six new MM risk loci, bringing the total number to 23. Integration of information from gene expression, epigenetic profiling and *in situ* Hi-C data for the 23 risk loci implicate disruption of developmental transcriptional regulators as a basis of MM susceptibility, compatible with altered B-cell differentiation as a key mechanism. Dysregulation of autophagy/apoptosis and cell cycle signalling feature as recurrently perturbed pathways. Our findings provide further insight into the biological basis of MM.

INTRODUCTION

Multiple myeloma (MM) is a malignancy of plasma cells primarily located within the bone marrow. Although no lifestyle or environmental exposures have been consistently linked to an increased risk of MM, the two- to four-fold increased risk observed in relatives of MM patients provides support for inherited genetic predisposition[1]. Our understanding of MM susceptibility has recently been informed by genome-wide association studies (GWAS), which have so far identified 17 independent risk loci for MM[2-5], with an additional locus being subtype-specific for t(11;14) translocation MM[6]. Much of the heritable risk of MM, however, remains unexplained and statistical modelling indicates that further common risk variants remain to be discovered[7].

To gain a more comprehensive insight into MM aetiology, we performed a new GWAS followed by a meta-analysis with existing GWAS and replication genotyping (totalling 9,974 cases and 247,556 controls). Here we report the identification of six new MM susceptibility loci as well as refined risk estimates for the previously reported loci. In addition, we have investigated the possible gene regulatory mechanisms underlying the associations seen at all 23 GWAS risk loci by analysing *in situ* promoter Capture Hi-C (CHi-C) in MM cells to characterize chromatin interactions between predisposition single nucleotide polymorphism (SNPs) and target genes, integrating these data with chromatin immunoprecipitation-sequencing (CHiP-seq) data generated in house and a range of publicly available genomics data. Finally, we have quantified the contribution of both new and previously discovered loci to the heritable risk of MM and implemented a likelihood-based approach to estimate sample sizes required to explain 80% of the heritability.

RESULTS

Association analysis

We conducted a new GWAS using the OncoArray platform[8] (878 MM cases and 7,083 controls from the UK), followed by a meta-analysis with six published MM GWAS data sets (totalling 7,319 cases and 234,385 controls) (**Fig. 1, Supplementary Tables 1-3**)[2-5]. To increase genomic resolution, we imputed data to >10 million SNPs. Quantile-quantile (*Q*-*Q*) plots for SNPs with minor allele frequency (MAF) >1% after imputation did not show evidence of substantive over-dispersion for the OncoArray GWAS (λ = 1.03, λ_{1000}=1.02, **Supplementary Fig. 1**). We derived joint odds ratios (ORs) under a fixed-effects model for each SNP with MAF >1%. Finally, we sought validation of nine SNPs associated at $P < 1 \times 10^{-6}$ in the meta-analysis, which did not map to known MM risk loci and displayed a consistent OR across all GWAS data sets, by genotyping an additional 1,777 cases and 6,088 controls from three independent series (Germany, Denmark and Sweden). After meta-analysis of the new and pre-existing GWAS data sets and replication series, we identified genome-wide significant associations (*i.e.* $P < 5 \times 10^{-8}$)[9] for six new loci at 2q31.1, 5q23.2, 7q22.3, 7q31.33, 16p11.2 and 19p13.11 (**Table 1, Supplementary Table 4** and **5, Fig. 2**). Additionally, borderline associations were identified at two loci with *P*-values of 5.93×10^{-8} (6p25.3) and 9.90×10^{-8} (7q21.11), which have corresponding Bayesian false-discovery probabilities[10] of 4% and 6%, respectively (**Supplementary Table 4** and **5**). We found no evidence for significant interactions between any of the 23 risk loci. Finally, we found no evidence to support the existence of the putative risk locus at 2p12.3 (rs1214346), previously proposed by Erickson *et al* [11](GWAS meta-analysis *P*-value = 0.32).

Risk SNPs and myeloma phenotype

We did not find any association between sex or age at diagnosis and the 23 MM risk SNPs using case-only analysis (**Supplementary Table 6** and **7**). Aside from previously reported relationships between the risk loci at 11q13.3 and 5q15 with t(11;14) MM[6] and hyperdiploid MM[12], respectively, we found no evidence for subtype-specific associations (**Supplementary Table 8-11**) or an impact on MM-specific survival (**Supplementary Table 12**). A failure to demonstrate additional relationships may, however, be reflective of limited study power. Collectively, these data suggest that the risk variants are likely to have generic effects on MM development.

Contribution of risk SNPs to heritability

Using Linkage Disequilibrium Adjusted Kinships (LDAK)[13], the heritability of MM ascribable to all common variation was 15.6% (±4.7); collectively the previously identified and new risk loci account for 15.7% of the GWAS heritability (13.6% and 2.1% respectively). To assess the collective impact of all identified risk SNPs we constructed polygenic risk scores (PRS) considering the combined effect of all risk SNPs modelled under a log-normal relative risk distribution[14]. Using this approach, an individual in the top 1% of genetic risk has a 3-fold increased risk of MM when compared to an individual with median genetic risk (**Supplementary Fig. 2**). We observed an enrichment of risk variants among familial MM compared with both sporadic MM cases and population-based controls comparable to that expected in the absence of a strong monogenic predisposition (respective *P*-values 0.027 and 1.60×10^{-5}; **Supplementary Fig. 3**). Undoubtedly, the identification of further risk loci through the analysis of larger GWAS are likely to improve the performance of any PRS model. To estimate the sample size required to explain a greater proportion of the GWAS heritability, we implemented a likelihood-based approach using association statistics in combination with LD information to model the effect-size distribution[15,16]. The effect-size distributions for susceptibility SNPs were best modelled using the three-component model (mixture of two normal distributions) (**Supplementary Fig. 4**). Under this model, to identify SNPs explaining 80% of the GWAS heritability is likely to require sample sizes in excess of 50,000 (**Supplementary Fig. 5**).

Functional annotation and biological inference of risk loci

To the extent that they have been studied, many GWAS risk SNPs localise to non-coding regions and influence gene regulation[17]. To investigate the functional role of previously reported and new MM risk SNPs we performed a global analysis of SNP associations using ChIP-seq data generated on the MM cell line KMS11 and publicly accessible naïve B-cell Blueprint Epigenome Project data[18]. We found enrichment of MM SNPs in regions of active chromatin, as indicated by the presence of H3K27ac, H3K4Me3 and H3K4Me1 marks (**Supplementary Fig. 6**). We also observed an enrichment of relevant B-cell transcription factor (TF) binding sites using ENCODE GM12878 lymphoblastoid cell line data (**Supplementary Fig. 7**). Collectively these data support the tenet that the MM predisposition loci influence risk through effects on *cis*-regulatory networks involved in transcriptional initiation and enhancement.

Since genomic spatial proximity and chromatin looping interactions are key to the regulation of gene expression, we interrogated physical interactions at respective genomic regions in KMS11 and naïve B-cells using CHi-C data[19]. We also sought to gain insight into the possible biological mechanisms for associations by performing an expression quantitative trait locus (eQTL) analysis using mRNA expression data on CD138-purified MM plasma cells; specifically, we used Summary data-based Mendelian Randomization (SMR) analysis[20] to test for pleiotropy between GWAS signal and *cis*-eQTL for genes within 1 Mb of the sentinel SNP to identify a causal relationship. We additionally annotated risk loci with variants mapping to binding motifs of B-cell specific TFs. Finally, we catalogued direct promoter variants and non-synonymous coding mutations for genes within risk loci (**Table 2 and Fig. 1**).

Although preliminary, and requiring functional validation, our analysis delineates four potential candidate disease mechanisms across the 23 MM risk loci (**Table 2**). Firstly, four of the risk loci contain candidate genes linked to regulation of cell cycle and genomic instability, as evidenced by Hi-C looping interactions in KMS11 cells to *MTAP* (at 9p21.3) and eQTL effects for *CEP120* (at 5q23.2). *CEP120* is required for microtubule assembly and elongation, with overexpression of *CEP120* leading to uncontrolled centriole elongation[21]. rs58618031 (7q31.33) maps 5′ of *POT1*, the protection of telomeres 1 gene. POT1 is part of the shelterin complex that functions to protect telomeres and maintain chromosomal stability[22,23]. While mutated *POT1* is not a feature of MM, it is commonly observed in B-cell chronic lymphocytic leukaemia[24-26]. The looping interaction from the rs58618031 annotated enhancer element implicates *ASB15*. Members of the ASB-family feature as protein components of the ubiquitin-proteasome system, intriguingly a therapeutic target in MM[27-29].

Second, candidate genes encoding proteins involved in chromatin remodelling were implicated at three of the MM risk loci, supported by promoter variants at 2q31.1, 7q36.1 and 22q13.1. The new locus at 2q31.1 implicates *SP3*, encoding a TF, which through promoter interaction, has a well-established role in B-cell development influencing the expression of germinal centre genes, including activation-induced cytidine deaminase AID[30,31].

Third, the central role *IRF4-MYC*-mediated apoptosis/autophagy in MM oncogenesis is supported by variation at five loci, including eQTL effects *WAC* (at 10p12.1) and Hi-C looping interactions (at 8q24.21 and 16q23.1). The 7p15.3 association ascribable to rs4487645 has been documented to

influence expression of *c-MYC*-interacting *CDCA7L* through differential IRF4 binding [32]. Similarly, the long-range interaction between *CCAT1* (colon cancer-associated transcript 1) and *MYC* provides an attractive biological basis for the 8q24.21 association, given the notable role of *MYC* in MM[33,34]. It is noteworthy that the promising risk locus at 6p25.3 contains *IRF4*. At the new locus 19p13.11, the missense variant (NP_057354.1:p.Leu104Pro) and the correlated promoter SNP rs11086029 implicates *KLF2* in MM biology. Demethylation by KDM3A histone demethylase sustains *KLF2* expression and influences IRF4-dependent MM cell survival[35]. The new 16p11.2 risk locus contains a number of genes including Proline-Rich Protein 14 (*PRR14*), which is implicated in PI3-kinase/Akt/mTOR signalling, a therapeutic target in myelomatous plasma cells [36].

Fourth, loci related to B-cell and plasma cell differentiation and function are supported by variation at three loci, including eQTL effects (*ELL2* at 5q15) and Hi-C looping interactions (at 6q21). As previously inferred from GM12878 cell line data, the region at 6q21 (rs9372120, *ATG5*) participates in intra-chromosome looping with the B-cell transcriptional repressor *PRDM1* (alias *BLIMP1*)[4]. Additionally, SNP rs34562254 at 17p11.2 is responsible for the amino acid substitution (NP_036584.1:p.Pro251Leu) in TNFRSF13B, a key regulator of normal B-cell homeostasis, which has an established role in MM biology[37-42].

DISCUSSION

Our meta-analysis of a new GWAS series in conjunction with previously published MM datasets has identified six novel risk loci. Together, the new and previously reported loci explain an estimated 16% of the SNP heritability for MM in European populations. Ancestral differences in the risk of developing MM are well recognised, with a greater prevalence of MM in African Americans as compared with those with European ancestry[43]. It is plausible that the effects of MM risk SNPs may differ between Europeans and non-Europeans and hence contribute to differences in prevalence rates. Thus far there has only been limited evaluation of this possibility with no evidence for significant differences[44].

Integration of Hi-C data with ChIP-seq chromatin profiling from MM and lymphoblastoid cell lines and naïve B-cells, and eQTL analysis, using patient expression data, has allowed us to gain preliminary insight into the biological basis of MM susceptibility. This analysis suggests a model of MM susceptibility based on transcriptional dysregulation consistent with altered B-cell differentiation, where dysregulation of autophagy/apoptosis and cell cycle signalling feature as recurrently modulated pathways. Specifically, our findings implicate mTOR-related genes *ULK4*, *ATG5* and *WAC*, and by virtue of the role of *IRF4-MYC* related autophagy, *CDCA7L*, *DNMT3A, CBX7* and *KLF2* in MM development (**Table 2**). Further investigations are necessary to decipher the functional basis of risk SNPs, nevertheless we highlight mTOR-signalling and the ubiquitin-proteasome pathway, targets of approved drugs in MM. As a corollary of this, genes elucidated via the functional annotation of GWAS discovered MM risk loci may represent promising therapeutic targets for myeloma drug discovery. Finally, our estimation of sample sizes required to identify a larger proportion of the heritable risk of MM attributable to common variation underscore the need for further international collaborative analyses.

METHODS

Ethics

Collection of patient samples and associated clinico-pathological information was undertaken with written informed consent and relevant ethical review board approval at respective study centres in accordance with the tenets of the Declaration of Helsinki. Specifically for the Myeloma-IX trial by the Medical Research Council (MRC) Leukaemia Data Monitoring and Ethics committee (MREC 02/8/95, ISRCTN68454111), the Myeloma-XI trial by the Oxfordshire Research Ethics Committee (MREC 17/09/09, ISRCTN49407852), HOVON65/GMMG-HD4 (ISRCTN 644552890; METC 13/01/2015), HOVON87/NMSG18 (EudraCTnr 2007-004007-34, METC 20/11/2008), HOVON95/EMN02 (EudraCTnr 2009-017903-28, METC 04/11/10), University of Heidelberg Ethical Commission (229/2003, S-337/2009, AFmu-119/2010), University of Arkansas for Medical Sciences Institutional Review Board (IRB 202077), Lund University Ethical Review Board (2013/54), the Norwegian REK 2014/97, the Danish Ethical Review Board (no: H-16032570) and Icelandic Data Protection Authority (2,001,010,157 and National Bioethics Committee 01/015).

The diagnosis of MM (ICD-10 C90.0) in all cases was established in accordance with World Health Organization guidelines. All samples from patients for genotyping were obtained before treatment or at presentation.

Primary GWAS

We analysed constitutional DNA (EDTA-venous blood derived) from 931 cases ascertained through the UK Myeloma XI trial; detailed in **Supplementary Table 1**. Cases were genotyped using the Illumina OncoArray (Illumina Inc. San Diego, CA 92122, USA). Controls were also genotyped using the OncoArray and comprised: (1) 2,976 cancer-free men recruited by the PRACTICAL Consortium - the UK Genetic Prostate Cancer Study (UKGPCS) (age <65 years), a study conducted through the Royal Marsden NHS Foundation Trust and SEARCH (Study of Epidemiology & Risk Factors in Cancer), recruited via GP practices in East Anglia (2003-2009), (2) 4,446 cancer-free women across the UK, recruited via the Breast Cancer Association Consortium (BCAC).

Standard quality-control measures were applied to the GWAS[45]. Specifically, individuals with low SNP call rate (<95%) as well as individuals evaluated to be of non-European ancestry (using the HapMap version 2 CEU, JPT/CHB and YRI populations as a reference) were excluded (**Supplementary Fig. 8**). For apparent first-degree relative pairs, we excluded the control from a case-control pair; otherwise, we excluded the individual with the lower call rate. SNPs with a call rate <95% were excluded as were those with a MAF <0.01 or displaying significant deviation from Hardy–Weinberg equilibrium ($P < 10^{-5}$). GWAS data were imputed to >10 million SNPs using IMPUTE2 v2.3[46] software in conjunction with a merged reference panel consisting of data from 1000 Genomes Project[47] (phase 1 integrated release 3 March 2012) and UK10K[48]. Genotypes were aligned to the positive strand in both imputation and genotyping. We imposed predefined thresholds for imputation quality to retain potential risk variants with MAF >0.01 for validation. Poorly imputed SNPs with an information measure <0.80 were excluded. Tests of association between imputed SNPs and MM was performed under an additive model in SNPTESTv2.5[49]. The adequacy of the case–control matching and possibility of differential genotyping of cases and controls was evaluated using a *Q-Q* plot of test statistics (**Supplementary Fig. 1**). The inflation λ was based on the 90% least-significant SNPs [50] and assessment of λ_{1000}. Details of SNP QC are provided in in **Supplementary Table 2**.

Published GWAS

The data from six previously reported GWAS[2-5] are summarized in **Supplementary Table 1.** All these studies were based on individuals with European ancestry and comprised: UK-GWAS (2,282 cases, 5,197 controls), Swedish-GWAS (1,714 cases, 10,391 controls), German-GWAS (1,508 cases, 2,107 controls), Netherlands-GWAS (555 cases, 2,669 controls), US-GWAS (780 cases, 1,857 controls) and Iceland (480 cases, 212,164 controls).

Replication studies and technical validation

To validate promising associations, we analysed three case-control series from Germany, Sweden and Denmark, summarised in **Supplementary Table 3**. The German replication series comprised 911 cases collected by the German Myeloma Study Group (Deutsche Studiengruppe Multiples Myeloma (DSMM)), GMMG, University Clinic, Heidelberg, and University Clinic, Ulm. Controls comprised 1,477 healthy German blood donors recruited between 2004 and 2007 by the Institute of Transfusion Medicine and Immunology, University of Mannheim, Germany. The Swedish replication series comprised 534 MM cases from the Swedish National Myeloma Biobank and the

Danish replication series comprised 332 MM cases from the University Hospital of Copenhagen. As controls, we analysed 2,382 Swedish blood donors and 2,229 individuals from Denmark and Skåne County, Sweden (the southernmost part of Sweden adjacent to Denmark). Replication genotyping of German and Scandinavian samples was performed using competitive allele-specific PCR KASPar chemistry (LGC, Hertfordshire, UK). Call rates for SNP genotypes were >95% in each of the replication series. To ensure the quality of genotyping in all assays, at least two negative controls and duplicate samples (showing a concordance of >99%) were genotyped at each centre. The fidelity of imputation was assessed by directly sequencing a set of 147 randomly selected samples from the UK OncoArray case series. Imputation was found to be robust; concordance was >90% (**Supplementary Table 13**). Genotyping and sequencing primers are detailed in **Supplementary Table 14** and **15,** respectively.

Meta-analysis

Meta-analyses were performed using the fixed-effects inverse-variance method using META v1.6 [51]. Cochran's *Q*-statistic to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation due to heterogeneity was calculated. Using the meta-analysis summary statistics and LD correlations from a reference panel of the 1000 Genomes Project combined with UK10K, we implemented Genome-wide Complex Trait Analysis[52] to perform conditional association analysis. Association statistics were calculated for all SNPs conditioning on the top SNP in each loci showing genome-wide significance. This was carried out step-wise.

For borderline associations, the Bayesian false-discovery probability (BFDP)[10] was calculated based on a plausible OR of 1.2 and a prior probability of association of 0.0001. For both promising associations, the BFDP was <10%.

Fluorescence *in situ* hybridization

Fluorescence *in situ* hybridization (FISH) and ploidy classification of UK and German samples were performed as previously described[53,54]. Logistic regression in case-only analyses was used to assess the relationship between SNP genotype and IgH translocations or tumour ploidy.

eQTL analysis

eQTL analyses were performed using CD138-purified plasma cells from 183 UK MyIX trial patients and 658 German GMMG patients[32]. Briefly, German and UK data were pre-processed separately,

followed by analysis using a Bayesian approach to probabilistic estimation of expression residuals to infer broad variance components, accounting for hidden determinants influencing global expression. The association between genotype of SNPs and expression of genes within 1 Mb either side of each MM risk locus was evaluated based on the significance of linear regression coefficients. We pooled data from each study under a fixed-effects model.

The relationship between SNP genotype and gene expression we carried out using Summary-data-based Mendelian Randomization (SMR) analysis as per Zhu *et al*[20]. Briefly, if b_{xy} is the effect size of *x* (gene expression) on *y* (slope of *y* regressed on the genetic value of *x*), b_{zx} is the effect of *z* on *x*, and b_{zy} be the effect of *z* on *y*, b_{xy} (b_{zy}/b_{zx}) is the effect of *x* on *y*. To distinguish pleiotropy from linkage where the top associated cis-eQTL is in LD with two causal variants, one affecting gene expression the other affecting trait we tested for heterogeneity in dependent instruments (HEIDI), using multiple SNPs in each *cis*-eQTL region. Under the hypothesis of pleiotropy b_{xy} values for SNPs in LD with the causal variant should be identical. For each probe that passed significance threshold for the SMR test, we tested the heterogeneity in the b_{xy} values estimated for multiple SNPs in the *cis*-eQTL region using HEIDI.

GWAS summary statistics files were generated from the meta-analysis. We set a threshold for the SMR test of $P_{SMR} < 1 \times 10^{-3}$ corresponding to a Bonferroni correction for 45 tests*, i.e*. 45 probes which demonstrated an association in the SMR test. For all genes passing this threshold we generated plots of the eQTL and GWAS associations at the locus, as well as plots of GWAS and eQTL effect sizes (*i.e*. input for the HEIDI heterogeneity test). HEIDI test *P*-values < 0.05 were considered as reflective of heterogeneity. This threshold is, however, conservative for gene discovery because it retains fewer genes than when correcting for multiple testing. SMR plots for significant eQTLs are shown in **Supplementary Fig. 9 and 10** and a summary of results are shown in **Supplementary Table 16.**

Promoter capture Hi-C

To map risk SNPs to interactions involving promoter contacts and identify genes involved in MM susceptibility, we analysed publicly accessible promoter capture Hi-C (CHi-C) data on the naïve B-cells downloaded from Blueprint Epigenome Project. Additionally, we also analysed promotor CHi-C data we have previously generated for the MM cell line KMS11[12]. Interactions were called using

the CHiCAGO pipeline to obtain a unique list of reproducible contacts[55] and those with a −log(weighted *P*) ≥5 were considered significant.

Chromatin state annotation

Variant sets (*i.e*. sentinel risk SNP and correlated SNPs, r^2>0.8) were annotated for putative functional effect based upon histone mark ChIP-seq data for H3K27ac, H3K4Me1, H3K27Me3 , H3K9Me3, H3K36Me3 and H3K27Me3 from KMS11 cell lines, generated in-house, and naïve B-cells from Blueprint Epigenome Project[56]. We used ChromHMM to infer chromatin states by integrating information on these histone modifications, training the model on three MM cell lines; KMS11, MM1S and JJN3. Genome-wide signal tracks were binarized (including input controls for ChIP-seq data), and a set of learned models were generated using ChromHMM software[57]. A 12-state model was suitable for interpretation and biological meaning was assigned to the states based on chromatin marks use putative rules as previously described (**Supplementary Fig. 11**).

TF and histone mark enrichment analysis

To examine enrichment in specific TF binding across risk loci, we adapted the method of Cowper-Sal lari *et al*.[58]. Briefly, for each risk locus, a region of strong LD (defined as r^2>0.8 and D'>0.8) was determined, and these SNPs were considered the associated variant set (AVS). Publically available data on TF ChIP-seq uniform peak data were obtained from ENCODE for the GM12878 cell line, including data for 82 TF and 11 histone marks[59]. In addition, ChIP-seq peak data for 6 histone marks from KMS11 cell line was generated in-house and naïve B-cell ChIP-seq data was downloaded from Blueprint Epigenome Project[56]. For each mark, the overlap of the SNPs in the AVS and the binding sites was assessed to generate a mapping tally. A null distribution was produced by randomly selecting SNPs with the same characteristics as the risk-associated SNPs, and the null mapping tally calculated. This process was repeated 10,000 times, and *P*-values calculated as the proportion of permutations where null mapping tally was greater or equal to the AVS mapping tally. An enrichment score was calculated by normalizing the tallies to the median of the null distribution. Thus, the enrichment score is the number of standard deviations of the AVS mapping tally from the median of the null distribution tallies. Enrichment plots are shown in **Supplementary Fig. 4** and **5**.

Functional annotation

For the integrated functional annotation of risk loci, variant sets (*i.e.* all SNPs in LD $r^2 > 0.8$ with the sentinel SNP) were annotated with: (i) presence of a Hi-C contact linking to a gene promoter, (ii) presence of an association from SMR analysis, (iii) presence of a regulatory ChromHMM state, (iv) evidence of transcription factor binding, (v) presence of a nonsynonymous coding change. Candidate causal genes were then assigned to MM risk loci using the target genes implicated in annotation tracks (i), (ii), (iiii) and (iv). If the data supported multiple gene candidates, the gene with the highest number of individual functional data points was considered as the candidate. Where multiple genes have the same number of data points, all genes are listed. Direct non-synonymous coding variants were allocated additional weighting. Competing mechanisms for the same gene (*e.g*. both coding and promoter variants) were permitted.

Heritability analysis

We used LDAK to estimate the polygenic variance (*i.e.* heritability) ascribable to all genotyped and imputed GWAS SNPs from summary statistic data. SNP specific expected heritability, adjusted for LD, MAF and genotype certainty was calculated from the UK10K and 1000 Genomes data. Samples were excluded with a call rate <0.99 or if individuals were closely related or of divergent ancestry from CEU. Individual SNPs were excluded if they showed deviation from HWE with $P < 1 \times 10^{-5}$, an individual SNP genotype yield <95%, MAF <1%, SNP imputation score <0.99 and the absence of the SNP in the GWAS summary statistic data. This resulted in a total 1,254,459 SNPs which were used to estimate the heritability of MM.

To estimate the sample size required for a given proportion of the GWAS heritability we implemented a likelihood-based approach to model the effect-size distribution[15], using association statistics from the MM meta-analysis, and LD information from individuals of European ancestry in the 1000 Genomes Project Phase 3. LD values were based on an r^2 threshold of 0.1 and a window size of 1 MB. The goodness of fit of the observed distribution of *P*-values against the expected from a two-component model (single normal distribution) and a three-component model (mixture of two normal distributions) were assessed[15], and a better fit was observed for the latter model (**Supplementary Figure 4**). The percentage of GWAS heritability explained for a projected sample size was determined using this model and is based on power calculations for the discovery of genome-wide significant SNPs. The genetic variance explained was calculated as the proportion of

total GWAS heritability explained by SNPs reaching genome-wide significance at a given sample size. The 95% confidence intervals were determined using 10,000 simulations.

Polygenic risk scores for familial MM (n=38) from 25 families were compared with sporadic MM (n=1,530) and population-based controls (n=10,171); first as a simple sum of risk alleles and secondly as sum of risk alleles weighted by their log-transformed odds ratios. Family member scores were averaged. A one-sided Student's t-test was used to assess difference between groups. The genetic data has been previously described[5,60] with the familial MM cases having been identified by linkages of Swedish registry information.

Data availability

SNP genotyping data that support the findings of this study have been deposited in Gene Expression Omnibus with accession codes GSE21349, GSE19784, GSE24080, GSE2658 and GSE15695; in the European Genome-phenome Archive (EGA) with accession code EGAS00000000001; in the European Bioinformatics Institute (Part of the European Molecular Biology Laboratory) (EMBL-EBI) with accession code E-MTAB-362 and E-TABM-1138; and in the database of Genotypes and Phenotypes (dbGaP) with accession code phs000207.v1.p1. Expression data that support the findings of this study have been deposited in GEO with accession codes GSE21349, GSE2658, GSE31161 and EMBL-EBI with accession code E-MTAB-2299. The remaining data are contained within the paper and Supplementary Files or available from the author upon request. KMS11 Hi-C data used in this manuscript are deposited in EGA under accession number EGAS00001002614. The accession number for the KMS11 ChIP-seq data reported in this paper is EGA: S00001002414. Naïve B-cell HiC data used in this work is publically available from Blueprint Blueprint Epigenome Project [https://osf.io/u8tzp/]. ChIP-seq data for H3K27ac, H3K4Me1, H3K27Me3, H3K9Me3, H3K36Me3 and H3K27Me3 from naïve B-cells is publically available and was obtained from Blueprint Epigenome Project [http://www.blueprint-epigenome.eu/].

ACKNOWLEDGMENTS

In the United Kingdom, Myeloma UK and Bloodwise provided principal funding. Additional funding was provided by Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund) and The Rosetrees Trust. M.W. is supported by funding from Mr Ralph Stockwell. A.S. is supported by a clinical fellowship from Cancer Research UK and charitable funds from the Royal Marsden Hospital. N.W. was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM125503 This study made use of genotyping data on the 1958 Birth Cohort generated by the Wellcome Trust Sanger Institute (http://www.wtccc.org.uk). We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of UK myeloma Oncoarray data. The BCAC study would not have been possible without the contributions of the following: Manjeet K. Bolla, Qin Wang, Kyriaki Michailidou and Joe Dennis. BCAC is funded by Cancer Research UK (C1287/A10118, C1287/A16563). For the BBCS study, we thank Eileen Williams, Elaine Ryder-Mills and Kara Sargus. The BBCS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the National Institute of Health Research (NIHR) Biomedical Research Centre (BRC), and the National Cancer Research Network (NCRN). We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (UK). We thank the SEARCH and EPIC teams which were funded by a programme grant from Cancer Research UK [A16561] and supported by the UK NIHR BRC at

the University of Cambridge. We thank Breast Cancer Now and The Institute of Cancer Research (ICR) for support and funding of the UKBGS, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR BRC. UKGPCS would like to thank The Institute of Cancer Research and The Everyman Campaign for funding support. The UKGPCS acknowledges The Prostate Cancer Research Foundation, Prostate Action, The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI), the NIHR funding to the NIHR Biomedical Research data managers and consultants for their work in the UKGPCS study and urologists and other persons involved in the planning, and data collection of the CAPS study. Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) [U19 CA 148537 for ELucidating Loci Involved in Prostate cancer SuscEptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201200008I]. Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher). The PRACTICAL consortium was supported by Cancer Research UK Grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, C16913/A6135, European Commission's Seventh Framework Programme grant agreement n° 223175 (HEALTH-F2-2009-223175), and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON initiative). We would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now Prostate Action), The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. The APBC BioResource, which forms part of the PRACTICAL consortium, consists of the following members: Wayne Tilley, Gail Risbridger, Renea Taylor, Judith A Clements, Lisa Horvath, Vanessa Hayes, Lisa Butler, Trina Yeadon, Allison Eckert, Pamela Saunders, Anne-Maree Haynes, Melissa Papargiris. We thank the staff of the CTRU University of Leeds and the NCRI haematology Clinical Studies Group. The US GWAS was supported by a grant from the National Institutes of Health (P01CA055819). The German study was supported by the Dietmar-Hopp-Stiftung, Germany, the German Cancer Aid (110,131), the German Ministry of Education and Science (CLIOMMICS 01ZX1309), The German Research Council (DFG; Project SI 236/81, SI 236/)-1, ER 155/6-1 and the DFG CRI 216), the Harald Huppert Foundation and the Multiple Myeloma Research Foundation. The patients were collected by the GMMG and DSMM studies. The German

GWAS made use of genotyping data from the population-based HNR study, which is supported by the Heinz Nixdorf Foundation (Germany). The genotyping of the Illumina HumanOmni-1 Quad BeadChips of the HNR subjects was financed by the German Center for Neurodegenerative Disorders (DZNE), Bonn. We are grateful to all investigators who contributed to the generation of this data set. The German replication controls were collected by Peter Bugert, Institute of Transfusion Medicine and Immunology, Medical Faculty Mannheim, Heidelberg University, German Red Cross Blood Service of Baden-Württemberg-Hessen, Mannheim, Germany. This work was supported by research grants from the Swedish Foundation for Strategic Research (KF10-0009), the Marianne and Marcus Wallenberg Foundation (2010.0112), the Knut and Alice Wallenberg Foundation (2012.0193), the Swedish Research Council (2012–1753), the Royal Swedish Academy of Science, ALF grants to the University and Regional Laboratories (Labmedicin Skåne), the Siv-Inger and Per-Erik Andersson Foundation, the Medical Faculty at Lund University, the Borås foundation for cancer research, and the Swedish Society of Medicine. We thank Jörgen Adolfsson, Tomas Axelsson, Anna Collin, Ildikó Frigyesi, Patrik Magnusson, Bertil Johansson, Jan Westin and Helga Ögmundsdóttir for their assistance. This work was supported by Center for Translational Molecular Medicine (BioCHIP), a clinical research grant from the European Hematology Association, an EMCR Translational Research Grant, a BMBF grant from CLIOMMICS (01ZX1309A) and FP7 grant MSCNET (LSHC-Ct-2006-037602). We thank the staff of the HOVON, as well as patients and physicians at participating sites. In addition, we also thank Jasper Koenders, Michael Vermeulen, André Uitterlinden and Nathalie van der Velde for their assistance. We are indebted to the clinicians who contributed samples to Swedish, Icelandic, Norwegian and Danish biobanks. We are indebted to the patients and other individuals who participated in the project.

AUTHOR CONTRIBUTIONS

M.W. and R.S.H. designed the study. M.W. and R.S.H. drafted the manuscript with contributions from K.H., B.N., A.S. and N.W. M.W. performed principal statistical and bioinformatics analyses. A.S., N.L., P.L., D.C.J. and G.O. performed additional bioinformatics analyses. S.K. performed *in situ* CHi-C. P.B. coordinated UK laboratory analyses. M.W and A.H. performed sequencing of UK samples. D.C.J. managed and prepared Myeloma IX and Myeloma XI Case Study DNA samples. M.K., G.J.M., F.E.D., W.A.G. and G.H.J. performed ascertainment and collection of Case Study samples. B.A.W. performed UK expression analyses. F.M.R. performed UK fluorescence *in situ* hybridization analyses. H.G., U.B., J.H., J.N., and N.W. coordinated and managed Heidelberg samples. C.L. and H.E. coordinated and managed Ulm/Wurzburg samples. A.F. coordinated German genotyping. C.C. and O.R.B performed German genotyping. P.H. and M.M.N. performed GWAS of German cases and controls. B.C. and M.I.d.S.F. carried out statistical analysis. K.H. coordinated the German part of the project. M.M.N. generated genotype data from the Heinz-Nixdorf recall study. K-H.J., contributed towards the Heinz-Nixdorf control data set. NN from Bonn and K-H.J. provided samples for the German GWAS. M.H. and B.N. coordinated the Swedish/Norwegian part of the project. M.A. and B.-M.H. performed data analysis. B.S., M.J., E.J., S.L., C.H., A.-K.W., U.-H.M., H.N., S.N., A.V., N.F.A., A.W., I.T. and U.G. performed sample acquisition, sample preparation, clinical data acquisition and additional data analyses of Sweden/Norway samples. In Iceland, G.T. and D.F.G performed statistical analysis. S.Y.K. provided clinical information. T.R. performed additional analyses. U.T. and K.S. performed project oversight. M.v.D., P.S., A.B. and R.K. coordinated and prepared HOVON65/GMMG-HD4, HOVON87/NMSG18 and HOVON95/EMN02 studies for participating in this study, and coordinated genotyping and pre-processing. At the Myeloma Institute, University of Arkansas for Medical Sciences, N.W. coordinated the US part of the project and performed statistical and eQTL analyses. O.W.S. and N.W managed Case Study samples. G.J.M. and F.E.D. performed ascertainment and collection of Case Study samples.

COMPETING INTERESTS

The authors declare no competing non-financial interests but the following competing financial interests: G.T., D.F.G., T.R., K.S. and U.T. are employed by deCode Genetics/Amgen Inc.

REFERENCES

1. Altieri, A., Chen, B., Bermejo, J.L., Castro, F. & Hemminki, K. Familial risks and temporal incidence trends of multiple myeloma. *Eur J Cancer* **42**, 1661-70 (2006).
2. Broderick, P. *et al.* Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet* **44**, 58-61 (2011).
3. Chubb, D. *et al.* Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nat Genet* **45**, 1221-1225 (2013).
4. Mitchell, J.S. *et al.* Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nat Commun* **7**, 12050 (2016).
5. Swaminathan, B. *et al.* Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. *Nat Commun* **6**, 7213 (2015).
6. Weinhold, N. *et al.* The CCND1 c.870G>A polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. *Nat Genet* **45**, 522-525 (2013).
7. Mitchell, J.S. *et al.* Implementation of genome-wide complex trait analysis to quantify the heritability in multiple myeloma. *Sci Rep* **5**, 12473 (2015).
8. Amos, C.I. *et al.* The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev* **26**, 126-135 (2017).
9. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* **273**, 1516-7 (1996).
10. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
11. Erickson, S.W. *et al.* Genome-wide scan identifies variant in 2q12.3 associated with risk for multiple myeloma. *Blood* **124**, 2001-3 (2014).
12. Li, N. *et al.* Genetic Predisposition to Multiple Myeloma at 5q15 Is Mediated by an ELL2 Enhancer Polymorphism. *Cell Rep* **20**, 2556-2564 (2017).
13. Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nat Genet* **49**, 986-992 (2017).
14. Frampton, M.J. *et al.* Implications of polygenic risk for personalised colorectal cancer screening. *Ann Oncol* **27**, 429-34 (2016).
15. Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits and implications for the future. *BioRxiv* (2017).
16. Chatterjee, N. *et al.* Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet* **45**, 400-5, 405e1-3 (2013).
17. Freedman, M.L. *et al.* Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* **43**, 513-8 (2011).
18. Javierre, B.M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384.e19 (2016).
19. Risca, V.I. & Greenleaf, W.J. Unraveling the 3D genome: genomics tools for multiscale exploration. *Trends Genet* **31**, 357-72 (2015).
20. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
21. Comartin, D. *et al.* CEP120 and SPICE1 cooperate with CPAP in centriole elongation. *Curr Biol* **23**, 1360-6 (2013).
22. Rice, C. *et al.* Structural and functional analysis of the human POT1-TPP1 telomeric complex. *Nat Commun* **8**, 14928 (2017).
23. Pinzaru, A.M. *et al.* Telomere Replication Stress Induced by POT1 Inactivation Accelerates Tumorigenesis. *Cell Rep* **15**, 2170-84 (2016).
24. Chang, S. Cancer chromosomes going to POT1. *Nat Genet* **45**, 473-5 (2013).
25. Speedy, H.E. *et al.* A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat Genet* **46**, 56-60 (2014).

26. Speedy, H.E. *et al.* Germ line mutations in shelterin complex genes are associated with familial chronic lymphocytic leukemia. *Blood* **128**, 2319-2326 (2016).
27. McDaneld, T.G., Hannon, K. & Moody, D.E. Ankyrin repeat and SOCS box protein 15 regulates protein synthesis in skeletal muscle. *Am J Physiol Regul Integr Comp Physiol* **290**, R1672-82 (2006).
28. McDaneld, T.G. & Spurlock, D.M. Ankyrin repeat and suppressor of cytokine signaling (SOCS) box-containing protein (ASB) 15 alters differentiation of mouse C2C12 myoblasts and phosphorylation of mitogen-activated protein kinase and Akt. *J Anim Sci* **86**, 2897-902 (2008).
29. Kile, B.T. *et al.* The SOCS box: a tale of destruction and degradation. *Trends Biochem Sci* **27**, 235-41 (2002).
30. Park, S.R. *et al.* HoxC4 binds to the promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA recombination and somatic hypermutation. *Nat Immunol* **10**, 540-50 (2009).
31. Steinke, J.W. *et al.* Identification of an Sp factor-dependent promoter in GCET, a gene expressed at high levels in germinal center B cells. *Mol Immunol* **41**, 1145-53 (2004).
32. Li, N. *et al.* Multiple myeloma risk variant at 7p15.3 creates an IRF4-binding site and interferes with CDCA7L expression. *Nat Commun* **7**, 13656 (2016).
33. Holien, T., Vatsveen, T.K., Hella, H., Waage, A. & Sundan, A. Addiction to c-MYC in multiple myeloma. *Blood* **120**, 2450-3 (2012).
34. Kuehl, W.M. & Bergsagel, P.L. MYC addiction: a potential therapeutic target in MM. *Blood* **120**, 2351-2 (2012).
35. Ohguchi, H. *et al.* The KDM3A-KLF2-IRF4 axis maintains myeloma cell survival. **7**, 10258 (2016).
36. Yang, M., Lewinska, M., Fan, X., Zhu, J. & Yuan, Z.M. PRR14 is a novel activator of the PI3K pathway promoting lung carcinogenesis. *Oncogene* **35**, 5527-5538 (2016).
37. Moreaux, J. *et al.* APRIL and TACI interact with syndecan-1 on the surface of multiple myeloma cells to form an essential survival loop. *Eur J Haematol* **83**, 119-29 (2009).
38. Moreaux, J. *et al.* The level of TACI gene expression in myeloma cells is associated with a signature of microenvironment dependence versus a plasmablastic signature. *Blood* **106**, 1021-30 (2005).
39. Ju, S. *et al.* Correlation of expression levels of BLyS and its receptors with multiple myeloma. *Clin Biochem* **42**, 387-99 (2009).
40. Novak, A.J. *et al.* Expression of BCMA, TACI, and BAFF-R in multiple myeloma: a mechanism for growth and survival. *Blood* **103**, 689-94 (2004).
41. Mackay, F. & Schneider, P. TACI, an enigmatic BAFF/APRIL receptor, with new unappreciated biochemical and biological properties. *Cytokine Growth Factor Rev* **19**, 263-76 (2008).
42. Moreaux, J. *et al.* TACI expression is associated with a mature bone marrow plasma cell signature and C-MAF overexpression in human myeloma cell lines. *Haematologica* **92**, 803-11 (2007).
43. Landgren, O. & Weiss, B.M. Patterns of monoclonal gammopathy of undetermined significance and multiple myeloma in various ethnic/racial groups: support for genetic factors in pathogenesis. *Leukemia* **23**, 1691-7 (2009).
44. Rand, K.A. *et al.* A Meta-analysis of Multiple Myeloma Risk Regions in African and European Ancestry Populations Identifies Putatively Functional Loci. *Cancer Epidemiol Biomarkers Prev* **25**, 1609-1618 (2016).
45. Turner, S. *et al.* Quality control procedures for genome-wide association studies. *Curr Protoc Hum Genet* **Chapter 1**, Unit1.19 (2011).
46. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
47. Abecasis, G.R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
48. Huang, J. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* **6**, 8111 (2015).
49. 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**, 906-13 (2007).
50. Clayton, D.G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* **37**, 1243-6 (2005).

51. Liu, J.Z. *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* **42**, 436-40 (2010).
52. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
53. 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**, 1150-7 (2010).
54. Walker, B.A. *et al.* A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. *Blood* **116**, e56-65 (2010).
55. Cairns, J. *et al.* CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol* **17**, 127 (2016).
56. Fernandez, J.M. *et al.* The BLUEPRINT Data Analysis Portal. *Cell Syst* **3**, 491-495.e5 (2016).
57. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-6 (2012).
58. Cowper-Sal lari, R. *et al.* Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nat Genet* **44**, 1191-8 (2012).
59. de Souza, N. The ENCODE project. *Nat Methods* **9**, 1046 (2012).
60. Halvarsson, B.M. *et al.* Direct evidence for a polygenic etiology in familial multiple myeloma. *Blood Advances* **1**, 619-623 (2017).
61. Via, M., Gignoux, C. & Burchard, E.G. The 1000 Genomes Project: new opportunities for research and social challenges. *Genome Med* **2**, 3 (2010).

END NOTES

[¥]The PRACTICAL Consortium (http://practical.icr.ac.uk/):

OncoArray (in addition to those named in the author list):

Brian E. Henderson[39*], Christopher A. Haiman[39], Sara Benlloch[1], Fredrick R. Schumacher[40,41], Ali Amin Al Olama[22,42], Sonja I. Berndt[43], David V. Conti[39], Fredrik Wiklund[44], Stephen Chanock[43], Victoria L. Stevens[45], Catherine M. Tangen[45], Jyotsna Batra[46,47], Judith Clements[46,47], Henrik Gronberg[44], Johanna Schleutker[48,49,50], Demetrius Albanes[43], Stephanie Weinstein[43], Alicja Wolk[51], Catharine West[52], Lorelei Mucci[53], Géraldine Cancel-Tassin[54,55], Stella Koutros[43], Karina Dalsgaard Sorensen[56,57], Eli Marie Grindedal[58], David E. Neal[59,60], Freddie C. Hamdy[61], Jenny L. Donovan[62], Ruth C. Travis[63], Robert J. Hamilton[64], Sue Ann Ingles[1], Barry Rosenstein[65,66], Yong-Jie Lu[67], Graham G. Giles[68,69], Adam S. Kibel[70], Ana Vega[71], Manolis Kogevinas[72,73,74,75], Kathryn L. Penney[76], Jong Y. Park[77], Janet L. Stanford[78,79], Cezary Cybulski[80], Børge G. Nordestgaard[81,82], Hermann Brenner[83,84,85], Christiane Maier[86], Jeri Kim[87], Esther M. John[88,89], Manuel R. Teixeira[90,91], Susan L. Neuhausen[92], Kim De Ruyck[93], Azad Razack[94], Lisa F. Newcomb[78,95], Davor Lessel[96], Radka Kaneva[97], Nawaid Usmani[98,99], Frank Claessens[100], Paul A. Townsend[101], Manuela Gago Dominguez[102,103], Monique J. Roobol[104], Florence Menegaux[105], Kay-Tee Khaw[106], Lisa Cannon-Albright[107,108], Hardev Pandha[109], Stephen N. Thibodeau[109]

[*]In memorium

39. Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA.
40. Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA.
41. Seidman Cancer Center, University Hospitals, Cleveland, OH, USA.
42. University of Cambridge, Department of Clinical Neurosciences, Cambridge, UK.
43. Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA.
44. Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden.
45. Epidemiology Research Program, American Cancer Society, 250 Williams Street, Atlanta, GA, USA.
46. Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Queensland, Australia.
47. Translational Research Institute, Brisbane, Queensland, Australia.
48. Department of Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, Finland.
49. Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, Finland.
50. BioMediTech, University of Tampere, Tampere, Finland.
51. Division of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Sweden.
52. Institute of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Radiotherapy Related Research, The Christie Hospital NHS Foundation Trust, Manchester, UK.
53. Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA.
54. CeRePP, Pitie-Salpetriere Hospital, Paris, France.
55. UPMC Univ Paris 06, GRC N°5 ONCOTYPE-URO, CeRePP, Tenon Hospital, Paris, France.

56. Department of Molecular Medicine, Aarhus University Hospital, Denmark.
57. Department of Clinical Medicine, Aarhus University, Denmark.
58. Department of Medical Genetics, Oslo University Hospital, Norway.
59. University of Cambridge, Department of Oncology, Addenbrooke's Hospital, Cambridge, UK.
60. Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK.
61. Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK, Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK.
62. School of Social and Community Medicine, University of Bristol, Bristol, UK.
63. Cancer Epidemiology, Nuffield Department of Population Health University of Oxford, Oxford, UK.
64. Dept. of Surgical Oncology, Princess Margaret Cancer Centre, Toronto, Canada.
65. Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
66. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New
67. Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, London, UK.
68. Cancer Epidemiology & Intelligence Division, The Cancer Council Victoria, Melbourne, Victoria, Australia.
69. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia.
70. Division of Urologic Surgery, Brigham and Womens Hospital, Boston, MA, USA.
71. Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica, CIBERER, IDIS, Santiago de Compostela, Spain.
72. Centre for Research in Environmental Epidemiology (CREAL), Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain.
73. CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain.
74. IMIM (Hospital del Mar Research Institute), Barcelona, Spain.
75. Universitat Pompeu Fabra (UPF), Barcelona, Spain.
76. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA, USA.
77. Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, USA.
78. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.
79. Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA.
80. International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland.
81. Faculty of Health and Medical Sciences, University of Copenhagen, Denmark.
82. Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark.
83. Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.
84. German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.
85. Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany.
86. Institute for Human Genetics, University Hospital Ulm, Ulm, Germany.
87. The University of Texas M. D. Anderson Cancer Center, Department of Genitourinary Medical Oncology, Houston, TX, USA.

88. Cancer Prevention Institute of California, Fremont, CA, USA.
89. Department of Health Research & Policy (Epidemiology) and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA , USA.
90. Department of Genetics, Portuguese Oncology Institute of Porto, Porto, Portugal.
91. Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal.
92. Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, CA, USA.
93. Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences, Gent, Belgium.
94. Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia.
95. Department of Urology, University of Washington, Seattle, WA, USA.
96. Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.
97. Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University, Sofia, Bulgaria.
98. Department of Oncology, Cross Cancer Institute, University of Alberta, Edmonton, Alberta, Canada.
99. Division of Radiation Oncology, Cross Cancer Institute, Edmonton, Alberta, Canada.
100. Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium.
101. Institute of Cancer Sciences, Manchester Cancer Research Centre, University of Manchester, Manchester Academic Health Science Centre, St Mary's Hospital, Manchester, UK.
102. Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigacion Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde, SERGAS, Santiago De Compostela, Spain.
103. University of California San Diego, Moores Cancer Center, La Jolla, CA, USA.
104. Department of Urology, Erasmus University Medical Center, Rotterdam, the Netherlands.
105. Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France.
106. Clinical Gerontology Unit, University of Cambridge, Cambridge, UK.
107. Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA.
108. George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT, USA.
109. The University of Surrey, Guildford, Surrey, UK.

						OncoArray		Previous data		Replication		Combined meta		
SNP	Locus	Chr.	Pos. (b37)	Risk Allele	RAF	OR	*P*	OR	P_{meta}	OR	P_{meta}	OR	P_{meta}	I^2
rs7577599	2p23.3	2	25613146	T	0.81	1.22	2.63×10^{-3}	1.24	1.24×10^{-16}	-	-	1.23	1.29×10^{-18}	0
rs4325816	2q31.1	**2**	**174808899**	**T**	**0.77**	**1.16**	**1.23×10^{-2}**	**1.11**	**1.30×10^{-5}**	**1.16**	**3.00×10^{-3}**	**1.12**	**7.37×10^{-9}**	**9**
rs6599192	3p22.1	3	41992408	G	0.16	1.24	1.35×10^{-3}	1.26	8.75×10^{-18}	-	-	1.26	4.96×10^{-20}	0
rs10936600	3q26.2	3	169514585	A	0.75	1.18	5.12×10^{-3}	1.20	5.94×10^{-15}	-	-	1.20	1.20×10^{-16}	0
rs1423269	5q15	5	95255724	A	0.75	1.09	0.125	1.17	1.57×10^{-11}	-	-	1.16	8.30×10^{-12}	23
rs6595443	5q23.2	**5**	**122743325**	**T**	**0.43**	**1.14**	**9.87×10^{-3}**	**1.10**	**4.69×10^{-6}**	**1.10**	**0.022**	**1.11**	**1.20×10^{-8}**	**0**
rs34229995	6p22.3	6	15244018	G	0.02	1.05	0.781	1.40	1.76×10^{-8}	-	-	1.36	5.60×10^{-8}	0
rs3132535	6p21.3	6	31116526	A	0.29	1.26	2.67×10^{-5}	1.20	2.97×10^{-17}	-	-	1.21	6.00×10^{-21}	0
rs9372120	6q21	6	106667535	G	0.21	1.18	7.74×10^{-3}	1.20	8.72×10^{-14}	-	-	1.19	2.40×10^{-15}	0
rs4487645	7p15.3	7	21938240	C	0.65	1.23	1.06×10^{-4}	1.24	5.30×10^{-25}	-	-	1.24	2.80×10^{-28}	0
rs17507636	7q22.3	**7**	**106291118**	**C**	**0.74**	**1.12**	**5.71×10^{-2}**	**1.12**	**5.54×10^{-7}**	**1.10**	**0.036**	**1.12**	**9.20×10^{-9}**	**50**
rs58618031	7q31.33	**7**	**124583896**	**T**	**0.72**	**1.17**	**7.61×10^{-3}**	**1.11**	**4.70×10^{-6}**	**1.10**	**0.061**	**1.12**	**2.73×10^{-8}**	**0**

Table 1 (continued on following page)

						OncoArray		Previous data		Replication		Combined meta		
SNP	**Locus**	**Chr.**	**Pos. (b37)**	**Risk Allele**	**RAF**	**OR**	***P***	**OR**	P_{meta}	**OR**	P_{meta}	**OR**	P_{meta}	I^2
rs7781265	7q36.1	7	150950940	A	0.12	1.33	3.23×10^{-4}	1.20	1.82×10^{-7}	-	-	1.22	4.82×10^{-10}	49
rs1948915	8q24.21	8	128222421	C	0.32	1.19	1.68×10^{-3}	1.14	3.14×10^{-10}	-	-	1.15	2.53×10^{-12}	26
rs2811710	9p21.3	9	21991923	C	0.63	1.13	1.76×10^{-2}	1.14	6.50×10^{-10}	-	-	1.14	3.64×10^{-11}	0
rs2790457	10p12.1	10	28856819	G	0.73	1.09	0.124	1.12	8.44×10^{-7}	-	-	1.11	2.66×10^{-6}	0
rs13338946	16p11.2	**16**	**30700858**	**C**	**0.26**	**1.17**	$\mathbf{7.90\times10^{-3}}$	**1.12**	$\mathbf{2.22\times10^{-7}}$	**1.26**	$\mathbf{2.5\times10^{-7}}$	**1.15**	$\mathbf{1.02\times10^{-13}}$	**26**
rs7193541	16q23.1	16	74664743	T	0.58	1.14	9.01×10^{-3}	1.12	1.14×10^{-8}	-	-	1.12	3.68×10^{-10}	34
rs34562254	17p11.2	17	16842991	A	0.10	1.32	7.63×10^{-4}	1.30	3.63×10^{-17}	-	-	1.30	1.18×10^{-19}	29
rs11086029	19p13.11	**19**	**16438661**	**T**	**0.24**	**1.26**	$\mathbf{1.02\times10^{-4}}$	**1.12**	$\mathbf{1.69\times10^{-6}}$	**1.15**	$\mathbf{5.00\times10^{-3}}$	**1.14**	$\mathbf{6.79\times10^{-11}}$	**42**
rs6066835	20q13.13	20	47355009	C	0.08	1.13	0.162	1.24	1.16×10^{-9}	-	-	1.23	6.58×10^{-10}	38
rs138747	22q13.1	22	35700488	A	0.66	-	-	1.21	2.58×10^{-8}	-	-	1.21	2.58×10^{-8}	0
rs139402	22q13.1	22	39546145	C	0.44	1.11	4.146×10^{-2}	1.23	4.98×10^{-26}	-	-	1.22	3.84×10^{-26}	56

Table 1

							Functional evidence					
SNP	**Locus**	**bp(b37)**	**Genes in LD block**	**Coding variant**	**Promoter variant**	**Promoter/ enhancer chromatin states**	**TF binding[1]**	**Hi-C contact(s) in KMS11 cells**	**Hi-C contact(s) in naïve B-cells**	**eQTL**	**Functional study**	**Candidate causal gene(s)**	**Candidate disease mechanism**
rs7577599	2p23.3	25613146	*DTNB*										
rs4325816	**2q31.1**	**174808899**	***SP3***		***SP3***	**active promoter, transcribed enhancer weakly acetylated, intermediate enhancer**	**BATF, CTCF, MAZ, NFIC, RAD21, YY1**					***SP3***	**Chromatin remodelling**
rs6599192	3p22.1	41992408	*ULK4*										
rs10936600	3q26.2	169514585	*ACTRT3 MYNN LRRC34*	*LRRC34*		active promoter, distal promoter	ATF2, EBF1, MAZ, MXI1, POL2RA, SIN3A, STAT5A, TAF1, TBLR1XR1 +21	*GPR160, SEC62-AS1*	*GPR160, LRRC31, MYNN, PDCD10, SERPINI1, SEC62, SAMD7, SEC62-AS1, SKIL, PHC3, PDCD10*			*LRRC34*	
rs1423269	5q15	95255724	*ELL2*	*ELL2*		intermediate enhancer, active enhancer, distal promoter	ATF2, BCLAF1, EBF1, IKZF1, MAZ, MEF2C, MXI1, SPI1, STAT5A, TBLR1XR1 +23	*VPS13C*				*ELL2*	B-cell development
rs6595443	**5q23.2**	**122743325**	***CEP120***	***CEP120***		**transcribed enhancer weakly acetylated**	**SPI1**	***SNX2, SNX24***	***SNX2, SNX24***	**CEP120**		***CEP120***	**cell cycle/ genomic stability**

Table 2 (continued on following page)

							Functional evidence					
SNP	**Locus**	**bp(b37)**	**Genes in LD block**	**Coding variant**	**Promoter variant**	**Promoter/ enhancer chromatin states**	**TF binding[1]**	**Hi-C contact(s) in KMS11 cells**	**Hi-C contact(s) in naïve B-cells**	**eQTL**	**Functional study**	**Candidate causal gene(s)**	**Candidate disease mechanism**
rs34229995	6p22.3	15244018	*JARID2*			intermediate enhancer, active promoter, distal promoter	FOXM1, IKZF1, MEF2A, NFIC, RELA, RUNX3, SPI1, YY1, ZNF143						
rs3132535	6p21.3	31116526	*PSORS1C1 CCHCR1*										
rs9372120	6q21	106667535	*ATG5*			intermediate enhancer, active enhancer		*PREP*	*PRDM1, PREP*			*PRDM1*	B-cell development
rs4487645	7p15.3	21938240	*DNAH11 CDCA7L*				IRF4, MYC, POLR2A, POU2F2, RUNX3, SPI1, TAF1, WRNIP1				CDCA7L	*CDCA7L*	apoptosis/autophagy
rs17507636	**7q22.3**	**106291118**	***CCDC71L***										
rs58618031	**7q31.33**	**124583896**	***POT1***			**distal promoter, active enhancer, intermediate enhancer**	**NFIC**		***ASB15, IQUB, WASL***				**cell cycle/ genomic stability**
rs7781265	7q36.1	150950940	*ABCF2 CHPF2 SMARCD3*		*ABCF2, CHPF2*	active promoter, poised promoter	EBF1, EZH2, POLR2A, SIN3A, TAF1, YY1	*ASIC3, ABCF2, ATG9B*				*ABCF2*	chromatin remodelling

Table 2 (continued on following page)

							Functional evidence						
SNP	**Locus**	**bp(b37)**	**Genes in LD block**	**Coding variant**	**Promoter variant**	**Promoter/ enhancer chromatin states**	**TF binding[1]**	**Hi-C contact(s) in KMS11 cells**	**Hi-C contact(s) in naïve B-cells**	**eQTL**	**Functional study**	**Candidate causal gene(s)**	**Candidate disease mechanism**
rs1948915	8q24.21	128222421					ATF2, BCLAF1, EBF1, MAZ, MXI1, POL2RA, SIN3A, SPI1, STAT5A +18		*CASC11, MYC*			*MYC*	apoptosis/ autophagy
rs2811710	9p21.3	21991923	*CDKN2A, MTAP, CDKN2B-AS1*	*CDKN2A*	*CDKN2A, CDKN2B-AS1*	active promoter		*MTAP*	*MTAP*			*CDKN2A, MTAP*	cell cycle/ genomic stability
rs2790457	10p12.1	28856819	*WAC*			intermediate enhancer	CTCF	*LYZL1*	*MASTL, YME1L1*	WAC		*WAC*	apoptosis/ autophagy
rs13338946	**16p11.2**	**30700858**	***PRR14 FBRS SRCAP***	***PRR14***	***FBRS***	**active promoter, distal promoter**	**EBF1, MAZ, MXI1, POL2RA, SIN3A, SPI1, TAF1 +11**	***DCTPP1, DOC2A, FBXL19, GDPD3, ITGAL, MYLPF, PPP4C, SEPHS2, SEPT1, TBC1D10B, ZNF48, ZNF771***	***FBRS, PRR14, DCTPP1, MYLPF, TBC1D10B, SEPHS2***			***PRR14***	**apoptosis/ autophagy**
rs7193541	16q23.1	74664743	*RFWD3 GLG1*	*RFWD3*	*RFWD3*	active promoter	PML, TBP	*GLG1, NPIPL2*	*GLG1, HSPE1P, CFDP1, PSMD7, RFWD3, GABARAPL2*			*RFWD3*	

Table 2 (continued on following page)

								Functional evidence					
SNP	**Locus**	**bp(b37)**	**Genes in LD block**	**Coding variant**	**Promoter variant**	**Promoter/ enhancer chromatin states**	**TF binding[1]**	**Hi-C contact(s) in KMS11 cells**	**Hi-C contact(s) in naïve B-cells**	**eQTL**	**Functional study**	**Candidate causal gene(s)**	**Candidate disease mechanism**
rs34562254	17p11.2	16842991	*TNFRSF13B*	*TNFRSF13B*		intermediate enhancer, distal promoter, active enhancer	CTCF, POL2RA, STAT5A					*TNFRSF13B*	B-cell development
rs11086029	**19p13.11**	**16438661**	***KLF2***	***KLF2***	***KLF2***	**poised promoter**	**CTCF, EGR1, IKZF1, NFYB, POLR2A, RFX5, SIN3A, SPI1**					***KLF2***	**apoptosis/ autophagy**
rs6066835	20q13.13	47355009	*PREX1*			poised promoter	ATF2, EBF1, IKZF1, MEF2C, POL2RA, SPI1, TBLR1XR1 +9		*ARFGEF2*				
rs138747	22q13.1	35700488	*HMGXB4 TOM1*	*HMGXB4*	*TOM1, HMGXB4*	active promoter, transcribed enhancer weakly acetylated, intermediate enhancer, distal promoter, active enhancer, transcribed weak enhancer weakly acetylated	BCLAF1, EBF1, MAZ, POL2RA, STAT5A +46	*CRYBB1, HMOX1, APOL3, TOM1, LARGE, HMGXB4*	*FBXO7, HMGXB4, RASD2, MB*				
rs139402	22q13.1	39546145	*CBX7*		*CBX7*	distal promoter, intermediate enhancer, active promoter, poised promoter	BCLAF1, CHD2, CTCF, EBF1, MAZ, NFYB, POLR2A, RELA, RFX5, TBP, TAF1, ZNF143		*APOBEC3B-AS1, RPL3*			*CBX7*	chromatin remodelling

Table 2

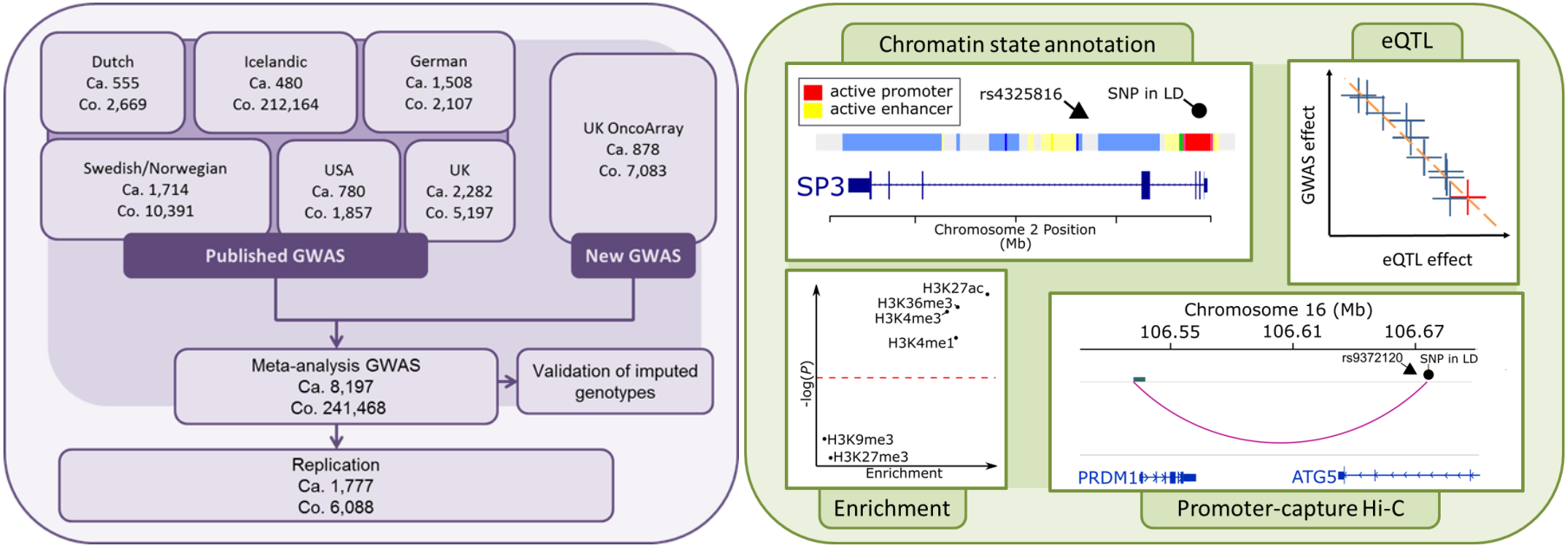

Dutch
Ca. 555
Co. 2,669
Icelandic
Ca. 480
Co. 212,164
German
Ca. 1,508
Co. 2,107
Swedish/Norwegian
Ca. 1,714
Co. 10,391
USA
Ca. 780
Co. 1,857
UK
Ca. 2,282
Co. 5,197
UK OncoArray
Ca. 878
Co. 7,083
Published GWAS
New GWAS
Meta-analysis GWAS
Ca. 8,197
Co. 241,468
Validation of imputed genotypes
Replication
Ca. 1,777
Co. 6,088
Chromatin state annotation
active promoter
active enhancer
rs4325816
SNP in LD
SP3
Chromosome 2 Position (Mb)
eQTL
GWAS effect
eQTL effect
H3K27ac
H3K36me3
H3K4me3
H3K4me1
-log(P)
•H3K9me3
•H3K27me3
Enrichment
Chromosome 16 (Mb)
106.55
106.61
106.67
rs9372120
SNP in LD
PRDM1
ATG5
Enrichment
Promoter-capture Hi-C

Figure 1

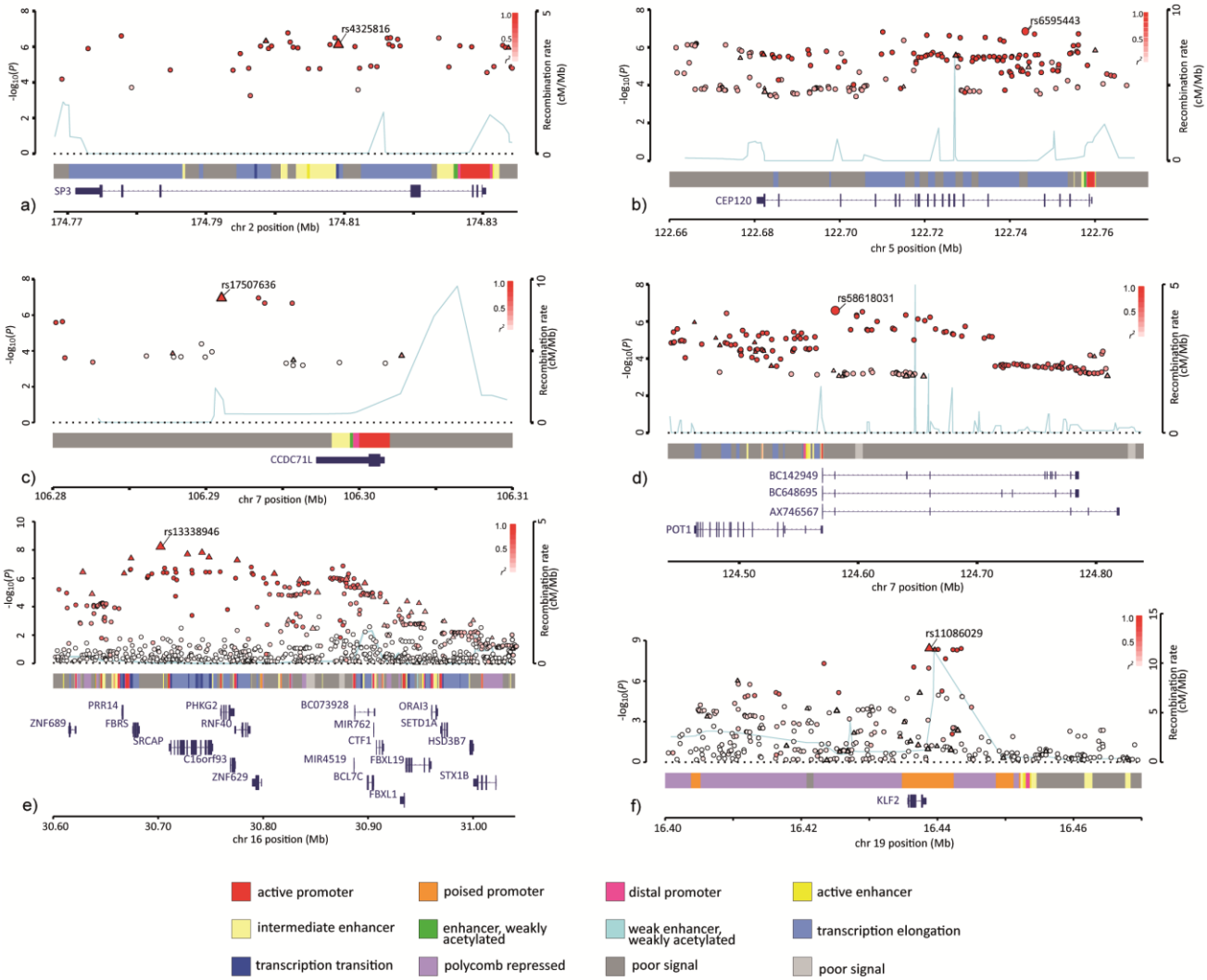

a)
rs4325816
SP3
chr 2 position (Mb)
b)
rs6595443
CEP120
chr 5 position (Mb)
c)
rs17507636
CCDC71L
chr 7 position (Mb)
d)
rs58618031
BC142949
BC648695
AX746567
POT1
chr 7 position (Mb)
e)
rs13338946
PRR14
PHKG2
BC073928
ORAI3
ZNF689
FBRS
RNF40
MIR762
SETD1A
SRCAP
CTF1
HSD3B7
C16orf93
MIR4519
FBXL19
ZNF629
BCL7C
STX1B
FBXL1
chr 16 position (Mb)
f)
rs11086029
KLF2
chr 19 position (Mb)
-log10(P)
Recombination rate (cM/Mb)
active promoter
poised promoter
distal promoter
active enhancer
intermediate enhancer
enhancer, weakly acetylated
weak enhancer, weakly acetylated
transcription elongation
transcription transition
polycomb repressed
poor signal
poor signal

Figure 2.

FIGURE AND TABLE LEGENDS

Figure 1: GWAS study design

Details of the new and existing GWAS samples, including recruitment centres or trials and quality control, are provided in **Supplementary Tables 1** and **2**. Trials or centres from which replication samples were recruited are detailed in **Supplementary Table 3**. Ca., cases; Co., controls; eQTL, expression quantitative trait loci; SNP, single nucleotide polymorphism; LD, linkage disequilibrium.

Figure 2: Regional plots of the six new risk loci

Regional plots of loci (a) 2q31.1, (b) 5q23.2, (c) 7q22.3, (d) 7q31.33, (e) 16p11.2 and (f) 19p13.11. Plots show results of the meta-analysis for both genotyped (triangles) and imputed (circles) single nucleotide polymorphisms (SNPs), and recombination rates. $-\log_{10}(P)$ (*y* axes) of the SNPs are shown according to their chromosomal positions (*x* axes). The sentinel SNP in each combined analysis is shown as a large circle or triangle and is labelled by its rsID. The colour intensity of each symbol reflects the extent of LD with the top SNP, white (r^2 = 0) through to dark red (r^2 = 1.0). Genetic recombination rates, estimated using 1000 Genomes Project samples, are shown with a light blue line. Physical positions are based on NCBI build 37 of the human genome. Also shown are the relative positions of genes and transcripts mapping to the region of association. Genes have been redrawn to show their relative positions; therefore maps are not to physical scale. The middle track represents the chromatin-state segmentation track (ChromHMM) for KMS11.

Table 1: Summary of genotyping results for all 23 risk SNPs

New loci discovered through this study are emboldened.

RAF, risk allele frequency; P_{trend}, *P*-value for trend, via logistic regression; P_{meta}, *P*-value for fixed effects meta-analysis; I^2, heterogeneity index (0–100). RAF are based on the UK cohort control series, with the exception of rs138747, which is sourced from the 1000 Genomes Project[61].

Table 2: Summary of functional annotation of the 23 risk loci

Newly identified risk loci are emboldened. [1] Where > 10 TF were implicated at a locus, only those that overlap with TF which demonstrated enrichment in GM12878 are shown here. A full list of TFs localising to loci are detailed in **Supplementary Table 17**.

SUPPLEMENTARY INFORMATION

Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma

Went *et al.*

		Trial/Recruitment Centre	Pre-QC	Sex discrepancy	Call rate fail	Heterozygosity rate	Related Individuals	Non-European Ancestary	Post-QC
UK	**Cases**	UK MRC MyIX, UK MRC MyXI	2,329	10	1	NA	2	34	2,282
	Controls	1958 Birth Cohort, National Blood Service	5,199	0	0	NA	2	0	5,197
Sweden/Norway	**Cases**	Swedish National Myeloma Biobank Norwegian Biobank for Myeloma							1,714
	Controls	TWINGENE							10,391
Germany	**Cases**	GMMG-HD3, GMMG-HD4, GMMG-HD5	1,512	1	0	NA	0	3	1,508
	Controls	Heinz Nixdorf Recall	2,107	0	0	NA	0	0	2,107
Netherlands	**Cases**	HOVON65/GMMG-HD4, HOVON95/EMN02, HOVON87/NMSG18	608	0	2	7	0	44	555
	Controls	B-PROOF	2,669	0	0	0	0	0	2,669
USA	**Cases**	Total Therapy II, Total Therapy III, Total Therapy 3B, Total Therapy 4	1,076	0	0	9	1	286	780
	Controls	Cancer Genetic Markers of Susceptibility	2,234	0	4	2	0	369	1,857
Iceland	**Cases**	Icelandic Cancer Registry							480
	Controls	deCODE							212,164
OncoArray	**Cases**	UK MRC MyXI	931	6	1	5	3	44	878
	Controls	PRACTICAL, BCAC	7,519	8	1	7	68	364	7,083

Supplementary Table 1: Details of the quality control filters applied to each GWAS. Samples were excluded due to call rate (<95% or failed genotyping), ancestry (principle components analysis or other samples reported to be not of white, European descent), relatedness (any individuals found to be duplicated or related within or between data sets through IBS) or sex discrepancy. Dutch, German, UK, USA, Sweden/Norway and Iceland: These studies have been previously reported in their entirety with comprehensive details on QC[1-4]. MyIX, Myeloma IX; MyXI, Myeloma XI; B-PROOF, B-vitamins for the prevention of osteoporotic fractures; UKGPCS, UK Genetic Prostate Cancer Study; BCAC, Breast Cancer Association Consortium.

	Pre-QC	Call rate fail	HWE fail	MAF < 0.01	Post-QC	Imputed (filtered)
UK	409,429	997	7	3	408,422	8,517,071
Sweden/Norway						7,182,761
Germany	401,405	113	0	1	401,291	8,282,831
Netherlands	646,124	6,523	18,104	0	621,497	8,628,799
USA	296,998	4	171	9,151	287,672	8,085,846
Iceland						10,291,845
OncoArray	459,068	6,851	12	73,239	378,966	8,309,850

Supplementary Table 2: Details of the quality control filters applied to each GWAS. For the OncoArray genotyped SNPs with a call rate <95% were excluded as were those with a MAF <0.01 or showing significant deviation from Hardy-Weinberg equilibrium (*i.e*. $P < 10^{-5}$). Imputed SNPs with information score <0.8 and MAF <0.01 were excluded. Dutch, German, UK, USA, Sweden/Norway and Iceland: These studies have been previously reported in their entirety with comprehensive details on QC[1-4].

	Samples	Trial/Recruitment Centre
Cases	911	German Myeloma Study Group
Controls	1,477	Institute of Transfusion Medicine and Immunology, University of Mannheim, Germany
Cases	534	Swedish National Myeloma Biobank
Controls	2,382	Swedish blood donors
Cases	332	University Hospital of Copenhagen
Controls	2,229	Individuals from Denmark and Skåne County

Supplementary Table 3: Details of the replication sample recruitment.

				OncoArray				Netherlands				German			
SNP	**Chr.**	**Pos. (b37)**	**Risk Allele**	**Cases RAF**	**Controls RAF**	**OR**	***P*- value**	**Cases RAF**	**Controls RAF**	**OR**	***P*- value**	**Cases RAF**	**Controls RAF**	**OR**	***P*- value**
rs4325816	2	174808899	T	0.79	0.76	1.16	0.012	0.78	0.76	1.10	0.24	0.79	0.75	1.21	0.002
rs6595443	5	122743325	T	0.47	0.44	1.14	0.010	0.44	0.43	1.04	0.52	0.47	0.45	1.10	0.087
rs17507636	7	106291118	C	0.77	0.75	1.12	0.057	0.76	0.73	1.16	0.04	0.75	0.71	1.14	0.028
rs58618031	7	124583896	T	0.76	0.73	1.17	0.008	0.73	0.72	1.08	0.30	0.75	0.72	1.16	0.012
rs13338946	16	30700858	C	0.28	0.25	1.17	0.008	0.26	0.25	1.09	0.24	0.29	0.27	1.11	0.078
rs11086029	19	16438661	T	0.29	0.25	1.26	1.01×10^{-4}	0.25	0.24	1.04	0.65	0.23	0.22	1.05	0.466
rs1050976	6	408079	T	0.53	0.51	1.09	0.080	0.50	0.45	1.21	3.7×10^{-3}	0.47	0.46	1.03	0.525
rs11629542	15	90098754	G	0.58	0.55	1.16	0.004	0.58	0.55	1.12	0.10	0.56	0.56	1.03	0.635
rs17501560	7	81415783	A	0.85	0.82	1.24	0.001	0.82	0.81	1.03	0.72	0.82	0.80	1.12	0.098

(Table continued on following page)

				Sweden				UK				USA			
SNP	**Chr.**	**Pos. (b37)**	**Risk Allele**	**Cases RAF**	**Controls RAF**	**OR**	***P*- value**	**Cases RAF**	**Controls RAF**	**OR**	***P*- value**	**Cases RAF**	**Controls RAF**	**OR**	***P*- value**
rs4325816	2	174808899	T	0.78	0.77	1.07	0.141	0.79	0.77	1.12	0.006	0.80	0.78	1.18	0.026
rs6595443	5	122743325	T	0.17	0.16	1.15	0.010	0.45	0.43	1.06	0.080	0.47	0.41	1.27	1.13×10^{-4}
rs17507636	7	106291118	C	0.78	0.76	1.09	0.044	0.76	0.74	1.08	0.048	0.79	0.74	1.30	2.03×10^{-4}
rs58618031	7	124583896	T	0.48	0.50	1.09	0.039	0.74	0.72	1.10	0.019	0.76	0.72	1.20	0.008
rs13338946	16	30700858	C	0.29	0.27	1.12	0.008	0.28	0.26	1.14	0.001	0.27	0.26	1.08	0.232
rs11086029	19	16438661	T	0.24	0.23	1.08	0.067	0.27	0.24	1.13	0.002	0.28	0.22	1.38	6.4×10^{-6}
rs1050976	6	408079	T	0.47	0.45	1.06	0.119	0.56	0.53	1.12	0.001	0.53	0.51	1.10	0.109
rs11629542	15	90098754	G	0.56	0.56	1.05	0.171	0.57	0.54	1.13	0.001	0.58	0.54	1.18	0.006
rs17501560	7	81415783	A	0.83	0.81	1.15	0.004	0.83	0.81	1.11	0.025	0.81	0.80	1.04	0.624

(Table continued on following page)

			Iceland					Meta $_{GWAS}$			Meta $_{GWAS + REP}$		
SNP	**Chr.**	**Pos. (b37)**	**Risk Allele**	**Cases RAF**	**Controls RAF**	**OR**	***P*- value**	**OR**	***P*- value**	I^2	**OR**	***P*- value**	I^2
rs4325816	2	174808899	T	-	-	0.94	0.50	1.12	6.44×10^{-7}	21	1.12	7.37×10^{-9}	9
rs6595443	5	122743325	T	-	-	1.07	0.39	1.11	1.76×10^{-7}	25	1.11	1.20×10^{-8}	0
rs17507636	7	106291118	C	-	-	1.06	0.49	1.12	8.52×10^{-8}	1	1.12	9.20×10^{-9}	50
rs58618031	7	124583896	T	-	-	1.01	0.87	1.11	1.66×10^{-7}	0	1.12	2.73×10^{-8}	0
rs13338946	16	30700858	C	-	-	1.19	0.03	1.13	7.14×10^{-9}	0	1.15	1.02×10^{-13}	26
rs11086029	19	16438661	T	-	-	1.08	0.38	1.14	4.17×10^{-9}	60	1.14	6.79×10^{-11}	42
rs1050976	6	408079	T	-	-	1.15	0.07	1.10	3.74×10^{-7}	0	1.10	5.93×10^{-8}	0
rs11629542	15	90098754	G	-	-	1.09	0.23	1.10	2.37×10^{-7}	0	1.08	2.75×10^{-6}	40
rs17501560	7	81415783	A	-	-	1.19	0.05	1.13	3.85×10^{-7}	0	1.12	9.90×10^{-8}	0

Supplementary Table 4: Summary statistics for novel variants showing an association with multiple myeloma risk in the GWAS meta-analysis at $P < 1.0 \times 10^{-6}$. These were taken forward for replication and those which showed association at $P < 5\times 10^{-8}$ were considered genome-wide significant. Odds ratios derived with respect to the risk allele. Cases RAF, risk allele frequency in discovery cases. Controls RAF, risk allele frequency in discovery controls. Shown are discovery association *P* values for individual studies. $Meta_{GWAS}$ shows a meta-analysis of previously published GWAS including new discovery OncoArray dataset. $Meta_{GWAS+REP}$ represents a meta-analysis of previously published GWAS, new discovery OncoArray dataset and replication series. Heterogeneity index, I^2 (0-100), quantifies the proportion of the total variation due to heterogeneity.

				German Replication				Swedish Replication				Danish Replication			
			Risk	Cases	Controls			Cases	Controls			Cases	Controls		
SNP	Chr.	Pos. (b37)	Allele	RAF	RAF	OR	*P*- value	RAF	RAF	OR	*P*- value	RAF	RAF	OR	*P*- value
rs4325816	2	174808899	T	0.79	0.76	1.19	0.014	0.78	0.77	1.06	0.461	0.83	0.79	1.25	0.037
rs6595443	5	122743325	T	0.47	0.45	1.08	0.202	0.45	0.42	1.15	0.038	0.44	0.43	1.04	0.639
rs17507636	7	106291118	C	-	-	-	-	0.79	0.76	1.19	0.036	0.74	0.75	0.93	0.485
rs58618031	7	124583896	T	-	-	-	-	0.75	0.72	1.18	0.032	0.72	0.71	1.03	0.761
rs13338946	16	30700858	C	0.32	0.28	1.24	0.001	0.29	0.27	1.13	0.112	0.37	0.28	1.51	9.0×10^{-6}
rs11086029	19	16438661	T	0.23	0.21	1.18	0.022	0.24	0.22	1.12	0.149	0.26	0.24	1.11	0.293
rs1050976	6	408079	T	0.49	0.47	1.07	0.268	0.47	0.45	1.10	0.188	0.48	0.46	1.08	0.371
rs11629542	15	90098754	G	0.44	0.46	0.92	0.205	0.54	0.54	1.01	0.910	0.57	0.54	1.16	0.091
rs17501560	7	81415783	A	0.81	0.81	1.00	0.954	0.83	0.80	1.22	0.016	0.82	0.81	1.08	0.461

Supplementary Table 5: Replication of top association signals. Showing SNPs which were taken forward for replication genotyping. Cases RAF, risk allele frequency of replication cases; Control RAF, risk allele frequency of replication controls. *P* values are shown for each replication series (logistic regression). rs17507636 had been previously replicated in the German cohort, with association values[4]; cases RAF: 0.760, controls RAF: 0.735, OR: 1.15, *P* value: 0.06. A meta-analysis of this with discovery cohorts and replication series was performed using R version 3.3.1 (R Development Core Team, Vienna, Austria).

	USA		Germany		OncoArray		UK		Meta				
RSID	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**	**OR**	**95% CIs**	***P*- value**	**I^2**	**P_{HET}**
rs34229995	-0.43	0.15	-0.08	0.69	-0.23	0.50	-0.05	0.79	0.87	0.69 - 1.10	0.23	0	0.73
rs9372120	0.16	0.20	-0.11	0.22	0.02	0.87	-0.01	0.85	0.99	0.90 - 1.08	0.84	6	0.36
rs7781265	-0.06	0.69	0.02	0.84	0.15	0.28	-0.11	0.22	0.98	0.87 - 1.10	0.72	0	0.44
rs1948915	0.02	0.86	0.05	0.53	0.07	0.49	0.04	0.58	1.04	0.96 - 1.13	0.30	0	0.99
rs2811710	0.13	0.24	0.03	0.74	-0.02	0.83	-0.03	0.67	1.01	0.93 - 1.10	0.80	0	0.65
rs2790457	0.04	0.75	0.06	0.45	-0.02	0.83	0.12	0.08	1.07	0.98 - 1.17	0.12	0	0.73
rs7193541	0.05	0.67	0.01	0.88	0.08	0.44	0.02	0.74	1.03	0.95 - 1.12	0.44	0	0.95
rs6066835	-0.60	0.00	-0.24	0.06	0.30	0.07	0.07	0.49	0.94	0.83 - 1.07	0.36	82	0.00
rs7577599	-0.04	0.79	-0.10	0.33	0.22	0.12	0.01	0.94	1.00	0.90 - 1.12	0.96	14	0.32
rs6599192	0.12	0.37	0.01	0.93	-0.13	0.28	-0.03	0.67	0.98	0.89 - 1.08	0.73	0	0.56
rs4487645	0.09	0.46	-0.03	0.79	0.04	0.73	0.07	0.29	1.05	0.96 - 1.16	0.27	0	0.89
rs34562254	0.19	0.24	-0.07	0.51	-0.01	0.93	0.05	0.60	1.02	0.91 - 1.14	0.72	0	0.58
rs3132535	0.03	0.82	0.11	0.17	-0.12	0.25	0.03	0.61	1.03	0.95 - 1.12	0.49	0	0.39
rs1423269	0.25	0.04	0.00	0.99	0.02	0.87	0.10	0.15	1.09	0.99 - 1.19	0.07	5	0.37
rs139402	-0.06	0.56	-0.09	0.24	-0.11	0.27	0.01	0.83	0.96	0.89 - 1.03	0.25	0	0.64
rs138747	-0.34	0.21	0.22	0.17	-0.46	0.04	-0.04	0.80	0.94	0.79 - 1.13	0.51	59	0.06
rs10936600	0.13	0.30	0.02	0.85	-0.10	0.40	-0.04	0.55	0.99	0.90 - 1.09	0.81	0	0.54
rs11086029	-0.11	0.40	0.15	0.21	-0.06	0.16	-0.10	0.41	0.96	0.88 - 1.05	0.37	27	0.25
rs13338946	0.08	0.51	0.07	0.31	0.07	0.49	-0.08	0.28	1.06	0.97 - 1.15	0.20	0	0.69
rs17507636	0.06	0.14	-0.04	0.50	0.09	0.75	-0.20	0.22	1.03	0.94 - 1.12	0.59	25	0.26
rs4325816	-0.02	0.91	0.01	0.80	-0.04	0.96	-0.01	0.56	0.97	0.89 - 1.07	0.60	0	0.99
rs58618031	0.05	0.19	0.01	0.56	-0.06	0.96	-0.16	0.48	0.97	0.89 - 1.07	0.56	0	0.53
rs6595443	0.13	0.36	0.07	0.09	0.09	0.47	-0.10	0.12	1.08	1.00 - 1.16	0.07	8	0.35

Supplementary Table 6: Relationship between SNP genotype and sex. Analysis based on beta values calculated from logistic regression on the discovery phase data sets from UK (2282 cases), Oncoarray (878 cases), German (1508 cases) and USA (780 cases) series. The meta-analysis was conducted using a fixed-effects model. This assumes that the underlying effect across all studies is the same. To test for potential heterogeneity, Cochran's Q-statistic was calculated such that P_{HET} >0.05 implied the presence of non-significant heterogeneity. The heterogeneity index, I^2 (0-100), was also measured; this quantifies the proportion of the total variation due to heterogeneity.

	USA		Germany		OncoArray		UK		Meta				
RSID	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**	**OR**	**95% CIs**	***P*- value**	**I^2**	**P_{HET}**
rs34229995	0.93	0.45	0.07	0.94	-1.34	0.43	0.44	0.67	1.23	0.40 - 3.79	0.71	0	0.74
rs9372120	-1.18	0.03	0.15	0.74	0.31	0.60	0.64	0.08	1.15	0.74 - 1.80	0.53	61	0.05
rs7781265	0.02	0.97	-0.41	0.43	0.45	0.53	-0.09	0.84	0.92	0.53 - 1.61	0.77	0	0.80
rs1948915	0.28	0.56	0.36	0.35	-0.22	0.68	-0.43	0.19	0.95	0.64 - 1.43	0.82	0.5	0.39
rs2811710	-0.02	0.97	-0.40	0.32	-0.01	0.98	-0.20	0.56	0.83	0.55 - 1.25	0.36	0	0.92
rs2790457	0.51	0.36	-0.24	0.57	-0.14	0.80	-0.36	0.31	0.86	0.55 - 1.33	0.50	0	0.62
rs7193541	-0.55	0.25	-0.22	0.56	-0.62	0.22	-0.02	0.94	0.77	0.53 - 1.14	0.19	0	0.69
rs6066835	-0.71	0.35	0.14	0.82	-1.45	0.08	-0.07	0.88	0.71	0.38 - 1.34	0.29	0	0.40
rs7577599	-0.50	0.46	0.22	0.63	0.87	0.21	0.13	0.75	1.19	0.71 - 1.97	0.51	0	0.56
rs6599192	-0.68	0.26	-0.20	0.63	-0.52	0.40	0.04	0.92	0.79	0.50 - 1.26	0.33	0	0.73
rs4487645	-0.68	0.19	0.33	0.56	-0.46	0.40	0.18	0.58	0.93	0.60 - 1.45	0.75	0	0.39
rs34562254	-1.00	0.16	0.35	0.45	0.35	0.62	-0.14	0.76	0.98	0.57 - 1.66	0.93	0	0.40
rs3132535	0.03	0.95	-0.28	0.40	-0.65	0.22	0.02	0.94	0.84	0.57 - 1.24	0.38	0	0.70
rs1423269	0.50	0.36	-0.54	0.16	-1.06	0.06	0.23	0.53	0.85	0.55 - 1.30	0.45	52	0.10
rs139402	-0.32	0.50	0.68	0.03	0.89	0.07	-0.30	0.32	1.22	0.85 - 1.75	0.29	63	0.04
rs138747	1.33	0.24	1.03	0.15	0.86	0.45	0.71	0.34	2.56	1.10 - 5.95	0.03	0	0.97
rs10936600	0.49	0.40	-0.48	0.22	-1.55	0.01	-0.53	0.17	0.61	1.00 - 2.47	0.03	51	0.11
rs11086029	-1.10	0.04	-0.27	0.48	-1.36	0.11	-0.17	0.61	0.64	0.41 - 0.99	0.04	15	0.32
rs13338946	-1.07	0.04	-0.30	0.41	0.25	0.65	0.50	0.14	0.95	0.64 - 1.43	0.81	59	0.06
rs17507636	0.60	0.30	0.59	0.17	0.54	0.36	0.12	0.73	1.48	0.95 - 2.31	0.09	0	0.80
rs4325816	0.08	0.90	-0.36	0.37	-0.01	0.99	-0.15	0.69	0.85	0.54 - 1.33	0.48	0	0.93
rs58618031	0.22	0.69	-0.76	0.07	-0.83	0.15	-0.16	0.67	0.68	0.44 - 1.07	0.10	0	0.39
rs6595443	0.65	0.16	-0.22	0.51	-0.43	0.38	-0.26	0.39	0.89	0.62 - 1.27	0.51	15	0.32

Supplementary Table 7: Relationship between SNP genotype and age at diagnosis. Analysis based on beta values calculated from linear regression on the discovery phase data sets from UK (2282 cases), Oncoarray (878 cases), German (1508 cases) and USA (780 cases) cohorts. The meta-analysis was conducted using a fixed-effects model. This assumes that the underlying effect across all studies is the same. To test for potential heterogeneity, Cochran's Q-statistic was calculated such that P_{HET} >0.05 implied the presence of non-significant heterogeneity. The heterogeneity index, I^2 (0-100), was also measured; this quantifies the proportion of the total variation due to heterogeneity.

	German		UK		OncoArray		Meta	
RSID	**Beta**	***P*-value**	**Beta**	***P*-value**	**Beta**	***P*-value**	**Beta**	***P*-value**
rs2790457	0.12	0.41	-0.18	0.18	0.44	0.14	0.002	0.99
rs13338946	-0.17	0.23	0.03	0.78	0.28	0.34	-0.03	0.76
rs7193541	0.13	0.29	-0.07	0.53	-0.17	0.52	0.002	0.98
rs34562254	-0.02	0.92	-0.26	0.12	-0.40	0.29	-0.17	0.14
rs11086029	-0.14	0.33	-0.11	0.39	0.11	0.70	-0.10	0.28
rs6066835	-0.13	0.54	-0.35	0.06	0.90	0.07	-0.16	0.22
rs138747	0.42	0.13	-0.18	0.52	1.00	0.10	0.20	0.28
rs139402	0.08	0.52	-0.07	0.56	-0.25	0.33	-0.02	0.78
rs4325816	0.02	0.90	-0.08	0.58	0.15	0.62	-0.01	0.88
rs6599192	0.26	0.11	-0.10	0.47	0.23	0.48	0.07	0.48
rs10936600	-0.03	0.83	0.11	0.46	0.03	0.91	0.04	0.69
rs1423269	0.12	0.44	0.05	0.70	-0.11	0.69	0.06	0.54
rs6595443	-0.30	0.02	-0.13	0.24	-0.07	0.78	-0.19	0.02
rs34229995	0.52	0.11	-0.39	0.29	-1.15	0.26	0.05	0.83
rs3132535	0.02	0.87	0.01	0.91	-0.11	0.69	0.005	0.95
rs9372120	-0.11	0.47	0.04	0.76	0.37	0.22	0.02	0.87
rs4487645	0.28	0.05	0.05	0.67	0.76	0.01	0.21	0.02
rs17507636	-0.003	0.98	0.02	0.86	-0.20	0.54	-0.01	0.95
rs58618031	-0.03	0.84	0.16	0.20	0.02	0.95	0.07	0.41
rs7781265	0.21	0.25	-0.02	0.90	0.01	0.99	0.08	0.49
rs1948915	0.03	0.83	0.15	0.20	0.02	0.94	0.09	0.30
rs2811710	0.17	0.22	-0.02	0.89	-0.16	0.56	0.04	0.63
rs7577599	0.29	0.10	-0.03	0.86	0.04	0.91	0.10	0.34

Supplementary Tables 8: Relationship between SNP genotype and t(4;14) subtype. German cases: 142, UK cases: 170, Oncoarray cases: 33, Meta: 345. Case-only analysis; Beta values obtained from logistic regression. FISH and ploidy classification of UK and German samples were determined as previously described[5,6].

	German		UK		OncoArray		Meta	
RSID	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**
rs2790457	-0.09	0.43	0.32	0.004	-0.06	0.82	0.10	0.19
rs13338946	-0.05	0.64	0.36	0.001	-0.24	0.34	0.11	0.12
rs7193541	0.02	0.82	-0.02	0.85	0.48	0.03	0.04	0.50
rs34562254	-0.05	0.68	-0.14	0.35	-0.46	0.15	-0.12	0.19
rs11086029	-0.04	0.74	0.12	0.27	-0.11	0.66	0.03	0.70
rs6066835	-0.13	0.41	0.17	0.30	-0.44	0.29	-0.02	0.88
rs138747	0.14	0.52	-0.42	0.06	-0.36	0.50	-0.15	0.33
rs139402	-0.17	0.07	-0.09	0.37	-0.27	0.23	-0.14	0.03
rs4325816	0.004	0.97	0.04	0.75	-0.13	0.62	0.01	0.93
rs6599192	0.13	0.27	-0.03	0.83	0.13	0.65	0.06	0.46
rs10936600	-0.14	0.22	0.01	0.94	-0.26	0.32	-0.09	0.27
rs1423269	-0.003	0.98	0.13	0.26	0.12	0.61	0.07	0.37
rs6595443	-0.04	0.71	-0.09	0.34	0.21	0.36	-0.04	0.53
rs34229995	-0.23	0.35	0.03	0.92	-0.44	0.62	-0.15	0.44
rs3132535	-0.05	0.60	0.11	0.31	0.27	0.24	0.04	0.53
rs9372120	0.09	0.43	0.19	0.10	0.10	0.68	0.14	0.08
rs4487645	0.13	0.23	-0.14	0.18	-0.06	0.80	-0.01	0.87
rs17507636	0.03	0.81	-0.13	0.25	-0.05	0.85	-0.05	0.52
rs58618031	0.02	0.83	-0.05	0.66	-0.19	0.46	-0.03	0.71
rs7781265	-0.08	0.54	-0.38	0.01	-0.30	0.34	-0.22	0.02
rs1948915	-0.20	0.05	0.08	0.42	0.60	0.01	-0.003	0.96
rs2811710	0.02	0.89	-0.12	0.26	-0.08	0.71	-0.05	0.44
rs7577599	-0.11	0.42	-0.09	0.49	0.04	0.89	-0.09	0.33

Supplementary Tables 9: Relationship between SNP genotype and t(11;14) subtype. German cases: 277, UK cases: 231, Oncoarray cases: 47, Meta: 555. Case-only analysis; Beta values obtained from logistic regression. FISH and ploidy classification of UK and German samples were determined as previously described[5,6].

	German		UK		OncoArray		Meta	
RSID	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**
rs2790457	0.28	0.37	0.34	0.31	1.06	0.07	0.41	0.05
rs13338946	0.20	0.49	-0.06	0.86	-0.24	0.68	0.05	0.82
rs7193541	-0.03	0.92	-0.08	0.80	-0.05	0.93	-0.05	0.80
rs34562254	-0.24	0.52	0.56	0.20	-0.07	0.92	0.08	0.76
rs11086029	-0.18	0.57	-0.24	0.45	0.43	0.44	-0.12	0.56
rs6066835	-0.49	0.28	-0.90	0.06	0.66	0.49	-0.54	0.09
rs138747	-0.06	0.93	-0.63	0.33	0.00	1.00	-0.29	0.49
rs139402	0.24	0.36	-0.62	0.04	-0.06	0.91	-0.13	0.48
rs4325816	0.45	0.17	0.01	0.98	0.10	0.87	0.23	0.31
rs6599192	-0.42	0.22	0.19	0.61	0.08	0.90	-0.11	0.64
rs10936600	-0.12	0.71	0.04	0.91	-0.54	0.36	-0.12	0.59
rs1423269	0.61	0.06	-0.12	0.72	1.12	0.04	0.41	0.06
rs6595443	-0.45	0.10	0.08	0.78	0.14	0.79	-0.16	0.40
rs34229995	0.38	0.59	1.67	0.08	-1.06	0.59	0.70	0.20
rs3132535	-0.09	0.75	0.15	0.63	0.13	0.81	0.03	0.88
rs9372120	-0.18	0.59	0.31	0.36	-0.95	0.10	-0.08	0.71
rs4487645	0.52	0.08	0.08	0.80	-0.24	0.67	0.24	0.24
rs17507636	0.29	0.35	0.05	0.87	-0.34	0.59	0.12	0.57
rs58618031	0.10	0.75	-0.04	0.90	-0.68	0.23	-0.06	0.77
rs7781265	0.20	0.59	-0.21	0.63	0.92	0.21	0.15	0.58
rs1948915	0.05	0.86	-0.14	0.66	0.07	0.89	-0.02	0.91
rs2811710	-0.24	0.42	0.19	0.54	0.43	0.40	0.03	0.87
rs7577599	0.34	0.36	-0.39	0.33	0.70	0.36	0.07	0.77

Supplementary Tables 10: Relationship between SNP genotype and t(14;16) subtype. German cases: 29, UK cases: 24, Oncoarray cases: 8, Meta: 61 Case-only analysis; Beta values obtained from logistic regression. FISH and ploidy classification of UK and German samples were determined as previously described[5,6].

	German		UK		OncoArray		Meta	
RSID	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**	**Beta**	***P*- value**
rs2790457	0.03	0.71	0.14	0.06	-	-	0.10	0.09
rs13338946	-0.04	0.59	-0.05	0.50	-0.02	0.91	-0.04	0.39
rs7193541	0.08	0.28	0.04	0.57	-	-	0.06	0.25
rs34562254	0.01	0.92	0.15	0.12	0.51	0.04	0.12	0.09
rs11086029	0.11	0.22	-0.01	0.92	-0.10	0.60	0.03	0.59
rs6066835	0.06	0.60	0.37	0.001	-0.40	0.21	0.20	0.01
rs138747	-0.32	0.04	-0.05	0.74	-0.28	0.49	-0.19	0.08
rs139402	0.07	0.33	0.18	0.01	0.22	0.19	0.14	0.003
rs4325816	-0.05	0.61	-0.06	0.43	-	-	-0.06	0.35
rs6599192	-0.15	0.13	-0.05	0.55	-	-	-0.09	0.15
rs10936600	0.04	0.63	0.01	0.86	-	-	0.03	0.64
rs1423269	-0.07	0.40	-0.13	0.08	-	-	-0.11	0.06
rs6595443	-0.05	0.50	-0.002	0.98	-0.15	0.39	-0.03	0.49
rs34229995	-0.15	0.43	0.03	0.89	-	-	-0.07	0.63
rs3132535	0.09	0.23	0.04	0.54	-	-	0.07	0.21
rs9372120	-0.04	0.68	-0.11	0.14	-0.14	0.48	-0.09	0.12
rs4487645	-0.08	0.33	-0.03	0.62	-	-	-0.05	0.32
rs17507636	-0.01	0.91	-0.001	0.99	0.08	0.69	0.001	0.98
rs58618031	-0.09	0.27	0.03	0.68	0.20	0.30	-0.01	0.92
rs7781265	0.01	0.94	-0.01	0.95	-	-	0.0001	1.00
rs1948915	0.08	0.30	-0.04	0.52	-0.34	0.06	-0.02	0.74
rs2811710	0.001	0.99	-0.02	0.78	-	-	-0.01	0.84
rs7577599	-0.061	0.56	0.08	0.35			0.02	0.74

Supplementary Tables 11: Relationship between SNP genotype and hyperdiploid subtype. German cases: 661, UK cases: 702, Oncoarray cases: 257, Meta: 1,620. Case-only analysis; Beta values obtained from logistic regression. FISH and ploidy classification of UK and German samples were determined as previously described[5,6].

		Germ-GMMG		UK-MyIX		UK-MyXI		USA-UAMS		Meta	
SNP	**Risk Allele**	**HR**	***P*- value**	**HR**	***P*- value**	**HR**	***P*- value**	**HR**	***P*- value**	**HR**	***P*- value**
rs4325816	T	1.06	0.69	0.88	0.12	0.86	0.27	0.85	0.11	0.89	0.03
rs6599192	G	1.27	0.14	1.01	0.95	1.08	0.54	0.98	0.86	1.04	0.50
rs10936600	A	1.06	0.69	1.00	0.95	1.10	0.45	0.98	0.83	1.02	0.70
rs1423269	A	1.07	0.61	1.05	0.49	0.97	0.80	0.97	0.78	1.02	0.69
rs6595443	T	1.14	0.27	0.98	0.78	0.92	0.39	1.03	0.66	1.00	0.91
rs34229995	G	1.12	0.71	0.68	0.03	1.24	0.56	0.99	0.96	0.88	0.30
rs3132535	A	0.87	0.27	0.91	0.18	0.89	0.27	1.07	0.40	0.95	0.21
rs9372120	G	0.99	0.92	1.22	0.01	0.84	0.12	1.05	0.63	1.06	0.21
rs4487645	C	1.04	0.77	0.94	0.39	1.10	0.38	0.99	0.90	0.99	0.89
rs17507636	C	1.06	0.70	1.03	0.73	1.10	0.42	1.04	0.67	1.05	0.36
rs58618031	T	1.00	0.99	1.01	0.89	1.14	0.23	1.17	0.08	1.07	0.14
rs7781265	A	0.82	0.20	1.20	0.07	0.90	0.46	-	-	1.02	0.77
rs1948915	C	0.96	0.73	1.02	0.83	0.96	0.68	1.04	0.59	1.01	0.89
rs2790457	G	0.88	0.40	0.91	0.21	0.92	0.48	0.94	0.55	0.92	0.08
rs13338946	C	0.76	0.03	1.02	0.75	1.07	0.54	0.91	0.25	0.96	0.32
rs7193541	T	1.01	0.94	0.94	0.34	1.05	0.61	1.06	0.43	1.00	0.96
rs34562254	A	0.91	0.51	1.07	0.46	1.40	0.05	1.13	0.33	1.09	0.16
rs11086029	T	0.98	0.89	0.82	0.01	1.06	0.57	1.05	0.62	0.94	0.17
rs6066835	C	0.92	0.63	1.03	0.77	1.13	0.48	0.90	0.41	0.99	0.87
rs139402	C	0.92	0.49	1.05	0.49	1.06	0.60	1.03	0.68	1.03	0.51

Supplementary Table 12: Relationship between genome-wide significant SNPs genotype and patient overall survival[7]. Data from: 1,165 cases from the UK MRC Myeloma-IX trial (UK-MyIX); 877 MM cases from the UK MRC Myeloma-XI trial (UK-MyXI); 511 of the patients recruited to the German-GWAS (GER-GMMG); 703 MM cases in the UAMS Myeloma Institute for Research and Therapy GWAS (US-UAMS). *P*-values calculated from Cox regression analysis. Data for SNPs rs2811710, rs7577599 and rs138747, or a correlated SNP (r^2 >0.6) to use as proxy, were not present in the survival analysis.

	UK Cases			
rsID	AA	Aa	aa	r^2
rs58618031	7/7	58/59	83/83	0.99
rs11629542	28/31	59/63	51/53	0.91
rs6595443	53/54	74/78	41/41	0.97

Supplementary Table 13: Concordance between directly sequenced and imputed genotype. Showing SNPs which were genome-wide significant after replication. These comprised 147 randomly selected samples from the Oncoarray case series. AA, major homozygote; Aa, heterozygote; aa, minor homozygote. r^2 indicates Pearson product-moment correlation coefficient between imputed and sequenced genotype.

rsID		KASP Primer Sequence	Conditions
rs4325816	KASP Primer A1	GAAGGTGACCAAGTTCATGCTAACCTAGGTTGCTGGGAGAATGAT	Std42
	KASP Primer A2	GAAGGTCGGAGTCAACGGATTCCTAGGTTGCTGGGAGAATGAC	
	KASP Common Primer	CATGTGACGTTGTTTTCATAAATCTCATAA	
rs6595443	KASP Primer A1	GAAGGTGACCAAGTTCATGCTCCATTTCTGATAGTGTGTGTTAAAGTCT	Std42plus5
	KASP Primer A2	GAAGGTCGGAGTCAACGGATTCCATTTCTGATAGTGTGTGTTAAAGTCA	
	KASP Common Primer	GTGAATGCACCTAACAGAGTATCAAAATA	
rs1050976	KASP Primer A1	GAAGGTGACCAAGTTCATGCTAAGTGATGTGTTTACATTTACTGAAATGC	Std42plus5
	KASP Primer A2	GAAGGTCGGAGTCAACGGATTCAAGTGATGTGTTTACATTTACTGAAATGT	
	KASP Common Primer	TTTTCTCTGTCTTCCAGCAAGACCTAAT	
rs17507636	KASP Primer A1	GAAGGTGACCAAGTTCATGCTTTCACTGTAGCCATCTGTATCCC	Std42plus5
	KASP Primer A2	GAAGGTCGGAGTCAACGGATTCTTTCACTGTAGCCATCTGTATCCT	
	KASP Common Primer	CCTGCTTCTTTAATTATGTATAGGGTAGAA	
rs17501560	KASP Primer A1	GAAGGTGACCAAGTTCATGCTCAAGATACAACAGGTGAGACCCAA	Std42plus5
	KASP Primer A2	GAAGGTCGGAGTCAACGGATTAAGATACAACAGGTGAGACCCAG	
	KASP Common Primer	TGTCCTTAATAGTTTAGTCTCCAAAATCAT	
rs58618031	KASP Primer A1	GAAGGTGACCAAGTTCATGCTAGGAGGCCTCAGGAAACTTACG	Std42plus15
	KASP Primer A2	GAAGGTCGGAGTCAACGGATTAAGGAGGCCTCAGGAAACTTACA	
	KASP Common Primer	CTGACATTTTCCCCACTGGCATTCAT	
rs11629542	KASP Primer A1	GAAGGTGACCAAGTTCATGCTAAGTACGTGCCTAAAAGATGGACAC	Std42plus10
	KASP Primer A2	GAAGGTCGGAGTCAACGGATTAAGTACGTGCCTAAAAGATGGACAG	
	KASP Common Primer	GCCATGTCTGGGGCACTATTTCTAA	
rs13338946	KASP Primer A1	GAAGGTGACCAAGTTCATGCTCGAGACTCTATCTCAATAAATGAATAAAATG	Std42
	KASP Primer A2	GAAGGTCGGAGTCAACGGATTGCGAGACTCTATCTCAATAAATGAATAAAATA	
	KASP Common Primer	CACCCCACTTCATTTTTTCATAACACGTA	
rs11086029	KASP Primer A1	GAAGGTGACCAAGTTCATGCTGTGGCCTCCTCTACGTTGAAAAAAAA	Std42plus5
	KASP Primer A2	GAAGGTCGGAGTCAACGGATTGTGGCCTCCTCTACGTTGAAAAAAAT	
	KASP Common Primer	GGCTTCCAGGAAGAGGTAAGTAGTT	

Supplementary Table 14: Details of genotyping primers and reaction conditions.

rsID		Sequencing Primer Sequence	Sequencing Direction	Sequencing Additives
rs6595443	Forward	AAGGAGTCAATTCTGCAAAAAG	Reverse	1M Betaine
	Reverse	TGCTGTTGTTGTTTGAAGTGG		
rs58618031	Forward	TGATAGTCATTTCTCACAAGAGCTG	Forward	1M Betaine
	Reverse	TCTCTGTCAAAATGAAACTTACCTTC		
rs11629542	Forward	CCAACCTCCTCATTGTAGGG	Forward	1M Betaine
	Reverse	AGCAAGAAACAAAGCACAGG		

Supplementary Table 15: Details of sequencing primers and reaction conditions.

KASPAR conditions

Std42

- Hot Start: 94ºC for 15 minutes
- Stage 1: 20 cycles
    - 94ºC for 10 seconds
    - 57ºC for 5 seconds
    - 72ºC for 10 seconds
- Stage 2: 22 cycles
    - 94ºC for 10 seconds
    - 57ºC for 20 seconds
    - 72ºC for 40 seconds

Std42plus5

- Hot Start: 94ºC for 15 minutes
- Stage 1: 20 cycles
    - 94ºC for 10 seconds
    - 57ºC for 5 seconds
    - 72ºC for 10 seconds
- Stage 2: 22 cycles
    - 94ºC for 10 seconds
    - 57ºC for 20 seconds
    - 72ºC for 40 seconds
- Stage 3: 5 cycles
    - 94ºC for 10 seconds
    - 57ºC for 1 minute

Std42plus10

- Hot Start: 94ºC for 15 minutes
- Stage 1: 20 cycles
    - 94ºC for 10 seconds
    - 57ºC for 5 seconds
    - 72ºC for 10 seconds
- Stage 2: 22 cycles
    - 94ºC for 10 seconds
    - 57ºC for 20 seconds
    - 72ºC for 40 seconds
- Stage 3: 10 cycles
    - 94ºC for 10 seconds
    - 57ºC for 1 minute

Std42plus15

- Hot Start: 94ºC for 15 minutes
- Stage 1: 20 cycles
    - o 94ºC for 10 seconds
    - o 57ºC for 5 seconds
    - o 72ºC for 10 seconds
- Stage 2: 22 cycles
    - o 94ºC for 10 seconds
    - o 57ºC for 20 seconds
    - o 72ºC for 40 seconds
- Stage 3: 15 cycles
    - o 94ºC for 10 seconds
    - o 57ºC for 1 minute

Sequencing conditions

Std

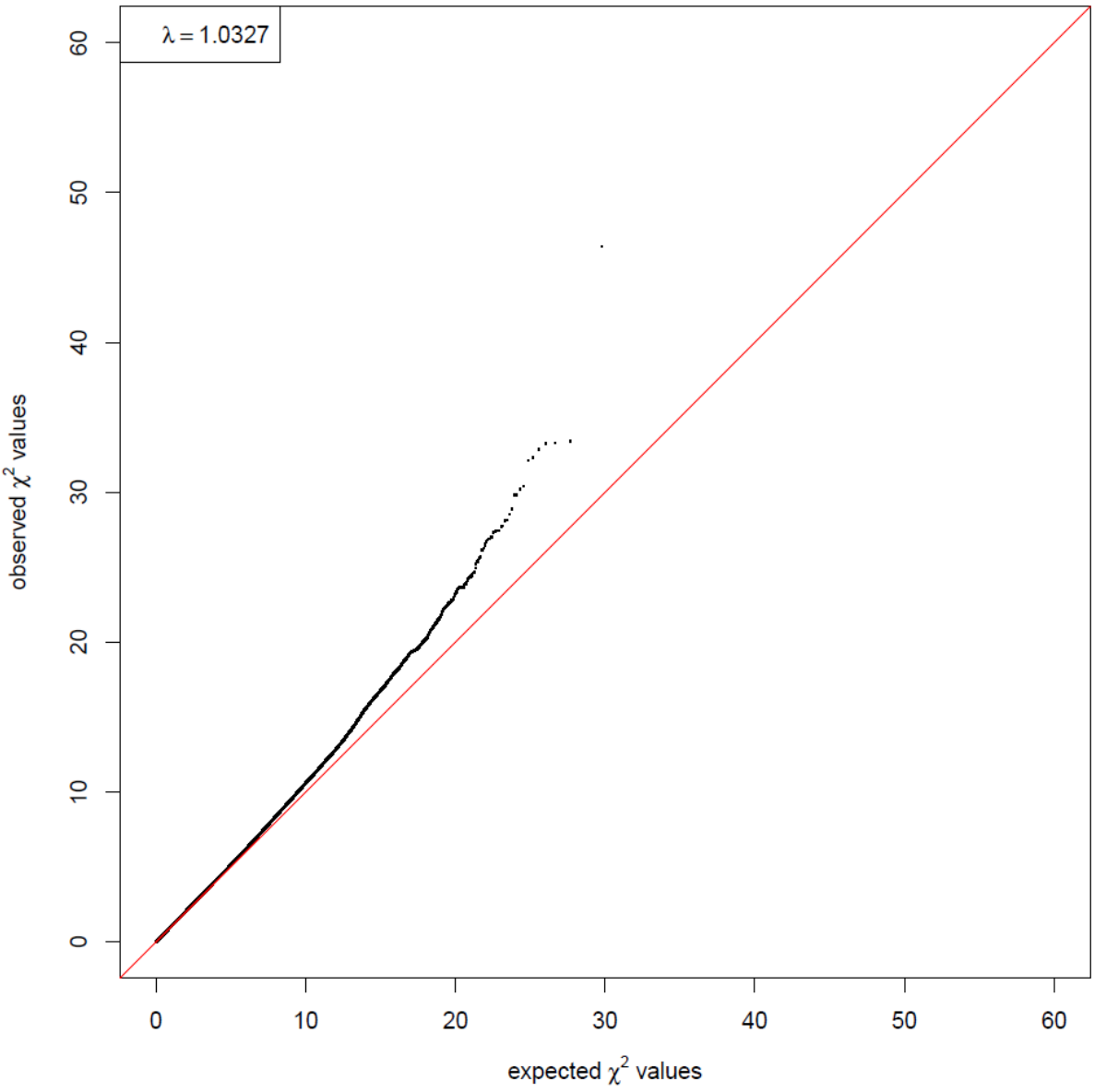
- 95ºC for 5 minutes
- 25 cycles
    - o 96ºC for 30 seconds
    - o 50ºC for 15 seconds
    - o 60ºC for 1 minute

rsID	Locus	Probe chromosome	Gene	P_{SMR}	P_{HEIDI}
rs6595443	5q23.2	5	CEP120	1.27×10^{-4}	6.60×10^{-2}
rs2807754	10p21.1	10	WAC	4.53×10^{-5}	6.28×10^{-1}
rs1423269	5q15	5	ELL2	7.08×10^{-7}	5.58×10^{-3}
rs4487645	7p15.3	7	CDCA7L	8.37×10^{-15}	1.08×10^{-2}
rs6090899	20q13.13	20	PREX1	4.01×10^{-4}	5.46×10^{-3}

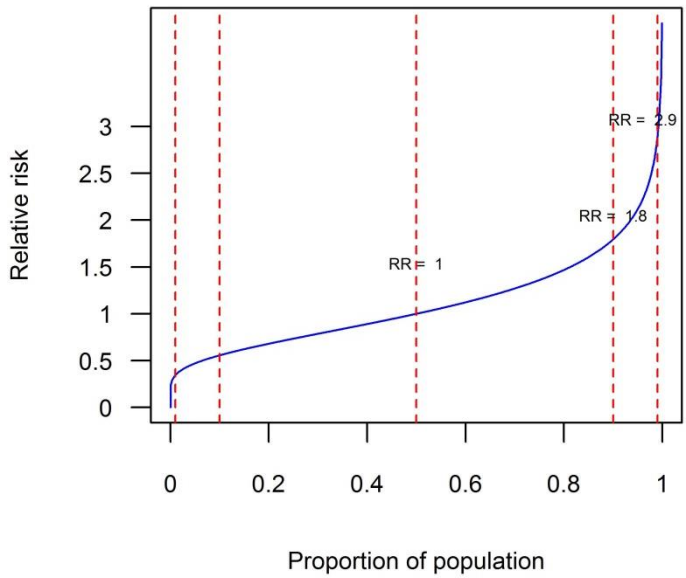
Supplementary Table 16: Summary of results from SMR analysis. We set a threshold for the SMR test of P_{SMR} $<1\times10^{-3}$ corresponding to a Bonferroni correction for 45 tests. For all genes passing this threshold we generated plots of the eQTL and GWAS associations at the locus, as well as plots of GWAS and eQTL effect sizes (i.e. corresponding to input for the HEIDI heterogeneity test). HEIDI test *P* values <0.05 were considered as being reflective of heterogeneity. This threshold is conservative for gene discovery because it retains fewer genes than when correcting for multiple testing. Probes which passed the HEIDI threshold are highlighted in grey.

Locus	Lead SNP	Transcription Factor																	
3q26.2	rs10936600	ATF2	BATF	CEBPB	CHD1	CTCF	EBF1	ELF1	ELK1	EP300	ETS1	FOXM1	IRF4	MAX	MAZ	MTA3	MXI1	NFIC	PML
		POLR2A	POLR3G	POU2F2	RUNX3	RXRA	SIN3A	STAT5A	TAF1	TBL1XR1	TBP	WRNIP1	YY1						
5q15	rs1423269	ATF2	BATF	BCL11A	BCL3	BCLAF1	BHLHE40	CEBPB	CHD2	EBF1	EP300	FOXM1	IKZF1	IRF4	JUND	MAZ	MEF2A	MEF2C	MTA3
		MXI1	NFATC1	NFIC	PML	POU2F2	RELA	RUNX3	SP1	SPI1	STAT3	STAT5A	TBL1XR1	TBP	TCF12	TCF3			
8q24	rs1948915	ATF2	BCL3	BCLAF1	CEBPB	CTCF	EBF1	EP300	FOXM1	JUND	MAZ	MEF2A	MEF2C	MTA3	MXI1	NFIC	PML	POLR2A	RAD21
		RELA	RUNX3	SIN3A	SPI1	STAT3	STAT5A	TBL1XR1	YY1	SMC3									
16p11	rs13338946	BCL3	CHD1	CHD2	CTCF	EBF1	MAZ	MXI1	NFIC	POLR2A	RELA	RUNX3	SIN3A	SP1	SPI1	TAF1	TCF12	WRNIP	YY1
20q13	rs6066835	ATF2	BCL11A	EBF1	ELF1	EP300	FOXM1	IKZF1	MEF2A	MEF2C	NFIC	POLR2A	RUNX3	SPI1	TBL1XR1	USF1	WRNIP1		
22q13	rs138747	ATF2	ATF3	BCL3	BHLHE40	CEBPB	CHD1	CHD2	EBF1	EGR1	ELF1	ELK1	EP300	FOS	FOXM1	GABPA	IKZF1	IRF4	MAX
		MAZ	MEF2C	MTA3	MXI1	NFATC1	NFE2	NFIC	NFYA	NFYB	NR2C2	PAX5	PBX3	PML	POLR2A	POU2F2	RELA	RUNX3	SIN3A
		SP1	SPI1	SRF	STAT5A	TAF1	TBL1XR1	TBP	TCF12	TCF3	USF1	USF2	WRNIP1	YY1	ZEB1	ZNF143			

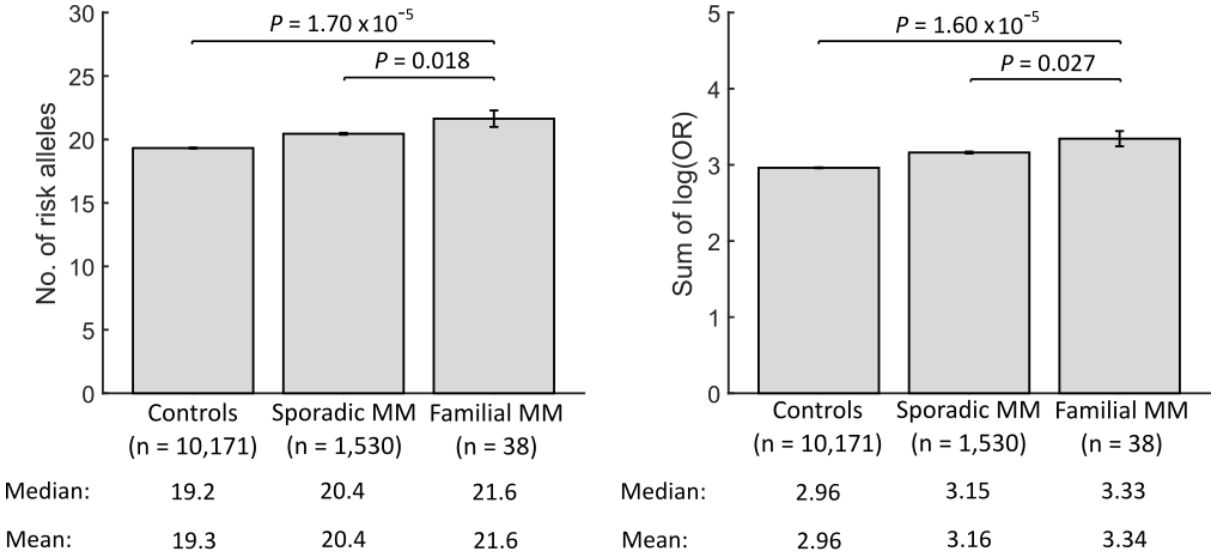
Supplementary Table 17: Full lists of TF binding at selected loci. TF ChIP-seq (161 factors) with Factorbook Motifs for GM12878 were downloaded from ENCODE[8].



Supplementary Figure 1: Quantile-Quantile (*Q-Q*) plots of observed and expected χ^2 values of association between SNP genotype and risk of multiple myeloma after imputation for the OncoArray cohort. λ=1.0327, λ_{1000}=1.0209**.** The red line represents the null hypothesis of no true association. *Q-Q* plots for the UK, Sweden/Norway, Germany, Iceland, USA and Netherlands sets have been previously reported[1-4].



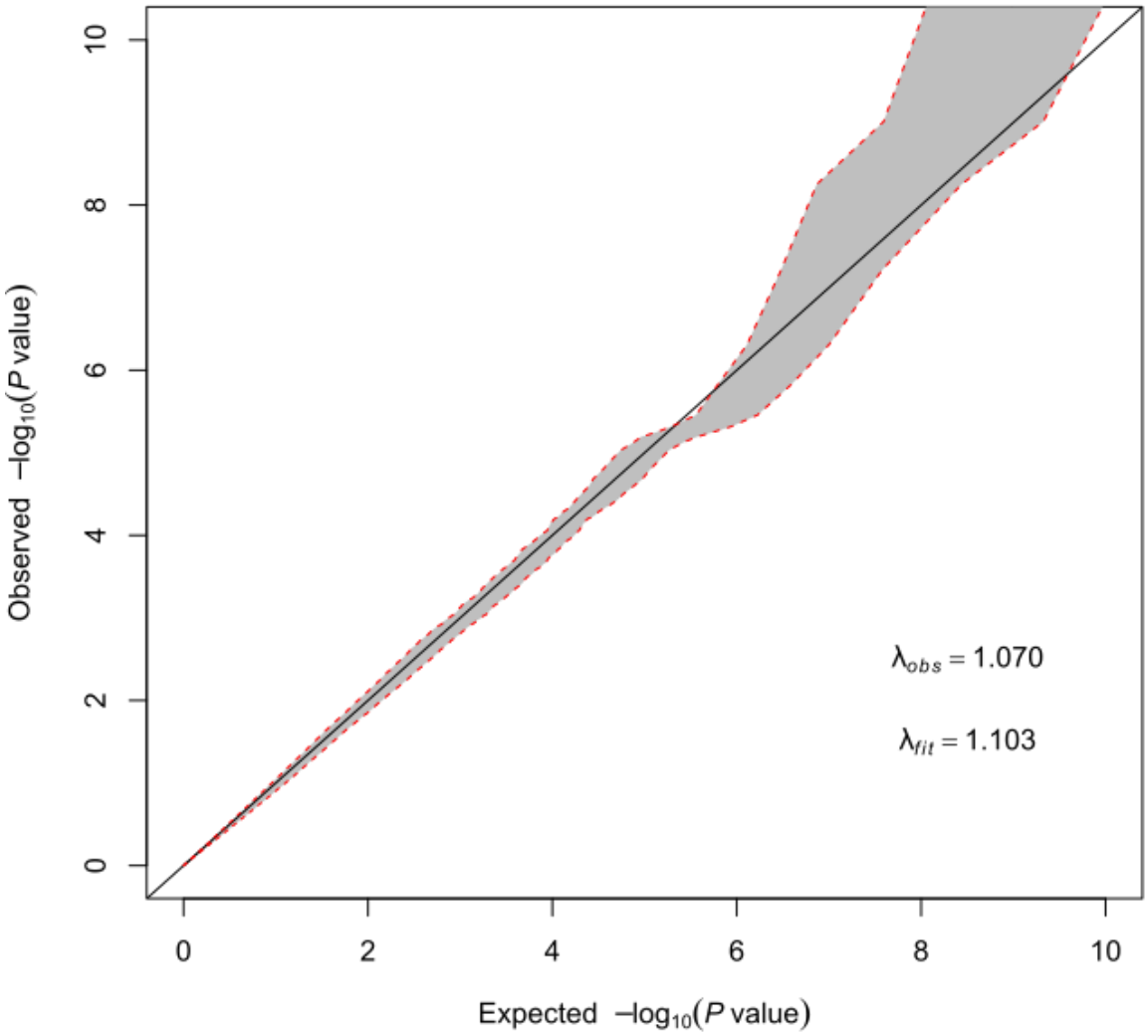
Supplementary Figure 2: Population distribution of polygenic risk score (PRS) ordered by relative risk (RR) (compared with population median risk). PRS is based on the 23 risk SNPs. Vertical red lines (left to right) correspond to 1%, 10%, 50%, 90%, and 99% centile, respectively.



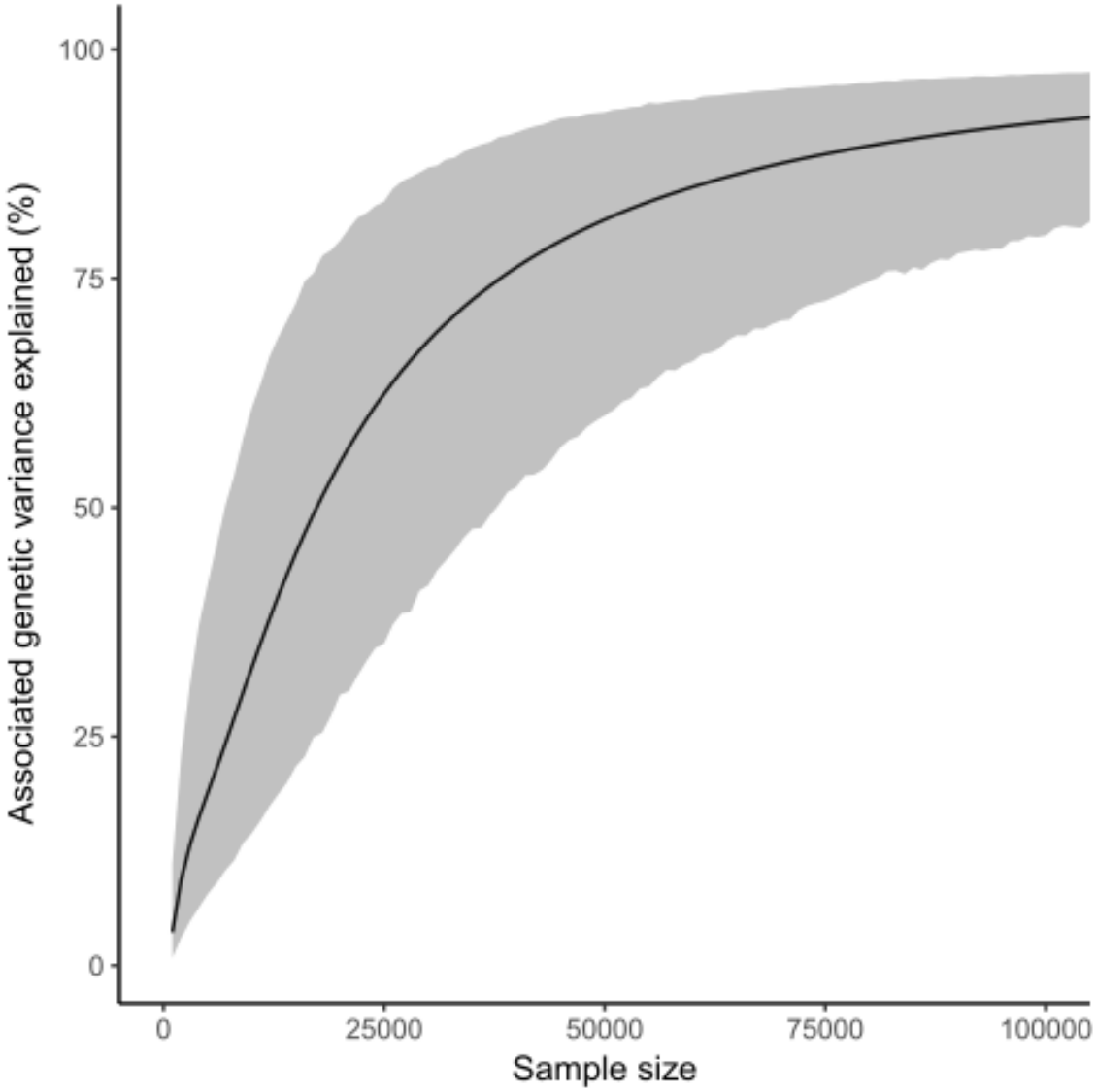
Supplementary Figure 3: Polygenic risk scores (PRS) for familial MM, sporadic MM and population-controls. A higher risk allele burden is seen in the familial MM compared with both sporadic MM and controls (difference in PRS score tested by one-sided Student's t-test). (a) Based on number of risk alleles carried; (b) Calculated as the sum log-transformed odds ratios. The observed 1.08-fold enrichment of PRS in familial over sporadic cases is entirely compatible the expected familial risk attributable to the 23 risk SNPs of 1.10 given by:

$$\prod_{i}^{n=23} \frac{p_i r_i^2 + q_i}{p_i r_i + q_i^2}$$

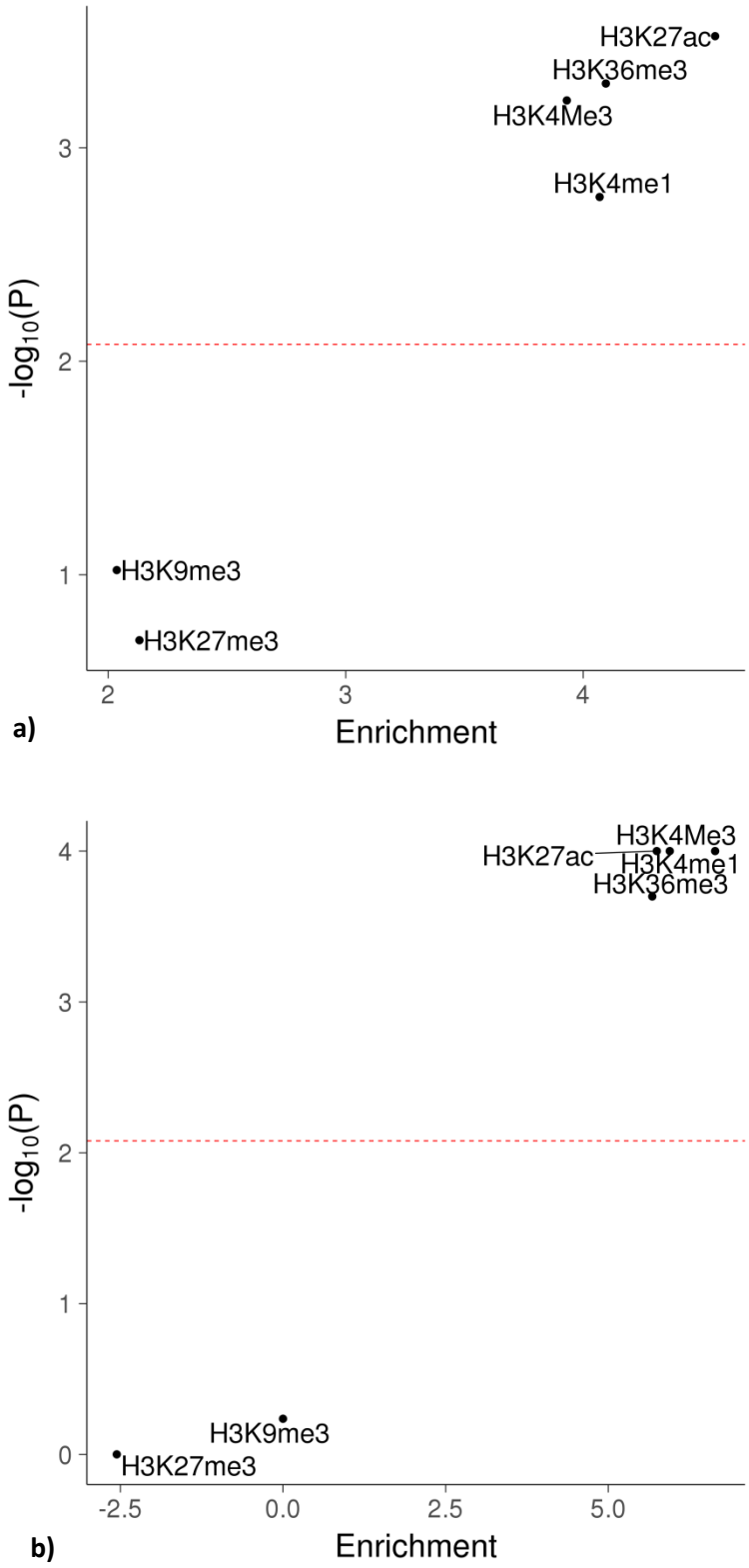
where p_i is the frequency of the risk allele for locus i, $q_i = 1 - p_i$, and r_i is the estimated per-allele OR.



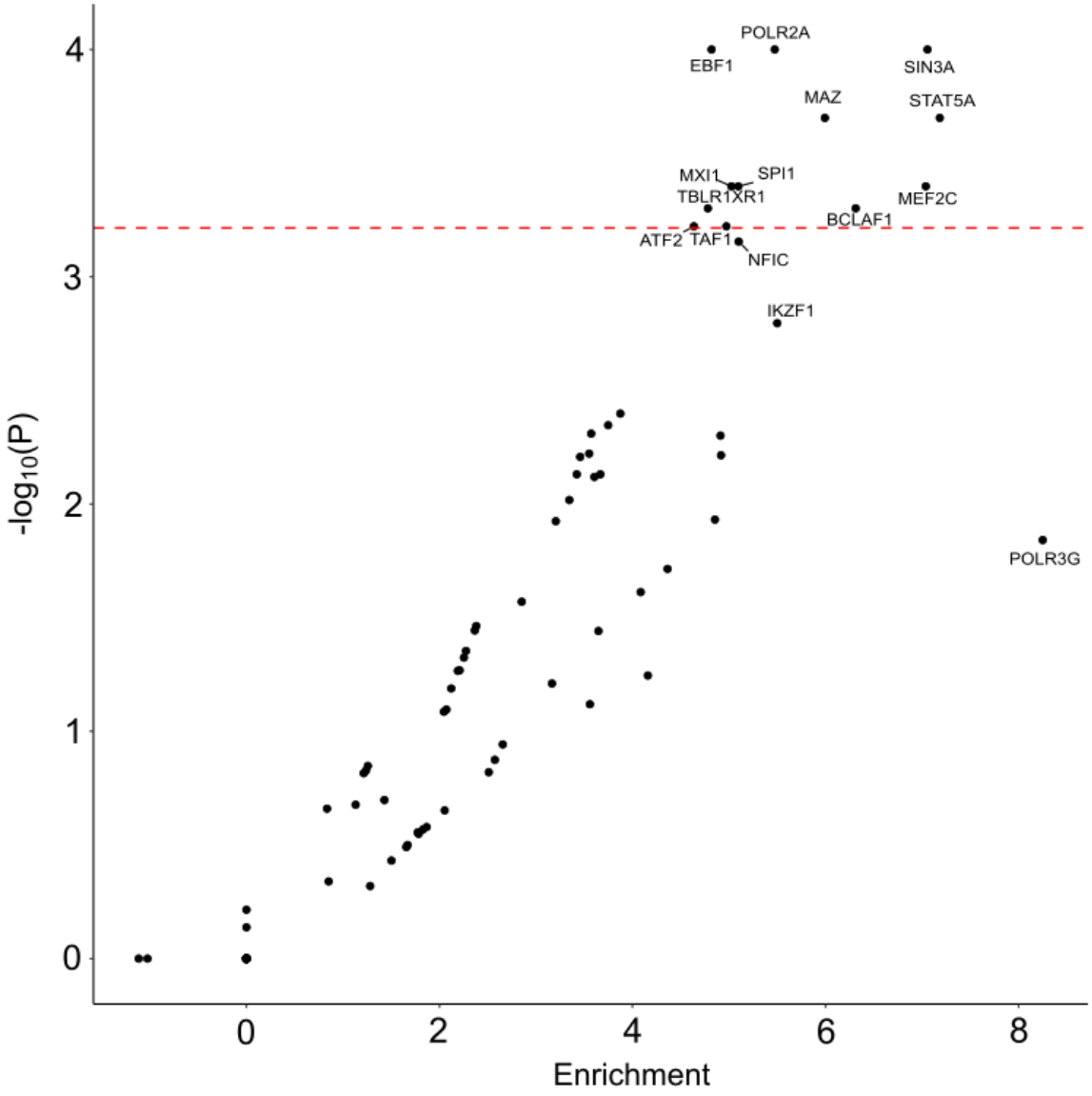
Supplementary Figure 4: *Q-Q* plot comparing observed distributions of association statistics against those expected under a three-component model. Grey shaded area represents the 80% confidence interval.



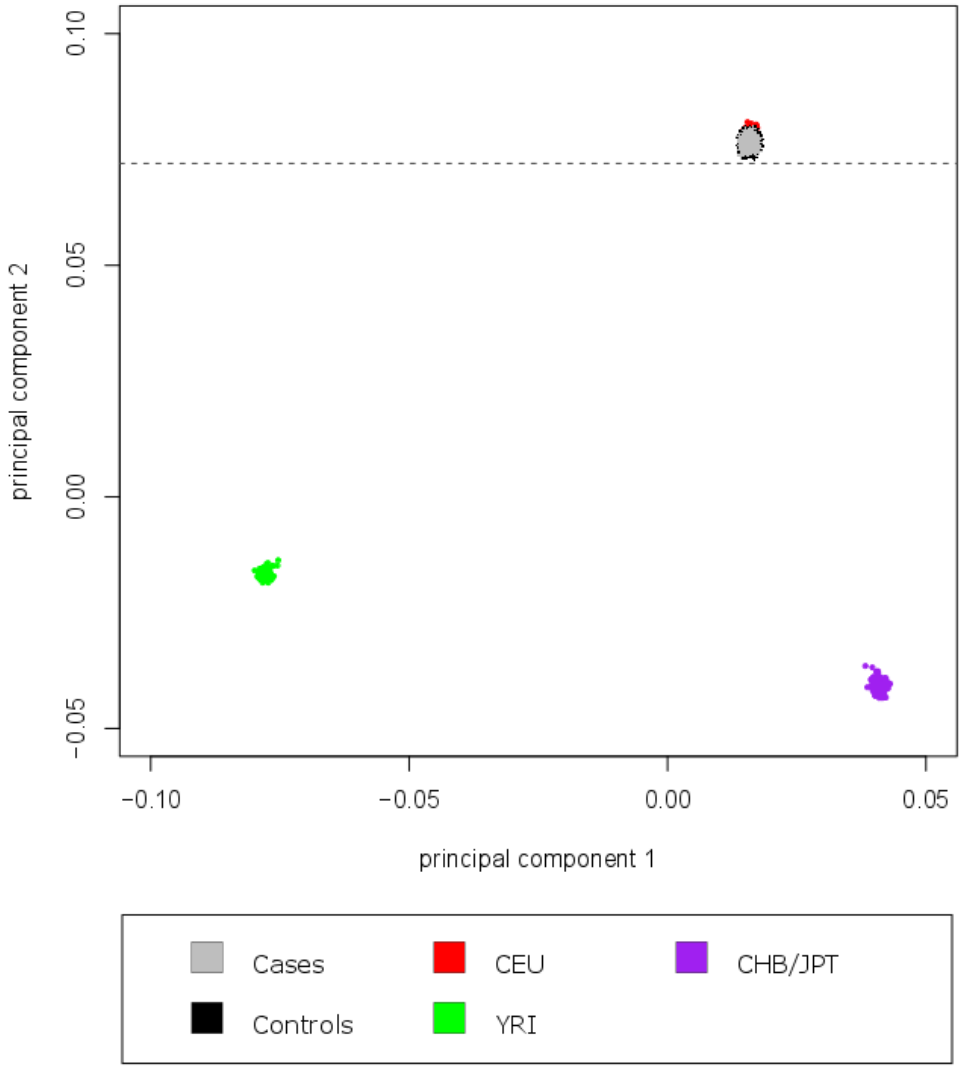
Supplementary Figure 5: Projected percentage of GWAS heritability explained for a given sample size. Results were obtained using a three-component model to estimate distribution of effect sizes. Grey shaded area represents the 95% confidence interval of the heritability estimate.



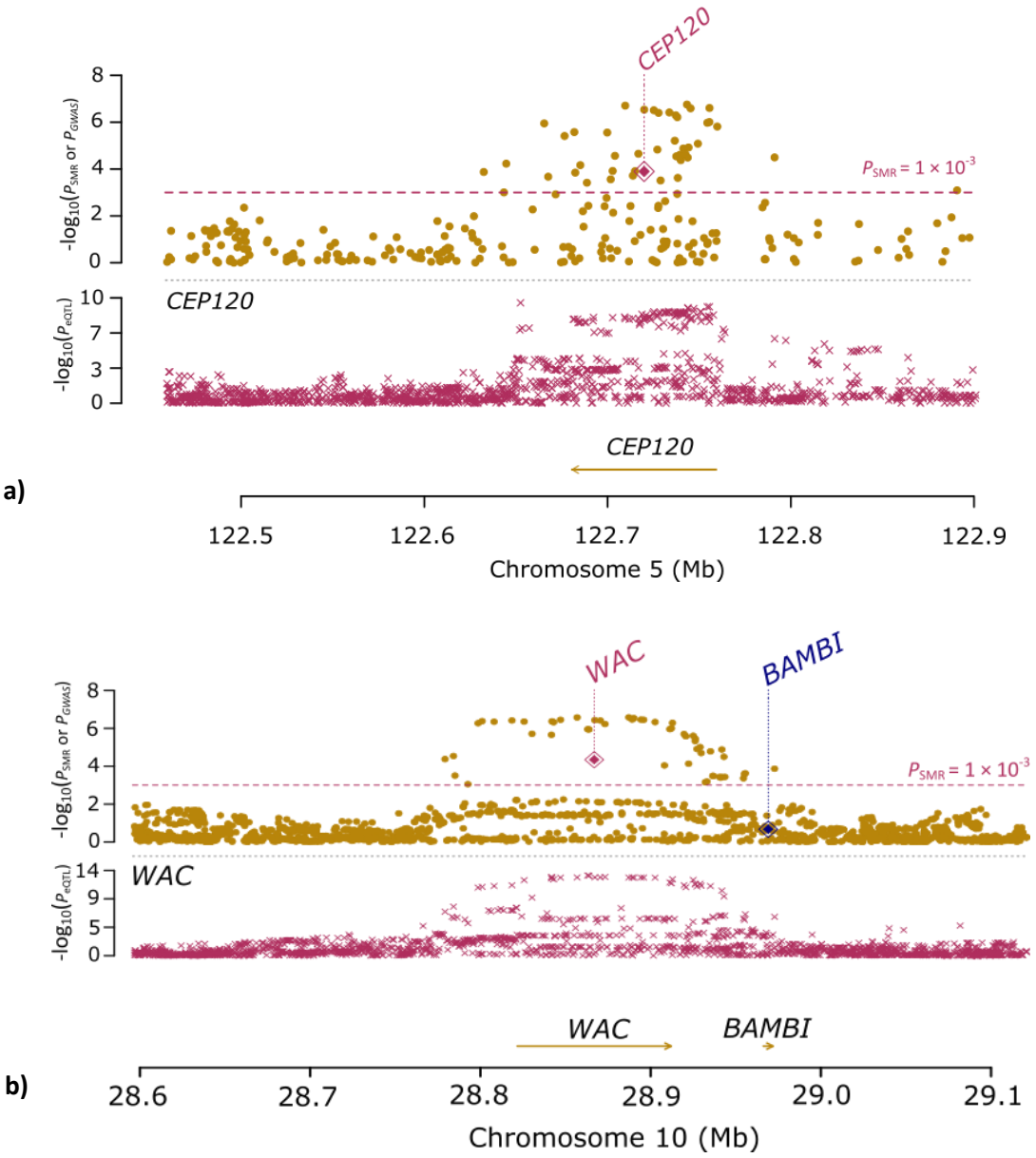
Supplementary Figure 6: The overrepresentation of histone marks in (a) naïve B and (b) KMS11 cells at the location of new and known MM risk SNPs demonstrates that risk SNPs are enriched in regions of open chromatin. The red line denotes the Bonferroni corrected *P*-value threshold.



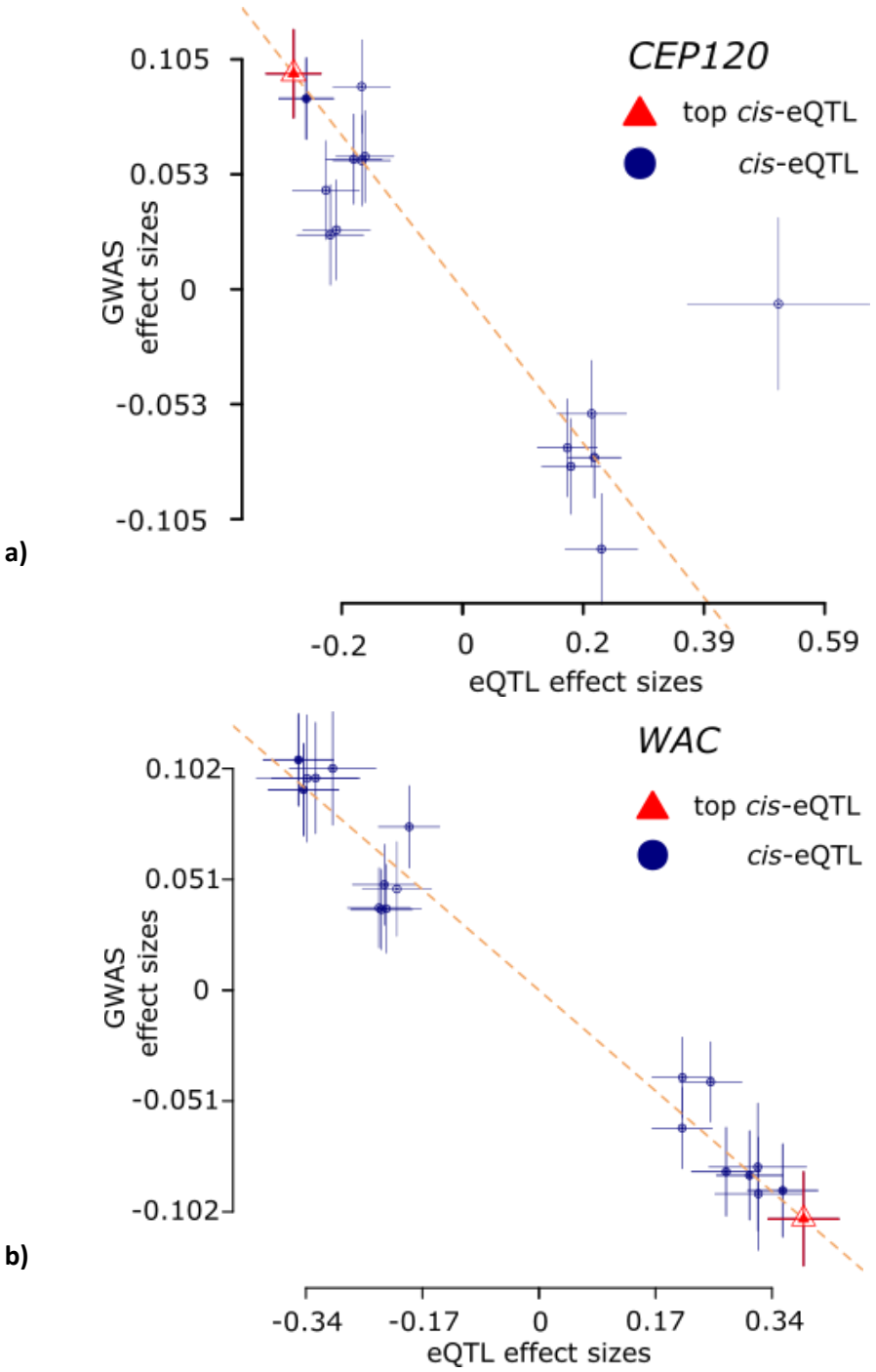
Supplementary Figure 7: The overrepresentation of transcription factor (TF) binding sites in GM12878 cells at the location of new and known MM risk SNPs demonstrates that risk SNPs are enriched in regions of B-cell relevant TF binding. The red line denotes the Bonferroni corrected *P*-value threshold.



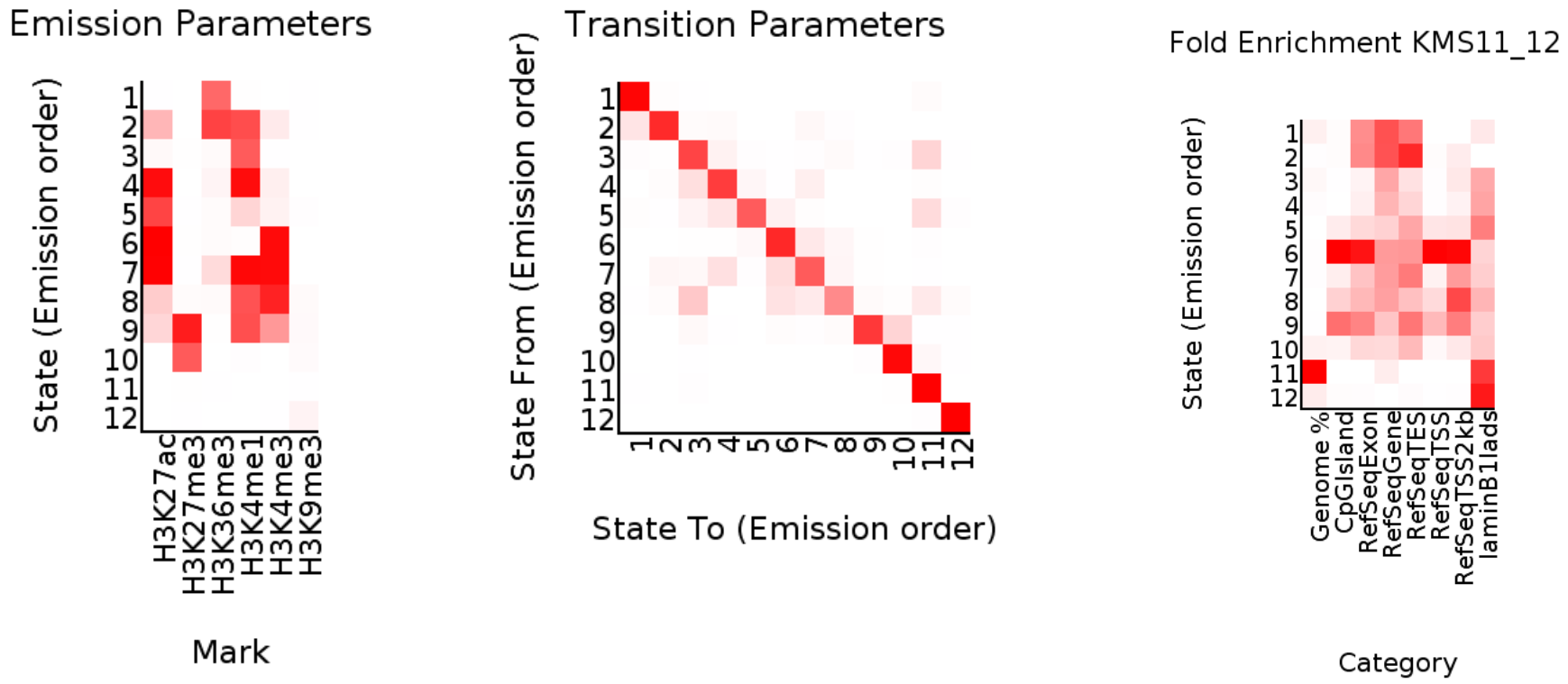
Supplementary Figure 8: Principal components analysis plot for the OncoArray cohort after removal of non-European cases. The first two principal components of the analysis are plotted. Cases and controls outside of the intervals $0.0155 \leq x \leq 0.019$, and $0.0735 \leq y \leq 0.079$ were excluded in order to remove individuals of non-European ancestry (grey dotted line shows the lower threshold of the second principal component). HapMap CEU individuals are plotted in red; CHB/JPT individuals are plotted in purple; YRI individuals are plotted in green. Cases are plotted in grey, controls plotted in black.



Supplementary Figure 9: Summary data-based Mendelian Randomization (SMR) analysis locus plot at a) 5q23.2 and b) 10p12.1. Upper panel - brown dots represent *P*-values for SNPs from the GWAS meta-analysis, diamonds represent *P*-values for probes from the SMR test; lower panel – crosses represent eQTL *P*-values of SNPs from MM plasma cells from 183 MRC MyIX trial patients (GEO: GSE21349) and 658 Heidelberg GMMG patients (EMBL-EBI: E-MTAB-2299), with genes passing the SMR (i.e. $P_{SMR} < 0.001$) and HEIDI (*i.e.* $P_{HEIDI} > 0.05$) tests highlighted in red. Probeset ID refers to Affymetrix U133 2.0 Plus Array custom chip definition file (CDF v.17) mapping to Entrez genes.



Supplementary Figure 10: Summary data-based Mendelian Randomization analysis effect plot at (a) 5q23.2 and (b) 10p12.1 Blue dots represent effect sizes of SNPs from the GWAS meta-analysis against those from the eQTL study of MM plasma cells from 183 MRC MyIX trial patients (GEO: GSE21349) and 658 Heidelberg GMMG patients (EMBL-EBI: E-MTAB-2299). The top *cis*-eQTL is highlighted by a red diamond. Error bars are the standard errors of the SNP effects. An estimate of b_{xy} at the top *cis*-eQTL is represented by the orange dotted line.



Supplementary Figure 11: Heat maps outputted by ChromHMM pipeline show a) emission parameters, b) transition parameters and c) state functional enrichments for the KMS11 MM cell line. Columns in **(c)** are labelled as follows: Genome % indicates the relative percentage of the genome represented by each state and relative fold enrichment for RefSeq transcription start sites (TSS); CpG Islands; 2000 base pair intervals around the TSS; exons; genes; transcript end sites (TES); evolutionary conservation; and nuclear lamina associated regions, respectively. Heat maps shown were used to assign states based on previously described rules[9-11]. The ChromHMM model was learned across 3 MM cell lines; JJN3, KMS11 and MM1S.

REFERENCES

1. Broderick, P. *et al.* Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet* **44**, 58-61 (2011).
2. Chubb, D. *et al.* Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nat Genet* **45**, 1221-1225 (2013).
3. Swaminathan, B. *et al.* Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. *Nat Commun* **6**, 7213 (2015).
4. Mitchell, J.S. *et al.* Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nat Commun* **7**, 12050 (2016).
5. Chiecchio, L. et al. Deletion of chromosome 13 detected by conventional cytogenetics is a critical prognostic factor in myeloma. *Leukemia* **20**, 1610-7 (2006).
6. 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**, 1150-7 (2010).
7. Johnson, D.C. et al. Genome-wide association study identifies variation at 6q25.1 associated with survival in multiple myeloma. *Nat Commun* **7**, 10290 (2016).
8. de Souza, N. The ENCODE project. *Nat Methods* **9**, 1046 (2012).
9. Hoffman, M.M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res* **41**, 827-41 (2013).
10. Fiziev, P. *et al.* Systematic Epigenomic Analysis Reveals Chromatin States Associated with Melanoma Progression. *Cell Rep* **19**, 875-889 (2017).
11. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet* **47**, 1179-1186 (2015).