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

Molly Went, Amit Sud, Asta Försti, Britt-Marie Halvarsson, Niels Weinhold et al.

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 9974 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.

Correspondence and requests for materials should be addressed to K.H. (email: K.Hemminki@dkfz-heidelberg.de) or to B.N. (email: bjorn.nilsson@med.lu.se) or to R.S.H. (email: richard.houlston@icr.ac.uk).

A full list of authors and their affiliations appears at the end of the paper.
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. 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, with an additional locus being subtype-specific for t(11;14) translocation MM. Much of the heritable risk of MM, however, remains unexplained and statistical modelling indicates that further common risk variants remain to be discovered.

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 9794 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 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 characterise 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 (878 MM cases and 7083 controls from the UK), followed by a meta-analysis with six published MM GWAS data sets (totalling 7319 cases and 234,385 controls) (Fig. 1, Supplementary Tables 1–3). 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 ($\lambda = 1.03, \lambda_{1000} = 1.02$, 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 1777 cases and 6088 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. $P < 5 \times 10^{-8}$) for six new loci at 2q31.1, 5q23.2, 7q22.3, 7q31.33, 16p11.2 and 19p13.11 (Table 1, Supplementary Table 4 and Fig. 2). Additionally, borderline associations were identified at two loci with $P$ values of $5.93 \times 10^{-6}$ (6p25.3) and $9.90 \times 10^{-8}$ (7q21.11), which have corresponding Bayesian false-discovery probabilities (BFDP) 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.

**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 and hyperdiploid MM, 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 (LDK), 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. Using this approach, an individual in the top 1% of genetic risk has a threefold increased risk of MM when compared to an individual with median genetic risk. 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...
We found enrichment of MM SNPs in regions of active and publicly accessible naive B-cell Blueprint Epigenome Project.

To investigate the functional role of previously reported and new (Supplementary Fig. 7). Collectively these data support the tenet sites using ENCODE GM12878 lymphoblastoid cell line data and an enrichment of relevant B-cell transcription factor-(TF) binding and H3K4Me1 marks (Supplementary Fig. 6). We also observed that the MM predisposition loci in KMS11 and naive B-cells using CHi-C data19. We also sought interrogation physical interactions at respective genomic regions interactions are key to the regulation of gene expression, we 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 Randomisation (SMR) analysis20 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 (Supplementary Data 1 and Fig. 1).

Although preliminary, and requiring functional validation, our analysis delineates four potential candidate disease mechanisms across the 23 MM risk loci (Supplementary Data 1). 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 2p31.1, 5q23.2) maps 5′ of POT1, the protection of telomeres 1 gene.

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 naive B-cell Blueprint Epigenome Project data. 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 naive B-cells using CHi-C data. 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 Randomisation (SMR) analysis20 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 (Supplementary Data 1 and Fig. 1).

Although preliminary, and requiring functional validation, our analysis delineates four potential candidate disease mechanisms across the 23 MM risk loci (Supplementary Data 1). 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 2p31.1, 5q23.2). CEP120 is required for microtubule assembly and elongation with overexpression of CEP120 leading to uncontrolled centriole elongation, 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.

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.

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 POT1 (at 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.

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.

Third, 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 POT1 (at 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.

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.
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 CDC47 through differential IRF4 binding. 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. 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 IRF4 in MM biology. Demethylation by KDM3A histone demethylase sustains IRF4 binding and is labelled by its rsID. The colour intensity of each symbol represents expression and implies a potential role in MM oncogenesis is 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). Additionally, SNP rs34562254 at 17p11.2 is responsible for the amino acid substitution (NP_036584.1:p.Pro251Leu) 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 KMS1.

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). 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 homoeostasis, which has an established role in MM biology.27–42.

Discussion

Our meta-analysis of a new GWAS series in conjunction with previously published MM data sets has identified six novel risk loci.
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. 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.

Integration of Hi-C data with ChIP-seq chromatin profiling from MM and lymphoblastoid cell lines and naive 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 ULK1, ATRX, and WAC, and by virtue of the role of IRF4- MYC related autophagy, CDCA7L, DNMT3A, CBX7 and KLF2 in MM development (Supplementary Data 1). 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 that 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 underscores 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 (MRC 02/8/95, IRScRN68454111), the Myeloma XI trial by the Oxfordshire Research Ethics Committee (MREC 17/09/09, IRScRN4967852), HOVON65/ GMMG-HD4 (IRScRN 644552890; METC 13/01/2013), HOVON87/NSMG18 (EudraCTn 2007-004007-34, METC 20/11/2008), HOVON95/EMN02 (EudraCT 2009-01793-28, METC 04/11/10), University of Heidelberg Ethics Committee (229/2003, 5-337/2009, 4Fnu-119/2010), University of Arkansas for Medical Sciences Institutional Review Board (IRB 202077), Lund University Ethics Review Board (2013/54), the Norwegian REK 2014/97, the Danish Ethical Review Commission (229/2003, S-337/2009, AFmu-119/2010), University of Arkansas for Medical Sciences Institutional Review Board, University of Arkansas for Medical Sciences Institutional Review Board and University of Copenhagen (EudraCTnr 2007-004007-34, METC 20/11/2008), HOVON95/EMN02 (EudraCT 2009-01793-28, METC 04/11/10), University of Heidelberg Ethics Committee (229/2003, 5-337/2009, 4Fnu-119/2010), University of Arkansas for Medical Sciences Institutional Review Board (IRB 202077), Lund University Ethics 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).

Replication studies and technical validation. To promote distinguishing potential associations from spurious genetic findings, we analysed three case–control series from Germany, Sweden and Denmark, comprising in Supplementary Table 1. All these studies were based on individuals with European ancestry and compared UK GWAS (2282 cases, 1978 controls), Swedish–GWAS (1714 cases, 10,391 controls), German–GWAS (1508 cases, 2107 controls), Netherlands–GWAS (555 cases, 2669 controls), US–GWAS (780 cases, 1857 controls) and Iceland (480 cases, 2116 controls).

Published GWAS. The data from six previously reported GWAS3–5 are summarised in Supplementary Table 1. All these studies were based on individuals with European ancestry and compared UK GWAS (2282 cases, 1978 controls), Swedish–GWAS (1714 cases, 10,391 controls), German–GWAS (1508 cases, 2107 controls), Netherlands–GWAS (555 cases, 2669 controls), US–GWAS (780 cases, 1857 controls) and Iceland (480 cases, 2116 controls).

Meta-analysis. Meta-analyses were performed using the fixed-effects inverse-variance method using META v1.6.2 and Cochran’s Q-statistic to test for heterogeneity and the I² 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 Analysis2 to perform conditional association analysis. Association statistics were calculated for all SNPs containing on the top SNP in each loci showing genome-wide significance. This was carried out step-wise.

For borderline associations, the BFBD10 was calculated based on a plausible OR of 1.2 and a prior probability of association of 0.0001. For both promising associations, the BFBD was <10%.

Fluorescence in situ hybridisation. Fluorescence in situ hybridisation (FISH) and ploidy classification of UK and German samples were performed as previously described33,34. Logistic regression in case–control 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 Myelx trial patients and 685 German GMMG patients32. 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 was carried out using SMR analysis as per Zhu et al.3. Briefly, if h is the effect size of x (gene expression) on y (slope of y regressed on the genetic value of x), h is the effect of x on y, h is the effect of x on y, h 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 and the other affecting a trait, we tested for heterogeneity in independent datasets (HEIDI), using multiple SNPs in each cis-eQTL region. Under the hypothesis of pleiotropy, h values for SNPs in LD with the causal variant should be identical. For each probe that passed significance

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-04989-w | www.nature.com/naturecommunications
threshold for the SMR test, we tested the heterogeneity in the \( h_{a} \) 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 \( P_{\text{SMR}} < 1 \times 10^{-8} \) corresponding to a Bonferroni correction for 45 tests, i.e. 45 probes which demonstrated an association in the SMR test. For all genes that passed this threshold, we generated plots of the eQTL and GWAS association at the locus, as well as plots of GWAS and eQTL effects (i.e. input for the HEIDI heterogeneity test). HEIDI test \( P < 0.05 \) 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 CHi-C. To map risk SNPs to interactions involving promoter contacts and identify genes involved in MM susceptibility, we analysed publicly accessible promoter CHi-C data on the naive B cells downloaded from Blueprint Epigenome Project. Additionally, we also analysed promoter CHi-C data that we have previously generated for the MM cell line KMS11\(^{2,2} \). Interactions were called using the CHiCAGO pipeline to obtain a unique list of reproducible contacts\(^{5,6} \) and those with a \(-\log\text{weighted} \ P \geq 0.05 \) 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 naive B cells from Blueprint Epigenome\(^{5,6} \). 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 JN3. Genome-wide signal tracks were binarized (including input controls for ChIP-seq data), and a set of learned models were generated using ChromHMM software\(^{5,6} \). A 12-state model was suitable for interpretation of these binary tracks. We assigned the states based on the genomic marks that 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.\(^{5,6} \). Briefly, for each risk locus, a region of strong LD (defined as \( r^2 > 0.8 \) and \( D^2 > 0.8 \) was determined, and these SNPs were considered the associated variant set (AVS)).

Publicly 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\(^{5,6} \). In addition, ChIP-seq peak data for six histone marks from KMS11 cell line were generated in-house, and naive B cell ChIP-seq data were downloaded from Blueprint Epigenome Project\(^{5,6} \). For each map, the overlap of the SNPs in the AVS and the binding sites was assessed to generate a mapping tally. A null dis-

tribution 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 were 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 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. 6 and 7.

**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 and (v) presence of a nonsynonymous change across risk loci. Candidate causal genes were then assigned to MM risk loci using

- (i) presence of a Hi-C contact linking to a gene promoter,
- (ii) presence of an

results of this study have been deposited in Gene Expression Omnibus with accession codes GSE13149, GSE19784, GSE24080, GSE2658 and GSE15895; in the European Genome-phenome Archive (EGA) with accession code EGA500008000001; 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 phs00207-v1.1. Expression data that support the findings of this study have been deposited in GEO with accession codes GSE21349, GSE2658, GSE3161 and EMBL-EBI with accession code E-MTAB-2299. The remaining data are contained within the paper and Supplementary Files or are available from the author upon request. KMS11 Hi-C data used in this manuscript are deposited in EGA under accession number EGA500001002814. The accession number for the KMS11 ChIP-seq data reported in this paper is EGA: S00001002414. Naive B-cell HiC data used in this work is publically available from Blueprint Blueprint Epigenome Project [https://osf.io/ubtpz/]. ChIP-seq data for H3K27ac, H3K4Me1, H3K27Me3, H3K9Me3, H3K36Me3 and H3K27Me3 from naive B cells are publically available and were obtained from Blueprint Epigenome Project [http://www.blueprint-epigenome.eu/].

Received: 12 December 2017 Accepted: 6 June 2018
Published online: 13 September 2018

References

1. Aliteri, A., Chen, B., Bermejo, J. L., Castro, F. & Hemminki, K. Familial risks and temporal incidence trends of multiple myeloma. *Eur. J. Cancer* 42, 1661–1670 (2006).
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 in *SWEN2* influence multiple myeloma risk. *Nat. Commun.* 6, 7213 (2015).
6. Weinhold, N. et al. The CCND1 c.870G>A polymorphism is a risk factor for 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. Biomark. Prev.* 26, 126–135 (2017).
9. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science* 273, 1516–1517 (1996).
10. Wakefield, J. A. Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* 81, 208–227 (2007).
11. Erickson, S. W. et al. Genome-wide scan identifies variant in 2q12.3 associated with an increased risk for multiple myeloma. *Blood* 124, 2001–2003 (2014).
12. Li, N. et al. Genetic predisposition to multiple myeloma at 5q15 is mediated by an ELL2 enhancer polymorphism. *Nat. Genet.* 45, 722–725 (2013).
13. Mitchell, J. S. et al. Implementation of genome-wide complex trait analysis to quantify the heritability in multiple myeloma. *Sci. Rep.* 5, 12473 (2015).
14. Alonso, G. et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat. Genet.* 44, 58–61 (2011).
15. Frampton, M. J. et al. Implications of polygenic risk for personalised colorectal cancer screening. *Ann. Oncol.* 27, 429–434 (2016).
16. 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. Preprint at [https://www.biorxiv.org/content/ early/2017/08/11/175406](https://www.biorxiv.org/content/early/2017/08/11/175406) (2017).
16. Chatterjee, N. et al. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat. Genet. 45, 490–495 (2013).
17. Freedman, M. L. et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat. Genet. 43, 513–518 (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, 357–372 (2015).
20. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487 (2016).
21. Comarint, D. et al. CEP120 and SPICEI cooperate with CPAP in centriole elongation. Curr. Biol. 23, 1360–1366 (2013).
22. Rice, C. et al. Structural and functional analysis of the human POTI-TTP1 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, 2170–2184 (2016).
24. Chang, S. Cancer chromosomes going to POT1. Nat. Genet. 45, 473–475 (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. McDanel, T. G., Hannon, K. & Moody, D. E. Ankyrin repeat and SOCS box protein 15 regulates protein synthesis in skeletal muscle. Am. J. Physiol. Regul. Integ. Comp. Physiol. 290, R1672–R1682 (2006).
28. McDanel, 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, 2897–2902 (2008).
29. Kile, B. T. et al. The SOCS box: a tale of destruction and degradation. Trends Biochem. Sci. 27, 235–241 (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, 540–550 (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, 1145–1153 (2004).
32. Li, N. et al. Multiple myeloma risk variant at 7p15.3 creates an IRF4-binding site and interferes with CDCA7L expression. Nat. Genet. 40, 86–92 (2008).
33. Holien, T., Vatsveen, T. K., Hella, H., Waage, A. & Sundan, A. Addiction to c-Myc in multiple myeloma cells to form an essential survival loop. Cell Rep. 5, 208–217 (2013).
34. Kuehl, W. M. & Bergsagel, P. L. MYC addiction: a potential therapeutic target for multiple myeloma. Blood 120, 1245–1246 (2012).
35. Speedy, H. E. 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, 1145–1153 (2004).
36. Li, N. et al. Multiple myeloma risk variant at 7p15.3 creates an IRF4-binding site and interferes with CDCA7L expression. Nat. Genet. 40, 86–92 (2008).
37. Holien, T., Vatsveen, T. K., Hella, H., Waage, A. & Sundan, A. Addiction to c-Myc in multiple myeloma cells to form an essential survival loop. Cell Rep. 5, 208–217 (2013).
38. Ohtsuki, H. et al. The KIDMA-KL2-IRF4 axis maintains myeloma cell survival. Nat. Commun. 7, 10258 (2016).
39. Yang, M., Lewis, I., Fan, X., Zhu, J. & Yuan, Z. M. PRR14 is a novel activator of the PI3K pathway promoting lung carcinogenesis. Oncogene 35, 5527–5538 (2016).
40. 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, 119–129 (2009).
41. 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, 1021–1030 (2005).
42. Xu, J. et al. Correlation of expression levels of Bls/5 and its receptors with multiple myeloma. Clin. Biochem. 42, 387–399 (2009).
43. Novak, A. J. et al. Expression of BCMA, TACI and BAFF-R in multiple myeloma: a mechanism for growth and survival. Blood 103, 689–694 (2004).
44. Macker, F. & Schneider, P. TACI, an enigmatic BAFF/APRIL receptor, with new unappreciated biochemical and biological properties. Cytokine Growth Factor Rev. 19, 263–276 (2008).
45. Moreaux, J. et al. TACI expression is associated with a mature bone marrow plasma cell signature and C-MAP overexpression in human myeloma cell lines. Haematologica 92, 803–811 (2007).
46. 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, 1691–1697 (2009).
47. 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. Biomark. Prev. 19, 1360–1366 (2010).
48. Anderson, C.A. et al. Data quality control in genetic case-control association studies. Nat. Protoc. 5, 1564–1573 (2010).
49. 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).
50. Abecasis, G. R. et al. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073 (2010).
51. Huang, J. et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat. Commun. 6, 8111 (2015).
52. Sach, M., McVean, G. & Donnelly, P. A multiple imputation method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906–913 (2007).
53. Clayton, D. G. et al. Population structure, differential bias and genome control in a large-scale, case-control association study. Nat. Genet. 37, 1243–1246 (2005).
54. Liu, J. Z. et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat. Genet. 42, 436–440 (2010).
55. 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).
56. Nében, 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 stem cell transplantation. Haematologica 95, 1150–1157 (2010).
57. Walker, B. A. et al. A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. Blood 116, e56–e65 (2010).
58. Cairns, J. et al. CHICAGO: robust detection of DNA looping interactions in Hi-C data. Genome Biol. 17, 127 (2016).
59. Fernandez, J. M. et al. The BLUEPRINT data analysis portal. Cell Syst. 3, 491–496 (2016).
60. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods 9, 215–216 (2012).
61. 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, 1191–1198 (2012).
62. de Souza, N. The ENCODE project. Nat. Methods 9, 1046 (2012).
63. Halvarsson, B. M. et al. Direct evidence for a polygenic etiology in familial multiple myeloma. Blood Adv. 1, 619–623 (2017).
64. Via, M., Gignoux, C. & Burchard, E. G. The 1000 genomes project: new opportunities for research and social challenges. Genome Med. 2, 112 (2010).
Molly Went, Amit Sud, Asta Försti, Britt-Marie Halvarsson, Niels Weinhold, Scott Kimber, Mark van Duin, Guðmar Thorleifsson, Amy Holroyd, David C. Johnson, Ni Li, Giulia Orlando, Philip J. Law, Mina Ali, Bowang Chen, Jonathan S. Mitchell, Daniel F. Gudbjartsson, Owen W. Stephens, Uta Bertsch, Peter Broderick, Chiara Campo, Obul R Bandapalli, Hermann Einsele, Walter A. Gregory, Urban Gullberg, Jens Hillengass, Per Hoffmann, Graham H. Jackson, Karl-Heinz Jöckel, Ellinor Johnsson, Sigurður Y. Kristinsson, Ulf-Henrik Mellqvist, Hareth Nahi, Douglas Easton, Paul Pharoah, Alison Dunning, Julian Peto.

**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., J.S.M. 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 hybridisation analyses. The PRACTICAL consortium, D.E., P.P., A.D., J.P., F.C., R.E., Z. K.-J., K.M. and N.P. provided control samples for the UK OncoArray GWAS. H.G., U.B., J.H., J.N. and N.W. coordinated and managed Heidelberg samples. C.L. and H.E. coordinated and managed UMe/Wurzburg samples. A.F. coordinated genomic typing. G.H. and M.M.N. performed GWAS of German cases and controls. H.T., B.C. and M.I.d.S.F. carried out statistical analysis. K.I.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. N.N. from Bonn and K.-H.I. 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. E.J., A.-K.W., U.-H.M., H.N., A.V., N.E.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/EMN202 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.

**Additional information**

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04989-w.

**Competing interests**

G.T., D.F.G., T.R., K.S. and U.T. are employed by deCode Genetics/Amgen Inc. The remaining authors declare no competing interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**Publisher’s note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2018
Federico Canzian, Anthony Swerdlow, Rosalind A. Eeles, ZSofia Kote-Jarai, Kenneth Muir, Nora Pashayan, The PRACTICAL consortium, Jolanta Nickel, Markus M. Nöthen, Thorunn Rafnar, Fiona M. Ross, Miguel Inacio da Silva Filho, Hauke Thomsen, Ingemar Turesson, Annette Vangsted, Niels Frost Andersen, Anders Waage, Brian A. Walker, Anna-Karin Wihlborg, Annemiek Broyl, Faith E. Davies, Unnur Thorsteinsdottir, Christian Langer, Markus Hansson, Hartmut Goldschmidt, Martin Kaiser, Pieter Sonnevald, Kari Stefansson, Gareth J. Morgan, Kari Hemminki, Björn Nilsson, Richard S. Houlston.

The PRACTICAL consortium

Brian E. Henderson, Christopher A. Haiman, Sara Benlloch, Fredrick R. Schumacher, Ali Amin Al Olama, Sonja I. Berndt, David V. Conti, Fredrik Wiklund, Stephen Chanock, Victoria L. Stevens, Catherine M. Tangen, Jyotsna Batra, Judith Clements, Henrik Gronberg, Johanna Schleutker, Demetrios Albanes, Stephanie Weinstein, Alicja Wolk, Catharine West, Lorelei Mucci, Géraldine Cancel-Tassin, Stella Koutros, Karina Dalsgaard Sorensen, Eli Marie Grindedal, David E. Neal, Freddie C. Hamdy, Jenny L. Donovan, Ruth C. Travis, Robert J. Hamilton, Sue Ann Ingles, Barry Rosenstein, Yong-Jie Lu, Graham G. Giles, Adam S. Kibel, Ana Vega, Manolis Kogevinas, Kathryn L. Penney, Jong Y. Park, Janet L. Stanford, Cezary Cybulski, Børge G. Nordestgaard, Hermann Brenner, Christiane Maier, Jeri Kim, Esther M. John, Manuel R. Teixeira, Susan L. Neuhausen, Kim De Ruyck, Azad Razack, Lisa F. Newcomb, Davor Lessel, Radka Kavea, Nawaid Usmani, Frank Claessens, Paul A. Townsend, Manuela Gago-Dominguez, Monique J. Roobol, Florence Menegaux, Kay-Tee Khaw, Lisa Cannon-Albright, Hardev Pandha & Stephen N. Thibodeau.
