



ARTICLE

<https://doi.org/10.1038/s42003-019-0329-2>

OPEN

Genome-wide interaction and pathway-based identification of key regulators in multiple myeloma

Subhayan Chattopadhyay ^{1,2}, Hauke Thomsen¹, Pankaj Yadav¹, Miguel Inacio da Silva Filho¹, Niels Weinhold^{3,4}, Markus M. Nöthen^{5,6}, Per Hoffman^{5,6,7}, Uta Bertsch³, Stefanie Huhn³, Gareth J. Morgan⁴, Hartmut Goldschmidt^{3,8}, Richard Houlston ^{9,10}, Kari Hemminki^{1,11} & Asta Försti^{1,11}

Inherited genetic susceptibility to multiple myeloma has been investigated in a number of studies. Although 23 individual risk loci have been identified, much of the genetic heritability remains unknown. Here we carried out genome-wide interaction analyses on two European cohorts accounting for 3,999 cases and 7,266 controls and characterized genetic susceptibility to multiple myeloma with subsequent meta-analysis that discovered 16 unique interacting loci. These risk loci along with previously known variants explain 17% of the heritability in liability scale. The genes associated with the interacting loci were found to be enriched in transforming growth factor beta signaling and circadian rhythm regulation pathways suggesting immunoglobulin trait modulation, T_H17 cell differentiation and bone morphogenesis as mechanistic links between the predisposition markers and intrinsic multiple myeloma biology. Further tissue/cell-type enrichment analysis associated the discovered genes with hemic-immune system tissue types and immune-related cell types indicating overall involvement in immune response.

¹Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany. ²Faculty of Medicine, University of Heidelberg, Heidelberg 69117, Germany. ³University Clinic Heidelberg, Internal Medicine V, Heidelberg 69117, Germany. ⁴Myeloma Institute, University of Arkansas for Medical Sciences, Little Rock 72205 AR, USA. ⁵Institute of Human Genetics, University of Bonn, Bonn 53127, Germany. ⁶Department of Genomics, Life & Brain Research Center, University of Bonn, Bonn 53127, Germany. ⁷Department of Biomedicine, University of Basel, Basel 4003, Switzerland. ⁸National Centre of Tumor Diseases, Heidelberg 69120, Germany. ⁹Division of Genetics and Epidemiology, The Institute of Cancer Research, London SW7 3RP, UK. ¹⁰Division of Molecular Pathology, The Institute of Cancer Research, London SW7 3RP, UK. ¹¹Center for Primary Health Care Research, Lund University, 205 02 Malmö, Sweden. Correspondence and requests for materials should be addressed to A.F. (email: A.Foersti@dkfz.de)

Multiple myeloma is the second most prevalent hematological malignancy with almost 31,000 estimated new diagnoses in the United States in 2018¹. Multiple myeloma, a B-cell neoplasm, is characterized by proliferation of clonal plasma cells in bone marrow. Familial aggregation of multiple myeloma suggests predisposition due to inherited genetic variation^{2,3}. Susceptibility to multiple myeloma and its genetic relationship with the related diseases, monoclonal gammopathy of unknown significance (MGUS), and amyloid light chain (AL) amyloidosis, have lately been established through genome-wide association studies (GWASs)^{4–6}. Although a total of 23 risk loci have been discovered predisposing to multiple myeloma, they are estimated to explain only about 16% of the heritability^{5,7}. Moreover, genetic heterogeneity among multiple myeloma tumors bears complication in characterization of genetic susceptibility to multiple myeloma and in understanding of clinical consequences^{8,9}.

In addition to the linear association analysis, we have recently identified several inherited risk loci predisposing to MGUS through genome-wide genetic interaction¹⁰. To gain ample insight into genetic predisposition of multiple myeloma, we performed here the first genome-wide interaction study using two patient cohorts comprising a total of 3999 cases and 7266 controls. We extended the investigation with a subsequent meta-analysis of the two cohorts to increase the statistical power of detection. We also examined enrichment of expression of the identified genes in several tissue and cell types. Additionally, we performed gene set enrichment and pathway analyses to confer a biological understanding to our investigation. Collectively, our analyses support the hypothesis that genetic interaction plays a crucial role in multiple myeloma predisposition. The sentinel genes thus discovered are often expressed in tissues and cell lineages of hematopoietic system responsible for immune-modulation and they also influence inherited susceptibility to multiple myeloma through regulation of circadian rhythm and Smad-dependent TGF β pathways.

Results

Interacting chromosomal loci. Two quality controlled sets of genotyped data consisting 2282 cases and 5197 controls from the UK and 1717 cases and 2069 controls from Germany were subjected to pairwise interaction analysis accounting for 0.43 million and 0.52 million single-nucleotide polymorphisms (SNPs), respectively. Meta-analysis of associative linear interaction on transformed correlation statistics rendered 16 unique SNP pairs belonging to 16 exclusive chromosomal regions reaching genome-wide threshold of 5.0×10^{-10} (Fig. 1 and Supplementary Data 1).

The strongest meta-analyzed signal was provided by an interaction between rs7048811 at 9q21.31 (associated gene *GNAQ*) and rs7204305 at 16q24.1 (*IRF8*) (OR_{Meta} = 1.22; 95% CI = 1.12–1.32; $P = 1.3 \times 10^{-10}$, Supplementary Data 1). This interaction was consistent in both cohorts with a conservative level of significance (UK cohort: OR = 1.20, 95% CI = 1.08–1.33, $P = 7.0 \times 10^{-6}$; German cohort: OR = 1.24, 95% CI = 1.09–1.41, $P = 7.6 \times 10^{-6}$). The highest statistically significant OR was observed for the second most strong interaction signal between rs2167453 at 11p15.2 (*PDE3B*) and rs2734459 at 19q13.31 (*ZNF229*) (OR_{Meta} = 1.52, 95% CI = 1.33–1.73, $P = 1.3 \times 10^{-10}$).

Biological inference of the interacting chromosomal loci. Many of the risk SNPs identified, although showing promising genotypic interactions, are mapped to non-coding regions of the genome and possibly contribute to multiple myeloma etiology by affecting gene expression¹¹. In order to gain biological understanding of the newly identified interacting risk loci, we

interrogated expression quantitative trait locus (eQTL) data generated on malignant plasma cells obtained from patients of the German multiple myeloma trials. Strongest eQTL signals were observed by rs2167453 at 11p15.2 for cytochrome P450, family 2, subfamily R, polypeptide 1 (*CYP2R1*) and by rs923934 at 3q29 for family with sequence similarity 43, member A (*FAM43A*), both with $P_{eQTL} = 4.40 \times 10^{-5}$ (Table 1). Also the interacting partners of these SNPs served as eQTLs with a moderate signal, rs2734459 for *CLASRP*, *ZNF224*, and *APOE* and rs13201167 for *AKAP12* and *C6orf211*.

Summary-data-based Mendelian randomization (SMR) was employed to analyze pleiotropic effects between the GWAS signal and the cis-eQTL for genes residing within 1 Mb window of the interacting SNP loci to identify causal relationship between variants and disease phenotype via instrumentation of gene regulation¹². The strongest pleiotropic signal was observed at 4p15.33 by rs17362130 for RAS oncogene family member 28, *RAB28* ($P_{SMR} = 4.84 \times 10^{-3}$) and at 6p25.2 by rs6918808 for receptor (TNFRSF)-interacting serine/threonine kinase 1, *RIPK1* ($P_{SMR} = 5.04 \times 10^{-3}$, Table 1 and Fig. 2), respectively. Oncogenic ras family members are frequently mutated in multiple myeloma^{13,14}. *RIPK1* interacts with *RIPK3* to activate the necrosome complex that is responsible for instigation of several death receptors, which can induce apoptosis, necroptosis, or cell proliferation¹⁵. rs17362130 is also an eQTL for *NKX3-2* with a moderate signal ($P_{eQTL} = 2.11 \times 10^{-3}$) and rs6918808 for *SERPINB9*. *NKX3-2* is involved in skeletal development¹⁵. *SERPINB9* is a known inhibitor of granzyme and may mediate tumor immune evasion by apoptosis inhibition^{16,17}.

We investigated shared biological and information driven connections between the genes annotated to the variants by creating a genetic network. Unique annotations from the 16 interaction-identified variants along with the SMR identified causally related genes were subjected to network enrichment and a single batch of first order interacting genes based on data-mined enrichment index were additionally added to increase confidence of network associations (Fig. 3).

We applied Data-Driven Expression-Prioritized Integration for Complex Traits (DEPICT) for in silico analyses of enrichment of expression of genes annotated to the associated loci in tissues and cell types. To this end interaction-identified SNPs were clustered to 12 unique loci and were tested for significant excess expression of the corresponding genes in 209 Medical Subject Heading (MeSH) annotations against 37,427 microarrays procured in backend. Twenty-seven tissue or cell type annotations were discovered significant at a suggestive level ($P < 0.05$); 16 of them belonged to the hemic and immune system, two to the musculoskeletal system and one to the stomatognathic system (Fig. 4a), as well as six cell types related to hematopoietic system (Fig. 4b and Supplementary Data 2).

Biological inference of the GWAS-identified loci. Next, we investigated functional relationships among the previous GWAS-identified loci using the pathway analysis tool PASCAL with the bottom-up approach. To avoid possible complications arising from statistical convergence of the test statistic, we used sum of chi-square statistic to test for functional association against pathways extracted from REACTOME, KEGG, and BIOCARTA libraries (Supplementary Data 3). A total of 12 enriched pathways reached a global threshold of 0.0025 for the combined P -value. Three of the pathways, thus, detected were signaling cascades reflecting the activation status of the SMAD family proteins, as signal transducers for receptors of the cytokine TGF β represented by *SMAD2*, *SMAD3*, *SMAD4* heterotrimer regulates transcription, $P_{combined} = 5.70 \times 10^{-3}$, TGF β receptor signaling

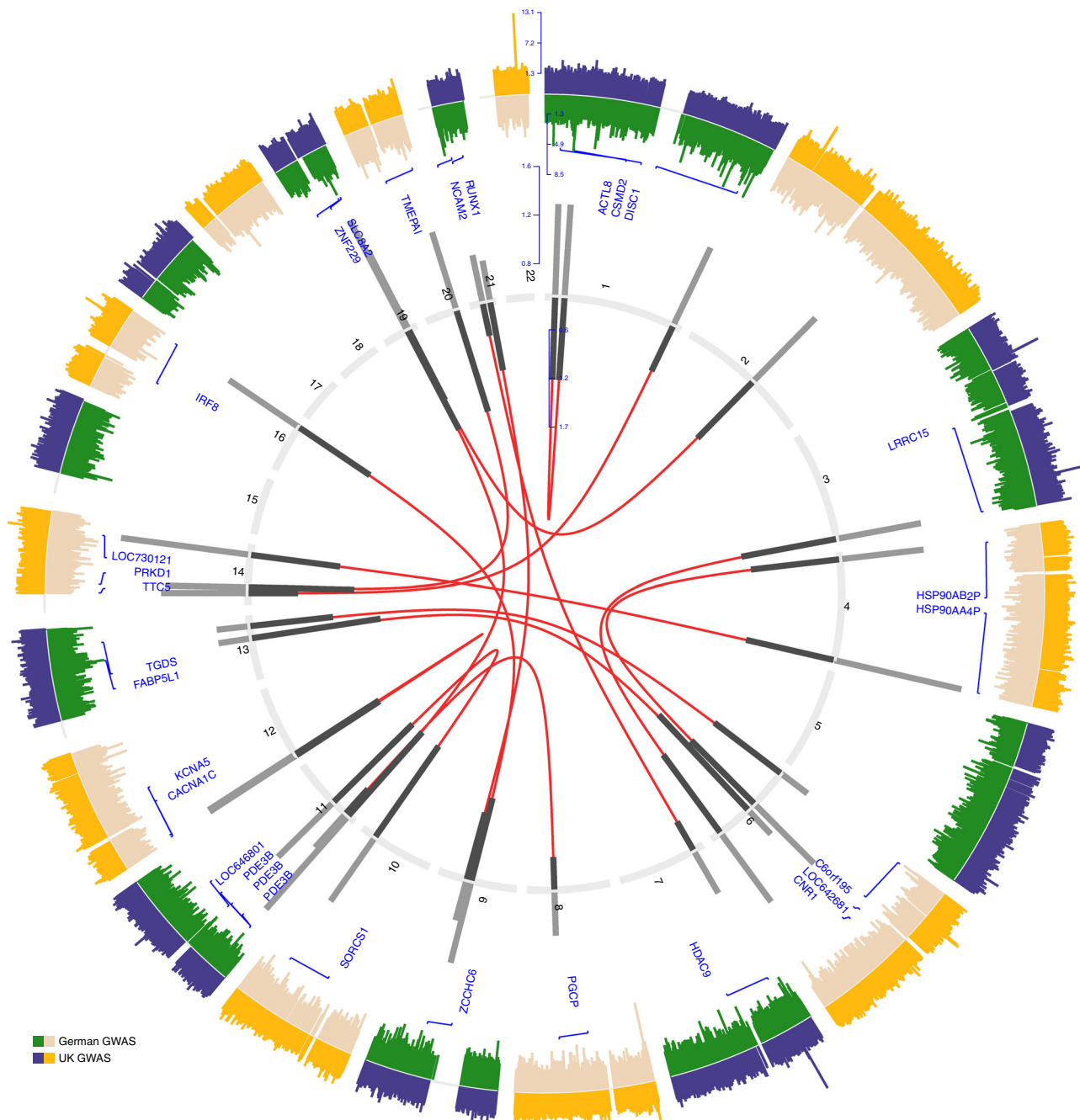


Fig. 1 Interaction analysis identifies 16 unique risk loci pairs. Circos plot of genome-wide association and significant interaction results for the identified paired risk loci. The two outer most panels display results from genome-wide association study on a Manhattan plot for autosomal variants on a negative log transformed scale. Inner numbered panel represents the chromosomes and effect-sizes of significant interacting pairs are plotted on bar charts from both samples (dark: German sample; light: UK sample). Interacting pairs are line joined in the inner most panel based on their chromosomal positions (NCBI build 19 human genome). Annotations of single-nucleotide polymorphisms to gene ids are displayed on the inner manhattan plot

activates *SMADs*, $P_{\text{combined}} = 8.60 \times 10^{-3}$ and transcriptional activity of *SMAD2*, *SMAD3*, *SMAD4* heterotrimer, $P_{\text{combined}} = 1.49 \times 10^{-2}$. Additionally, two pathways related to the regulation of circadian rhythms mediated by two nuclear receptor proteins retinoic acid receptor-related orphan receptor A (*RORA*) and Rev-Erba were identified: Circadian repression of expression by REV-ERBA, $P_{\text{combined}} = 5.52 \times 10^{-4}$ (Table 2) and *RORA* activates circadian expression. $P_{\text{combined}} = 2.13 \times 10^{-3}$. Also, modulation of *ALK* receptor tyrosine kinase activity was indicated with *ALK* pathway, $P_{\text{combined}} = 2.82 \times 10^{-3}$.

Heritability estimation. The previously identified 23 multiple myeloma risk SNPs were shown to account for 15.7% of the GWAS heritability, a relatively small fraction of the estimated 15.6% (± 4.7) net heritability explained by all common SNPs^{5,7,18}. The identified interacting loci explain an additional 1.3% of the GWAS heritability in the UK cohort (1.5% in the German cohort) in the liability scale, which brings the collectively explained GWAS heritability to a modest 17% (± 2.4). However, as the heritability estimates are based on individual SNP associations and do not take into account the interaction

Table 1 Genome-wide association study (GWAS) summary-data-based Mendelian randomization (SMR)

Probe	Gene name	Gene ID	Single-nucleotide polymorphism (SNP) ID	eQTL P-value	GWAS P-value	SMR P-value
9364_at	RAB28, member RAS oncogene family	RAB28	rs17362130	1.14E-03	3.68E-05	4.84E-03
8737_at	Receptor (TNFRSF)-interacting serine-threonine kinase 1	RIP1	rs6918808	1.23E-03	4.01E-05	5.04E-03
7289_at	Tubby like protein 3	TULP3	rs2238087	1.14E-03	2.58E-04	1.27E-02
808_at	Calmodulin 3 (phosphorylase kinase, delta)	CALM3	rs4802363	1.76E-03	1.99E-03	1.28E-02
11133_at	Kaptn (actin binding protein)	KPTM	rs4802363	1.98E-03	2.91E-03	1.30E-02
8605_at	Phospholipase A2, group IVC (cytosolic, calcium-independent)	PLA2G4C	rs4802363	1.72E-03	4.62E-03	1.33E-02
57820_at	Cyclin B1 interacting protein 1, E3 ubiquitin protein ligase	CCNB1IP1	rs10130942	1.41E-03	3.98E-03	1.86E-02
10082_at	Glypican 6	GPC6	rs17181808	1.06E-03	6.41E-04	1.86E-02
1690_at	coagulation factor C homolog, coxlin (Limulus polyphemus)	COCH	rs12436395	3.52E-04	1.88E-02	2.03E-02
120227_at	Cytochrome P450, family 2, subfamily R, polypeptide 1	CYP2R1	rs2167453	4.40E-05	2.56E-02	2.10E-02
579_at	NK3 homeobox 2	NKX3-2	rs17362130	2.11E-03	1.18E-03	2.72E-02
80759_at	KH homology domain containing 1	KHDC1	rs4706511	1.01E-03	5.49E-03	3.47E-02
10553_at	HIV-1 Tat interactive protein 2, 30 kDa	HTATIP2	rs10766743	1.85E-03	2.11E-03	3.60E-02
79624_at	Chromosome 6 open reading frame 211	C6orf211	rs13201167	4.40E-04	2.47E-03	3.65E-02
160897_at	G protein-coupled receptor 180	GPR180	rs17181808	4.40E-04	5.04E-03	3.66E-02
5272_at	Serpin peptidase inhibitor, clade B (ovalbumin), member 9	SERPINB9	rs6918808	1.01E-03	3.04E-03	3.80E-02
23483_at	TDP-glucose 4,6-dehydratase	TGDS	rs17181808	4.84E-04	6.32E-03	3.82E-02
440145_at	Mitotic spindle organizing protein 1	MZT1	rs17089906	2.64E-04	9.85E-03	4.20E-02
9590_at	A kinase (PRKA) anchor protein 12	AKAP12	rs13201167	2.16E-03	3.63E-03	4.27E-02
688_at	Kruppel-like factor 5 (intestinal)	KLF5	rs17089906	1.76E-04	2.01E-02	4.54E-02
7767_at	Zinc finger protein 224	ZNF224	rs2734459	7.04E-04	3.59E-03	4.66E-02
348_at	Apolipoprotein E	APOE	rs2734459	1.23E-03	5.19E-03	5.55E-02
81029_at	Wingless-type MMTV integration site family, member 5B	WNT5B	rs2238087	1.98E-03	6.45E-03	5.65E-02
404550_at	Chromosome 16 open reading frame 74	C16orf74	rs7204305	1.98E-03	6.74E-03	5.67E-02
11129_at	CLK4-associating serine/arginine rich protein	CLASRP	rs2734459	2.64E-04	1.32E-02	8.43E-02
131583_at	Family with sequence similarity 43, member A	FAM43A	rs923934	4.40E-05	1.90E-02	8.89E-02

term, the scope of interaction-identified heritability remains unanswered.

Discussion

We have performed the first genome-wide interaction study on multiple myeloma to date. We discovered 16 unique multiple myeloma risk locus pairs. Several of the discovered SNPs depicted eQTL effects for nearby genes and they were implicated in the networks and pathways relevant for multiple myeloma biology. We also demonstrated that genes annotated to the loci are highly expressed in tissues and cells of the hemic-immune system.

Interferon regulatory factor 8 (*IRF8*) together with G protein subunit alpha Q (*GNAQ*) were discovered to be the paired risk loci with highest statistical significance. *IRF8* is reported to be a risk locus for immunoglobulin trait modulation, whereas *GNAQ* is a guanine nucleotide-binding protein that regulates B-cell development and survival^{19,20}. *IRF8* has many functions in regulating innate immunity and immune cell development, including B- and T-cells, dendritic cells, and myeloid cells²¹. In early development, *IRF8* and *IRF4* function redundantly to regulate transition of pre-B-cells to maturing B-cells. In the germinal center development, the roles of these IRFs are complementary: *IRF8* directs early centroblast development, which is taken over by *IRF4* as centrocytes mature into plasma cells. *IRF8* induces activation-induced cytidine deaminase, which is a key enzyme catalyzing somatic hypermutations of plasma cells²¹. Similar to *IRF4*, *IRF8* transcriptional activity in multiple myeloma may also be related to differentiation of T helper (T_H) 17 cells, which have a regulatory effect on bone morphogenesis-related onset of multiple myeloma²². *IRF8* has been reported to act as an intrinsic transcriptional inhibitor of T_H 17 cells, at least partly through its physical interaction with retinoic acid receptor-related orphan receptor ROR γ ²³. As a confirmation of this signal, we identified enrichment of two circadian rhythm pathways regulated by two nuclear receptors, *RORA* and Rev-Erba, which are crucial for the development of T_H 17 cells²⁴. These findings together with our

previous identification of rs4487645 at 7p15.3, as a modulator of *IRF4* binding at an enhancer element of the c-Myc-interacting gene *CDCA7L* in multiple myeloma^{25–27}, support the role of the genetic variants in *IRF8* and its interacting partner in *GNAQ* in multiple myeloma susceptibility.

Another signaling cascade affecting immunoglobulin trait modulation, T_H 17 cell differentiation, and bone morphogenesis is the TGF β pathway²⁸, which was represented by three enriched pathways in our analysis. In multiple myeloma, enhanced bone resorption releases and activates TGF β , which is a potent inhibitor of osteoblast differentiation and mineralization²⁹. Our interaction analysis identified rs2834882 annotated to runt related transcription factor 1 (*RUNX1*) in interaction with rs2860107 at zinc finger CCHC-type containing 6 (*ZCCHC6*, alias *TUT7*). *RUNX* family transcriptional activities have been linked to TGF β -induced IgA class switching, which is involved in multiple myeloma pathogenesis^{19,30}. *RUNX* proteins also play a major role in cell differentiation, and *RUNX1* is specifically regulating hematopoiesis³¹. Germline mutations in *RUNX1* cause familial platelet disorder with propensity to myeloid malignancies and somatic loss of *RUNX1* function is related to several hematologic malignancies^{29,32}. *RUNX* transcription factors are integral components of signaling pathways enforced by TGF β family members including bone morphogenic proteins (BMPs). *RUNX1* and *RUNX2* are known modulators of BMP-2/7/9-induced osteoblast differentiation. *RUNX1* along with *RUNX2* is often found co-expressed in skeletal elements that regulate expression of BMP-2 and BMP-9³³. *RUNX2* regulatory activity in osteoblast differentiation is also regulated by transcription factor NKX3–2, whose expression was modulated by the sentinel SNP rs17362130 (Table 1)¹⁶. Additionally, *ZCCHC11* and *ZCCHC6* TUTase inhibitors are being investigated as potential agents for targeted therapy³⁴.

Contextually in multiple myeloma, TGF β induces differentiation arrest in osteoblasts, increases osteoclastogenesis, promotes angiogenesis, and suppresses host immunity in bone marrow microenvironment to create the so called multiple myeloma

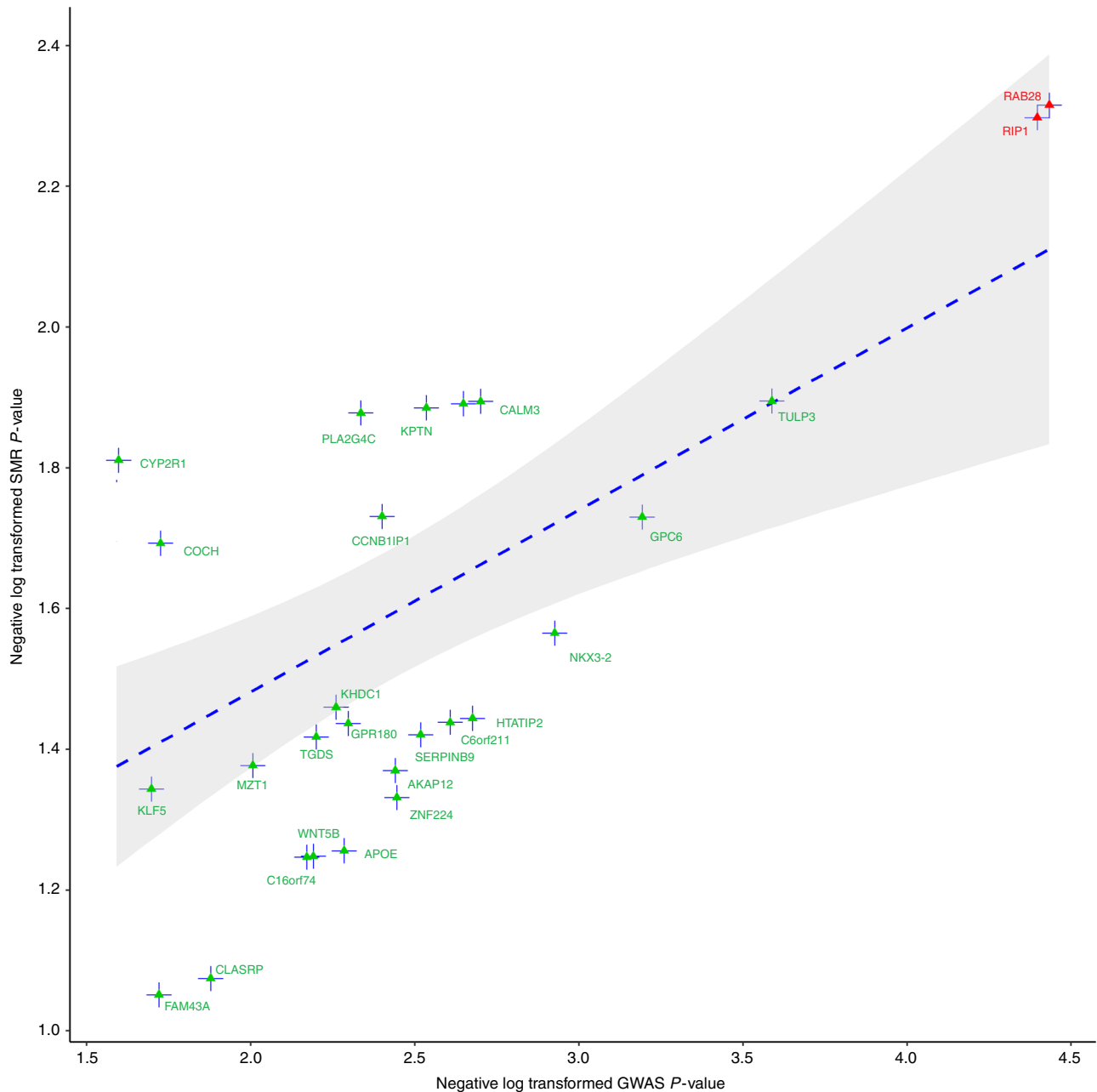


Fig. 2 Summary-data-based Mendelian randomization analysis of interaction detected multiple myeloma risk loci and gene expression in plasma cells. Negative log transformed P -values are plotted from GWAS against that of SMR identified causal cis-eQTLs at suggestive level. Top two significant elements are annotated in red. The blue line represents fitted linear regression representing linear association and the shaded region encompasses 95% confidence interval

niche, thus enhancing multiple myeloma cell growth and survival²⁹. TGF β -activated transcription factors, SMADs also interact with chromatin binding proteins HDAC1 and HDAC2. *HDAC1* is a class I histone deacetylase gene and multiple myeloma patients with high protein levels of HDAC1 were shown to have poor progression-free and overall survival³⁵. Moreover, inhibition of HDAC1 is demonstrated to induce multiple myeloma cell death³⁶. We noted a significant interaction between a class II HDAC family member, *HDAC9* and neural cell adhesion molecule 2, *NCAM2*. Deregulation of *HDAC9* in cells of lymphoid lineage is believed to induce B-cell lymphoproliferative disorders including Waldenström macroglobulinemia and is associated with general poor prognosis in cancer^{37,38}. *HDAC9* is also hypothesized to be responsible for lymphomagenesis by

regulating growth and survival related pathways and by modulating of BCL6 and p53 tumor suppressor activity³⁸. In germinal cells, it is shown to be co-expressed with BCL6, a therapeutic target for multiple myeloma³⁹. Controlled cell cycle is critical for normal cellular growth, and its deregulation may possibly stimulate carcinogenesis.

HDACs are also shown to have role in transcriptional activity of NKX3-2, one of the eQTL targets of our study. It has been shown that BMP and SMAD signaling modulates the activity of NKX3-2 in a BMP-dependent manner by promoting NKX3-2 binding with SMAD1/4 and HDAC/SIN3A corepressor complex⁴⁰. As HDAC inhibitors in general pose a vital role in cell cycle arrest induction and activation of intrinsic apoptotic mechanism, our observation leads to speculation that a common

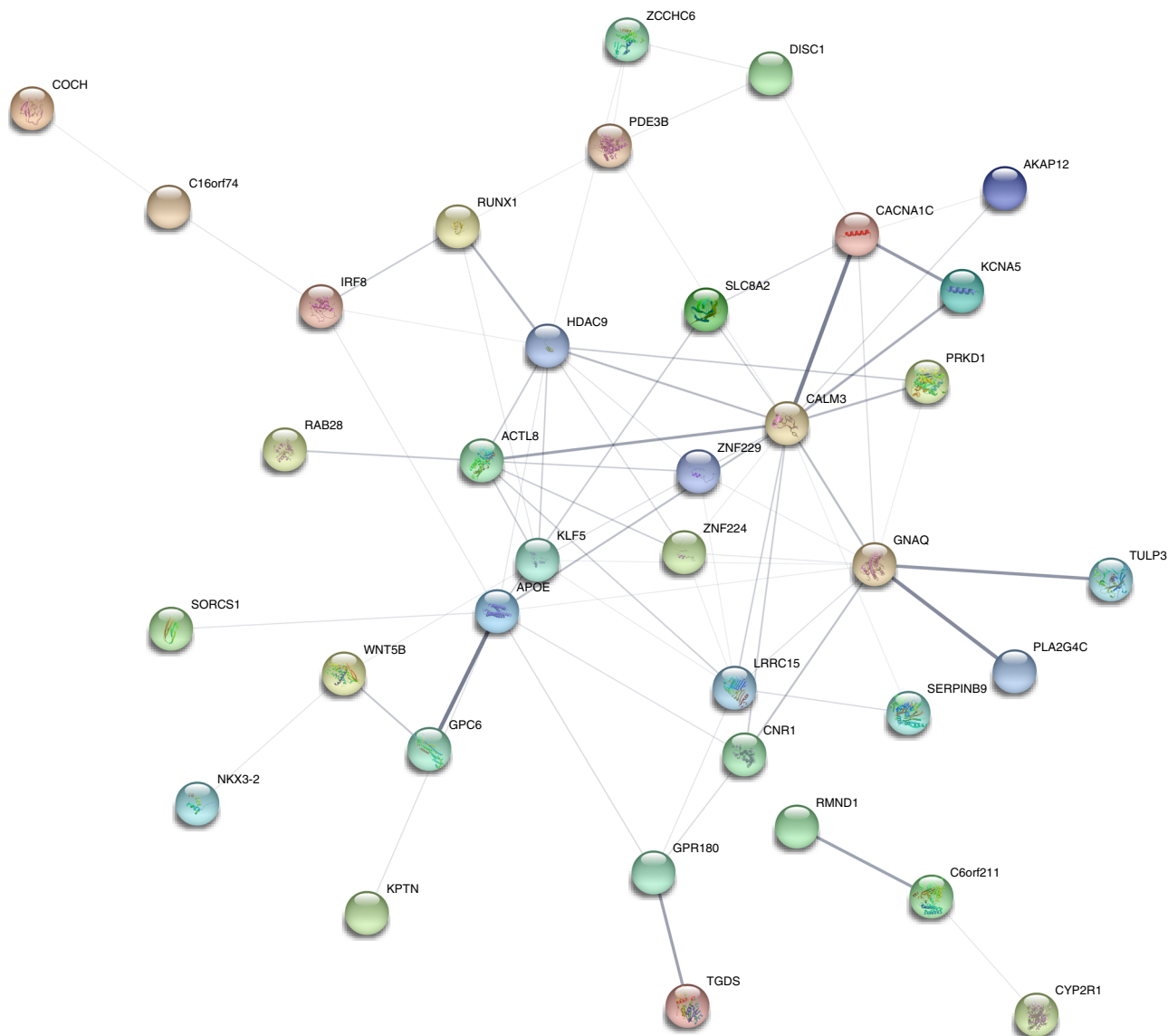


Fig. 3 Genetic network enrichment plot. All nodes represent direct annotations of interaction-identified elements or first order interactions. Thickness of the edges represent enrichment robustness between connecting nodes based on empirical evidence gathered from curated database (cyan), experimentally determined (magenta), gene neighborhood (forest green), gene fusion (red), gene co-occurrence (navy blue), text mining (lawn green), co-expression (black), protein homology (lavender indigo). Node color signifies different/shared protein functionality. Additional nodes are considered based on prediction score ≥ 0.99

variation in 7p21.1 may predispose to multiple myeloma progression.

The recent meta-analyses have pointed out apoptosis and autophagy, B-cell and plasma cell development, cell cycle regulation and genomic stability, chromatin remodeling and immunoglobulin production as the main pathways deregulated by the identified multiple myeloma susceptibility loci^{5,7}. We identified causally related genes implicated in apoptosis, such as *RIPK1* and *SERPINB9*. Among the interacting loci we identified genes involved in B-cell development and immunoglobulin production, such as *GNAQ* and *IRF8* and the TGF β pathway and genes modifying the chromatin state, such as *HDAC9*. As TGF β signaling is modified by ubiquitination and deubiquitination⁴¹, our study also support the importance of ubiquitin-proteasome signaling in multiple myeloma, which was highlighted by the meta-analysis together with the mechanistic target of rapamycin (mTOR) signaling as targets for already approved drugs in multiple myeloma⁷.

In conclusion, our findings provide further evidence that multiple myeloma is a genetically heterogeneous disease with inherited genetic susceptibility loci contributing excess risk via regulation of an assortment of regulatory networks and pathways. The two major signaling cascades we identified, TGF β signaling through its signal transducers SMADs and circadian rhythm regulation by *RORA* and Rev-Erba, integrate immunoglobulin trait modulation, T_H17 cell differentiation, and bone morphogenesis, and may provide a mechanistic link between the predisposition markers and intrinsic multiple myeloma biology.

Methods

Ethics. Patient samples and relevant clinico-pathological information was obtained conditional on written informed consent and was subject to approval of corresponding ethical review boards at respective study centers in accordance with the tenets of Declaration of Helsinki including Myeloma-IX trial by the Medical Research Council (MRC) Leukemia 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) and Ethical

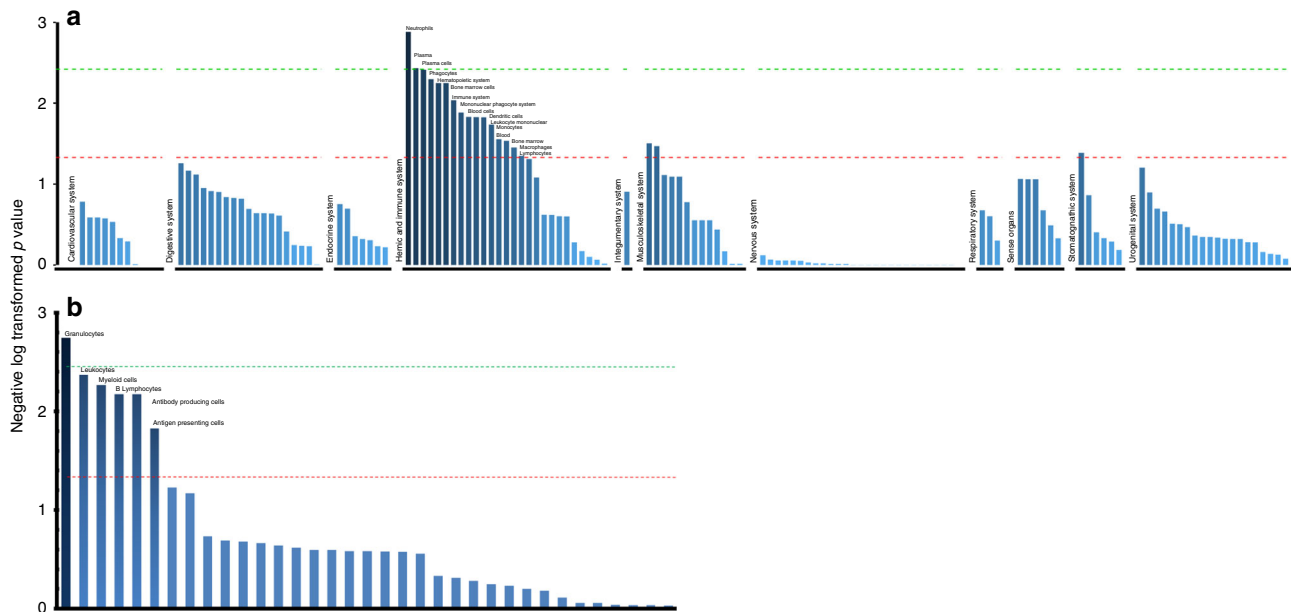


Fig. 4 Tissue and cell type enrichment plots. **a** Tissue enrichment identifies significant tissue types mostly affected with interaction-identified genes. **b** Cell type enrichment analysis identifies cells with observed expression regulation of the same candidates

Table 2 Pathway enrichment analysis with PASCAL detects 12 putative pathways related to multiple myeloma. Combined P -values are obtained with Brown's method. P_X denotes P -value obtained from interaction test of set X

Database	Pathway	P_{Ger}	P_{UK}	P_{Meta}	$P_{Combined}$
REACTOME	Circadian repression of expression by REV-ERBA	3.50E-04	1.45E-01	4.16E-03	5.52E-04
	APOBEC3G mediated resistance to HIV infection	5.79E-02	1.74E-03	2.09E-03	1.02E-03
	RORA activates circadian expression	1.24E-03	1.83E-01	1.20E-02	2.13E-03
	Deposition of new CENP-A containing nucleosomes as the centromere	7.00E-02	7.49E-03	3.82E-03	4.48E-03
	SMAD2 SMAD3 SMAD4 heterotrimer regulates transcription	8.83E-02	7.81E-03	1.88E-02	5.70E-03
	TGF β receptor signaling activates SMADs	1.73E-02	6.39E-02	4.38E-03	8.60E-03
	GABAA receptor activation	2.36E-02	6.27E-02	1.62E-02	1.11E-02
	Iron uptake and transport	4.84E-02	4.20E-02	8.91E-03	1.46E-02
	Transcriptional activity of SMAD2 SMAD3 SMAD4 heterotrimer	9.53E-02	2.18E-02	4.15E-02	1.49E-02
	Purine salvage	8.82E-02	2.51E-02	3.71E-02	1.57E-02
	Apoptosis induced DNA fragmentation	1.76E-02	1.29E-01	2.32E-02	1.60E-02
	BIOCARTA	ALK pathway	9.49E-03	3.28E-02	3.12E-02

Commission of medical faculty, University of Heidelberg (229/2003, S-337/2009, AFmu-119/2010).

Genome-wide association studies. Diagnosis of multiple myeloma (ICD-10 C90.0) adhered to the guidelines established by World Health Organization. Samples retrieved from all subjects were either before treatment or at presentation.

The UK GWAS⁵ consisted of 2282 cases (1755 male (post quality control (QC)) recruited through the UK MRC Myeloma-IX and Myeloma-XI trials (ISRCTN68454111: Myeloma- X <http://www.isrctn.com/search?q=ISRCTN68454111> and ISRCTN49407852: Myeloma- XI <http://www.isrctn.com/search?q=ISRCTN49407852>). DNA was extracted from EDTA-venous blood samples (90% before chemotherapy) and genotyped using Illumina Human OmniExpress-12 v1.0 arrays (Illumina). Controls were recruited from publicly accessible data generated by the Welcome Trust Case Control Consortium (WTCCC) from the 1958 Birth Cohort (58C; also known as the National Child Development Study) and National Blood Service. The control population comprised of 5197 individuals (2628 male (post QC)). Genotyping of these controls was conducted using Illumina Human 1-2 M-Duo Custom_v1 Array chips (www.wtccc.org.uk).

The German GWAS⁵ comprised 1717 cases (981 male (post QC)); mean age at diagnosis: 59 years). The cases were ascertained by the German-Speaking Multiple Myeloma Multicenter Study Group (GMMG) coordinated by the University Clinic, Heidelberg (ISRCTN06413384: GMMG-HD3 <http://www.isrctn.com/search?q=ISRCTN06413384>; ISRCTN64455289: GMMG-HD4 <http://www.isrctn.com/search?q=ISRCTN64455289>; and ISRCTN05745813: GMMG-MM5 <http://www.isrctn.com/search?q=ISRCTN05745813>).

www.isrctn.com/search?q=ISRCTN05745813). DNA was prepared from EDTA-venous blood or CD-138-negative bone marrow cells (< 1% tumor contamination). Genotyping of these samples was performed using Illumina Human OmniExpress-12 v1.0 arrays (Illumina). For controls, we used genotype data on 2107 healthy individuals, enrolled into the Heinz Nixdorf Recall (HNR) study genotyped using either Illumina HumanOmni1-Quad_v1 or OmniExpress-12 v1.0 arrays. Out of the whole recruited control population, 2069 (1028 male) remained after QC.

Analysis of GWAS. Quality control of the GWAS data was performed according to predetermined benchmarks already published⁵. In summary, inclusion of samples was initially liable to successful genotyping of $\geq 95\%$ of the SNPs. Duplicates, first-degree relatives, and closely related individuals were removed with an identity-based-test (IBS) score ≥ 0.80 . Genetic heterogeneity was assessed with principal component analysis using dissimilarity measure calculated with our SNP panel and genome-wide IBS distances in reference to the HapMap samples. In each of the samples, SNPs having a minor allele frequency < 0.01 and call rate of $< 95\%$ were filtered out. SNPs were also excluded subject to departure from Hardy-Weinberg equilibrium at $P < 1 \times 10^{-5}$ in controls.

Genome-wide interaction study. Analyses were primarily undertaken using PLINK (v1.09), CASSI (v3), METAINTER, and R (v3.4.0) software. The interaction between each SNP pair and the risk of multiple myeloma was assessed with Pearsonian product moment correlation coefficient-based test inspired by Welk-

Ziegler statistics given by the formula⁴²:

$$T_{WZ_{\text{case/control}}} = \frac{(r_A - r_N)^2}{\text{Var}(r_A) + \text{Var}(r_N)}$$

Where r is Pearsonian correlation coefficient statistics defined by Wellek and Ziegler⁴³. r_A and r_N , respectively, represent the statistics calculated among cases and controls separately. To this end, we used CASSI software. Genomic resolution of the whole interaction test space was deflated with default predefined control option where all the variants having weaker signal (single marker association $P > 1.0 \times 10^{-3}$) were excluded¹⁰. We performed the interaction test in the German and UK cohorts separately and meta-analyzed the results to strengthen the signals from co-occurring interacting pairs. To conduct meta-analysis METAINTER software was employed assuming a fixed effects model. Gamma approximated negative sum of log transformed interaction statistics from each of the two sets were considered as the test statistic for each variant pair and was tested with a weighted Chi-square statistic with four degrees of freedom⁴⁴. Odds ratio and associated 95% confidence intervals were calculated with unconditional logistic regression with independence assumption among each component of SNP pairs.

Expression quantitative trait loci analysis. Investigation of true regulatory effects of the SNPs identified with the interaction study was undertaken by analyzing eQTL data obtained from malignant plasma cells of 665 multiple myeloma patients (389 male) enrolled in the German multiple myeloma trials conducted in Heidelberg University clinic. CD-138 purified plasma cells were used for gene expression profiling using Affymetrix U133 2.0 plus arrays. The expression data was submitted to Gene Expression Omnibus (E-MTAB-2299). All analyses were undertaken with R software. GC-RMA was used to normalize the expression data and genes with transformed \log_2 expression < 3.5 in at least 95% of the samples were excluded from further consideration. With exclusive consideration of autosomal genes, 9722 genes remained after QC. We investigated the correlative relationship between the identified individual risk SNPs within 1 Mb window (cis-eQTL analysis), which narrowed the candidates to a set of 239 genes. A Holm-Sidak corrected level of significance for discovery was determined at < 0.0002 i.e., $\left[1 - (1 - 0.05)^{1/239} \right]$ on 239 probes corresponding to all the variants. Robust regression on a transformed Huber function was employed to model the qualitative traits as it warrants higher detection power in moderately contaminated sample⁴⁵. To avoid singularity of the argument space, variants in high linkage disequilibrium were discarded from consideration.

To extend the investigation of relation between SNP genotype and expression levels of genes and to identify causal candidates rather than mere associative pairings, we adapted SMR analysis as per Zhu et al.¹². In summary, if we nominate effect size of a differentially expressed gene X on coherence of a phenotype Y to be β_{XY} and consider the SNP genotypes to be the instrumental variable actively regulating both gene expression and the phenotype, then we can linearly estimate β_{XY} by comparative effect-sizes.

$$\hat{\beta}_{XY} = \frac{\hat{\beta}_{ZY}}{\hat{\beta}_{ZX}}$$

Where $\hat{\beta}_{ZY}$ is the estimated effect size of genetic factor on the phenotype, which is assessed as GWAS effect size and $\hat{\beta}_{ZX}$ is that of the genetic factor of the expression levels of the genes, i.e., the eQTL effect size. We need not distinguish pleiotropic effect from high linkage co-occurrence since the SNPs in linkage disequilibrium demonstrated equal effect size. Thus, reliability of causal genes was tested with the approximated SMR statistic against χ^2_1 .

Network enrichment. A protein–protein interaction confidence network was formulated with STRING (v10.5, 04/18/2018). Interactions between two proteins were calculated based on the likelihood confidence of an edge between the two nodes and was transposed to a scale of 0 to 1 (1 representing high confidence). We built our network with the genes annotated by the interaction-discovered SNPs and eQTL analysis; in addition to that, first batch of first-degree predicted interactive nodes were included given a confidence score > 0.99 . Erroneous discovery was restricted at 10%, which rendered a protein–protein interaction network index $P < 0.0054$ (observed number of interactions were tested against expected number of interactions with chi-square statistic with one degree of freedom).

Pathway enrichment. Initial in silico pathway enrichment was performed with the PASCAL tool interrogating the GWAS obtained summary statistics⁴⁶. To create mapping of genes and single entity gene-fusions with PASCAL, we considered all genes within 20 kb upstream and downstream to an index SNP and fused all the corresponding/flanking genes together when the genes were found affecting same pathway(s). Sum of chi-square statistics with individual one degree of freedom was computed by summing over association statistics corresponding to each pathway. Enrichment scores of individual pathways were subsequently obtained by a test

assuming chi-square distribution with degrees of freedom equal to the cardinality of fused gene sets.

Tissue and cell type enrichment. DEPICT was employed to analyze tissue and cell type enrichment that predicts differential regulation of the selected loci on any of the Medical Subject Heading (MeSH) annotations⁴⁷. To this end, 209 such annotations were tested for 37,427 inbuilt backend human microarrays on the Affymetrix HGU133a2.0 array platform. The tissue/cell type enrichment is thus performed on the normalized expression matrix after subjecting it to user selected dimension reduction criteria. SNP pairs discovered with interaction test represented 12 unique mapped regions against which the enrichment was tested, hence we tested against a conservative threshold of significance at negative log transformed P -value of 2.37 correcting for multiple testing, which retains the false discovery rate at $< 5\%$ ⁴⁸.

Heritability analysis. As hypothesized by earlier studies, heritability estimates of complex diseases with polygenic origin are more robust with lifetime risk compared to population prevalence⁴⁹. Following this notion, lifetime risk of multiple myeloma was assumed (0.007 for UK and 0.006 for German population; <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/myeloma>; https://www.krebsdaten.de/Krebs/EN/Home/homepage_node.html) to ascertain heritability of multiple myeloma explained by the risk SNPs discovered in the two different cohorts separately. Principal components were included to adjust for inflation as covariates. Genome-wide Complex Trait Analysis was used to estimate the genetic variance ascribable to the identified loci at a liability scale^{50,51}.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

SNP genotyping data that support the findings of this study have been deposited in Gene Expression Omnibus with accession codes GSE21349, GSE19784, and GSE15695; in the European Genome-phenome Archive (EGA) with accession code EGAS00000000001; 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 EMBL-EBI with accession code E-MTAB-2299. The remaining data are contained within the paper and Supplementary Data or are available from the author upon request.

Received: 23 October 2018 Accepted: 29 January 2019

Published online: 04 March 2019

References

- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **68**, 7–30 (2018).
- Frank, C. et al. Search for familial clustering of multiple myeloma with any cancer. *Leukemia* **30**, 627 (2015).
- Frank, C., Fallah, M., Sundquist, J., Hemminki, A. & Hemminki, K. Population landscape of familial cancer. *Sci. Rep.* **5**, 12891 (2015).
- Thomsen, H. et al. Genomewide association study on monoclonal gammopathy of unknown significance (MGUS). *Eur. J. Haematol.* **99**, 70–79 (2017).
- Mitchell, J. S. et al. Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nat. Commun.* **7**, 12050 (2016).
- da Silva Filho, M. I. et al. Genome-wide association study of immunoglobulin light chain amyloidosis in three patient cohorts: comparison with myeloma. *Leukemia* **31**, 1735–1742 (2017).
- Went, M. et al. Identification of multiple risk loci and potential regulatory mechanisms influencing susceptibility to multiple myeloma. *Nat. Commun.* **9**, 3707 (2018).
- Merz, M. et al. Prognostic significance of cytogenetic heterogeneity in patients with newly diagnosed multiple myeloma. *Blood Adv.* **2**, 1–9 (2018).
- Lohr, J. G. et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell.* **25**, 91–101 (2014).
- Chattopadhyay, S. et al. Enrichment of B cell receptor signaling and epidermal growth factor receptor pathways in monoclonal gammopathy of undetermined significance: a genome-wide genetic interaction study. *Mol. Med.* **24**, 30 (2018).
- Khurana, E. et al. Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* **17**, 93–108 (2016).
- Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
- Neri, A. et al. Ras oncogene mutation in multiple myeloma. *J. Exp. Med.* **170**, 1715–1725 (1989).

14. Walker, B. A. et al. Mutational spectrum, copy number changes, and outcome: Results of a sequencing study of patients with newly diagnosed myeloma. *J. Clin. Oncol.* **33**, 3911–3920 (2015).
15. Nuges, A. L. et al. RIP3 is downregulated in human myeloid leukemia cells and modulates apoptosis and caspase-mediated p65/RelA cleavage. *Cell Death Dis.* **5**, e1384 (2014).
16. Caron, M. M. J. et al. Hypertrophic differentiation during chondrogenic differentiation of progenitor cells is stimulated by BMP-2 but suppressed by BMP-7. *Osteoarthr. Cartil.* **21**, 604–613 (2013).
17. Fritsch, K., Finke, J. & Grullich, C. Suppression of granzyme B activity and caspase-3 activation in leukaemia cells constitutively expressing the protease inhibitor 9. *Ann. Hematol.* **92**, 1603–1609 (2013).
18. Mitchell, J. S. et al. Implementation of genome-wide complex trait analysis to quantify the heritability in multiple myeloma. *Sci. Rep.* **5**, 12473 (2015).
19. Jonsson, S. et al. Identification of sequence variants influencing immunoglobulin levels. *Nat. Genet.* **49**, 1182–1191 (2017).
20. Misra, R. S. et al. G alpha q-containing G proteins regulate B cell selection and survival and are required to prevent B cell-dependent autoimmunity. *J. Exp. Med.* **207**, 1775–1789 (2010).
21. Zhao, G. N., Jiang, D. S. & Li, H. Interferon regulatory factors: at the crossroads of immunity, metabolism, and disease. *Biochim. Biophys. Acta* **1852**, 365–378 (2015).
22. Di Lullo, G., Marcatti, M. & Protti, M. P. Non-redundant roles for Th17 and Th22 cells in multiple myeloma clinical correlates. *Oncoimmunology* **5**, e1093278 (2016).
23. Ouyang, X. et al. Transcription factor IRF8 directs a silencing programme for TH17 cell differentiation. *Nat. Commun.* **2**, 314 (2011).
24. Kojetin, D. J. & Burris, T. P. REV-ERB and ROR nuclear receptors as drug targets. *Nat. Rev. Drug. Discov.* **13**, 197–216 (2014).
25. Broderick, P. et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat. Genet.* **44**, 58–61 (2011).
26. Weinhold, N. et al. The 7p15.3 (rs4487645) association for multiple myeloma shows strong allele-specific regulation of the MYC-interacting gene CDCA7L in malignant plasma cells. *Haematologica* **100**, e110–e113 (2015).
27. 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).
28. David, C. J. & Massague, J. Contextual determinants of TGFbeta action in development, immunity and cancer. *Nat. Rev. Mol. Cell Biol.* **19**, 419–435 (2018).
29. Takeuchi, K. et al. Tgf-Beta inhibition restores terminal osteoblast differentiation to suppress myeloma growth. *PLoS ONE* **5**, e9870 (2010).
30. Watanabe, K. et al. Requirement for runx proteins in IgA class switching acting downstream of TGF-β1 and retinoic acid signaling. *J. Immunol.* **184**, 2785 (2010).
31. Ichikawa, M. et al. A role for RUNX1 in hematopoiesis and myeloid leukemia. *Int. J. Hematol.* **97**, 726–734 (2013).
32. Sood, R., Kamikubo, Y. & Liu, P. Role of RUNX1 in hematological malignancies. *Blood* **129**, 2070–2082 (2017).
33. Ji, C. et al. RUNX1 plays an important role in mediating bmp9-induced osteogenic differentiation of mesenchymal stem cells line C3H10T1/2, murine multi-lineage cells lines C2C12 and MEFs. *Int. J. Mol. Sci.* **18**, 1348 (2017).
34. Lin, S. & Gregory, R. I. Identification of small molecule inhibitors of Zcchc11 TUTase activity. *RNA Biol.* **12**, 792–800 (2015).
35. Mithraprabhu, S., Kalfi, A., Chow, A., Khong, T. & Spencer, A. Dysregulated Class I histone deacetylases are indicators of poor prognosis in multiple myeloma. *Epigenetics* **9**, 1511–1520 (2014).
36. Mithraprabhu, S., Khong, T., Jones, S. S. & Spencer, A. Histone deacetylase (HDAC) inhibitors as single agents induce multiple myeloma cell death principally through the inhibition of class I HDAC. *Br. J. Haematol.* **162**, 559–562 (2013).
37. Sun, J. Y. et al. Histone deacetylase inhibitors demonstrate significant preclinical activity as single agents, and in combination with bortezomib in Waldenstrom's macroglobulinemia. *Clin. Lymphoma Myeloma & Leuk.* **11**, 152–156 (2011).
38. Gil, V. S. et al. Deregulated expression of HDAC9 in B cells promotes development of lymphoproliferative disease and lymphoma in mice. *Dis. Model Mech.* **9**, 1483–1495 (2016).
39. Hideshima, T. et al. Bcl6 as a novel therapeutic target in multiple myeloma (MM). *Blood* **114**, 124–124 (2009).
40. Kim, D. W. & Lassar, A. B. Smad-dependent recruitment of a histone deacetylase/Sin3A complex modulates the bone morphogenetic protein-dependent transcriptional repressor activity of Nkx3.2. *Mol. Cell. Biol.* **23**, 8704–8717 (2003).
41. Imamura, T., Oshima, Y. & Hikita, A. Regulation of TGF-β family signalling by ubiquitination and deubiquitination. *J. Biochem.* **154**, 481–489 (2013).
42. Ueki, M. & Cordell, H. J. Improved statistics for genome-wide interaction analysis. *PLoS Genet.* **8**, e1002625 (2012).
43. Wellek, S. & Ziegler, A. A genotype-based approach to assessing the association between single nucleotide polymorphisms. *Hum. Hered.* **67**, 128–139 (2009).
44. Vaitisakovich, T., Drichel, D., Herold, C., Lacour, A. & Becker, T. METAINTER: meta-analysis of multiple regression models in genome-wide association studies. *Bioinform. (Oxf., Engl.)* **31**, 151–157 (2015).
45. Rantalainen, M., Lindgren, C. M. & Holmes, C. C. Robust Linear Models for Cis-eQTL Analysis. *PLoS ONE* **10**, e0127882 (2015).
46. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
47. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
48. Geller, F. et al. Genome-wide association analyses identify variants in developmental genes associated with hypospadias. *Nat. Genet.* **46**, 957–963 (2014).
49. Lu, Y. et al. Most common 'sporadic' cancers have a significant germline genetic component. *Hum. Mol. Genet.* **23**, 6112–6118 (2014).
50. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Human Genet.* **88**, 76–82 (2011).
51. Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519 (2011).

Acknowledgements

In the UK, 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. This study made use of genotyping data on the 1958 Birth Cohort generated by the Wellcome Trust Sanger Institute (<http://www.wtccc.org.uk>). The German study was supported by the Dietmar-Hopp-Stiftung, Germany, the German Cancer Aid (110, 131), the German Ministry of Education and Science (CLIO-MMICS 01ZX1309), the German Research Council (DFG; Project SI 236/81, SI 236/9–1, 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.

Author contributions

S.C., K.H. and A.F. designed the study; S.C., H.T., P.Y. and M.L.d.S.F. analyzed the data; N.W., U.B., S.H. and H.G. provided the German MM samples and patient data; G.J.M. and R.H.S. provided the UK samples and patient data; P.H. and M.M.N. were responsible for the GWAS; S.C. wrote the first draft of the manuscript; A.F. critically reviewed the manuscript; all authors read and accepted the final version of the manuscript.


Additional information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s42003-019-0329-2>.

Competing interests: The 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.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

