

## ARTICLE OPEN



# Haplotype analysis identifies functional elements in monoclonal gammopathy of unknown significance

Hauke Thomsen<sup>1</sup>, Subhayan Chattopadhyay<sup>2</sup>, Niels Weinhold<sup>3</sup>, Pavel Vodicka<sup>4,5,6</sup>, Ludmila Vodickova<sup>4,5,6</sup>, Per Hoffmann<sup>7,8</sup>, Markus M. Nöthen<sup>7</sup>, Karl-Heinz Jöckel<sup>9</sup>, Borge Schmidt<sup>9</sup>, Roman Hajek<sup>10</sup>, Göran Hallmans<sup>11</sup>, Ulrika Pettersson-Kymmer<sup>12</sup>, Florentin Späth<sup>13</sup>, Hartmut Goldschmidt<sup>13,14</sup>, Kari Hemminki<sup>15</sup> and Asta Försti<sup>16,17</sup>✉

© The Author(s) 2024

Genome-wide association studies (GWASs) based on common single nucleotide polymorphisms (SNPs) have identified several loci associated with the risk of monoclonal gammopathy of unknown significance (MGUS), a precursor condition for multiple myeloma (MM). We hypothesized that analyzing haplotypes might be more useful than analyzing individual SNPs, as it could identify functional chromosomal units that collectively contribute to MGUS risk. To test this hypothesis, we used data from our previous GWAS on 992 MGUS cases and 2910 controls from three European populations. We identified 23 haplotypes that were associated with the risk of MGUS at the genome-wide significance level ( $p < 5 \times 10^{-8}$ ) and showed consistent results among all three populations. In 10 genomic regions, strong promoter, enhancer and regulatory element-related histone marks and their connections to target genes as well as genome segmentation data supported the importance of these regions in MGUS susceptibility. Several associated haplotypes affected pathways important for MM cell survival such as ubiquitin-proteasome system (*RNF186*, *OTUD3*), PI3K/AKT/mTOR (*HINT3*), innate immunity (*SEC14L1*, *ZBP1*), cell death regulation (*BID*) and NOTCH signaling (*RBPJ*). These pathways are important current therapeutic targets for MM, which may highlight the advantage of the haplotype approach homing to functional units.

*Blood Cancer Journal* (2024)14:140; <https://doi.org/10.1038/s41408-024-01121-8>

## INTRODUCTION

Germline disease genetics has historically applied linkage studies between family members to find susceptibility genes. With time families have become smaller which has reduced the statistical power of family-based linkage studies [1]. However, with increasing understanding of the human genome organization, genetic variants, including single nucleotide polymorphisms (SNPs), have been identified at specific locations throughout the genome. SNPs are inherited together as haplotypes containing a linked sets of alleles and allele-specific biological functions, including genes and their cis-regulatory elements [2]. In the study on global reference for human genetic variation 3.53 million SNPs were used for the European population to define haplotypes [3]. One goal for the development of dense linkage map with well-defined haplotypes over the human genome has been the possibility to map susceptibility genes assuming that some marker SNPs would be in high linkage with the functional variants, even if the SNPs lacked independent functions [4]. This has been the

driving idea behind the genome-wide association studies (GWASs) for which the SNP coverage increased hugely in a decade. The result has been that the current GWASs target increasing numbers of haplotypes and may identify functional variants in even rare haplotypes [5].

The importance of haplotypes has been well known in some cancers, particularly in lymphoma, where immune related haplotypes at the HLA locus are associated with disease risk or protection [6]. Other cancer-related applications of haplotypes include dating the origins of mutations based on conserved haplotypes [7]. However, direct use of haplotypes has not been popular in human germline genetics, in marked contrast to its role in animal breeding [8, 9]. Nevertheless haplotype-based genetic mapping has been used in some recent Swedish cancer studies where novel candidate genes have been detected [10, 11].

In the present study we apply a haplotype-based gene mapping approach to monoclonal gammopathy of unknown significance (MGUS) which is a precursor condition for multiple myeloma (MM)

<sup>1</sup>MSB Medical School Berlin, Hochschule für Gesundheit und Medizin, Berlin, Germany. <sup>2</sup>Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, Lund, Sweden. <sup>3</sup>Department of Internal Medicine V, University of Heidelberg, Heidelberg, Germany. <sup>4</sup>Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Prague, Czech Republic. <sup>5</sup>Institute of Biology and Medical Genetics, 1st Medical Faculty, Charles University, Prague, Czech Republic. <sup>6</sup>Faculty of Medicine and Biomedical Center in Pilsen, Charles University in Prague, Pilsen, Czech Republic. <sup>7</sup>Institute of Human Genetics, University of Bonn, Bonn, Germany. <sup>8</sup>Department of Biomedicine, University of Basel, Basel, Switzerland. <sup>9</sup>Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany. <sup>10</sup>Department of Hematooncology, University Hospital Ostrava and Faculty of Medicine, University of Ostrava, Ostrava, Czech Republic. <sup>11</sup>Department of Public Health and Clinical Medicine, Umea University, Umea, Sweden. <sup>12</sup>Clinical Pharmacology, Department of Pharmacology and Clinical Neuroscience, Umea University, Umea, Sweden. <sup>13</sup>Department of Diagnostics and Intervention, Cancer Center, Hematology, Umeå University, Umeå, Sweden. <sup>14</sup>National Centre of Tumor Diseases, Heidelberg, Germany. <sup>15</sup>Department of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany. <sup>16</sup>Hopp Children's Cancer Center (KITZ), Heidelberg, Germany. <sup>17</sup>Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany. ✉email: a.foersti@kitz-heidelberg.de

Received: 22 May 2024 Revised: 2 August 2024 Accepted: 6 August 2024

Published online: 20 August 2024

and other plasma cell malignancies such as Waldenström macroglobulinemia and immunoglobulin light chain (AL) amyloidosis. The MGUS populations included Swedish, German and Czech individuals and local controls who have been genotyped in previous GWASs [12–15]. For haplotype analysis diverse populations may be a disadvantage as the haplotype structure may show subtle differences. However, the advantage is the increase in sample size and the possibility to internally replicate the findings in three populations.

## METHODS

### Populations and original GWASs

We included three independent GWAS data sets of MGUS patients and controls from Germany, the Czech Republic and Sweden, including a total of 992 MGUS cases and 2910 controls, as described elsewhere [15]. Collection of patient samples and associated clinical information was undertaken with informed consent in accordance with the tenets of the Declaration of Helsinki. The study was approved by the Ethical Committees of the universities of Heidelberg, Ostrava and Umea. The diagnosis of MGUS was based on the internationally accepted criteria of a monoclonal protein concentration <30 g/L, <10% monoclonal plasma cells in the bone marrow, normal plasma calcium and normal renal function, no bone destruction and no anemia.

Details of the individual GWASs and imputation have been described earlier [15]. Shortly, the German MGUS set was genotyped with Illumina Human OmniExpress-12 v1.1 arrays and the control set with Illumina Human OmniExpress-12 v1.0 arrays. The Czech individuals were genotyped using Illumina HumanOmniExpressExome8v1.3 arrays. The Swedish samples were genotyped with Illumina Human Omni1-Quad BeadChips or OmniExpress-12 v1.0 arrays. General quality control assessment of genotyping was performed as previously described by Broderick et al. [16] and Chubb et al. [17]. SNPs with a minor allele frequency (MAF) < 5% were excluded due to statistical uncertainty at lower allele frequencies using the current population sizes. After imputation using data from the combined UK10K - 1000 Genomes Project (Phase 3, Oct 2014) with IMPUTE2 v2.3.2 [18], the genotyped and imputed SNP sets consisted of 4 015 889 SNPs for the Czech population, 5,305,950 SNPs for the German population, and 4,455,885 SNPs for the Swedish population. All genomic positions are given in NCBI Build 37/UCSC hg19 coordinates.

### Construction of haplotype blocks

Haplotype blocks were created for each population separately based on 13,777,724 SNPs for a total of 3902 subjects. For construction of the haplotype blocks, two steps were followed: 1) phasing of chromosomes through SHAPEIT.v2 [19] and 2) construction of haplotype blocks through the “Ghap” package in R [20].

For phasing, the genotype data was segregated based on chromosomes with creation of individual chromosome binary ped and map files by using the PLINK program [21]. Chromosome-wide genotypes were then phased again using SHAPEIT.v2.12 with 200 states and window size of 0.5 Mb. SHAPEIT has been acknowledged to be most efficient for estimating haplotypes from the genotypes and for creating the consecutive haps and sample file which were used to generate haplotype blocks through GHap.

Haplotype blocks were constructed based on linkage disequilibrium (LD), Length (Len), and the number of SNPs (nsnp) and corresponding phase, sample, and marker files for each chromosome were created. Genomic positions in haplotype blocks were based on the distance from the first SNP. After phasing, haplotype blocks were constructed using the GHap function, which generates HapBlocks based on sliding windows. Windows and the step size can be specified in markers. For each window, block coordinates are generated. Based on the fact that the average length of linkage disequilibrium (LD) blocks in human populations is about 20 kb [22, 23] the window size was defined to be 15 SNPs in order to cover a similar span. The sliding block along the genome was defined to be 2 SNPs during the call haplotypes. For each window, block coordinates were generated as the mean position of the start and end position of the corresponding haplotype window.

Overall, a total of 99,518,698 haplotype blocks were created using the above-mentioned methods based on LD, the length and the number of SNPs. In a consecutive step, we used GHap package in R to derive haplotype “alleles” from these haplotypes’ blocks. Haplotype “alleles” were then exported in the PLINK ped file format, where haplotype allele counts

0, 1, and 2 are recoded as NN, NH, and HH genotypes (H = haplotype allele and N = NULL = all other alleles), as if haplotypes were bi-allelic markers. A regular ped file was obtained with PLINK and the haplotypes presenting a MAF < 0.01 were filtered out. These data sets were further used to perform a haplotype-based association study.

### GWAS of the haplotypes

Haplotype alleles determined by GHap were used for association studies conducted by PLINK using logistic regression models with the covariate sex. The association analysis was first performed for each single population based on the haplotypes derived in the steps above.

### Meta-analysis

Meta-analysis was performed using PLINK by pooling the beta values and standard errors for haplotypes from each single population data set. The meta-analysis was performed with a random effects inverse-variance weighted logistic model. Cochran’s Q-statistic was calculated, to test for heterogeneity, and the I<sup>2</sup> statistic measured, to quantify the proportion of the total variation due to heterogeneity.

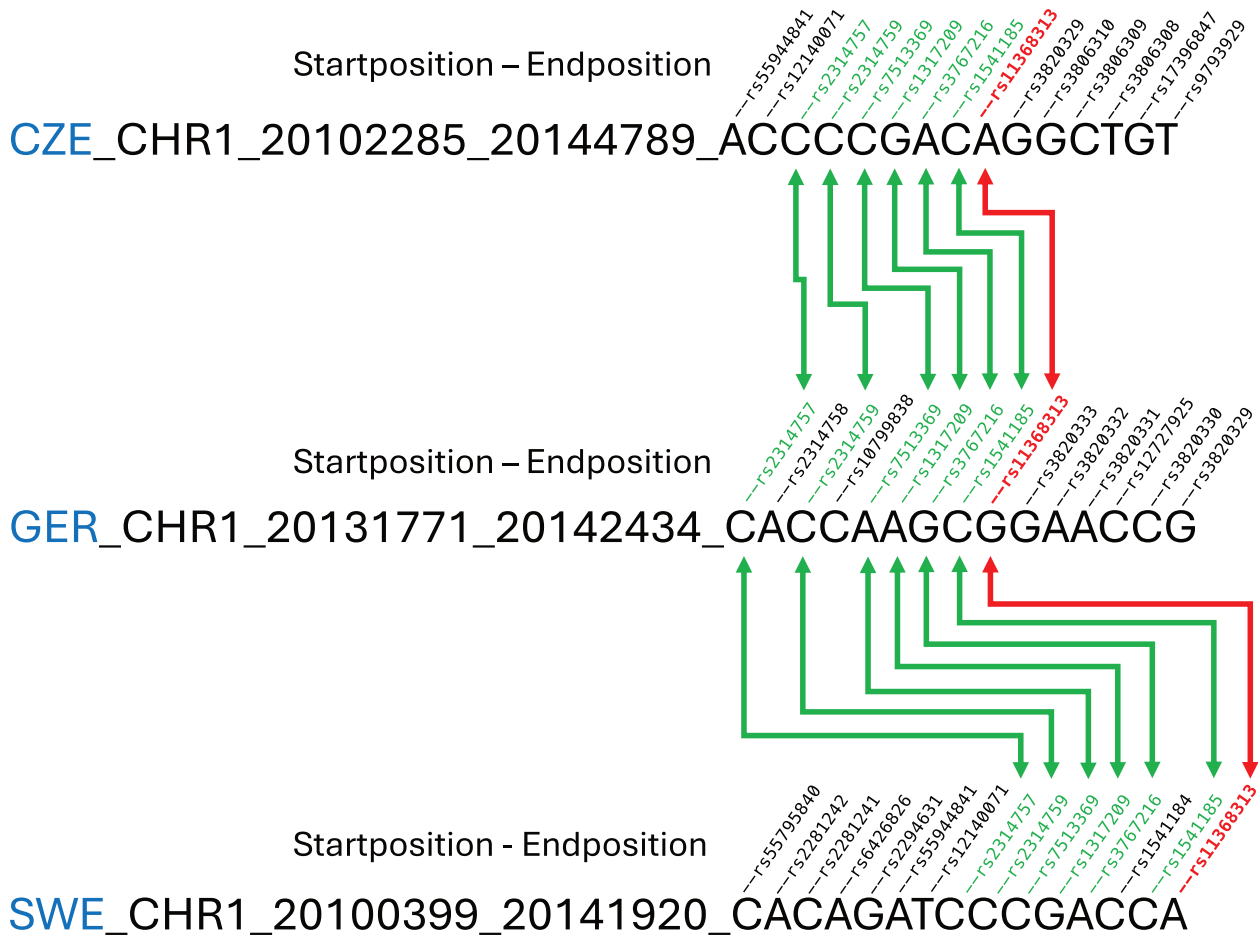
While normally the meta-analysis of GWAS data from different populations refers to common SNPs in the different data sets, this approach does not work for haplotypes, since only in rare cases exactly identical haplotypes could be derived in all populations. As an alternative, we therefore used a common SNP in the corresponding haplotypes in the three populations as an anchor point for the analysis (from now on referred to as “a joint SNP”). For a haplotype in each population, all SNPs forming the haplotype were assigned and a meta-analysis was performed for the joint SNP to represent each haplotype (Fig. 1). However, the joint SNP is only used as a common anchor and it represents the P value, odds ratio and frequency of the haplotype. In case of a large number of common SNPs in a haplotype, the most significant one was assigned as the joint SNP. Similarly, if common SNPs along the chromosome were assigned to several overlapping haplotypes, the haplotype represented by a joint SNP with the highest significance was selected.

### Gene identification in significant haplotype blocks

Haplotypes that after meta-analysis of the three populations were associated with the risk of MGUS at the genome-wide significance level  $p < 5 \times 10^{-8}$  and showed increased risk of MGUS in all three populations were investigated further using UCSC browser’s GRCh37/hg19 assembly (<https://genome.ucsc.edu>) [24, 25]. Overlapping haplotypes were defined by a joint SNP and the start and end positions of the haplotype regions were based on the lowest start and highest end positions, among the overlapping haplotypes. We included tracks of UCSC genes, Genome Segmentations and Histone modifications by ChIP-seq from ENCODE [26], CpG Islands, Transcription Factor ChIP-seq Clusters from ENCODE, Enhancers and promoters from GeneHancer (Double Elite) [27] and Interactions between GeneHancer regulatory elements and genes (Double Elite). For Histone modification tracks we included H3K27Ac as an enhancer marker, H3K4me3 as a promoter and transcribed region marker and H3K4Me1 as regulatory region marker from the GM12878 lymphoblastoid cell line from ENCODE. Further information about the associated GeneHancer promoters and enhancers were extracted via GeneCards (<https://www.genecards.org>). Our focus was on the Double Elite regulatory elements and associated genes, i.e. they were confirmed by several sources used by GeneHancer. We used the Roadmap Epigenomics project data on histone marks to evaluate the regions in different immune cell populations, including primary B-cells, monocytes, hematopoietic stem cells, neutrophils, T-cells and natural killer cells (<https://epigenomegateway.wustl.edu/>) [28]. We also evaluated the histone marks in other adult cells and tissues, including ovary, adipose nuclei, osteoblasts, lung, breast myoepithelial cells, adult liver, thymus, skeletal muscle, spleen and gastrointestinal tract. We checked the function of the genes located within the haplotype regions and genes interacting with the regulatory elements within the haplotype regions using GeneCards (<https://www.genecards.org>) and UniProt (<https://www.uniprot.org>). We used PubMed (<https://pubmed.ncbi.nlm.nih.gov>) to look for the relationship between the genes with cancer, especially with multiple myeloma or other hematological malignancies.

## RESULTS

We used a joint SNP as an anchor point to create overlapping haplotypes for the three populations of our study (Fig. 1). We



**Fig. 1** Meta-analysis of the haplotypes from the three populations, originating from the Czech Republic (CZE), Germany (GER) and Sweden (SWE), represented by the joint SNP rs11368313 (in red). Other SNPs common in all three populations are shown in green. For each population, the chromosome start and end position and the allele of the haplotype are shown.

identified 23 haplotypes that were associated with the risk of MGUS at the genome-wide significance level  $p < 5 \times 10^{-8}$  and showed increased risk of MGUS in all three populations (Table 1). One region represented by the SNP rs111335312 was located within the *ULK4* gene, in a known MM risk locus. Altogether 13 haplotype regions were overlapping with genes or regulatory elements interacting with the neighboring genes (Tables 1, 2). In these haplotype regions most of the regulatory elements were active in B-cells and/or immune cells; some of them were even recognized as super-enhancers. All regulatory elements had several target genes and most of them contained tens to a few hundred transcription factor binding sites.

In five haplotype regions, strong promoter, enhancer and regulatory element related histone marks and regulatory element-target gene association scores as well as ChromHMM and Segway genome segmentation data supported the importance of the regions in MGUS susceptibility (Fig. 2, Table 2). On chromosome 1, the region represented by rs11368313 overlapped with *TMCO4* (transmembrane and coiled-coil domains 2) and *RNF186* (ring finger protein 186) and showed interactions between promoter/enhancers of these two genes. Additionally, *RNF186* interacted with *OTUD3* (OTU deubiquitinase 3). Both RNF186 and OTUD3 are involved in (de)ubiquitination. A haplotype on chromosome 6 represented by rs10658790 overlapped with the promoter/enhancer of *HINT3* (histidine triad nucleotide binding protein 3), which interacted with promoter/enhancers of *TRMT11* (tRNA methyltransferase 11 homolog) and *NCOA7* (Nuclear

Receptor Coactivator 7). *HINT3* may be involved in PI3K/AKT/mTOR pathway.

On chromosome 17, the most interesting haplotype region represented by rs10163481 contained three regions with strong promoter and enhancer histone marks (Fig. 2, Table 2). These were related to GeneHancer promoter/enhancers and showed interactions between *SEC14L1* (SEC14 like lipid binding 1) and *SNHG20* (small nucleolar RNA host gene 20) and *SRSF2* (serine and arginine rich splicing factor 2) which is involved in RNA splicing. *SEC14L1* encodes a signal transduction inhibitor involved in innate immunity.

The region on chromosome 20 represented by rs111797554 overlapped with the *ZBP1* (Z-DNA binding protein 1) gene and its promoter and enhancer that interact with each other (Fig. 2, Table 2). *ZBP1* is involved in innate immunity responses. The region on chromosome 22, represented by rs1045588, involved two genes, *BID* (BH3 interacting domain death agonist) and *MICAL3* (microtubule associated monooxygenase, calponin and LIM domain containing 3), and showed several regions with very strong promoter and enhancer histone marks and promoter-enhancer interactions. *BID* belongs to the BCL-2 family of cell death regulators and *MICAL3* is involved in cell cycle regulation.

Further five haplotype regions with either weaker histone marks, genome segmentation data or regulatory element-target gene associations included a region on chromosome 4 represented by rs10006825 which contained an enhancer associated with *RBPJ* (recombination signal binding protein for

**Table 1.** Haplotype regions associated with the risk of MGUS.

Joint SNP	CHR	BP	OR	95% CI	p-value	Associated genes	Gene name	Function	GeneHancer interactions yes/no
rs11368313	1	20141882	3.59	2.37–5.43	$1.56 \times 10^{-9}$	TMCO2	Transmembrane and coiled-coil domains 2	Unknown; integral component of membrane	Yes
rs111335312	3	41777743	1.69	1.41–2.03	$1.13 \times 10^{-8}$	RNF186	Ring finger protein 186	Apoptosis; ubiquitination	
rs111723274	3	151675513	1.63	1.38–1.92	$8.25 \times 10^{-9}$	OTUD3	OTU deubiquitinase 3	Deubiquitination	
rs10938733	4	8676148	3.48	2.22–5.44	$4.68 \times 10^{-8}$	ULK4	unc-51 like kinase 4	Serine/threonine kinase	No
rs10006825	4	26153727	2.22	1.68–2.92	$1.37 \times 10^{-8}$	no gene			No
rs10006825	4	26153727	2.22	1.68–2.92	$1.37 \times 10^{-8}$	RBPJ	Recombination signal binding protein for immunoglobulin kappa J region	Regulator of Notch signaling; chromatin remodeling; binds to immunoglobulin kappa-type J segment recombination signal sequence	Yes
rs10940110	5	66794814	2.22	1.74–2.83	$1.18 \times 10^{-10}$	no gene			No
rs10514061	5	75435368	1.45	1.27–1.65	$4.19 \times 10^{-8}$	SV2C	Synaptic vesicle glycoprotein 2C	Neurotransmitter transport	No
rs10658790	6	126275488	1.59	1.36–1.86	$6.29 \times 10^{-9}$	HINT3	Histidine triad nucleotide binding protein 3	Nucleotide hydrolase and transferase	Yes
						TRMT11	tRNA methyltransferase 11 homolog	tRNA methylation	
						NCOA7	Nuclear receptor coactivator 7	Positive regulation of transcription by RNA polymerase II; coactivation of different nuclear receptors	
rs10253191	7	67372454	2.47	1.78–3.41	$4.76 \times 10^{-8}$	no gene			No
rs112048508	10	30443336	1.72	1.42–2.08	$2.81 \times 10^{-8}$	no gene			No
rs10567579	10	134807240	1.71	1.41–2.06	$2.62 \times 10^{-8}$	LINC01166; LINC01168	Long intergenic non-protein coding RNA	lncRNA	Yes
rs10628082	11	134612753	1.78	1.47–2.15	$2.39 \times 10^{-9}$	LOC105369584	Unknown locus		Yes
rs10840622	12	122288202	1.74	1.44–2.10	$6.40 \times 10^{-9}$	SETD1B	SET domain containing 1B; histone lysine methyltransferase	Chromatin remodeling; regulation of the transcriptional programming of multipotent hematopoietic progenitor cells and lymphoid lineage specification during hematopoiesis	No
						HPD	4-hydroxyphenylpyruvate dioxygenase	Tyrosine catabolism	
rs111792251	13	40895225	2.06	1.63–2.62	$2.07 \times 10^{-9}$	no gene			No
rs111282782	16	84564132	2.18	1.65–2.88	$3.87 \times 10^{-8}$	no gene			No
rs111362005	17	831933	3.84	2.38–6.20	$3.44 \times 10^{-8}$	NXN	Nucleoredoxin	Redox-dependent regulator of the Wnt signaling pathway; cell growth and differentiation	Yes

Table 1. continued

Joint SNP	CHR	BP	OR	95% CI	p-value	Associated genes	Gene name	Function	GeneHancer interactions yes/no
rs1024819	17	55607271	2.98	2.08–4.28	$3.02 \times 10^{-9}$	MSI2	Musashi RNA binding protein 2	Transcription regulation of genes involved in development and cell cycle	Yes
rs10163481	17	75148741	1.83	1.55–2.15	$6.37 \times 10^{-13}$	SEC14L1	SEC14 like lipid binding 1	Signal transduction inhibitor; innate immunity	Yes
						SNHG20	Small nucleolar RNA host gene 20	lncRNA	
						SRSF2	Serine and arginine rich splicing factor 2	mRNA splicing	
rs10543240	18	42189616	1.69	1.41–2.01	$5.06 \times 10^{-9}$	no gene			No
rs10420324	19	54469051	4.40	2.68–7.22	$4.87 \times 10^{-9}$	MYADM	Myeloid associated differentiation marker	Negative regulation of heterotypic cell-cell adhesion, macromolecule metabolic process and protein kinase C signaling	Yes
						PRKCG	Protein kinase C gamma	Expressed solely in the brain and spinal cord neurons	
						CACNG7	Calcium voltage-gated channel auxiliary subunit gamma 7	Regulation of the activity of L-type calcium channels	
						CACNG8	Calcium voltage-gated channel auxiliary subunit gamma 8	Regulation of the activity of L-type calcium channels	
						ZNF765	Zinc finger protein 765	Transcription factor	
						ZNF761	Zinc finger protein 761	Transcription factor	
						ZNF813	Zinc finger protein 813	Transcription factor	
						CNOT	CCR4-NOT transcription complex subunit 3	RNA-mediated gene silencing	
rs1051904	20	4163302	1.76	1.46–2.12	$3.22 \times 10^{-9}$	SMOX	Spermine oxidase	Oxidation of spermine to spermidine; determinant of cellular sensitivity to the antitumor polyamine analogs	Yes
rs111797554	20	56168743	2.18	1.66–2.85	$1.52 \times 10^{-8}$	ZBP1	Z-DNA binding protein 1	Innate immunity; apoptosis; necrosis	Yes
rs1045588	22	18271078	3.12	2.11–4.63	$1.46 \times 10^{-8}$	BID	BH3 interacting domain death agonist	BCL-2 family of cell death regulators	Yes
						MICAL3	Microtubule associated monoxygenase, calponin and LIM domain containing 3	Cell cycle; cell division; exocytosis	

**Table 2.** GeneHancer promoters and enhancers in the haplotype regions associated with the risk of MGUS, details extracted via GeneCards.

Haplotype chr: SNP <sup>a</sup>	GeneHancer (GH) Identifier	GH Type	GH Score	GH Sources	Gene Association Score	ENCODE transcription factor (TF) binding sites	Gene Targets <sup>c</sup>	Cells and tissues <sup>d</sup>	RoadMap Epigenomics <sup>e</sup>
chr1: rs11368313	GH01J019798	Promoter/Enhancer	1.9 <sup>b</sup>	RefSeq, EPDnew, Ensembl, ENCODE, CraniofacialAtlas	260.60 <sup>b</sup>	110 TFs	6 genes; TCMO4	B-cells, immune cells, many other cells and tissues	B-cells, immune cells, many other cells
	GH01J019812	Promoter/Enhancer	1.7 <sup>b</sup>	RefSeq, EPDnew, Ensembl, ENCODE, dbSUPER	259.30 <sup>b</sup>	49 TFs	7 genes; TCMO4, RNF186, OTUD3		
chr4: rs10006825	GH04J026131	Enhancer	1.1 <sup>b</sup>	FANTOM5, Ensembl, ENCODE	17.60 <sup>b</sup>	46 TFs	6 genes; RBPJ	Immune cells	B-cells, T-cells, natural killer cells
chr6: rs10658790	GH06J125955	Promoter/Enhancer	2.6 <sup>b</sup>	RefSeq, EPDnew, FANTOM5, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	250.70 <sup>b</sup>	313 TFs	7 genes; HINT3, TRMT11	B-cells, immune cells, many other cells and tissues	B-cells, other immune cells, many other cells
chr10: rs10567579	GH10J132962	Promoter/Enhancer	1.4	RefSeq, Ensembl, ENCODE, CraniofacialAtlas	250.70 <sup>b</sup>	20 TFs	7 genes; LINC01168, LINC01166	B-cells, immune cells, many other cells and tissues	B-cells, other immune cells, many other cells
	GH10J133008	Enhancer	1.3 <sup>b</sup>	RefSeq, ENCODE, CraniofacialAtlas, dbSUPER	29.7	43 TFs	12 genes; LINC01168	Some immune cells	Monocytes, neutrophils, natural killer cells
	GH10J133012	Enhancer	1.3 <sup>b</sup>	RefSeq, FANTOM5, ENCODE, dbSUPER	28.1	15 TFs	8 genes; LINC01168	Monocytes	Monocytes, neutrophils, natural killer cells
chr11: rs106282082	GH11J134729	Enhancer	0.8 <sup>b</sup>	RefSeq, ENCODE, dbSUPER	256.80 <sup>b</sup>	4 TFs	5 genes; LINC02714		
chr17: rs111362005	GH17J000914	Enhancer	1.5 <sup>b</sup>	RefSeq, FANTOM5, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	6.40 <sup>b</sup>	17 TFs	7 genes; NXN	Neutrophils, some other cells and tissues	Neutrophils, other cells
	GH17J000924	Promoter/Enhancer	1.8 <sup>b</sup>	RefSeq, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	17.90 <sup>b</sup>	162 TFs	12 genes; NXN	Some cells and tissues	Some cells
	GH17J000931	Promoter/Enhancer	1.6 <sup>b</sup>	Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	29.60 <sup>b</sup>	122 TFs	12 genes; NXN	Some cells and tissues	
chr17: rs1024819	GH17J057521	Enhancer	1.6 <sup>b</sup>	RefSeq, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	0.06	114 TFs	9 genes; MS12	B-cells, immune cells; SE_13320 (CD34 primary cells); many other cells and tissues	B-cells, some other immune cells, some other cells
chr17: rs10163481	GH17J077084	Enhancer	0.4 <sup>b</sup>	Ensembl, ENCODE, dbSUPER	255.60 <sup>b</sup>	no	6 genes; SNHG20		

Table 2. continued

Haplotype chr: SNP <sup>a</sup>	GeneHancer (GH) Identifier	GH Type	GH Score	GH Sources	Gene Association Score	ENCODE transcription factor (TF) binding sites	Gene Targets <sup>c</sup>	Cells and tissues <sup>d</sup>	RoadMap Epigenomics <sup>e</sup>
	GH17J077087	Promoter/ Enhancer	2.5 <sup>b</sup>	RefSeq, EPDnew, FANTOM5, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	250.70 <sup>b</sup>	263TFs	10 genes; SEC14L1, SNHG20	B-cells, immune cells	B-cells, other immune cells, many other cells
	GH17J077095	Enhancer	1.3 <sup>b</sup>	FANTOM5, Ensembl, ENCODE, dbSUPER	26.60 <sup>b</sup>	83 TFs	8 genes; SEC14L1, SNHG20	Immune cells	Immune cells, many other cells
	GH17J077106	Promoter/ Enhancer	1.9 <sup>b</sup>	RefSeq, FANTOM5, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	18.50 <sup>b</sup>	121 TFs	10 genes; SEC14L1	B-cells, lymphocyte of B lineage, immune cells	B-cells, other immune cells, many other cells
	GH17J077114	Promoter/ Enhancer	2.0 <sup>b</sup>	RefSeq, FANTOM5, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	10.20 <sup>b</sup>	233 TFs	9 genes; SEC14L1	B-cells, immune cells	B-cells, immune cells, many other cells
	GH17J077137	Promoter/ Enhancer	2.6 <sup>b</sup>	RefSeq, EPDnew, FANTOM5, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	259.8 <sup>b</sup>	300 TFs	12 genes; SEC14L1	B-cells, immune cells, other cells	B-cells, immune cells, many other cells
	GH17J077150	Enhancer	0.3 <sup>b</sup>	ENCODE, dbSUPER	7.20 <sup>b</sup>	no	7 genes; SEC14L1		
	GH17J077152	Enhancer	0.3 <sup>b</sup>	ENCODE, dbSUPER	7.20 <sup>b</sup>	no	6 genes; SEC14L1		
	GH17J077159	Promoter/ Enhancer	1.2 <sup>b</sup>	RefSeq, ENCODE, dbSUPER	7.20 <sup>b</sup>	54 TFs	7 genes; SEC14L1	B-cells	B-cells, immune cells, some other cells
chr19: rs10420324	GH19J053864	Promoter/ Enhancer	2.6 <sup>b</sup>	RefSeq, EPDnew, FANTOM5, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	265.10 <sup>b</sup>	255 TFs	39 genes; MYADM, ZNF765, ZNF813, ZNF761, CNOT3, PRKCG	B-cells, immune cells; SE_20401 (CD56 primary cells), SE in many immune cells	B-cells, immune cells, many other cells
	GH19J053874	Enhancer	0.6 <sup>b</sup>	ENCODE, dbSUPER	11.00 <sup>b</sup>	5 TFs	6 genes; MYADM	B-cells, immune cells; SE_20401 (CD56 primary cells), SE in many immune cells	B-cells, immune cells
	GH19J053877	Enhancer	0.6 <sup>b</sup>	ENCODE, dbSUPER	6.90 <sup>b</sup>	5 TFs	10 genes; PRKCG	B-cells; SE in some immune cells	Many cells
	GH19J053889	Enhancer	0.8 <sup>b</sup>	ENCODE, dbSUPER	18.50 <sup>b</sup>	24 TFs	6 genes; MYADM	Immune cells	B-cells, monocytes, natural killer cells

Table 2. continued

Haplotype chr: SNP <sup>a</sup>	GeneHancer (GH) Identifier	GH Type	GH Score	GH Sources	Gene Association Score	ENCODE transcription factor (TF) binding sites	Gene Targets <sup>c</sup>	Cells and tissues <sup>d</sup>	RoadMap Epigenomics <sup>e</sup>
	GH119J053907	Promoter/Enhancer	1.2 <sup>b</sup>	FANTOM5, Ensembl, ENCODE	260.40 <sup>b</sup>	31 TFs	8 genes; CACNG7, PRKCG, CACNG8		
	GH119J053956	Promoter/Enhancer	1.8 <sup>b</sup>	RefSeq, EPDnew, FANTOM5, Ensembl, ENCODE	260.40 <sup>b</sup>	54 TFs	23 genes; CACNG8, CACNG7		
	GH119J053977	Promoter/Enhancer	1.5 <sup>b</sup>	RefSeq, Ensembl, ENCODE	18.50 <sup>b</sup>	149 TFs	27 genes; CACNG8	B-cells, immune cells	Some cells
chr20: rs1051904	GH20J004133	Enhancer	1.7 <sup>b</sup>	efSeq, FANTOM5, Ensembl, ENCODE, dbSUPER	3.00 <sup>b</sup>	149TFs	10 genes; SMOX	Monocytes, neutrophils, other cells and tissues	Some cells
	GH20J004148	Promoter/Enhancer	2.0 <sup>b</sup>	RefSeq, EPDnew, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	268.70 <sup>b</sup>	83 TFs	6 genes; SMOX		Immune cells, many other cells
	GH20J004159	Enhancer	1.2 <sup>b</sup>	RefSeq, FANTOM5, ENCODE, dbSUPER	8.60 <sup>b</sup>	20 TFs	6 genes; SMOX	Monocytes	B-cells, other immune cells, many other cells
	GH20J004165	Enhancer	1.5 <sup>b</sup>	RefSeq, FANTOM5, ENCODE, CraniofacialAtlas, dbSUPER	20.7 <sup>b</sup>	36 TFs	9 genes; SMOX		Some cells
	GH20J004169	Promoter/Enhancer	2.1 <sup>b</sup>	RefSeq, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	7.90 <sup>b</sup>	190 TFs	7 genes; SMOX	B-cells, immune cells	B-cells, other immune cells, many other cells
chr20: rs111797554	GH20J057617	Promoter/Enhancer	2.0 <sup>b</sup>	RefSeq, EPDnew, FANTOM5, Ensembl, ENCODE, dbSUPER	264.30 <sup>b</sup>	67 TFs	4 genes; ZBP1	B-cells, immune cells; SE_20522 (CD56 primary cells), SE_16712 (CD4 naive primary cells)	B-cells, immune cells, gastrointestinal tract
	GH20J057627	Enhancer	1.1 <sup>b</sup>	FANTOM5, ENCODE, CraniofacialAtlas, dbSUPER	11.90 <sup>b</sup>	13 TFs	8 genes; ZBP1	B-cells, immune cells; SE_20522 (CD56 primary cells), SE_16712 (CD4 naive primary cells)	B-cells, other immune cells, many other cells
chr 22: rs1045588	GH22J017767	Enhancer	1.2 <sup>b</sup>	RefSeq, ENCODE, dbSUPER	4.90 <sup>b</sup>	49 TFs	4 genes; BID	Immune cells; SE_09461 (CD14+ monocytes)	B-cells, other immune cells, some other cells



Table 2. continued

Haplotype chr: SNP <sup>a</sup>	GeneHancer (GH) Identifier	GH Type	GH Score	GH Sources	Gene Association Score	ENCODE transcription factor (TF) binding sites	Gene Targets <sup>c</sup>	Cells and tissues <sup>d</sup>	RoadMap Epigenomics <sup>e</sup>
	GH22J017769	Promoter/ Enhancer	2.5 <sup>b</sup>	RefSeq, EPDnew, FANTOM5, Ensembl, ENCODE, CraniofacialAtlas, dbSUPER	264.80 <sup>b</sup>	155 TFs	7 genes; BID	B-cells, immune cells; SE_09461 (CD14+ monocytes); many cells and tissues	B-cells, other immune cells, many other cells
	GH22J017784	Enhancer	1.8 <sup>b</sup>	RefSeq, VISTA, ENCODE, CraniofacialAtlas, dbSUPER	0.41	153 TFs	8 genes	B-cells, immune cells; SE_09461 (CD14+ monocytes)	Immune cells, many other cells
	GH22J017794	Enhancer	1.6 <sup>b</sup>	RefSeq, FANTOM5, Ensembl, ENCODE, dbSUPER	15.80 <sup>b</sup>	83 TFs	10 genes; BID, MICAL3	B-cells, immune cells; SE_09461 (CD14+ monocytes)	B-cells, monocytes, hematopoietic stem cells, natural killer cells

<sup>a</sup>Haplotype identified by the chromosome and the joint SNP in the haplotype region.

<sup>b</sup>Double Elite GeneHancer score and Gene association score.

<sup>c</sup>Number of gene targets; Gene targets in the haplotype region.

<sup>d</sup>Examples of tissues, in which the promoter/enhancer is active in GeneHancer sources; Superenhancer (SE) in B-cells or immune cells.

<sup>e</sup>Cell populations with active chromatin state in 6 immune cell populations and over 10 adult cell and tissue types were extracted via Roadmap Epigenomics.

immunoglobulin kappa J region) (Table 2, Supplementary Fig. 1). RBPJ is regulating NOTCH-signaling. In the region on chromosome 17 represented by rs11362005 three enhancers within an intron of *NXN* (nucleoredoxin) showed interactions with an enhancer region flanking the promoter of *NXN*, which regulates Wnt signaling. In another region on chromosome 17, represented by rs1024819, a weak enhancer interaction within *MSI2* (musashi RNA binding protein 2) may affect the function of *MSI2* as a transcriptional regulator of genes involved in development and cell cycle.

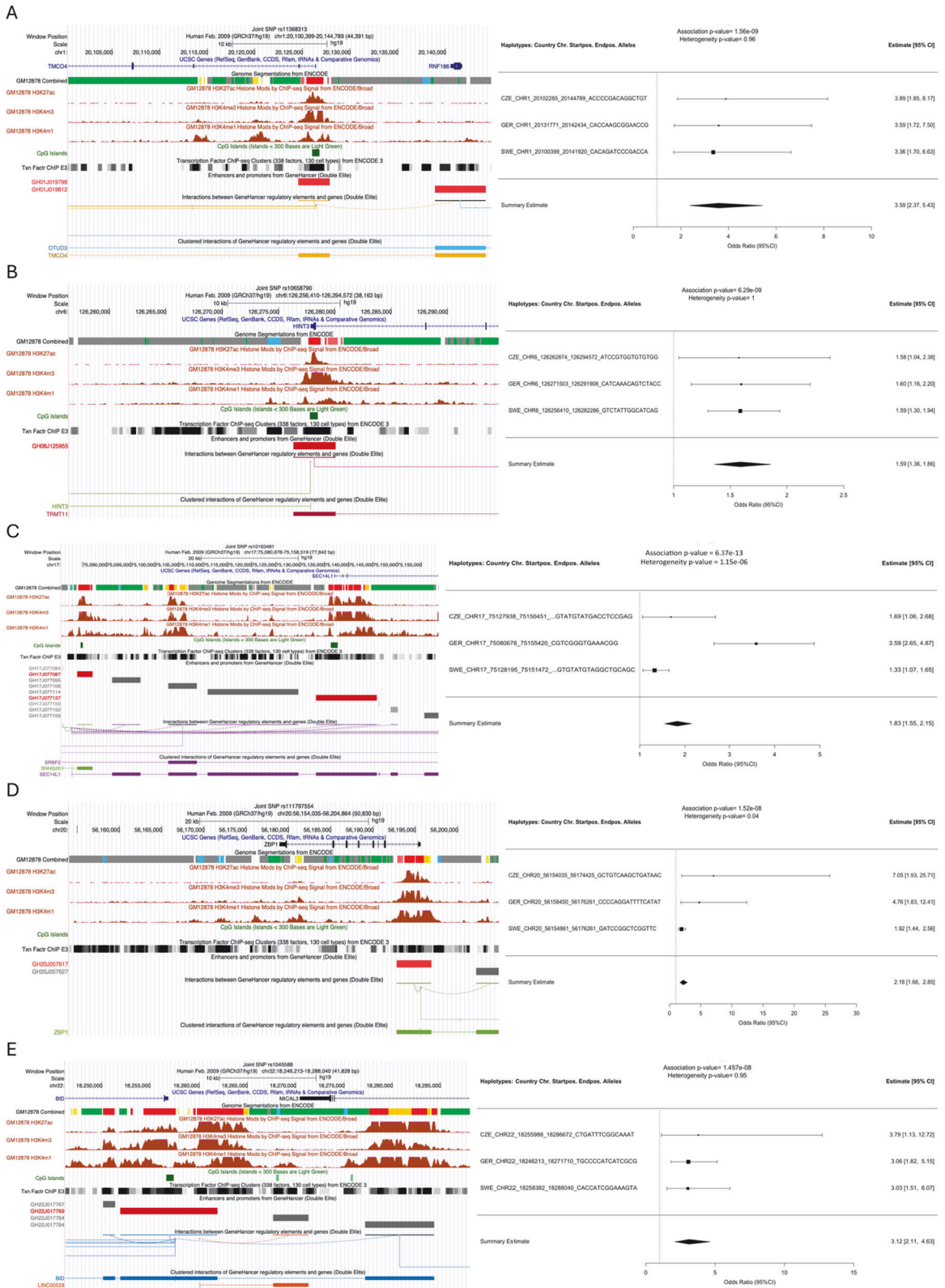
A large region on chromosome 19 represented by rs10420324 hosted *MYADM* (myeloid associated differentiation marker), *PRKCG* (protein kinase C gamma), *CACNG7* and *CACNG8* (calcium voltage-gated channel auxiliary subunit gamma 7 and 8, respectively) that interacted with each other as well as with three zinc finger proteins and *CNO3* (CCR4-NOT Transcription Complex Subunit 3), which is involved in RNA-mediated gene silencing (Table 2, Supplementary Fig. 1). *MYADM* is a negative regulator of heterotypic cell-cell adhesion and protein kinase C signaling. On chromosome 20 represented by rs1051904 many promoter-enhancer interactions were implicated between the *SMOX* (spermine oxidase) gene and long-distance enhancers. *SMOX* may act as a determinant of cellular sensitivity to antitumor polyamine analogs.

Furthermore, on chromosomes 10 and 11, GeneHancer regulatory elements were connected to long non-coding RNAs, however with unknown functional consequences (Table 2, Supplementary Fig. 1). On chromosome 12, a region represented by rs10840622 overlapped with the *SETD1B* (SET Domain Containing 1B, Histone Lysine Methyltransferase) and *HPD* (4-Hydroxyphenylpyruvate Dioxygenase) genes and showed strong promoter-related histone marks at the 3'end of *HPD*, but no interactions with any GeneHancer regulatory elements.

## DISCUSSION

We conducted a haplotype-based analysis on MGUS to complement the earlier GWAS analyses where individual SNPs were considered [13–15, 29]. A previous meta-analysis of the GWAS data from MGUS, MM and AL-amyloidosis identified 17 independent regions with genome-wide significance [13]. In the present study we identified 23 haplotypes that were associated with the risk of MGUS at the genome-wide significance level  $p < 5 \times 10^{-8}$  and showed increased risk of MGUS in all three study populations. Notably, only the *ULK4* containing haplotype on chromosome 3 shared significance (OR 1.70,  $p = 1.13 \times 10^{-8}$ ) with the previous study, which may suggest the importance of this locus. *ULK4* is a serine/threonine kinase. Several associated haplotype regions affected pathways important for MM cell survival and genes encoding important current therapeutic targets for MM, which may highlight the advantage of the haplotype approach homing to functional units.

We identified five haplotype regions, in which associations with regulatory elements and their connections to target genes supported by genome segmentation highlighted the possible role of the related haplotypes in MGUS. None of these regulatory elements were specific for B-cells, but were active also in other immune cells, and other adult cells and tissues, as implicated from different sources from GeneHancer and Roadmap Epigenomics. The haplotype on chromosome 1 included two (de)ubiquitination-related genes, *RNF186* and *OTUD3*, which interacted with each other. As MM cells produce high amounts of monoclonal antibodies, maintaining protein homeostasis is crucial for MM cells [30]. The ubiquitin–proteasome system plays an important role in this process and MM cell killing can be caused by blocking or interfering this system. Thus, it has been one of the primary targets in MM treatment starting in early 2000 with proteasome inhibitors and immune modulators [30]. Commonly used



**Fig. 2 Haplotype regions and Forest plots from the five most interesting haplotypes associated with the risk of MGUS.** Haplotype regions are shown using UCSC Genome browser's GRCh37/hg19 assembly and annotation tracks from ENCODE and GeneHancer. Forest plots show the overlapping haplotypes, represented by a joint SNP, of the three study populations from the Czech Republic (CZE), Germany (GER) and Sweden (SWE). For each population the odds ratio and the corresponding 95% confidence interval (CI) are shown as well as the summary estimate of the meta-analysis. **A** Chromosome 1, rs11368313, **B** Chromosome 6, rs10658790, **C** Chromosome 17, rs10163481, **D** Chromosome 20, rs111797554, **E** Chromosome 22, rs1045588.

proteasome inhibitors include bortezomib, carfilzomib and the oral drug ixazomib [31]. Commonly used immune modulators include thalidomide, lenalidomide and pomalidomide [31]. E3 ubiquitin ligase and deubiquitination inhibitors are currently being tested in preclinical studies and clinical trials [30, 31].

A haplotype on chromosome 6 covered the promoter/enhancer of *HINT3*, probably involved in PI3K/AKT/mTOR pathway, and interacting with promoter/enhancers of *NCOA7* which encodes an estrogen receptor associated protein. PI3K/AKT/mTOR pathway is one of the signaling pathways in the bone marrow microenvironment that promotes signaling events in MM cells and enhances their survival [32]. The pathway shows aberrant activation in many MM patients. Targeting PI3K/AKT/mTOR pathway has shown promising results in preclinical studies, however, clinical trials have been disappointing showing limited clinical efficiency in MM and severe side effects [32].

The haplotype on chromosome 17 encoding the *SEC14L1* gene and another haplotype on chromosome 20 associated with the *ZBP1* gene both are involved in innate immunity. *SEC14L1* was one of the genes that was mutated in the germline of one German MM family [33]. *ZBP1* is integral part of host defense against pathogens. It responds to a variety of stimuli, such as viral infection and homeostatic perturbations, leading to the formation of the PANoptosome complex and release of numerous cytokines and chemokines [34]. It is also known to be active in MM [35].

The haplotype on chromosome 22 covered two genes, *BID* and *MICAL3*. *MICAL3* is involved in cell cycle regulation and it has been implicated in various cancers [36]. *BID* belongs to the BCL-2 family of cell death regulators which have recently become extremely promising therapeutic targets in hematological malignancies, including MM, with the prime drug venetoclax [31, 37, 38]. Overexpression of Bcl-2 has been found in MM harboring a common translocation (11;14) which is a special indication for venetoclax [39]. This translocation is also common in MGUS [40]. Unfortunately, the MGUS samples of the present study have not been screened for cytogenetic alterations and we could not investigate this aspect further.

A further haplotype region with a possible association with immune functions was found on chromosome 4 represented with an enhancer associated with *RBPJ*. This enhancer seemed to be specific for immune cells, especially for B-cells, T-cells and natural killer cells. *RBPJ* is a key translational transducer of NOTCH-signaling and it plays a role in cellular immune response [41]. Similar to PI3K/AKT/mTOR pathway, NOTCH-signaling promotes communication between adjacent cells in the bone marrow microenvironment and has been shown to be dysregulated in the MM tumor niche [42].

The challenges of our haplotype-based association study included three different study populations and their different sample sizes as well as different genotyping arrays used for the sample sets. This also illustrates some limitations of our study. In general, use of several populations increases the sample size and allows replication of the findings. However, as the frequencies of the SNPs and their LD vary between populations, it is difficult to find exactly the same haplotypes in all populations. The use of different genotyping arrays with different numbers of genotyped SNPs affected the imputation and the final number of SNPs included in the study, and at the end the composition of the haplotypes. We aimed to solve these problems by identifying the common SNPs in all populations and defining a joint SNP as a representative for the haplotype as shown in Fig. 1. However, the different composition of the haplotypes in the study populations makes it also difficult to evaluate the effect of the haplotypes on the expression of genes connected to these haplotypes. We used GeneHancer and Roadmap Epigenomics data to investigate the activity of the regulatory elements in B-cells, other immune cells and other adult cells and tissues as a potential indicator for the effect of the haplotypes. The problem of different sample sizes in

the meta-analysis of GWAS data has been considered by Cook et al. [43]. They demonstrated that linear/logistic models can be used for meta-analysis of GWASs of binary phenotypes, without loss of power, even in the presence of extreme sample size and case-control imbalances, provided that inverse-variance weighting of allelic effect sizes after conversion onto the log-odds scale has been performed, as was done in our study.

In conclusion, our haplotype-based genetic association study identified several novel loci associated with the risk of MGUS. Genes and regulatory elements connected to neighboring genes were related to pathways dysregulated in MM, which serve as targets for already existing therapies, or may serve as targets for drugs that are currently tested in preclinical studies or clinical trials. Interestingly, these loci with exception of the *ULK4* locus, have not been found in large GWASs on MM which may show the strength of the haplotype-based approach testing functional chromosomal units rather than individual SNPs. Whether these MGUS-associated loci are important in the disease progression to MM as high-risk markers remains to be investigated.

## DATA AVAILABILITY

The GWAS data are available at the NHGRI-EBI Catalog of human genome-wide association studies, accession number GCST007824. The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

## REFERENCES

- Rahman N. Realizing the promise of cancer predisposition genes. *Nature*. 2014;505:302–8.
- Hauser E, Cremer N, Hein R, Deshmukh H. Haplotype-based analysis: a summary of GAW16 Group 4 analysis. *Genet Epidemiol*. 2009;33:524–8.
- Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
- International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437:1299–320.
- Sud A, Kinnerley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer*. 2017;17:692–704.
- Zhong C, Cozen W, Bolanos R, Song J, Wang SS. The role of HLA variation in lymphoma aetiology and survival. *J Intern Med*. 2019;286:154–80.
- Bergman A, Einbeigi Z, Olofsson U, Taib Z, Wallgren A, Karlsson P, et al. The western Swedish BRCA1 founder mutation 3171ins5; a 3.7 cM conserved haplotype of today is a reminiscence of a 1500-year-old mutation. *Eur J Hum Genet*. 2001;9:787–93.
- Wiggins GR, Cole JB, Hubbard SM, Sonstegard TS. Genomic selection in dairy cattle: The USDA experience. *Annu Rev Anim Biosci*. 2017;5:309–27.
- Bello SF, Lawal RA, Adeola AC, Nie Q. The study of selection signature and its applications on identification of candidate genes using whole genome sequencing data in chicken—a review. *Poult Sci*. 2023;102:102657.
- Barnekow E, Hasslow J, Liu W, Bryant P, Thutkawkorapin J, Wendt C, et al. A Swedish familial genome-wide haplotype analysis identified five novel breast cancer susceptibility loci on 9p24.3, 11q22.3, 15q11.2, 16q24.1 and Xq21.31. *Int J Mol Sci*. 2023;24:4468.
- Barnekow E, Liu W, Helgadottir HT, Michailidou K, Dennis J, Bryant P, et al. A Swedish genome-wide haplotype association analysis identifies a novel breast cancer susceptibility locus in 8p21.2 and characterizes three loci on chromosomes 10, 11 and 16. *Cancers*. 2022;14:1206.
- Chattopadhyay S, Thomsen H, da Silva Filho MI, Weinhold N, Hoffmann P, Nothen MM, 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*. 2018;24:30.
- Chattopadhyay S, Thomsen H, Weinhold N, Meziane I, Huhn S, da Silva Filho MI, et al. Eight novel loci implicate shared genetic etiology in multiple myeloma, AL amyloidosis, and monoclonal gammopathy of unknown significance. *Leukemia*. 2020;34:1187–91.
- Clay-Gilmour A, Chattopadhyay S, Hildebrandt MAT, Thomsen H, Weinhold N, Vodicka P, et al. Genome-wide meta-analysis of monoclonal gammopathy of undetermined significance (MGUS) identifies risk loci impacting IRF-6. *Blood Cancer J*. 2022;12:60.
- Thomsen H, Chattopadhyay S, Weinhold N, Vodicka P, Vodickova L, Hoffmann P, et al. Genome-wide association study of monoclonal gammopathy of unknown

- significance (MGUS): comparison with multiple myeloma. *Leukemia*. 2019;33:1817–21.
16. Broderick P, Chubb D, Johnson DC, Weinhold N, Forsti A, Lloyd A, et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet*. 2012;44:58–61.
  17. Chubb D, Weinhold N, Broderick P, Chen B, Johnson DC, Forsti A, et al. Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nat Genet*. 2013;45:1221–5.
  18. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 2010;11:499–511.
  19. Delaneau O, Coulonges C, Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics*. 2008;9:540.
  20. Utsunomiya YT, Milanese M, Utsunomiya AT, Ajmone-Marsan P, Garcia JF. GHap: an R package for genome-wide haplotyping. *Bioinformatics*. 2016;32:2861–2.
  21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
  22. Curtis D, Amos W. The human genome harbours widespread exclusive yin yang haplotypes. *Eur J Hum Genet*. 2024;32:691–6.
  23. Ribas G, Milne RL, Gonzalez-Neira A, Benítez J. Haplotype patterns in cancer-related genes with long-range linkage disequilibrium: no evidence of association with breast cancer or positive selection. *Eur J Hum Genet*. 2008;16:252–60.
  24. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*. 2002;12:996–1006.
  25. Raney BJ, Barber GP, Benet-Pagès A, Casper J, Clawson H, Cline MS, et al. The UCSC genome browser database: 2024 update. *Nucleic Acids Res*. 2024;52:D1082–8.
  26. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res*. 2013;41:D56–63.
  27. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*. 2017;2017.
  28. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–30.
  29. Clay-Gilmour AI, Hildebrandt MAT, Brown EE, Hofmann JN, Spinelli JJ, Giles GG, et al. Coinherited genetics of multiple myeloma and its precursor, monoclonal gammopathy of undetermined significance. *Blood Adv*. 2020;4:2789–97.
  30. Wirth M, Schick M, Keller U, Krönke J. Ubiquitination and ubiquitin-like modifications in multiple myeloma: biology and therapy. *Cancers*. 2020;12:3764.
  31. Bobin A, Liuu E, Moya N, Gruchet C, Sabirou F, Lévy A, et al. Multiple myeloma: an overview of the current and novel therapeutic approaches in 2020. *Cancers*. 2020;12:2885.
  32. Ramakrishnan V, Kumar S. PI3K/AKT/mTOR pathway in multiple myeloma: from basic biology to clinical promise. *Leuk Lymphoma*. 2018;59:2524–34.
  33. Catalano C, Paramasivam N, Blocka J, Giangioffe S, Huhn S, Schlesner M, et al. Characterization of rare germline variants in familial multiple myeloma. *Blood Cancer J*. 2021;11:33.
  34. Chen W, Gullett JM, Tweedell RE, Kanneganti TD. Innate immune inflammatory cell death: PANoptosis and PANoptosomes in host defense and disease. *Eur J Immunol*. 2023;53:e2250235.
  35. Ponnusamy K, Tzioni MM, Begum M, Robinson ME, Caputo VS, Katsarou A, et al. The innate sensor ZBP1-IRF3 axis regulates cell proliferation in multiple myeloma. *Haematologica*. 2022;107:721–32.
  36. Tominaga K, Minato H, Murayama T, Sasahara A, Nishimura T, Kiyokawa E, et al. Semaphorin signaling via MICAL3 induces symmetric cell division to expand breast cancer stem-like cells. *Proc Natl Acad Sci USA*. 2019;116:625–30.
  37. Kaloni D, Diepstraten ST, Strasser A, Kelly GL. BCL-2 protein family: attractive targets for cancer therapy. *Apoptosis*. 2023;28:20–38.
  38. Raab MS. Venetoclax in myeloma: to B, or not to B. *Blood*. 2024;143:4–5.
  39. Khan WJ, Ali M, Hashim S, Nawaz H, Hashim SN, Safi D, et al. Use of venetoclax in t(11;14) positive relapsed/refractory multiple myeloma: A systematic review. *J Oncol Pharm Pract*. 2023;10781552231218999.
  40. Weinhold N, Johnson DC, Chubb D, Chen B, Forsti A, Hosking FJ, et al. The CCND1 G870A polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. *Nat Genet*. 2013;45:522–5.
  41. Yan W, Menjivar RE, Bonilla ME, Steele NG, Kemp SB, Du W, et al. Notch Signaling Regulates Immunosuppressive Tumor-Associated Macrophage Function in Pancreatic Cancer. *Cancer Immunol Res*. 2024;12:91–106.
  42. Sabol HM, Delgado-Calle J. The multifunctional role of Notch signaling in multiple myeloma. *J Cancer Metastasis Treat*. 2021;7:20.
  43. Cook JP, Mahajan A, Morris AP. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur J Hum Genet*. 2017;25:240–5.

## ACKNOWLEDGEMENTS

This work was supported by the Dietmar Hopp Foundation, Black Swan Research Initiative/International Myeloma Foundation, and Transcan ERA-NET funding from the German Federal Ministry of Education and Research (BMBF). KH was supported by National Institute for Cancer Research—NICR (Programme EXCELES, ID Project No. LX22NPO5102), funded by the European Union—Next Generation EU, Grant Agency of Czech Republic 23-05609 S, GACR grant 23-05609S, the SALVAGE project, reg.no: CZ.02.01.01/00/22\_008/0004644.

## AUTHOR CONTRIBUTIONS

Conception and responsible for overall content: HT, KH, AF; Sample and data collection: NW, PV, LV, MMN, K-HJ, BS, RH, GH, UP-K, FS, HG; Study design: HT, KH, AF; Statistical analysis: HT; Manuscript drafting: HT, KH, AF; Manuscript editing and approval: All.

## FUNDING

Open Access funding enabled and organized by Projekt DEAL.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41408-024-01121-8>.

**Correspondence** and requests for materials should be addressed to Asta Försti.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024