Letter to the Editor

GENOME-WIDE ASSOCIATION STUDY OF MONOCLONAL GAMMOPATHY OF UNKNOWN SIGNIFICANCE (MGUS): COMPARISON WITH MULTIPLE MYELOMA

Hauke Thomsen[1], Subhayan Chattopadhyay[1], Niels Weinhold[2,3], Pavel Vodicka[4,5,6], Ludmila Vodickova[4,5,6], Per Hoffmann[7,8], Markus M Nöthen[7,9], Karl-Heinz Jöckel[10], Christian Langer[11], Roman Hajek[12], Göran Hallmans[13], Ulrika Pettersson-Kymmer[14], Claes Ohlsson[15], Florentin Späth[16], Richard Houlston[17,18], Hartmut Goldschmidt[2,19], Kari Hemminki[1]*, Asta Försti[1]*

*Shared senior authorship.

1. Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, D-69120, Heidelberg, Germany,
2. Department of Internal Medicine V, University of Heidelberg, Heidelberg, Germany,
3. Myeloma Institute, University of Arkansas for Medical Sciences, Little Rock, AR, USA,
4. Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, Videnska 1083, 142 00 Prague, Czech Republic,
5. Institute of Biology and Medical Genetics, 1st Medical Faculty, Charles University, Albertov 4, 128 00 Prague, Czech Republic,
6. Faculty of Medicine and Biomedical Center in Pilsen, Charles University in Prague, 30605 Pilsen, Czech Republic,
7. Institute of Human Genetics, University of Bonn, Bonn, Germany,
8. Department of Biomedicine, University of Basel, Basel, Switzerland,
9. Department of Genomics, Life & Brain Research Center, University of Bonn, Bonn, Germany,
10. Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University of Duisburg-Essen, Germany,
11. Department of Internal Medicine III, University of Ulm, Ulm, Germany,
12. Department of Hematooncology, University Hospital Ostrava, 17. listopadu 1790, 708 52 Ostrava, Czech Republic,
13. Department of Medical Biosciences/Pathology, University of Umea, Umea, Sweden,
14. Clinical Pharmacology, Department of Pharmacology and Clinical Neuroscience, Umea University, Umea, Sweden,

15. Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden,
16. Department of Radiation Sciences, Oncology, Umeå University, Umeå Sweden,
17. Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK,
18. Division of Molecular Pathology, The Institute of Cancer Research, London, UK,
19. National Centre of Tumor Diseases, Heidelberg, Germany.

Correspondence: Asta Försti, Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, Heidelberg 69120, Germany.
Telephone: +496221421803
Fax: +496221421810
Email: asta.foersti@dkfz.de

Word count: 1482 (text).

Key words: susceptibility, germline, low-risk genes, myeloma.

Running title: MGUS susceptibility

Monoclonal gammopathy of undetermined significance (MGUS) is a condition in which immunoglobulin derived serum M-protein is produced by a plasma cell clone and is detectable in the blood [1]. MGUS resembles multiple myeloma (MM), but antibody levels and number of plasma cells in the bone marrow are lower, and it is generally asymptomatic and often diagnosed idiosyncratically. However, since MGUS can lead to MM which develops at the rate of 0.5-1.5% a year, yearly monitoring is generally recommended [2]. The prevalence of MGUS increases from age 50 onwards (1.7% at 50-59 years), reaching 6.6% by age over 80 years according to a literature review [2]. As the cumulative incidence of MM by age 75 years in Sweden is 0.4% for men and 0.3% for women, it is clear that the prevalence of MGUS far exceeds that of MM (NORDCAN database http://www-dep.iarc.fr/NORDCAN/FI/frame.asp).

Genome-wide association studies (GWASs) have so far identified common genetic variants at 23 loci associated with MM risk [3]. Thus far the role of genetic variation influencing MGUS has only been studied to a limited extent [4]. Some of the first reported MM risk loci have in two early studies been shown to be at least weakly associated with MGUS [5]. To more comprehensively address the genetics of MGUS and its relationship to MM we conducted a GWAS of MGUS by analyzing 992 patients and 2,900 controls.

Detailed methods are described in the Supplement. We studied three independent sets of MGUS patients and controls. The German GWAS comprised 243 MGUS cases and 1,285 controls [4]. The Czech GWAS was based on 288 cases and 600 controls. The Swedish MGUS included 461 patients and 1025 controls. Odds ratios (ORs) for SNP associations in the three populations were meta-analyzed using PLINK software v1.90. Differences in the associations between MGUS and MM was tested using ASSET as previously [5]. The most promising MGUS associations were tested in the meta-analysis of the German population of 1717 MM cases and 2069 controls and the UK population of 2,282 MM cases and 5,197 controls; a recent study describing 23 genome-wide significant loci for MM [3, 6].

A Manhattan plot of GWAS on MGUS highlights 10 loci reaching a meta p-value below 10^{-5} (Supplementary Figure 1). The meta ORs for the risk allele ranged from 1.28 to 1.48 and the smallest p-values reached $5.6x10^{-7}$ for locus 3p22.1 (rs9848754, *ULK4)* (Table 1). With the exception of this locus and the locus at 17p11.2 (rs74998556, TNFRSF13B), the other 8 loci

were unique to MGUS when compared to MM (95%CIs were non-overlapping). A similar conclusion was reached by the ASSET analysis, displaying MGUS on 8 SNPs as positive and MM data as null or negative (Supplementary Figure 2). None of the signals from the 3 MGUS sample sets showed significant heterogeneity (p_{het} and I^2 statistic, Supplementary Table 1). Neither was heterogeneity observed for meta ORs for MM.

We compared GWAS associations for MGUS at the 23 loci for which genome-wide significant associations have been reported for MM (Table 2) [3]. ORs for 10 MGUS SNPs were nominally significant ($p<0.05$) and the risk allele was the same as for MM. For 9 of these loci the OR for MGUS (considering also the best SNP for MGUS in high LD with the MM SNP) was equal or higher than it was for MM and these included loci (marked by genes [3]): 2p23.3 (*DTNB)*, 2q31.1 (*SP3*), 3p22.1 (*ULK4)*, 6p22.3 *(JARID2)*, 7q36.1 (*ABCF2)*, 8q24.21 (*MYC*), 9p21.3 (*MTAP)*, 17p11.2 (*TNFRSF13B)* and 19p13.11 (*KLF2)*. The 95%CIs of MGUS ORs included the OR of MM at each of these loci. For 10 other SNPs the ORs were at least marginally higher for MM than for MGUS and the risk alleles were identical but the upper 95%CI of the MGUS SNPs covered the OR for MM. For 4 SNPs (16q23.1, 20q13.13, 22q13 and 22q13.1 marked by gene *CBX7*) the ORs for MM were substantially higher than those for MGUS for which the ORs were close to unity (1.00).

The data from Table 2 were subjected to ASSET analysis, and as expected all MM associations were classified as positive (Supplementary Figure 3). Among MGUS SNPs, 16 were also classified as positive, 4 as null (rs6595443, rs17507636, rs13338946, rs877529) and 3 as negative (rs7193541, rs6066835, rs138740) (Table 2, column 'MGUS OR'). Only for rs6066835 and rs138740 the 95%CIs did not overlap between MGUS and MM.

The results are summarized in Supplementary Figure 4 for each of the 31 loci. The 8 MGUS-specific loci are marked with stars. Functional considerations of the associated SNPs are based on data from the available annotation tools (Supplement).

In the interpretation of the genetic profiles between the two plasma cell dyscrasias one needs to consider the prevalence of these conditions, MGUS being more than 10 times more common than MM. Thus a higher OR for a SNP in MGUS compared to MM may imply that the related

gene predisposes to MGUS but does not contribute to progression to the end disease. If the OR for the SNP is increased to an equal extent in MGUS and MM, the SNP is likely to be predisposing equally to MGUS and to the end disease. If the risk is increased only in MM the related gene is likely to be predisposing to it. A large difference between the risk for MGUS and the end disease may indicate selection towards the risk genotype. A limitation of the study is that we have longitudinal follow-up data for a relatively short time not allowing flagging of the MGUS cases developing MM.

Applying these guidelines, and keeping in mind the caveat of limited sample sizes, the data suggest 8 loci to be specific to MGUS. These included 4 loci which marked genes *SGMS2, RIMS2* and *TSNARE1*, and one with limited biological data. The 4 other loci marked genes that had known functions in cancer: *PROX1, SFMBT, RAD51B* and *CSNK1G1*. PROX1 interacts with GATA2, an important regulator of hematopoiesis and roles in predisposition to myelodysplastic syndrome/acute myeloid leukemia [7]. SFMBT is a polycomb group epigenetic regulator with suggested functions in prostate cancer. The DNA repair gene *RAD51C* has been associated with germline mutations in common cancers [8, 9]. Casein kinase 1 gamma 1 isoforms contribute plasma cell survival [10-12]. According to the above convention these genes may be important in MGUS but the variants are selected against during progression to the end diseases.

The second group of SNPs was shared by the two dyscrasias and ORs tended to be higher in MGUS than in MM. These included SNPs marking *DTNB*, a gene encoding a component of the dystrophin-associated protein complex, *ULK4*, encoding a key regulator of mTOR-mediated autophagy [13], and *TNFRSF13B* encoding a key regulator of B- and T-cell functions with involvement in pathophysiology of MM [14]. The association of rs4487645 (IRF4 binding site) was weaker for MGUS than for MM while for rs11086029 (19p13.11, *KLF2*) the association was slightly higher. The data suggest that the underlying gene functions are vital for MGUS and MM but they appear most important for the end disease.

As the final group, the SNPs marked by genes, *PREX1* and *TOM1*, were not MGUS related. However, it is intriguing that *PREX1* appeared as a major interacting partner when genome-wide two gene-interactions were tested in MGUS [15]; in the same study *TOM1L1*, an analogue of *TOM1* was also a significant interaction partner. Individually, neither *PREX1* nor *TOM1*

associated with MGUS in this or in the earlier study [4]. If the role of *PREX1* or *TOM1* could be replicated in a larger interaction study the combined results could be explained with a model of higher enrichments of functional variants in the end disease. Additionally SNPs marked by genes *CEP120, CBX7* and *RFWD3* were enriched in MM compared to MGUS.

In summary, the present GWAS on MGUS appears to be capable of delineating distinctions between MGUS and MM germlines. Associations with the *PROX1, SFMBT, RAD51B and CSNK1G1* loci were only found for MGUS which may suggest that they are less important in the course of progression to MM. These genes have known functions in plasma cells and/or carcinogenesis, including homeobox transcription factor (PROX1) interacting with GATA2, chromatin remodeling through histone modification (SFMBT), double-stranded DNA repair (RAD51B) and cell cycle checkpoint arrest, DNA repair and Wnt signaling (CSNK1G1). These are the functions that have been proposed to the SNPs identified in MM (chromatin remodeling, B-cell development and cell cycle/genomic stability); additionally apoptosis/autophagy pathways were suggested for MM for which we did not find evidence in MGUS [3]. The association with *TNFRSF13* was stronger in MGUS compared to MM but the reverse was the case for the SNP forming the IRF4 binding site. *PREX1* and *TOM1* associations were only found in MM. If such distinctions can be verified in independent studies on MGUS they advance molecular understanding of the progression process, of the related prognostic markers and of the possible targets for intervention.

ACKNOWLEDGEMENT**S**

Supported by Multiple Myeloma Research Foundation, the German Ministry of Education and Science (01ZX1309B), the Harald Huppert Foundation and Deutsche Krebshilfe. Research by the Houlston Group is supported by grants from myeloma UK and BLOODWISE. PV and LV are recipients of a support from UNCE-MED006 and PROGRESS Q28.

CONFLICTS OF INTEREST

None.

Supplementary Information can be found online at the Leukemia website.

REFERENCES

1. Merlini G, Palladini G. Differential diagnosis of monoclonal gammopathy of undetermined significance. *Hematology / the Education Program of the American Society of Hematology American Society of Hematology Education Program* 2012; **2012:** 595-603.

2. Wadhera RK, Rajkumar SV. Prevalence of monoclonal gammopathy of undetermined significance: a systematic review. *Mayo Clinic proceedings* 2010 Oct; **85**(10)**:** 933-942.

3. Went M, Sud A, Forsti A, Halvarsson BM, Weinhold N, Kimber S, *et al.* Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma. *Nature communications* 2018 Sep 13; **9**(1)**:** 3707.

4. Thomsen H, Campo C, Weinhold N, Filho MI, Pour L, Gregora E, *et al.* Genome-wide association study on monoclonal gammopathy of unknown significance (MGUS). *European journal of haematology* 2017 Apr 04; **99:** 70-79.

5. Weinhold N, Johnson DC, Rawstron AC, Forsti A, Doughty C, Vijayakrishnan J, *et al.* Inherited genetic susceptibility to monoclonal gammopathy of unknown significance. *Blood* 2014 Jan 21; **123:** 2513-2517.

6. 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. *Nature Genet* 2013 Oct; **45**(10)**:** 1221-1225.

7. Kazenwadel J, Betterman KL, Chong CE, Stokes PH, Lee YK, Secker GA, *et al.* GATA2 is required for lymphatic vessel valve development and maintenance. *J Clin Invest* 2015 Aug 3; **125**(8)**:** 2979-2994.

8. Golmard L, Castera L, Krieger S, Moncoutier V, Abidallah K, Tenreiro H, *et al.* Contribution of germline deleterious variants in the RAD51 paralogs to breast and ovarian cancers. *Eur J Hum Genet* 2017 Dec; **25**(12)**:** 1345-1353.

9. Bennett JA, Braga AC, Pinto A, Van de Vijver K, Cornejo K, Pesci A, *et al.* Uterine PEComas: A Morphologic, Immunohistochemical, and Molecular Analysis of 32 Tumors. *Am J Surg Pathol* 2018 Oct; **42**(10)**:** 1370-1383.

10. Bjorklund CC, Ma W, Wang ZQ, Davis RE, Kuhn DJ, Kornblau SM, *et al.* Evidence of a role for activation of Wnt/beta-catenin signaling in the resistance of plasma cells to lenalidomide. *The Journal of biological chemistry* 2011 Apr 1; **286**(13)**:** 11009-11020.

11. Manni S, Carrino M, Manzoni M, Gianesin K, Nunes SC, Costacurta M, *et al.* Inactivation of CK1alpha in multiple myeloma empowers drug cytotoxicity by affecting AKT and beta-catenin survival signaling pathways. *Oncotarget* 2017 Feb 28; **8**(9)**:** 14604-14619.

12. Hu Y, Song W, Cirstea D, Lu D, Munshi NC, Anderson KC. CSNK1alpha1 mediates malignant plasma cell survival. *Leukemia* 2015 Feb; **29**(2)**:** 474-482.

13. Lebovitz CB, Robertson AG, Goya R, Jones SJ, Morin RD, Marra MA*, et al.* Cross-cancer profiling of molecular alterations within the human autophagy interaction network. *Autophagy* 2015; **11**(9)**:** 1668-1687.

14. Hengeveld PJ, Kersten MJ. B-cell activating factor in the pathophysiology of multiple myeloma: a target for therapy? *Blood cancer journal* 2015 Feb 27; **5:** e282.

15. 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. *Molecular medicine (Cambridge, Mass)* 2018 Jun 11; **24**(1)**:** 30.

Table 1 Meta-analyzed odds ratios for the most significant MGUS SNPs ($P < 10^{-5}$) compared with meta-analyzed odds ratios for MM

				MGUS			MM			
Chromosomal band	**SNP**	**Base pair (hg19)**	**Risk allele**	**OR $_{Meta}$**	**95% CI**	***P*-value**	**OR $_{Meta}$**	**95% CI**	***P*-value**	**GENCODE gene**
1q32.3	rs3009934	214301323	T	**1.32**	**1.18-1.48**	**2.0×10^{-06}**	0.99	0.93-1.05	7.9×10^{-01}	86kb 3' of PROX1
3p22.1	rs9848754	41753647	T	**1.44**	**1.25-1.67**	**5.6×10^{-07}**	**1.26**	**1.17-1.35**	**2.3×10^{-10}**	ULK4 intron
3q13.11	rs73180532	104051156	C	**1.28**	**1.15-1.42**	**6.8×10^{-06}**	1.00	0.95-1.06	9.6×10^{-01}	278kb 5' of AC016970.1
4q25	rs72888948	108802381	T	**1.48**	**1.25-1.76**	**5.1×10^{-06}**	1.07	0.97-1.18	1.6×10^{-01}	SGMS2 intron
8q22.3	rs9656789	105068489	A	**1.37**	**1.19-1.56**	**3.4×10^{-06}**	1.04	0.97-1.12	2.6×10^{-01}	RIMS2 intron
8q24.3	rs4928692	143466597	G	**1.38**	**1.21-1.59**	**2.5×10^{-06}**	1.10	1.02-1.18	1.2×10^{-02}	TSNARE1 intron
10p14	rs7920332	7250346	C	**1.27**	**1.14-1.42**	**7.1×10^{-06}**	1.00	0.93-1.08	9.1×10^{-01}	SFMBT2 intron
14q24.1	rs12436964	69108086	T	**1.31**	**1.17-1.47**	**2.4×10^{-06}**	1.06	1.00-1.13	4.0×10^{-02}	RAD51B
15q22.31	rs4561409	64535700	C	**1.30**	**1.16-1.45**	**6.3×10^{-06}**	0.97	0.92-1.03	4.0×10^{-01}	CSNK1G1 intron
17p11.2	rs74998556	16839782	T	**1.46**	**1.25-1.69**	**9.0×10^{-07}**	1.18	1.09-1.28	5.7×10^{-05}	TNFRSF13B

I^2 values were consistent with homogeneity of all meta-analyzed ORs
P_{het} was >0.5 for all GWAS ORs.
Bolding indicate significance at suggestive threshold of 10^{-5}
Abbreviations:
SNP, single nucleotide polymorphism; hg19, human genome NCBI build 19; OR $_{Meta}$, meta-analyzed odds ratio; CI, confidence interval

Table 2 Published multiple myeloma risk loci in MGUS meta-analysis

Published SNP	Best MGUS SNP*	CHR	Associated gene**	Base-pair position (hg19)	Published MM OR§	MGUS OR&	95% CI	Meta *P*	Risk allele	Other Allele	D'	r^2	P_{het}***	I^2***
rs6746082		2p23.3	DTNB	25659244	1.23#	1.19	1.04-1.35	**$8.2x10^{-03}$**	A	C			0.46	0.00
	rs201567111			25745570		1.31	1.15-1.49	**$3.3x10^{-05}$**	CT	C	0.71	0.36	0.63	0.00
rs4325816		2q31.1	**SP3**	174808899	1.12	1.19	1.05-1.35	**$6.0x10^{-03}$**	T	C			0.79	0.00
rs1052501		3p22.1	ULK4	41925398	1.26#	1.38	1.19-1.59	**$1.1x10^{-05}$**	C	T			0.30	18.07
	rs9848754			41753647		1.44	1.25-1.67	**$5.6x10^{-07}$**	T	C	1.00	1.00	0.32	12.07
rs10936599		3q26.2	ACTRT3, MYNN, **LRRC34**	169492101	1.20#	1.10	0.98-1.25	$9.6x10^{-02}$	C	T			0.69	0.00
rs56219066		5q15	**ELL2**	95242931	1.16#	1.10	0.97-1.25	$1.1x10^{-01}$	T	C			0.32	13.19
rs6595443		5q23.2	**CEP120**	122743325	1.11	1.03 O	0.92-1.14	$6.4x10^{-01}$	A	T			0.24	29.18
rs34229995		6p22.3	JARID2	15244018	1.36	1.37	1.03-1.81	**$2.9x10^{-02}$**	G	C			0.39	0.00
rs2285803		6p21.3	PSORS1C1, CCHCR1	31107258	1.21#	1.11	0.99-1.24	$5.4x10^{-02}$	T	C			0.47	0.00
rs9372120		6q21	ATG5, **PRDM1**	106667535	1.19	1.11	0.97-1.25	$1.2x10^{-01}$	G	T			0.64	0.00
rs4487645		7p15.3	DNAH11, **CDCA7L**	21938240	1.24	1.17	1.04-1.31	**$6.1x10^{-03}$**	C	A			0.59	0.00
	rs56249828			21944607		1.19	1.06-1.34	**$1.8x10^{-03}$**	T	C	0.95	0.87	0.85	0.00
rs17507636		7q22.3	CCDC71L	106291118	1.12	1.02 O	0.90-1.15	$7.4x10^{-01}$	C	T			0.73	0.00
rs58618031		7q31.33	POT1	124583896	1.12	1.07	0.95-1.20	$2.4x10^{-01}$	T	C			0.08	60.92
rs7781265		7q36.1	**ABCF2**, CHPF2, SMARCD3	150950940	1.22	1.22	1.03-1.43	**$1.9x10^{-02}$**	A	G			0.09	59.05
	rs219228			150939396		1.25	1.08-1.43	**$1.7x10^{-03}$**	C	A	0.94	0.48	0.6	0.00
rs1948915		8q24.21	**MYC**	128222421	1.15	1.20	1.07-1.33	**$1.5x10^{-03}$**	C	T			0.40	0.00
rs2811710		9p21.3	CDKN2A, **MTAP**, CDKN2B-AS1	21991923	1.14	1.15	1.02-1.28	**$1.6x10^{-02}$**	C	T			0.44	0.00
rs2790457		10p12.1	**WAC**	28856819	1.11	1.10	0.97-1.24	$1.1x10^{-01}$	G	A			0.40	0.00
rs13338946		16p11.2	PRR14, **FBRS**, SRCAP	30700858	1.15	1.04 O	0.92-1.17	$4.7x10^{-01}$	C	T			0.04	68.05
rs7193541		16q23.1	**RFWD3**, GLG1	74664743	1.12	0.98 N	0.87-1.09	$7.2x10^{-01}$	T	C			0.57	0.00
rs4273077		17p11.2	**TNFRSF13B**	16849139	1.30#	1.40	1.19-1.64	**$2.8x10^{-05}$**	G	A			0.25	27.45
	rs74998556			16839782		1.46	1.25-1.69	**$9.0x10^{-07}$**	T	A	1.00	0.66	0.23	30.25
rs11086029		19p13.11	**KLF2**	16438661	1.14	1.17	1.03-1.33	**$1.5x10^{-02}$**	T	A			0.63	0.00
	rs71178685			16443718		1.21	1.06-1.39	**$4.3x10^{-03}$**	TA	T	0.97	0.90	0.76	0.00
rs6066835		20q13.13	PREX1	47355009	1.23	0.92 N	0.74-1.12	$4.2x10^{-01}$	C	T			0.46	0.00
rs138740		22q13	HMGXB4, TOM1	35699582	1.21#	0.96 N	0.85-1.06	$4.1x10^{-01}$	C	T			0.51	0.00
rs877529		22q13.1	**CBX7**	39542292	1.22#	1.08 O	0.96-1.19	$1.7x10^{-01}$	A	G			0.78	0.00

[*] in case published SNP did not give the strongest signal
[**] Candidate causal gene, suggested by Went et. al. 2018, is indicated in bold
[§] Went et. al. 2018
[&] O=null in Asset analysis; N=negative in Asset analysis
[***] P_{het} and I^2 values were calculated for heterogeneity between the MGUS populations
[#] In Went et. al. the best SNP in the risk locus differs from the previously published one