Human Molecular Genetics, 2021, Vol. 30, No. 12 1142–1153

doi: 10.1093/hmg/ddab066
Advance Access Publication Date: 5 March 2021
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Sequencing at lymphoid neoplasm susceptibility loci maps six myeloma risk genes

Rosalie Griffin Waller[1], Robert J. Klein[2], Joseph Vijai[3,4], James D. McKay[5], Alyssa Clay-Gilmour[6,7], Xiaomu Wei[4], Michael J. Madsen[8], Douglas W. Sborov[8], Karen Curtin[8], Susan L. Slager[9], Kenneth Offit[3,4], Celine M. Vachon[6], Steven M. Lipkin[4], Charles Dumontet[10] and Nicola J. Camp[8,*]

[1]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT 84112, USA, [2]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, Icahn Institute for Data Science and Genomic Technology, New York, NY 10029-5674, USA, [3]Department of Medicine, Clinical Genetics Service, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA, [4]Department of Medicine, Weill Cornell Medical College, New York, NY 10065, USA, [5]Genetic Cancer Susceptibility, International Agency for Research on Cancer, 69372 Lyon Cedex 08, France, [6]Department of Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN 55905, USA, [7]Department of Epidemiology & Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, USA, [8]Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA, [9]Department of Health Sciences, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN 55905, USA and [10]INSERM 1052, CNRS 5286, University of Lyon, 69361 Lyon Cedex 07, France

*To whom correspondence should be addressed at: Huntsman Cancer Institute, University of Utah, 2000 Circle of Hope, Rm 4757, Salt Lake City, UT 84112, USA. Tel: +1 8015879351; Fax: 801-587-9351; Email: nicola.camp@hci.utah.edu

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

Abstract

Inherited genetic risk factors play a role in multiple myeloma (MM), yet considerable missing heritability exists. Rare risk variants at genome-wide association study (GWAS) loci are a new avenue to explore. Pleiotropy between lymphoid neoplasms (LNs) has been suggested in family history and genetic studies, but no studies have interrogated sequencing for pleiotropic genes or rare risk variants. Sequencing genetically enriched cases can help discover rarer variants. We analyzed exome sequencing in familial or early-onset MM cases to identify rare, functionally relevant variants near GWAS loci for a range of LNs. A total of 149 distinct and significant LN GWAS loci have been published. We identified six recurrent, rare, potentially deleterious variants within 5 kb of significant GWAS single nucleotide polymorphisms in 75 MM cases. Mutations were observed in *BTNL2*, *EOMES*, *TNFRSF13B*, *IRF8*, *ACOXL* and *TSPAN32*. All six genes replicated in an independent set of 255 early-onset MM or familial MM or precursor cases. Expansion of our analyses to the full length of these six genes resulted in a list of 39 rare and deleterious variants, seven of which segregated in MM families. Three genes also had significant rare variant burden in 733 sporadic MM cases compared with 935 control individuals:

Received: November 16, 2020. **Revised:** February 22, 2021. **Accepted:** February 23, 2021

© The Author(s) 2021. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

IRF8 ($P = 1.0 \times 10^{-6}$), *EOMES* ($P = 6.0 \times 10^{-6}$) and *BTNL2* ($P = 2.1 \times 10^{-3}$). Together, our results implicate six genes in MM risk, provide support for genetic pleiotropy between LN subtypes and demonstrate the utility of sequencing genetically enriched cases to identify functionally relevant variants near GWAS loci.

Introduction

Multiple myeloma (MM) is the second most common hematological cancer after non-Hodgkin lymphoma and has the worst 5-year survival (1). Risk factors for MM include a personal history of monoclonal gammopathy of undetermined significance (MGUS) (2,3), older age, male sex, obesity, Black race and a family history of cancer (4–7).

Inherited germline susceptibility for MM is consistently supported. Yet, the molecular basis is only partly understood (8). Excess MM risk among relatives has been observed in family aggregation (9,10), epidemiologic case-control (11–15), and registry-based (16,17) studies. Genome-wide association studies (GWASs) have identified 36 common single nucleotide polymorphism (SNPs) at 24 loci, which significantly associate with MM (18–23); however, this polygenic background does not explain familial cases (24). Recent exome studies have identified rare germline variants predisposing to MM in *ARID1A* and *USP45* (25), *KDM1A* (26) and *DIS3* (27). However, considerable missing heritability remains.

Rare risk variants at GWAS loci have not been explored in MM. Such variants could provide insight into the identity of the target genes at GWAS loci or provide additional new risk factors. High-throughput sequencing allows for the detection of rare and deleterious variants that genotype arrays and fine mapping cannot detect. However, sequencing studies in a GWAS design of population cases and controls would require thousands of cases for the detection of rare risk variants (28). Cost, lower incidence and poor survival in MM make such large sample sizes a challenge to attain, and so far, such resources do not exist. A focus on genetically enriched cases, such as early-onset or familial cases, holds promise and has previously been used successfully in mapping genes (29).

Familial co-aggregation of MM with other lymphoid neoplasms (LNs) provides evidence for genetic pleiotropies (30). A few studies have begun to explore methods to identify pleiotropy in the GWAS design (31,32), and converging SNP signals between MM and other LNs are starting to appear (33,34). Chronic lymphocytic leukemia (CLL) and MM have been shown to have genetic correlation using cross-trait linkage disequilibrium (LD) score regression (34). In a study specific to MM, CLL and Hodgkin's lymphoma (HL), significant pleiotropies were found between MM and CLL at 2p23.3, 2q13 and 11q24.1, between MM and HL at 3p22.1, and between MM and both CLL and HL at 3q26.2 (33). At the 11q24.1 locus, separate GWAS studies identified the same GWAS SNP in CLL (35–38) and follicular lymphoma (FL) (39). No studies to date have interrogated sequencing for pleiotropic genes or rare risk variants in MM.

We hypothesize that rare risk variants co-exist at GWAS loci and pleiotropic susceptibility exists between MM and other LN subtypes. We leverage MM exome sequencing resources focused on genetically enriched cases (early-onset or familial), restrict attention to previously published significant LN GWAS loci and use a discovery-replication design to increase robustness.

Results

Discovery

We investigated coding regions within 5 kb of the 149 LN GWAS SNPs (18–23,35–52). Six rare, recurrent and deleterious variants were identified in 15 of the 75 early-onset or familial MM cases (Table 1). The six variants were near 7 LN GWAS SNPs and include: (1) a butyrophilin-like 2 (*BTNL2*) frameshift variant (rs79379254) 139 bp from marginal zone lymphoma (MZL) GWAS SNP rs9461741 carried by two male, early-onset, African American MM cases; (2) an eomesodermin (*EOMES*) missense variant (rs200789175, 853 bp from HL GWAS SNP rs3806624) carried by three female MM cases; (3) a tumor necrosis factor receptor superfamily member 13B (*TNFRSF13B*) missense variant (rs34557412, 3048 bp from MM GWAS SNP rs4273077) carried by three MM cases; (4) an interferon regulatory factor 8 (*IRF8*) missense variant (rs28368114, 3391 bp from CLL GWAS SNP rs1044873) carried by three MM cases; (5) an acyl-coenzyme A oxidase-like (*ACOXL*) stop-gain variant (rs189429375, 3482 bp from CLL GWAS SNP rs1439287) carried by four MM cases and (6) a tetraspanin 32 (*TSPAN32*) missense variant (rs61744929, 3777 bp and 4332 bp from CLL GWAS SNPs rs2651823 and rs2521269, respectively) carried by two male, early-onset MM cases. Empirical assessment for significance using bootstrap simulations demonstrated the discovery findings were significant ($P = 0.012$). None of the six rare discovery variants were in significant LD with their corresponding GWAS SNP in any 1000Genomes population.

Replication

We investigated coding regions within 5 kb of the seven LN GWAS SNPs from the discovery phase (Table 1). Seven recurrent, rare and deleterious variants were identified in 19 of 255 early-onset MM or familial MM or MGUS cases (Table 1). Four variants replicated the discovery variants exactly. Three additional risk variants were identified: (1) a *BTNL2* missense variant (rs375897659, 2473 bp from the discovery MZL GWAS SNP) carried by two familial, African American MM cases; (2) an *EOMES* missense variant (rs371656694, 4308 bp from the discovery HL GWAS SNP) carried by two MM cases and (3) an *IRF8* missense variant (rs144424711, 3648 base pairs from a discovery CLL SNP) carried by two MGUS and one MM case. All six genes harbored recurrent, rare and deleterious variants in replication cases and became potential target risk genes.

Gene extension

In total, the discovery and replication phases identified nine rare, recurrent and potentially deleterious variants close to (<5 kb) LN GWAS SNPs in six target genes: *BTNL2*, *EOMES*, *TNFRSF13B*, *IRF8*, *ACOXL* and *TSPAN32*. In the extension phase, we interrogated the full exome sequencing of these six genes in the combined 330 discovery and replication cases to identify additional rare and potentially deleterious variants beyond the 5 kb restriction. An additional 30 variants were identified (Table 2, Fig. 1). In total,

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

Table 1. Discovery and replication risk variants near LN GWAS SNPs

Locus	Exome variant							BP	LN GWAS SNP	
	Gene	Variant ID	Change	Type	DIS N	REP N	AAF		SNP ID	Trait
6p21.32	*BTNL2*	6-32 370 726-AC-A[a]	V232X	FS	2		0.00200	139	rs9461741	MZL
		6-32 373 060-T-A[b]	D28V	MS		2	0.00003	2473		
3p24.1	*EOMES*	3-27 763 770-G-C[a,b]	Q6E	MS	3	2	0.00384	853	rs3806624	HL
		3-27 760 315-C-T[b]	V410M	MS		2	0.00005	4308		
17p11.2	*TNFRSF13B*	17-16 852 187-A-G[a,b]	C104R	MS	3	3	0.00336	3048	rs4273077	MM
16q24.1	*IRF8*	16-85 952 280-G-A[a]	V287M	MS	3		0.00046	3391	rs1044873	CLL
		16-85 952 023-C-T[b]	A201V	MS		3	0.00276	3648		
2q13	*ACOXL*	2-111 875 379-G-T[a,b]	G577*	SG	4	4	0.00525	3482	rs1439287	CLL
11p15.5	*TSPAN32*	11-2 325 427-T-C[a,b]	M91T	MS	2	4	0.00855	3777	rs2651823	CLL
								4332	rs2521269	

Key: Variant ID positions GRCh37; Type, variant consequence predicted by VEP; FS, frameshift variant; MS, missense variant; SG, stop-gain variant; DIS N, number of discovery cases carrying the variant (total $N = 75$); REP N, number of replication cases carrying the variant (total $N = 255$); AAF, alternate allele frequency in gnomAD non-cancer samples; BP, base pairs between the exome variant and the GWAS SNP.
[a]Discovery variant.
[b]Replication variant.

we found 72 of 330 MM/MGUS cases carried at least one of the 39 rare risk variants in the six genes (Supplementary Material, Table S1). More than half of the rare variants identified had a high combined annotation-dependent depletion (CADD) score. Twenty-two were in the top 1% (CADD ≥20) and four were in the top 0.1% (CADD ≥30.0) of the most deleterious mutations in the human genome (Table 2).

Inheritance in familial cases. Seven variants in five of the target genes segregated in an MM family (Table 2, Fig. 2). In *BTNL2*, an MM sib-pair in an African American family share a missense variant (Fig. 2A); the father also has MM but was not sequenced. Two families harbor segregating *EOMES* variants. First, an MM–MGUS sib-pair share a missense variant; another sibling has MM but was not sequenced (Fig. 2B). Second, an uncle–nephew MM pair and an obligate CLL parent share a missense variant (Fig. 2E). In *IRF8*, an MM–MGUS parent–offspring pair share a missense variant; a sibling of the parent also has MGUS but does not carry the variant (Fig. 2D). Two families harbor segregating *ACOXL* variants. A stop-gain variant is carried by two early-onset MM cases and an obligate CLL carrier separated by eight meiosis in a large Utah pedigree (Fig 2G). An missense variant in *ACOXL* is shared by three siblings, two with MGUS and one with MM; a fourth sibling has lymphoma but was not sequenced (Fig. 2C). In *TSPAN32*, an missense variant is shared by an MM–MGUS sib-pair and a cousin with MM (Fig. 2F).

Population extension

Variant burden in sporadic MM. Eighty-five variants were observed in the six target genes in the 733 sporadic MM cases (Supplementary Material, Tables S1 and S2) of which 29 were replications of variants seen in the discovery or replication cases (Supplementary Material, Table S2). In total, 174 of the 733 (23.7%) sporadic MM cases carried at least one of the 85 variants. By gene, the results were as follows: *ACOXL*, 16 variants and 57 MM carriers (22 of African descent); *BTNL2*, 18 variants and 24 MM carriers (19 of African descent); *EOMES*, 11 variants and 21 MM carriers (5 of African descent); *IRF8*, 14 variants and 30 MM carriers (4 of African descent); *TNFRSF13B*, 9 variants and 26 MM carriers (4 of African descent); and *TSPAN32*, 17 variants and 40 MM carriers (5 of African descent). Most variants were missense; two were framshift, seven were stop-gain or -loss, and two were splice-acceptor or -donor variants (Supplementary Material, Table S2). The average CADD score was highest in *ACOXL* at 22.5, followed by *EOMES* at 19.4. To test for the overall variant burden in each target gene in the sporadic MM cases versus controls, we performed a c-alpha test. Accounting for multiple testing our significance threshold was $\alpha = 0.008$ (0.05/6 genes). Three genes had significant variant burden in the sporadic MM cases: *IRF8* ($P = 1.0 \times 10^{-6}$), *EOMES* ($P = 6.0 \times 10^{-6}$) and *BTNL2* ($P = 2.1 \times 10^{-3}$).

Discussion

Our findings implicate six genes in MM risk, provide support for genetic pleiotropy between LN subtypes, suggest genes at GWAS loci are important for familial and early-onset risk and demonstrate the utility of sequencing in a limited number of genetically enriched cases to map functionally relevant variants near GWAS SNPs.

Myeloma is a relatively rare malignancy, and the sample sizes necessary to explore sequence-based exome-wide association for rare variants are very large and currently do not exist. Hence, we chose a multi-phase strategy for discovery which exploited three hypotheses. First, pleiotropic genetic risk factors exist across LNs. Second, early-onset and familial MM cases are enriched for genetic risk. Third, both common and rare risk variants exist for the same susceptibility genes. Relying on these hypotheses as our foundation, we leveraged MM resources that were focused on genetically enriched cases, restricted our attention to regions of the exome very close to previously established genome-wide significant susceptibility loci for common risk across a range of LN subtypes and constructed phases and tests to best utilize the sequencing that exists in myeloma. While not all myeloma risk genes will fit these hypotheses, our strategy optimized our ability to prioritize variants and identify MM risk genes that do.

Consistent with our hypothesis of genetic pleiotropy between MM and other LN subtypes, five of our six resulting MM risk genes are located near GWAS SNPs for other LNs (Table 1). These findings are in line with prior work supporting a shared etiology, including family studies (30) and converging GWAS signals between LNs (32–34,53). Significant pleiotropies between MM and other LNs have been observed at 2p23.3 (with CLL), 2q13 (CLL), 3p22.1 (HL), 3q26.2 (CLL, HL) and 11q24.1 (CLL, FL) (33). We provide additional evidence at 2q13 and are the first to observe LN pleiotropy at 3p24.1, 6p21.32, 11p15.5 and 16q24.1 (Table 1).

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

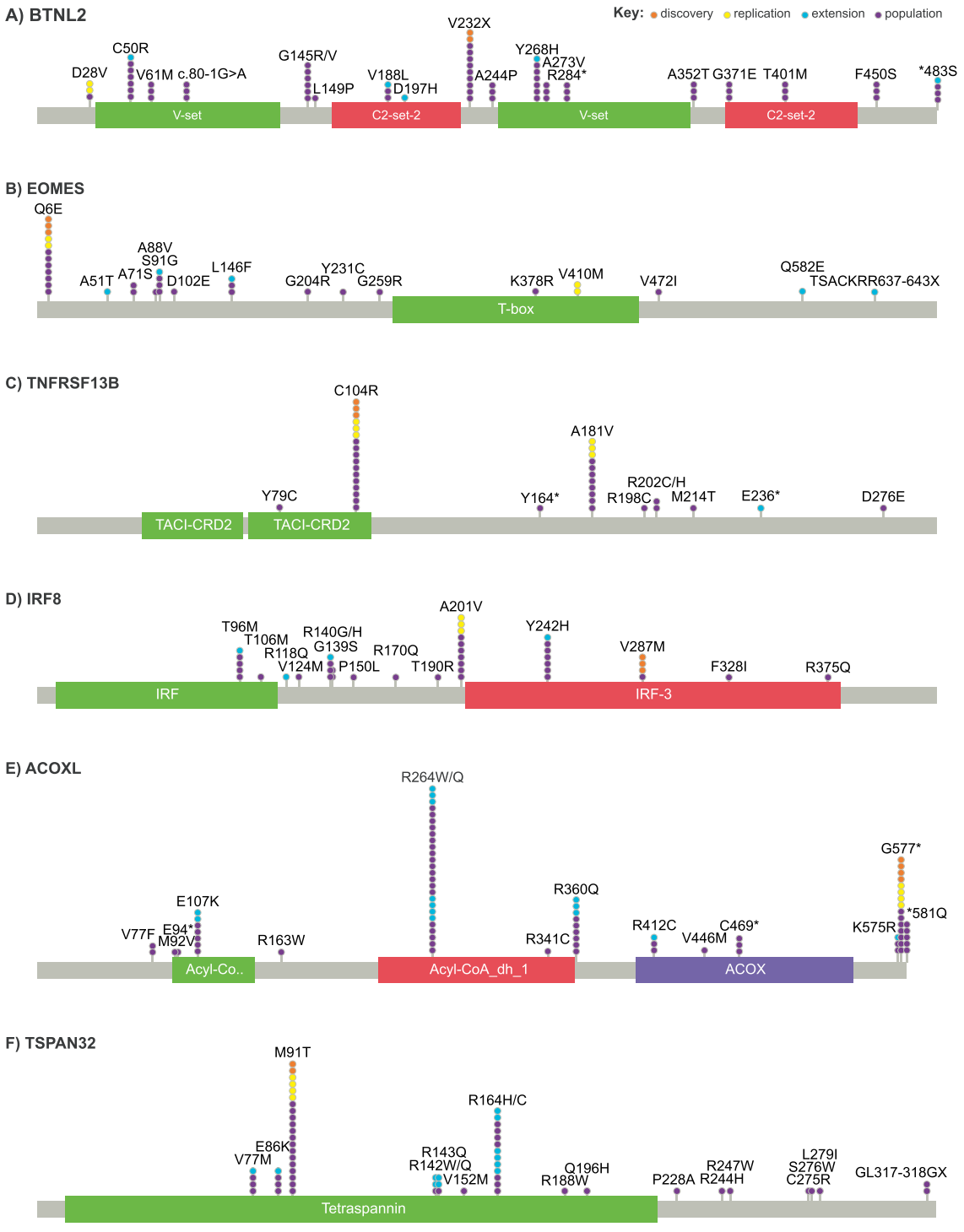

Figure 1. Variants observed in the six target genes. Each dot represents an MM or MGUS case with a variant resulting in an amino acid change at that position. Color corresponds to the study phase. Functional domains are annotated. We observed (**A**) 19 *BTNL2* variants, (**B**) 15 *EOMES* variants, (**C**) 10 *TNFRSF13B* variants, (**D**) 15 *IRF8* variants, (**E**) 18 *ACOXL* variants (splice variants not shown), and (**F**) 18 *TSPAN32* variants.

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

Table 2. Exome variants in the six genes of interest carried by individuals with MM or MGUS

Gene	Variant ID	Change	Type	CADD	N	AAF	Family (N carriers)
BTNL2	rs375897659[a]	D28V	MS	23.8	2	0.00003	MSKCC-5885 (3[c])
	rs79379254[b]	V232X	FS	18.0	2	0.00200	
	rs150942326	C50R	MS	24.0	1	0.00054	
	rs143816072	D197H	MS	14.6	1	0.00048	
	rs116715584	Y268H	MS	1.2	1	0.00054	
	rs9461742	V188L	MS	0.0	1	0.00118	
	rs186450060	*483S	SL	0.0	1	0.00038	
EOMES	rs200789175[a,b]	Q6E	MS	13.5	5	0.00384	Mayo-500 (2)
	rs371656694[a]	V410M	MS	28.0	2	0.00005	MSKCC-5066 (3[c])
	rs770923392	TSACKRR637-643X	FS	35.0	1	0.00015	
	rs200215171	L146F	MS	23.1	1	0.00085	
	rs373565897	A51T	MS	20.3	1	0.00012	
	rs770027537	Q582E	MS	17.7	1	0.00003	
	rs529375594	S91G	MS	8.1	1	0.00165	
TNFRSF13B	rs34557412[a,b]	C104R	MS	25.9	6	0.00336	
	rs72553883	A181V	MS	2.0	3	0.00541	
	rs201021960	E236*	SG	35.0	1	0.00015	
IRF8	rs28368114[b]	V287M	MS	25.7	3	0.00046	
	rs144424711[a]	A201V	MS	0.3	3	0.00276	INSERM-0608 (2)
	rs145048966	T96M	MS	33.0	1	0.00077	
	rs142267779	Y242H	MS	22.8	1	0.00179	
	rs200724492	R118Q	MS	11.6	1	0.00024	
	rs138032891	G139S	MS	7.9	1	0.00063	
ACOXL	rs189429375[a,b]	G577*	SG	37.0	8	0.00525	Utah-20245 (3[c])
	rs116012262	R264Q	MS	23.7	4	0.00150	INSERM-0909 (3)
	rs61730481	R264W	MS	27.6	3	0.00369	
	rs139505303	R360Q	MS	18.5	3	0.00428	
	rs150003283	E107K	MS	25.1	2	0.00695	
	rs150734728	R412C	MS	24.7	1	0.00244	
	rs142497134	—	SD	23.6	1	0.00030	
	rs113297147	—	SA	23.2	1	0.00007	
	rs574186205	K575R	MS	20.4	1	0.00024	
TSPAN32	rs61744929[a,b]	M91T	MS	19.4	6	0.00855	INSERM-0108 (3)
	rs148601311	R164C	MS	25.1	4	0.00295	
	rs142056890	R164H	MS	23.3	2	0.00201	
	rs2234302	R143Q	MS	10.2	2	0.00140	
	rs138129469	R142Q	MS	6.1	2	0.00027	
	rs747759649	E86K	MS	25.0	1	0.00006	
	rs2234296	V77M	MS	2.4	1	0.00178	

Key: SL, stop loss; SG, stop gain; SD, splice donor; SA, splice acceptor; CADD scores are PHRED-like [−10 x log10(rank/total)] scaled and rank each variant relative to all possible substitutions of the human genome (8.6×10^9). A score >10 indicates the variant is in the top 10% most deleterious substitutions in the human genome, a score >20 indicates the top 1% most deleterious and a score >30 indicates the top 0.1%; N, number discovery and replication MM/MGUS cases carrying the variant; AAF, allele frequency in gnomAD non-cancer samples; Family, family with segregating variant (number of MM/MGUS cases carrying the variant in the family).
[a]Discovery variant.
[b]Replication variant.
[c]includes an obligate MM or CLL carrier.

We also provide the first rare and deleterious variants in support of genetic pleiotropy, specifically between MM and CLL, MZL or HL.

Evidence for genetic pleiotropy was most pronounced between CLL and MM. Three genes in CLL susceptibility loci (*IRF8*, *ACOXL* and *TSPAN32*) harbored MM risk variants, many within 5 kb of four CLL SNPs (Table 1). Further, we identified parent–child CLL–MM variant sharing for *ACOXL* (Fig. 2G). Prior literature also supports pleiotropy between CLL and MM with converging evidence at known or novel GWAS loci (33,34). Our findings provide additional evidence for CLL–MM genetic pleiotropy at 2q13 (*ACOXL*), extending to rare risk variants and provide novel evidence at 11p15.5 (*TSPAN32*) and 16q24.1 (*IRF8*).

We are the first to observe evidence for pleiotropy between MM and MZL with *BTNL2* at 6p21.32. Of note, our discovery variant was only 139 base pairs from an MZL GWAS SNP (Table 1). This gene is especially intriguing because most carriers of *BTNL2* variants were of African ancestry (Supplementary Material, Table S1), including evidence for segregation in an MM family with African ancestry (Fig. 2A). It is well established that incidence of MM is twice as common in people of African descent (6). Our observations suggest *BTNL2* may be of particular interest to African MM risk. Conversely, MZL is more common in non-Hispanic Whites (54). Given the European predominance in the MZL GWAS (52), it is unknown if *BTNL2* may also be important in African MZL risk. *BTNL2* is in the HLA region—an area difficult

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

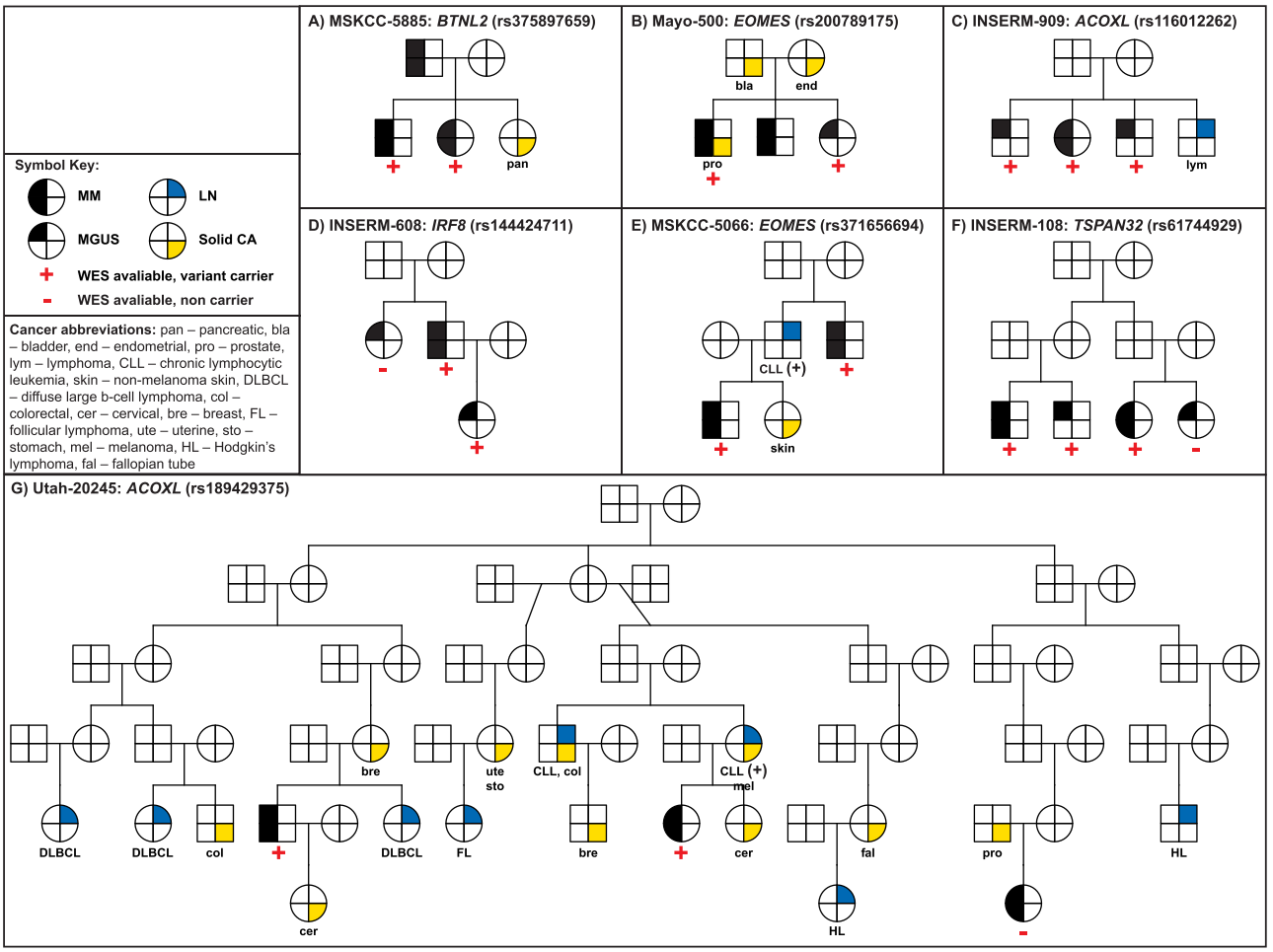

Figure 2. Variants shared with the disease in high-risk MM families. Unaffected relatives not shown. +/- indicates germline exome sequencing available and the individual carries or does not carry the variant, respectively. Circle indicates female and square indicates male. (**A**) An MM sib-pair share rs375897659 in *BTNL2*; the father has MM but was not sequenced. (**B**) An MM-MGUS sib-pair share rs200789175 in *EOMES*; another sibling has MM but was not sequenced. (**C**) rs116012262 in *ACOXL* is shared by three siblings, two with MGUS and one with MM; a fourth sibling has lymphoma but was not sequenced. (**D**) An MM-MGUS parent-offspring pair share a rs144424711 in *IRF8*; a sibling of the parent has MGUS but does not carry the variant. (**E**) An uncle-nephew MM pair and an obligate CLL parent share rs371656694 in EOMES. (**F**) rs61744929 in *TSPAN32* is carried by an MM-MGUS sib-pair and a cousin with MM. (**G**) rs189429375 in *ACOXL* is carried by two early-onset MM cases and an obligate CLL carrier.

to sequence and associated with many LN subtypes. However, we only considered variants passing high-quality standards and observed in gnomAD. Additional research into the biology of *BTNL2* and its role in MM risk for African cases and genetic pleiotropy between MM and MZL is warranted.

Our study is the first to provide support for pleiotropy between MM and HL at 3p24.1 (*EOMES*). A GWAS has also indicated 3p24.1 in CLL risk (42), and *EOMES* has previously been suggested as pleiotropic between CLL and HL (40). Our gene extension observed an MM family with parent–child CLL–MM variant sharing in *EOMES* (Fig. 2E). Hence, our findings support *EOMES* pleiotropies that extend to MM, HL and CLL. Additional research into the role of *EOMES* in MM, HL and CLL is needed to better understand the biology behind our results supporting pleiotropy.

The rarity of LNs together with sequencing costs continue to make the assembly of large cohorts of sequenced MM cases a challenge. Cases thought to have a high genetic component of disease have been successful in gene mapping in the face of limited data availability. Previous studies have also illustrated the advantage of a focus on genetically enriched cases and support the relevance of GWAS findings in familial and early-onset cases (24). Our design invoked the power of genetically enriched cases and the existence of both rare and common variants in risk genes. Consistent with this, we found that genes could be identified at GWAS loci (regions identified by common risk variants) that harbored a significant excess of rare, deleterious variants in MM cases with a familial and early-onset disease. This speaks to the likelihood that cancer susceptibility genes harbor risk alleles across the allele frequency spectrum and that complementary designs that search for common and/or rare variants will be needed to fully describe genetic etiologies.

Our results demonstrate the utility of sequencing in a limited number of early-onset and familial cases to identify functional variants near GWAS loci. However, the rare coding variants were not found to be in high LD with their respective GWAS SNPs. This would suggest that the rare variants we discovered are additional risk alleles that affect the same target genes and are not functional rare variants on haplotypes with their common counterparts to explain the GWAS signals. We acknowledge that LD measures are imprecise for rare exome variants and may have limited our ability to detect high levels of LD. The future

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

existence of immense sequenced populations will increase the power to measure LD in rare variants, and our results may need to be reassessed in this regard. Additional signals may have been missed if the common variant regulates a gene more than 5 kb away or falls in a very large intron. Nonetheless, our study suggests that GWAS loci may help guide discovery of rare risk variation, which will be especially valuable as the field turns to less well-annotated and non-coding regions of the genome.

Existing literature supports a role for the genes in immune disease and cancers. Proteins encoded by five of the genes we propose have been shown to have a role in immune regulation. BTNL2 decreases T-cell proliferation and cytokine release, acting as a negative T-cell regulator (55). EOMES regulates differentiation of CD8+ T-cells and is a critical regulator of tumor immune responses mediated by T-cells (56). TNFRSF13B plays a role in B-cell survival and maturation, is involved in antibody production and promotes cell signaling (57–59). IRF8 regulates expression of viral-induced genes and is a critical lineage-specific transcription factor for myeloid cell differentiation (60,61). TSPAN32 is involved in hematopoietic cell function (62,63) and has biased expression in the bone marrow (64).

IRF8 is focally deleted across many tumors, is likely a tumor suppressor in humans (65) and was recently shown to have frequent alterations in primary mediastinal large B-cell lymphoma (66). Patients with high levels of IRF8 expression experienced prolonged overall survival in clear cell renal sarcoma (65). Rare *BTNL2* variants have been shown to increase susceptibility to familial and sporadic prostate cancer (67). Notably, prostate cancer is also more prevalent in men of African descent (68), supporting the role of *BTNL2* in African cancer risk. Mutations in *TNFRSF13B* have also been associated with common variable immune deficiency (CVID), a condition that impairs the immune system (69–71). Patients with CVID have increased risk of recurrent infections, autoimmune disorders and non-Hodgkin lymphoma (72). High levels of TNFRSF13B have been shown to promote MM cell growth (73,74).

Substantial effort was made to design a study that played to the strengths of existing resources. Nonetheless, our results and the study are limited by the observational nature. Further studies are needed to demonstrate the functional relevance of specific variants in the genes we identified and the mechanisms for risk in MM and to better understand the biology behind the shared genetic etiology that our results suggest. The study was limited to genome-wide significant GWAS loci. Additional rare variants may exist at suggestive GWAS loci and could be explored in future work. Further, evidence for genetic pleiotropy between MM and other cancers and non-cancer phenotypes exists, and these could be pursued using this design.

In summary, we implicate six genes in MM risk, provide support for a shared genetic etiology between LN subtypes and demonstrate the utility of sequencing genetically enriched cases to identify functional variants near GWAS loci. Our findings are some of the first to identify putative functional variants near GWAS loci for LNs and to provide opportunities for focused functional studies to establish causation and better understand the disease biology. A better understanding of LN pleiotropy will help design studies to gain insight into the biology behind subtypes and uncover the remaining missing heritability. To date, only a handful of genes have been implicated in familial or early-onset myeloma risk (25–27). The identification six additional genes provide opportunities for focused functional studies to better understand the biology behind myeloma risk and may eventually lead to opportunities for genetic testing and counseling.

Materials and Methods

We evaluated exome sequencing near 149 published LN GWAS SNPs in 75 familial or early-onset MM cases for rare deleterious variants, replicated our findings in 255 familial or early-onset MM or precursor disease cases MGUS and extended interrogation of replicated genes for overall variant burden in 733 sporadic MM cases.

Ethics statement

Ethics committees at the University of Utah, Memorial Sloan Kettering, Mayo Clinic, Weill-Cornell and Comite´ de Protection des Personnes-SUD EST IV approved this research. All participants provided written informed consent.

Exome sequencing and processing

For all phases, DNA extracted from whole blood or saliva was sequenced using an Agilent SureSelect Human All Exon V4 or V5 + UTRs capture kit. Sequencing from all participants was jointly called at the Icahn School of Medicine at Mt. Sinai based on the genomic analysis toolkit (GATK) best practices (75). Briefly, FASTQ files were aligned to genome build 37 using Burrows-Wheeler Aligner version 0.7.8, indels were realigned using GATK, duplicates were removed using Picard MarkDuplicates and base quality scores were recalibrated using GATK. HaplotypeCaller generated individual genomic variant call format (GVCF) files for each individual, and GenotypeGVCFs produced the final joint calling. Variant quality score recalibration (VQSR) removed likely false-positive variant calls. The jointly called variant call format (VCF) file was annotated with variant effect predictor (VEP) (76) and queried using BCFtools (77).

Risk variant identification

In all phases, variants of interest were restricted to high-quality, rare and deleterious variants that exist in gnomAD. GnomAD version 2.1.1 exome VCFs were downloaded and utilized to filter participant variants based on quality and rarity (78). High-quality variants were defined as those that: passed variant quality filters (GATK VQSR excluded tranches 99.00-100 and gnomAD filters); had no more than two alternate alleles in gnomAD; were not in a known segmental duplication and were not within a larger indel (gnomAD star annotation). Rare was defined as an alternate allele frequency (AAF) ≤ 0.01 across all gnomAD non-cancer samples and AAF ≤ 0.05 in any non-cancer gnomAD subpopulation. Deleterious was defined as a high or moderate predicted impact based on VEP annotation. Additional criteria were required for each phase-based variant location and recurrence, as described in the following.

Discovery phase

The focus of this phase was to identify recurrently (observed at least twice), rare and deleterious variants within 5 kb of known LN GWAS SNPs in 75 discovery cases. The goal was to narrow attention to target genes harboring rare deleterious variants at GWAS loci.

Discovery cases. Seventy-five individuals from Utah with an MM diagnosis were selected based on membership in an extended high-risk pedigree or early-onset disease (≤45 years). Pedigrees were identified using the Utah Population Database and were selected for excess disease clustering ($P < 0.05$) through

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

many generations. Forty-three cases resided in a high-risk pedigree. Forty-one cases were early onset. Most cases were of northwest European descent ($N = 72$), three were of African descent. All self-reported race matched genetic ancestry.

LN GWAS regions. The GWAS Catalog (https://www.ebi.ac.uk/gwas/) was accessed on 12 March 2019, and SNPs that attained genome-wide significance ($P < 5 \times 10^{-8}$) were considered. We identified 149 significant SNPs from 24 LN studies (Supplementary Material, Table S3). LN subtypes with significant associations included: seven CLL studies with 78 SNPs (35–38,40–42), six MM studies with 36 SNPs (18–23), five HL studies with 20 SNPs (43–47), 10 FL studies with 10 SNPs (39,48–50), a diffuse large B-cell lymphoma (DLBCL) study with 5 SNPs (51) and an MZL study with 2 SNPs (52). Two SNPs were independently associated with two LN subtypes: rs735665 in CLL (35–38) and FL (39), and rs10936599 in CLL (35) and MM (22). We created 5 kb regions on either side of the 149 SNPs, which together defined 1 414 997 bp (some regions overlapped) of which 81 442 bp intersected with the exome sequencing captures. Hence, the discovery focus was on these 81.4 kb of exome capture (0.1% of the total 81 762 734 bp).

Significance assessment. To investigate the likelihood of observing the discovery results by chance, we repeated the process in local controls with individual-level sequence data, which were jointly called with the discovery cases. Whole-genome sequences were available on 134 unrelated, unaffected individuals from the Utah Centre d'Etude du Polymorphisme Humain project. We restricted the genomes to the 81.4 kb of exome capture used on the discovery cases, and variants of interest were filtered using the same criteria as in the discovery cases. To determine significance, we used a bootstrap approach. A random set of 75 controls were selected from the 134 individuals using VCFTOOLS (79), and the number and level of recurrence of rare, deleterious variants were determined. This process was repeated 1 000 000 times with sample replacement. An empirical *P*-value was determined as the proportion of control sets that attained simulated results equal to or exceeding the results observed in the discovery cases. This is a single, global test across the entire 81.4 kb region, and our discovery results were assessed at the $\alpha = 0.05$ level for significance.

LD between LN GWAS SNPs and rare risk variants. We assessed LD between the discovery rare risk variants and their corresponding GWAS SNPs using LDlink (80). Given the discrepancy in allele frequencies, we used D′ which indicates the proportion of maximum possible LD based on the allele frequencies. A statistically significant D′ close to 1.0 would indicate that the rare risk variant and GWAS risk SNP may lie on a haplotype and represent the same association signal (i.e. rare risk variant could drive the known common SNP association). Otherwise, different genetic risk factors may be influencing the same target gene. LD in all 1000 Genomes Project sub-populations was investigated.

Replication phase

The focus of this phase was to reproduce the discovery of recurrent (observed at least twice), rare and deleterious variants within 5 kb of known LN GWAS SNPs and in a target gene in 255 replication cases. The goal was to further narrow attention to target genes harboring rare deleterious variants which were observed in independent sample sets at GWAS loci.

Replication cases. An independent set of 255 individuals diagnosed with MM or MGUS with early onset MM (≤45 years) or a family history of MM (self-reported in close relatives: 70 MM, 46 MGUS). Those with a family history fall into 44 densely clustered families: 25 were ascertained by Institut National de la Santé et de la Recherche Médicale (INSERM), France (36 MM, 38 MGUS); 9 by Mayo Clinic, Minnesota (10 MM, 8 MGUS and 10 unaffected family members); 6 by Memorial Sloan Kettering Cancer Center (MSKCC), New York (14 MM); 3 by International Agency for Research on Cancer (IARC), France (8 MM) and one by Weill Cornell, New York (2 MM). Most of the families contained both MM and MGUS cases (32 families total), and six families contained a sequenced unaffected relative. Three families had African ancestry; the remainder were of northwest European descent.

Replication regions. Variants of interest at the replication phase were required to reside within the intersection of the 10 kb regions surrounding LN GWAS SNPs and the specific genes harboring recurrent rare risk variants in the discovery phase.

Extension phase

The full length of a gene may stretch beyond the 5 kb from an LN GWAS SNP. In this phase, we interrogated the full exome sequencing of the genes with variants in the discovery and replication phases. First, in *Gene Extension*, we identified all rare, deleterious variants in the early-onset or familial discovery and replication cases ($N = 330$) in the target genes and investigated co-sharing within families where applicable. Second, in *Population Extension*, we identified all variants within the target genes carried by MM cases and controls available in Database of Genotypes and Phenotypes (dbGaP). Sequence data from the CoMMpass Study (81) were downloaded: 733 sporadic MM (115 of African descent). Data from several non-MM studies in dbGAP were downloaded: 935 controls with no known MM or MGUS diagnosis (219 of African descent). African ancestry was determined based on principal component analysis of the sequence data and overlay with the 1000Genomes populations. Variant burden for each target gene was assessed using a c-alpha test (82) with 1 000 000 permutations in the GEMINI software (83). For variant burden in sporadic MM, we continued to require high or moderate impact variants but loosened the rarity criteria to AAF ≤ 0.05. Multiple testing was addressed by adjusting the significance threshold by the number of genes tested.

Data sharing statement

All variants within the regions investigated in this paper along with the scripts used to select the variants can be found on GitHub: https://github.com/rosaliegwaller/rare-variants-gwas-loci. For discovery and replication sequencing data in the early-onset and familial cases, please contact nicola.camp@hci.utah.edu. All sequence data for the sporadic MM cases and controls are available at dbGaP under accession numbers: phs000348.v2.p1, phs000748.v4.p3, phs000209.v13.p3, phs000276.v2.p1, phs000179.v5.p2, phs000298.v3.p2, phs000424.v6.p1, phs000653.v2.p1, phs000687.v1.p1, phs000814.v1.p1 and phs000806.1.p1.

Supplementary Material

Supplementary Material is available at *HMG* online.

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

Acknowledgements

We thank the DNA Sequencing Core Facility and Genomics Core Facility at the University of Utah, and the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

Data collection was made possible, in part, by the Utah Population Database and the Utah Cancer Registry. We thank the participants and their families who make this research possible. The Utah Cancer Registry is funded by the National Cancer Institute's SEER Program (Contract No. HHSN261201800016I) and the US Center for Disease Control and Prevention's National Program of Cancer Registries (Cooperative Agreement No. NU58DP0063200), with additional support from the University of Utah and Huntsman Cancer Foundation. Partial support for all datasets within the Utah Population Database is provided by the University of Utah, Huntsman Cancer Institute, and the Huntsman Cancer Institute Cancer Center Support grant (P30CA42014) from the National Cancer Institute.

Data from dbGaP study phs000209 comes from the Multi-Ethnic Study of Atherosclerosis (MESA). MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-RR-025005, and UL1-TR-000040. This exome sequencing from MESA is part of the NHLBI Grand Opportunity Exome Sequencing Project (GO-ESP). Funding for GO-ESP was provided by NHLBI grants RC2 HL103010 (HeartGO), RC2 HL102923 (LungGO) and RC2 HL102924 (WHISP). The exome sequencing was performed through NHLBI grants RC2 HL102925 (BroadGO) and RC2 HL102926 (SeattleGO). HeartGO gratefully acknowledges MESA for providing biological samples or data for this study.

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Funding

This work was supported by the National Cancer Institute of the National Institutes of Health [F99CA234943 to R.G.W., R21CA209533 to J.V., R01CA167824 to S.M.L.]; and a grant from the French National Cancer Institute (INCA) and the Fondation Française pour la Recherche contre le Myélome et les Gammapathies (FFRMG). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* doi: 10.3322/caac.21660.
2. Weiss, B.M., Abadie, J., Verma, P., Howard, R.S. and Kuehl, W.M. (2009) A monoclonal gammopathy precedes multiple myeloma in most patients. *Blood*, **113**, 5418–5422.
3. Landgren, O., Kyle, R.A., Pfeiffer, R.M., Katzmann, J.A., Caporaso, N.E., Hayes, R.B., Dispenzieri, A., Kumar, S., Clark, R.J., Baris, D., Hoover, R. and Rajkumar, S.V. (2009) Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood*, **113**, 5412–5417.
4. Morgan, G.J., Davies, F.E. and Linet, M. (2002) Myeloma aetiology and epidemiology. *Biomed. Pharmacother.*, **56**, 223–234.
5. Greenberg, A.J., Rajkumar, S.V. and Vachon, C.M. (2012) Familial monoclonal gammopathy of undetermined significance and multiple myeloma: epidemiology, risk factors, and biological characteristics. *Blood*, **119**, 5359–5366.
6. Waxman, A.J., Mink, P.J., Devesa, S.S., Anderson, W.F., Weiss, B.M., Kristinsson, S.Y., McGlynn, K.A. and Landgren, O. (2010) Racial disparities in incidence and outcome in multiple myeloma: a population-based study. *Blood*, **116**, 5501–5506.
7. Teras, L.R., DeSantis, C.E., Cerhan, J.R., Morton, L.M., Jemal, A. and Flowers, C.R. (2016) US lymphoid malignancy statistics by World Health Organization subtypes: 2016 US lymphoid malignancy statistics by World Health Organization subtypes. *CA Cancer J. Clin.*, **66**, 443–459.
8. Morgan, G.J., Johnson, D.C., Weinhold, N., Goldschmidt, H., Landgren, O., Lynch, H.T., Hemminki, K. and Houlston, R.S. (2014) Inherited genetic susceptibility to multiple myeloma. *Leukemia*, **28**, 518–524.
9. Grosbois, B., Jego, P., Attal, M., Payen, C., Rapp, M.J., Fuzibet, J.G., Maigre, M. and Bataille, R. (1999) Familial multiple myeloma: report of fifteen families. *Br. J. Haematol.*, **105**, 768–770.
10. Hemminki, K., Li, X. and Czene, K. (2004) Familial risk of cancer: data for clinical counseling and cancer genetics. *Int. J. Cancer*, **108**, 109–114.
11. Bourguet, C.C., Grufferman, S., Delzell, E., DeLong, E.R. and Cohen, H.J. (1985) Multiple myeloma and family history of cancer. A case–control study. *Cancer*, **56**, 2133–2139.
12. Eriksson, M. and Hållberg, B. (1992) Familial occurrence of hematologic malignancies and other diseases in multiple myeloma: a case–control study. *Cancer Causes Control*, **3**, 63–67.
13. Schinasi, L.H., Brown, E.E., Camp, N.J., Wang, S.S., Hofmann, J.N., Chiu, B.C., Miligi, L., Beane Freeman, L.E., de Sanjose, S., Bernstein, L. *et al.* (2016) Multiple myeloma and family history of lymphohaematopoietic cancers: results from the International Multiple Myeloma Consortium. *Br. J. Haematol.*, **175**, 87–101.
14. Landgren, O., Linet, M.S., McMaster, M.L., Gridley, G., Hemminki, K. and Goldin, L.R. (2006) Familial characteristics of autoimmune and hematologic disorders in 8,406 multiple myeloma patients: a population-based case–control study. *Int. J. Cancer*, **118**, 3095–3098.
15. VanValkenburg, M.E., Pruitt, G.I., Brill, I.K., Costa, L., Ehtsham, M., Justement, I.T., Innis-Shelton, R.D., Salzman, D., Reddy, E.S., Godby, K.N. *et al.* (2016) Family history of hematologic malignancies and risk of multiple myeloma: differences by race and clinical features. *Cancer Causes Control*, **27**, 81–91.
16. Kristinsson, S.Y., Björkholm, M., Goldin, L.R., Blimark, C., Mellqvist, U.H., Wahlin, A., Turesson, I. and Landgren, O. (2009) Patterns of hematologic malignancies and solid tumors among 37,838 first-degree relatives of 13,896 patients with multiple myeloma in Sweden. *Int. J. Cancer*, **125**, 2147–2150.
17. Altieri, A., Chen, B., Bermejo, J.L., Castro, F. and Hemminki, K. (2006) Familial risks and temporal incidence trends of multiple myeloma. *Eur. J. Cancer*, **42**, 1661–1670.

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

18. Mitchell, J.S., Li, N., Weinhold, N., Försti, A., Ali, M., van Duin, M., Thorleifsson, G., Johnson, D.C., Chen, B., Halvarsson, B.M. *et al.* (2016) Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nat. Commun.*, **7**, 12050. doi: 10.1038/ncomms12050.
19. Weinhold, N., Johnson, D.C., Chubb, D., Chen, B., Försti, A., Hosking, F.J., Broderick, P., Ma, Y.P., Dobbins, S.E., Hose, D. *et al.* (2013) The CCND1 c.870G>A polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. *Nat. Genet.*, **45**, 522–525.
20. Went, M., Sud, A., Försti, A., Halvarsson, B.M., Weinhold, N., Kimber, S., van Duin, M., Thorleifsson, G., Holroyd, A., Johnson, D.C. *et al.* (2018) Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma. *Nat. Commun.*, **9**, 3707. doi: 10.1038/s41467-018-04989-w.
21. Swaminathan, B., Thorleifsson, G., Jöud, M., Ali, M., Johnsson, E., Ajore, R., Sulem, P., Halvarsson, B.M., Eyjolfsson, G., Haraldsdottir, V. *et al.* (2015) Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. *Nat. Commun.*, **6**, 7213. doi: 10.1038/ncomms8213.
22. Chubb, D., Weinhold, N., Broderick, P., Chen, B., Johnson, D.C., Försti, A., Vijayakrishnan, J., Migliorini, G., Dobbins, S.E., Holroyd, A. *et al.* (2013) Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nat. Genet.*, **45**, 1221–1225.
23. Broderick, P., Chubb, D., Johnson, D.C., Weinhold, N., Försti, A., Lloyd, A., Olver, B., Ma, Y., Dobbins, S.E., Walker, B.A. *et al.* (2011) Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat. Genet.*, **44**, 58–61.
24. Halvarsson, B.M., Wihlborg, A.K., Ali, M., Lemonakis, K., Johnsson, E., Niroula, A., Cibulskis, C., Weinhold, N., Försti, A., Alici, E. *et al.* (2017) Direct evidence for a polygenic etiology in familial multiple myeloma. *Blood Adv.*, **1**, 619–623.
25. Waller, R.G., Darlington, T.M., Wei, X., Madsen, M.J., Thomas, A., Curtin, K., Coon, H., Rajamanickam, V., Musinsky, J., Jayabalan, D. *et al.* (2018) Novel pedigree analysis implicates DNA repair and chromatin remodeling in multiple myeloma risk. *PLoS Genet.*, **14**, e1007111. doi: 10.1371/journal.pgen.1007111.
26. Wei, X., Calvo-Vidal, M.N., Chen, S., Wu, G., Revuelta, M.V., Sun, J., Zhang, J., Walsh, M.F., Nichols, K.E., Joseph, V. *et al.* (2018) Germline lysine-specific demethylase 1 (LSD1/KDM1A) mutations confer susceptibility to multiple myeloma. *Cancer Res.*, **78**, 2747–2759.
27. Pertesi, M., Vallée, M., Wei, X., Revuelta, M.V., Galia, P., Demangel, D., Oliver, J., Foll, M., Chen, S., Perrial, E. *et al.* (2019) Exome sequencing identifies germline variants in DIS3 in familial multiple myeloma. *Leukemia*, **33**, 2324–2330.
28. Gibson, G. (2012) Rare and common variants: twenty arguments. *Nat. Rev. Genet.*, **13**, 135–145.
29. Hall, J.M., Lee, M.K., Newman, B., Morrow, J.E., Anderson, L.A., Huey, B. and King, M.C. (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*, **250**, 1684–1689.
30. Sud, A., Chattopadhyay, S., Thomsen, H., Sundquist, K., Sundquist, J., Houlston, R.S. and Hemminki, K. (2019) Analysis of 153 115 patients with hematological malignancies refines the spectrum of familial risk. *Blood*, **134**, 960–969.
31. Bhattacharjee, S., Rajaraman, P., Jacobs, K.B., Wheeler, W.A., Melin, B.S., Hartge, P., GliomaScan Consortium, Yeager, M., Chung, C.C., Chanock, S.J. and Chatterjee, N. (2012) A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *Am. J. Hum. Genet.*, **90**, 821–835.
32. Wu, Y.H., Graff, R.E., Passarelli, M.N., Hoffman, J.D., Ziv, E., Hoffmann, T.J. and Witte, J.S. (2018) Identification of pleiotropic cancer susceptibility variants from genome-wide association studies reveals functional characteristics. *Cancer Epidemiol. Biomark. Prev.*, **27**, 75–85.
33. Law, P.J., Sud, A., Mitchell, J.S., Henrion, M., Orlando, G., Lenive, O., Broderick, P., Speedy, H.E., Johnson, D.C., Kaiser, M. *et al.* (2017) Genome-wide association analysis of chronic lymphocytic leukaemia, Hodgkin lymphoma and multiple myeloma identifies pleiotropic risk loci. *Sci. Rep.*, **7**, 41071. doi: 10.1038/srep41071.
34. Went, M., Sud, A., Speedy, H., Sunter, N.J., Försti, A., Law, P.J., Johnson, D.C., Mirabella, F., Holroyd, A., Li, N. *et al.* (2018) Genetic correlation between multiple myeloma and chronic lymphocytic leukaemia provides evidence for shared aetiology. *Blood Cancer J.*, **9**, 1. doi: 10.1038/s41408-018-0162-8.
35. Speedy, H.E., Di Bernardo, M.C., Sava, G.P., Dyer, M.J., Holroyd, A., Wang, Y., Sunter, N.J., Mansouri, L., Juliusson, G., Smedby, K.E. *et al.* (2014) A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. *Nat. Genet.*, **46**, 56–60.
36. Berndt, S.I., Skibola, C.F., Joseph, V., Camp, N.J., Nieters, A., Wang, Z., Cozen, W., Monnereau, A., Wang, S.S., Kelly, R.S. *et al.* (2013) Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat. Genet.*, **45**, 868–876.
37. Slager, S.L., Skibola, C.F., Di Bernardo, M.C., Conde, L., Broderick, P., McDonnell, S.K., Goldin, L.R., Croft, N., Holroyd, A., Harris, S. *et al.* (2012) Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. *Blood*, **120**, 843–846.
38. Di Bernardo, M.C., Crowther-Swanepoel, D., Broderick, P., Webb, E., Sellick, G., Wild, R., Sullivan, K., Vijayakrishnan, J., Wang, Y., Pittman, A.M. *et al.* (2008) A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat. Genet.*, **40**, 1204–1210.
39. Conde, L., Halperin, E., Akers, N.K., Brown, K.M., Smedby, K.E., Rothman, N., Nieters, A., Slager, S.L., Brooks-Wilson, A., Agana, L. *et al.* (2010) Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat. Genet.*, **42**, 661–664.
40. Law, P.J., Berndt, S.I., Speedy, H.E., Camp, N.J., Sava, G.P., Skibola, C.F., Holroyd, A., Joseph, V., Sunter, N.J., Nieters, A. *et al.* (2017) Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. *Nat. Commun.*, **8**, 14175. doi: 10.1038/ncomms14175.
41. Slager, S.L., Rabe, K.G., Achenbach, S.J., Vachon, C.M., Goldin, L.R., Strom, S.S., Lanasa, M.C., Spector, L.G., Rassenti, L.Z., Leis, J.F. *et al.* (2011) Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. *Blood*, **117**, 1911–1916.
42. Berndt, S.I., Camp, N.J., Skibola, C.F., Vijai, J., Wang, Z., Gu, J., Nieters, A., Kelly, R.S., Smedby, K.E., Monnereau, A. *et al.* (2016) Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat. Commun.*, **7**, 10933. doi: 10.1038/ncomms10933.
43. Frampton, M., da Silva Filho, M.I., Broderick, P., Thomsen, H., Försti, A., Vijayakrishnan, J., Cooke, R., Enciso-Mora, V., Hoffmann, P., Nöthen, M.M. *et al.* (2013) Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's lymphoma. *Nat. Commun.*, **4**, 2549. doi: 10.1038/ncomms3549.

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

44. Sud, A., Thomsen, H., Law, P.J., Försti, A., Filho, M., Holroyd, A., Broderick, P., Orlando, G., Lenive, O., Wright, L. *et al.* (2017) Genome-wide association study of classical Hodgkin lymphoma identifies key regulators of disease susceptibility. *Nat. Commun.*, **8**, 1892. doi: 10.1038/s41467-017-00320-1.
45. Cozen, W., Timofeeva, M.N., Li, D., Diepstra, A., Hazelett, D., Delahaye-Sourdeix, M., Edlund, C.K., Franke, L., Rostgaard, K., Van Den Berg, D.J. *et al.* (2014) A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. *Nat. Commun.*, **5**, 3856. doi: 10.1038/ncomms4856.
46. Enciso-Mora, V., Broderick, P., Ma, Y., Jarrett, R.F., Hjalgrim, H., Hemminki, K., van den Berg, A., Olver, B., Lloyd, A., Dobbins, S.E. *et al.* (2010) A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). *Nat. Genet.*, **42**, 1126–1130.
47. Urayama, K.Y., Jarrett, R.F., Hjalgrim, H., Diepstra, A., Kamatani, Y., Chabrier, A., Gaborieau, V., Boland, A., Nieters, A., Becker, N. *et al.* (2012) Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. *J. Natl. Cancer Inst.*, **104**, 240–253.
48. Skibola, C.F., Bracci, P.M., Halperin, E., Conde, L., Craig, D.W., Agana, L., Iyadurai, K., Becker, N., Brooks-Wilson, A., Curry, J.D. *et al.* (2009) Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. *Nat. Genet.*, **41**, 873–875.
49. Skibola, C.F., Berndt, S.I., Vijai, J., Conde, L., Wang, Z., Yeager, M., de Bakker, P.I., Birmann, B.M., Vajdic, C.M., Foo, J.N. *et al.* (2014) Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the HLA region. *Am. J. Hum. Genet.*, **95**, 462–471.
50. Smedby, K.E., Foo, J.N., Skibola, C.F., Darabi, H., Conde, L., Hjalgrim, H., Kumar, V., Chang, E.T., Rothman, N., Cerhan, J.R. *et al.* (2011) GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. *PLoS Genet.*, **7**, e1001378. doi: 10.1371/journal.pgen.1001378.
51. Cerhan, J.R., Berndt, S.I., Vijai, J., Ghesquières, H., McKay, J., Wang, S.S., Wang, Z., Yeager, M., Conde, L., de Bakker, P.I. *et al.* (2014) Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nat. Genet.*, **46**, 1233–1238.
52. Vijai, J., Wang, Z., Berndt, S.I., Skibola, C.F., Slager, S.L., de Sanjose, S., Melbye, M., Glimelius, B., Bracci, P.M., Conde, L. *et al.* (2015) A genome-wide association study of marginal zone lymphoma shows association to the HLA region. *Nat. Commun.*, **6**, 5751. doi: 10.1038/ncomms6751.
53. Vijai, J., Kirchhoff, T., Schrader, K.A., Brown, J., Dutra-Clarke, A.V., Manschreck, C., Hansen, N., Rau-Murthy, R., Sarrel, K., Przybylo, J. *et al.* (2013) Susceptibility loci associated with specific and shared subtypes of lymphoid malignancies. *PLoS Genet.*, **9**, e1003220. doi: 10.1371/journal.pgen.1003220.
54. Khalil, M.O., Morton, L.M., Devesa, S.S., Check, D.P., Curtis, R.E., Weisenburger, D.D. and Dores, G.M. (2014) Incidence of marginal zone lymphoma in the United States, 2001-2009 with a focus on primary anatomic site. *Br. J. Haematol.*, **165**, 67–77.
55. Nguyen, T., Liu, X.K., Zhang, Y. and Dong, C. (2006) BTNL2, a butyrophilin-like molecule that functions to inhibit T cell activation. *J. Immunol.*, **176**, 7354–7360.
56. Zhu, Y., Ju, S., Chen, E., Dai, S., Li, C., Morel, P., Liu, L., Zhang, X. and Lu, B. (2010) T-bet and eomesodermin are required for T cell-mediated antitumor immune responses. *J. Immunol.*, **185**, 3174–3183.
57. Gross, J.A., Dillon, S.R., Mudri, S., Johnston, J., Littau, A., Roque, R., Rixon, M., Schou, O., Foley, K.P., Haugen, H. *et al.* (2001) TACI-Ig neutralizes molecules critical for B cell development and autoimmune disease. *Immunity*, **15**, 289–302.
58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., Ashkenazi, A. and Dixit, V.M. (2000) Identification of a receptor for BLyS demonstrates a crucial role in humoral immunity. *Nat. Immunol.*, **1**, 37–41.
59. Seyler, T.M., Park, Y.W., Takemura, S., Bram, R.J., Kurtin, P.J., Goronzy, J.J. and Weyand, C.M. (2005) BLyS and APRIL in rheumatoid arthritis. *J. Clin. Invest.*, **115**, 3083–3092.
60. Taniguchi, T., Ogasawara, K., Takaoka, A. and Tanaka, N. (2001) IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.*, **19**, 623–655.
61. Salem, S., Langlais, D., Lefebvre, F., Bourque, G., Bigley, V., Haniffa, M., Casanova, J.L., Burk, D., Berghuis, A., Butler, K.M. *et al.* (2014) Functional characterization of the human dendritic cell immunodeficiency associated with the IRF8(K108E) mutation. *Blood*, **124**, 1894–1904.
62. Koi, M., Johnson, L.A., Kalikin, L.M., Little, P.F., Nakamura, Y. and Feinberg, A.P. (1993) Tumor cell growth arrest caused by subchromosomal transferable DNA fragments from chromosome 11. *Science*, **260**, 361–364.
63. Robb, L., Tarrant, J., Groom, J., Ibrahim, M., Li, R., Borobakas, B. and Wright, M.D. (2001) Molecular characterisation of mouse and human TSSC6: evidence that TSSC6 is a genuine member of the tetraspanin superfamily and is expressed specifically in haematopoietic organs. *Biochim. Biophys. Acta*, **1522**, 31–41.
64. Fagerberg, L., Hallström, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K. *et al.* (2014) Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics*, **13**, 397–406.
65. Muhitch, J.B., Hoffend, N.C., Azabdaftari, G., Miller, A., Bshara, W., Morrison, C.D., Schwaab, T. and Abrams, S.I. (2019) Tumor-associated macrophage expression of interferon regulatory Factor-8 (IRF8) is a predictor of progression and patient survival in renal cell carcinoma. *J. Immunother. Cancer*, **7**, 155. doi: 10.1186/s40425-019-0630-0.
66. Mottok, A., Hung, S.S., Chavez, E.A., Woolcock, B., Telenius, A., Chong, L.C., Meissner, B., Nakamura, H., Rushton, C., Viganò, E. *et al.* (2019) Integrative genomic analysis identifies key pathogenic mechanisms in primary mediastinal large B-cell lymphoma. *Blood*, **134**, 802–813.
67. Fitzgerald, L.M., Kumar, A., Boyle, E.A., Zhang, Y., McIntosh, L.M., Kolb, S., Stott-Miller, M., Smith, T., Karyadi, D.M., Ostrander, E.A. *et al.* (2013) Germline missense variants in the BTNL2 gene are associated with prostate cancer susceptibility. *Cancer Epidemiol. Biomark. Prev.*, **22**, 1520–1528.
68. Rebbeck, T.R. (2017) Prostate cancer genetics: variation by race, ethnicity, and geography. *Semin. Radiat. Oncol.*, **27**, 3–10.
69. Castigli, E., Wilson, S.A., Garibyan, L., Rachid, R., Bonilla, F., Schneider, L. and Geha, R.S. (2005) TACI is mutant in common variable immunodeficiency and IgA deficiency. *Nat. Genet.*, **37**, 829–834.
70. Salzer, U., Chapel, H.M., Webster, A.D., Pan-Hammarström, Q., Schmitt-Graeff, A., Schlesier, M., Peter, H.H., Rockstroh, J.K., Schneider, P., Schäffer, A.A. *et al.* (2005) Mutations in TNFRSF13B encoding TACI are associated with common

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024

variable immunodeficiency in humans. *Nat. Genet.*, **37**, 820–828.
71. Salzer, U., Bacchelli, C., Buckridge, S., Pan-Hammarström, Q., Jennings, S., Lougaris, V., Bergbreiter, A., Hagena, T., Birmelin, J., Plebani, A. *et al.* (2009) Relevance of biallelic versus monoallelic TNFRSF13B mutations in distinguishing disease-causing from risk-increasing TNFRSF13B variants in antibody deficiency syndromes. *Blood*, **113**, 1967–1976.
72. Resnick, E.S., Moshier, E.L., Godbold, J.H. and Cunningham-Rundles, C. (2012) Morbidity and mortality in common variable immune deficiency over 4 decades. *Blood*, **119**, 1650–1657.
73. Moreaux, J., Cremer, F.W., Reme, T., Raab, M., Mahtouk, K., Kaukel, P., Pantesco, V., De Vos, J., Jourdan, E., Jauch, A. *et al.* (2005) The level of TACI gene expression in myeloma cells is associated with a signature of microenvironment dependence versus a plasmablastic signature. *Blood*, **106**, 1021–1030.
74. Tai, Y.T., Lin, L., Xing, L., Cho, S.F., Yu, T., Acharya, C., Wen, K., Hsieh, P.A., Dulos, J., van Elsas, A. *et al.* (2019) APRIL signaling via TACI mediates immunosuppression by T regulatory cells in multiple myeloma: therapeutic implications. *Leukemia*, **33**, 426–438.
75. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.*, **43**, 491–498.
76. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. and Cunningham, F. (2016) The ensembl variant effect predictor. *Genome Biol.*, **17**, 122. doi: 10.1186/s13059-016-0974-4.
77. Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.
78. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.
79. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
80. Machiela, M.J. and Chanock, S.J. (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, **31**, 3555–3557.
81. Keats, J.J., Craig, D.W., Liang, W., Venkata, Y., Kurdoglu, A., Aldrich, J., Auclair, D., Allen, K., Harrison, B., Jewell, S. *et al.* (2013) Interim analysis of the MMRF Commpass trial, a longitudinal study in multiple myeloma relating clinical outcomes to genomic and immunophenotypic profiles. *Blood*, **122**, 532–532.
82. Neale, B.M., Rivas, M.A., Voight, B.F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S.M., Roeder, K. and Daly, M.J. (2011) Testing for an unusual distribution of rare variants. *PLoS Genet.*, **7**, e1001322.
83. Paila, U., Chapman, B.A., Kirchner, R. and Quinlan, A.R. (2013) GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput. Biol.*, **9**, e1003153. doi: 10.1371/journal.pcbi.1003153.

Downloaded from https://academic.oup.com/hmg/article/30/12/1142/6159376 by guest on 10 April 2024