

LYMPHOID NEOPLASIA

Exome sequencing reveals recurrent germ line variants in patients with familial Waldenström macroglobulinemia

Aldo M. Roccaro,^{1,2,*} Antonio Sacco,^{1,*} Jiantao Shi,^{3,*} Marco Chiarini,^{1,2} Adriana Perilla-Glen,¹ Salomon Manier,¹ Siobhan Glavey,¹ Yosra Aljawai,¹ Yuji Mishima,¹ Yawara Kawano,¹ Michele Moschetta,¹ Mick Correll,¹ Ma. Reina Improgo,¹ Jennifer R. Brown,¹ Luisa Imberti,² Giuseppe Rossi,² Jorge J. Castillo,¹ Steven P. Treon,¹ Matthew L. Freedman,¹ Eliezer M. Van Allen,¹ Winston Hide,⁴ Elaine Hiller,¹ Irene Rainville,¹ and Irene M. Ghobrial¹

¹Dana-Farber Cancer Institute, Department of Medical Oncology, Harvard Medical School, Boston, MA; ²Department of Hematology, Centro per la Ricerca Emato-oncologica AIL, Spedali Civili di Brescia, Brescia, Italy; ³Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA; and ⁴Department of Biostatistics, Harvard School of Public Health, Boston, MA

Key Points

- *LAPTM5*^{c403t} and *HCLS1*^{g496a} are potentially novel contributors for the genetic predisposition to familial WM.
- *LAPTM5*^{c403t} and *HCLS1*^{g496a} represent possible candidates for screening in familial WM.

Familial aggregation of Waldenström macroglobulinemia (WM) cases, and the clustering of B-cell lymphoproliferative disorders among first-degree relatives of WM patients, has been reported. Nevertheless, the possible contribution of inherited susceptibility to familial WM remains unrevealed. We performed whole exome sequencing on germ line DNA obtained from 4 family members in which coinheritance for WM was documented in 3 of them, and screened additional independent 246 cases by using gene-specific mutation sequencing. Among the shared germ line variants, *LAPTM5*^{c403t} and *HCLS1*^{g496a} were the most recurrent, being present in 3/3 affected members of the index family, detected in 8% of the unrelated familial cases, and present in 0.5% of the nonfamilial cases and in <0.05 of a control population. *LAPTM5* and *HCLS1* appeared as relevant WM candidate genes that characterized familial WM individuals and were also functionally relevant to the tumor clone. These findings highlight potentially novel contributors for the genetic predisposition to familial WM and indicate that *LAPTM5*^{c403t} and *HCLS1*^{g496a} may represent predisposition alleles in patients with familial WM. (*Blood*. 2016;127(21):2598-2606)

Introduction

The evaluation of cancer occurrence within families is important for unraveling the molecular events that drive tumorigenesis. Waldenström macroglobulinemia (WM) represents a B-cell lymphoproliferative disorder, classified as a lymphoplasmacytic lymphoma, according to the World Health Organization classification.¹ WM represents a rare B-cell malignancy that accounts for 1% to 2% of all hematologic neoplasms, with an incidence rate of 3 to 4 cases per million people, per year.^{2,3} Evidence also exists that immunoglobulin M-monoclonal gammopathy of undetermined significance (IgM-MGUS) is associated with an increased risk of developing WM.^{4,5} Most recently, whole genome sequencing studies have demonstrated the occurrence of *MYD88* and *CXCR4*/WHIM-like somatic variants in more than 90% and 30% to 35% of WM patients, respectively.⁶⁻⁹

Previous studies have identified familial aggregation of WM cases and the clustering of B-cell lymphoproliferative disorders among first-degree relatives of patients with WM,¹⁰⁻¹⁶ including chronic lymphocytic leukemia (CLL), non-Hodgkin lymphoma (nHL), multiple myeloma (MM), IgM-MGUS, and IgG/IgA-MGUS. These findings support a possible contribution of inherited susceptibility to familial WM. Nevertheless, genetic linkage studies have failed to clearly identify rare, highly penetrant alleles underlying a subset of B-cell

lymphoproliferative disorders, yielding to an oligogenic model whereby relatively common, low-penetrance alleles would contribute to the LP phenotype in familial cases.^{17,18} In fact, genome-wide association studies have identified common variants associated with MM.¹⁹⁻²² Case control studies have similarly identified significant association with other lymphoproliferative disorders.^{20,23} Whether relatively common germ line variants may contribute specifically to familial WM cases remains unexplored.

We therefore performed whole exome sequencing on germ line DNA obtained from 4 members of a single family with documented coinheritance of WM (3 affected; 1 unaffected) and applied bioinformatics tools to identify candidate germ line variants likely to have a biological role in WM signaling pathways. We screened an additional 246 independent, unrelated WM cases (50 probands from familial cases and 196 individuals from nonfamilial cases) for the identified variants. We report that *LAPTM5*^{c403t} and *HCLS1*^{g496a} represent the most recurrent variants, present in 3/3 affected members of the index family. Each of these variants was present in 8% of the unrelated familial cases. Each variant was present in 0.5% of the nonfamilial cases and in <0.05 of a control population (1000 Genomes; www.1000genomes.org). These findings highlight potentially

Submitted November 5, 2015; accepted February 12, 2016. Prepublished online as *Blood* First Edition paper, February 22, 2016; DOI 10.1182/blood-2015-11-680199.

*A.M.R., A.S., and J.S. contributed equally to this work.

The online version of this article contains a data supplement.

The data reported in this article have been deposited in the Gene Expression Omnibus database (accession number: SRP053196).

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2016 by The American Society of Hematology

novel contributors for the genetic predisposition to familial WM and suggest possible candidates for screening in familial WM.

Methods

Study oversight

Approval for these studies was obtained from the Dana-Farber Cancer Institute Institutional Review Board. Informed consent was obtained from all patients in accordance with the Declaration of Helsinki. Written consent was obtained from all study participants.

Study participants

We studied a family with 3 first-degree relatives affected with WM, and a healthy unaffected family member who was considered to be a control, and performed whole exome sequencing. An additional 246 participants representing individual probands affected with WM from 246 independent families were included in these studies: 50 of the participants had a family history of WM or other B-cell lymphoproliferative disorders, whereas the remaining 196 were nonfamilial patients. Based on previous identification of familial aggregation of WM cases and clustering of B-cell lymphoproliferative disorders (including CLL, MM, and nHL) among the first-degree relatives of WM patients,¹¹⁻¹⁶ we defined a familial case as a patient with first-degree relatives who are also affected by a B-cell lymphoproliferative disorder, including WM, CLL, nHL, MM, IgM-MGUS, and IgG/IgA-MGUS; nonfamilial cases are individuals whose family members are either healthy or are diagnosed with a solid tumor, but not a B-cell lymphoproliferative malignancy.

Sequence data generation

Genomic DNA was isolated from buccal cells collected from each study participant, including the 246 independent WM cases, and was subjected to library construction, according to standard methods, followed by shearing, end repair, phosphorylation, and ligation to barcoded sequencing adapters. The DNA was size-selected to exonic hybrid capture using SureSelect v2 Exome bait (Agilent, Santa Clara, CA). Samples were multiplexed and sequenced on Illumina HiSeq flow cells with the goal of an average depth of coverage of 100× (Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Boston, MA). Reads were aligned to GRCh37 using Burrows-Wheeler aligner²⁴ with default parameters, and the resulting SAM files were converted to BAM files using Picard (<http://picard.sourceforge.net>).

Quality control of sequencing data

To evaluate the overall quality of sequenced samples, we used BamUtil (<http://genome.sph.umich.edu/wiki/BamUtil>) to calculate various statistics, including the total number of reads, mapping rate, percentage of proper pairs, and duplication rate. Given that the SureSelectQXT v4 platform covers ~51 M, a mean coverage was calculated for each sample. We evaluated the distribution of the mean coverage across all targeted regions. The DepthOfCoverage function from GATK (v2.74) was used with the “-mmq 10” parameter.²⁵ All unmapped reads, duplicated reads, and reads with low mapping quality (<10) were removed. Finally, a more comprehensive callable analysis on all targeted bases was adopted by considering simultaneously sequencing quality, mapping quality, and coverage, according to the CallableLoci tool (GATK, Broad Institute, Cambridge, MA).

Variant calling

Single nucleotide variants (SNVs) were called by GATK,²⁶ based on best practice workflow. Briefly, GATK was used for base quality score recalibration and indel realignment, followed by variant calling with use of UnifiedGenotyper with *-stand_call_conf* = 30.

Selection of potential WM variants

We first selected variants that occurred in affected members but were absent in the unaffected member. The resulting SNVs were annotated by snpEff²⁷ and

Oncotator (<http://www.broadinstitute.org/oncotator/>); variants located in exonic regions were considered for further analysis. Synonymous variants were filtered out, and the resulting SNVs were annotated with dbNSFP, a database that was developed for functional prediction and annotation of all potential nonsynonymous SNVs in the human genome.²⁵ Allele frequencies in 1000 Genomes, as well as Polymorphism Phenotyping v2 (PolyPhen-2) prediction, were used to quantify the deleteriousness of SNVs with allele frequencies >0.05 in 1000 Genomes were filtered out, and the remaining SNVs were defined as potential familial WM-associated variants.

Differential expression analysis

Differential expression was analyzed using the Bioconductor (www.bioconductor.org) package limma, in the R statistical computing environment (www.r-project.org)²⁸; using empirical Bayes-moderated *t* statistics to calculate *P* values for 2-class unpaired samples. Differentially expressed genes were identified using a false discovery rate (FDR) cutoff of 1%. Among the differentially expressed genes, those with more than a 2-fold change were defined as a signature.

Gene expression data

GSE12668 was used to define genes that were differentially expressed between bone marrow-derived primary WM cells and their normal counterparts. Tissue specificity for genes of interest in normal tissues was retrieved directly from Gene Enrichment Profiler (GEP; <http://xavierlab2.mgh.harvard.edu/EnrichmentProfiler/index.html>). Expression levels for genes of interest in primary cancer tissues were downloaded from the cbiportal (<http://cbiportal.org>),²⁹ which integrates data generated by The Cancer Genome Atlas (TCGA) Research Network (<http://cancergenome.nih.gov/>) and by many other sources. Gene expression modules were assessed using an independent messenger RNA (mRNA) data set (GSE6691).

Disease gene prioritization

Given the selected potential familial WM-associated variants/genes, we used Gene Relationships Across Implicated Loci (GRAIL)³⁰ coupled with global coexpressed gene database (COXPRESdb)³¹ to assess the functional relatedness between these genes and those that were differentially expressed between bone marrow-derived primary WM cells compared with their normal counterparts, using GSE12668.

Sanger sequencing

The observed *LAPTM5* and *HCLS1* variants were validated by Sanger sequencing. The Fisher exact test was used to assess the significance of variants observed in familial cases, compared with those in nonfamilial cases.

Three-dimensional protein modeling

Three-dimensional *LAPTM5* protein reconstruction was obtained using the Phyre2 server, as described previously.³²

Statistics

All analyses of raw sequencing metrics were performed using the limma package in an R/Bioconductor computational environment; differentially expressed genes were identified using an FDR cutoff of 1%. Fisher exact test was used to assess the significance of variants observed in familial cases compared with those in nonfamilial cases.

Results

Identification of *LAPTM5*^{c403t} and *HCLS1*^{g496a} variants in familial WM

We studied a kindred in which 3 members are affected by familial WM (Figure 1A). The diagnosis of WM was confirmed in all cases by

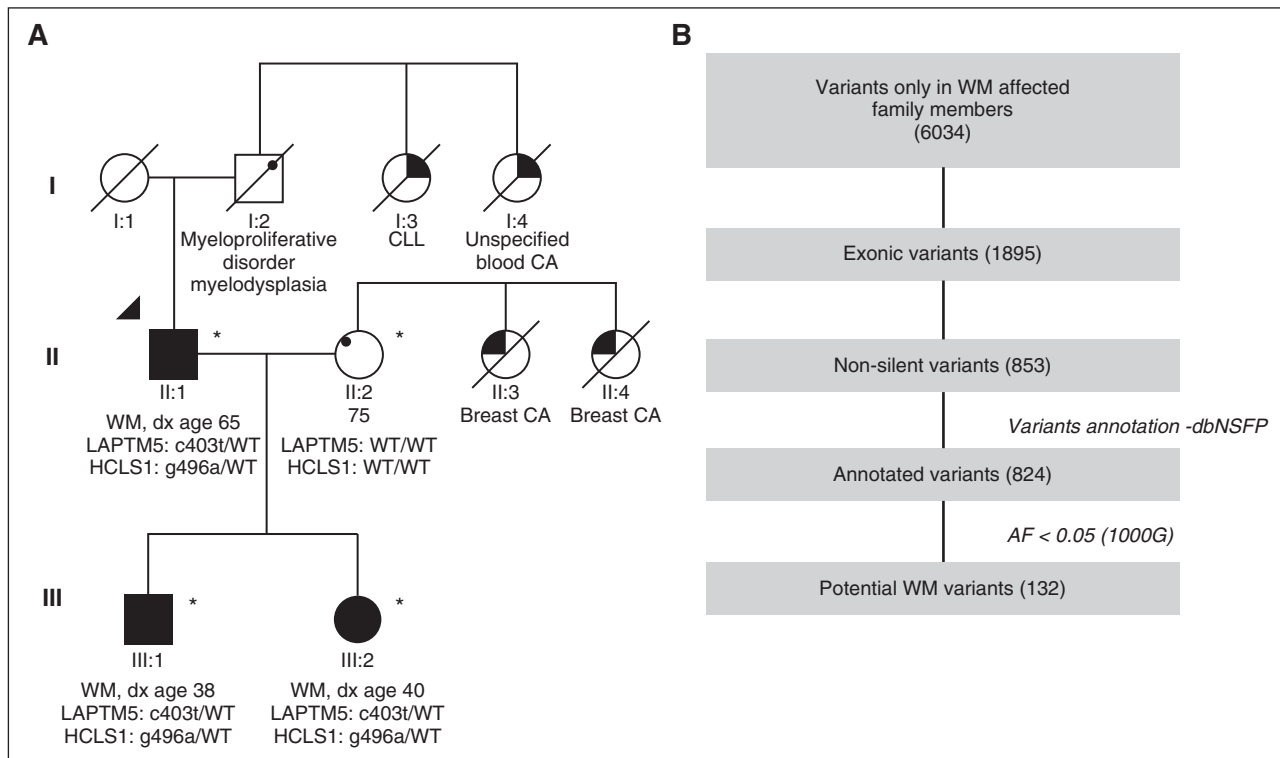


Figure 1. Family with documented coinheritance of WM. (A) Squares indicate male family members, circles indicate female family members, and slashes indicate a deceased family member. In the family we studied, 3 members were identified as affected with WM (II:1, III:1, III:2). Arrowhead indicates the proband (first member diagnosed with WM). *Sequenced samples. Age of diagnosis is indicated for the 3 affected members with WM. The sequenced unaffected member was aged 75 years at the time of sampling. Other diseases are specified for all the remaining members. (B) Workflow for selecting potential WM-associated variants. The number of variants identified in each step is shown. CA, cancer; dx, diagnosis; 1000G, 1000 Genomes; WT, wild-type.

histology and immunohistochemistry³³ (clinical features are summarized in supplemental Table 1, available on the *Blood* Web site). Whole exome sequencing was performed on the 3 affected members and 1 unaffected member. The criteria to identify candidate WM-associated variants were nonsynonymous SNVs that were present only in the affected family members and absent in the unaffected member and an allele frequency <0.05 using 1000 Genomes. This initial screen identified 132 candidate exonic, nonsynonymous familial WM variants, mapping to 127 genes (Figure 1B; supplemental Table 1; supplemental Figure 1). We next performed a gene/variant prioritization to select the most significant WM-relevant variants by using GRAIL,³⁰ which integrates text mining or coexpression databases for gene prioritization. GRAIL takes a group of seed genes to build a subnetwork and test whether a query gene is functionally related to the seeds. Because the seeds should be disease-related genes, we used a WM gene expression signature as seeds and then queried the potential WM variants identified in this study for prioritization. The publically available gene expression data set (GSE12668)³⁴ was evaluated to define a WM cell-mRNA signature by comparing primary bone marrow-derived CD19⁺ WM tumor cells obtained from newly diagnosed untreated WM patients with their normal counterparts obtained from healthy individuals (HDs); we found 393 genes that were significantly different between WM patients and HDs, thus confirming the presence of a specific WM mRNA signature (FDR cutoff $<1\%$), (Figure 2A; supplemental Table 3). We next assessed the functional relatedness between potential familial WM-associated genes and the observed WM mRNA signature using the GRAIL algorithm³⁰ coupled with a global coexpression network COXPRESdb, as reported.³¹ Using an FDR cutoff of 5%, 13 genes were predicted to be functionally related to the WM mRNA

signature (Figure 2B). In addition, the observed germ line variants were annotated and interpreted by implementing PolyPhen-2, as described,³⁵ thus allowing us to prioritize the deleteriousness of SNVs shared among all 3 affected WM family members.

It has been reported that disease-related genes tend to be tissue-specific³⁶; we therefore assessed the tissue specificity of the selected genes for further prioritization. The GEP database contains expression profiles and tissue specificity scores for ~12 000 genes across 126 normal primary human tissues and 23 cancers.³⁷ WM is a B-cell lymphoproliferative disorder and is classified as an IgM-secreting lymphoplasmacytic lymphoma¹; we thus used the GEP database to assign a score to each of the selected genes, which was the average of its tissue specificity score across normal B cells. We reasoned that genes containing deleterious variants that are significantly related to the WM mRNA expression signature, and are also highly specific to B cells, may be the most promising familial WM-associated genes. Among the 13 potential candidate genes obtained as the result of the network-based gene prioritization algorithm independently of any B-cell tissue specificity filtering criteria, the most significant gene candidates were found to be *HCLS1* and *LAPT5* (Figure 2C; supplemental Table 4). Variants in these 2 genes cosegregated within the affected family members, being absent in the unaffected family member and present in less than 1% of the control population, where allele frequency and heterozygote frequency were 0.38%/0.7% and 0.39%/0.7% for *LAPT5* and *HCLS1*, respectively, using 1000 Genomes (supplemental Table 5). Both genes are highly B-cell tissue-specific (supplemental Figure 2A-C). In contrast, the 2 control genes *HRC* and *IL22RA1* (which are not functionally related to WM expression signature, but contain benign variants) are not B cell–tissue-specific (supplemental

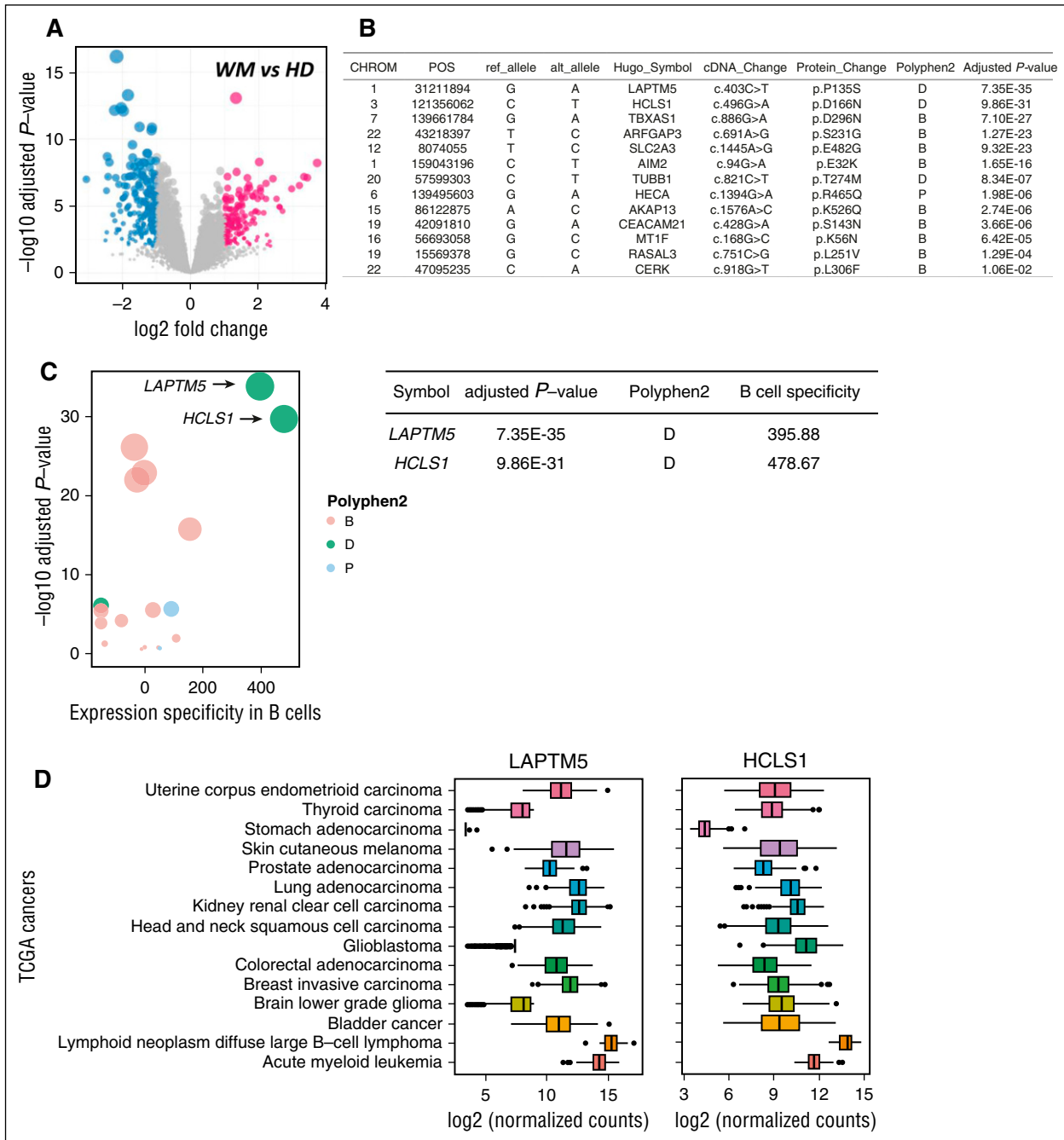


Figure 2. Identifying potential WM-associated variants and prioritization of potential WM genes. (A) Differential expression analysis between WM tumor cells and their normal counterparts using a public data set (GSE12668). The significantly differentially expressed genes are plotted in light blue, with an adjusted *P* value cutoff of 1% and 2-fold change; nonsignificant genes are shown in gray. (B) Genes containing potential WM variants were prioritized with use of GRAIL coupled with coexpression networks. The significant ones (adjusted *P* values of 5%) are shown. Genomic positions (POS; GRCh37) as well as complementary DNA (cDNA) and protein changes are listed. Deleteriousness of a variant was predicted by PolyPhen2 either as benign (B), damaging (D), or possibly damaging (P). CHROM, chromosome. (C) Gene prioritization, with use of coexpression networks and tissue specificity in B cells. GRAIL was used to assess the significance of functional relatedness; adjusted *P* values are shown in log10 scale. The mean tissue specificity score for each gene was calculated as the average of its tissue specificity scores across different B-cell samples in the GEP database. Two genes with the highest functional significance and highest tissue specificity scores in B cells are indicated. Genes containing B, D, and P variants are shown in red, green, and blue, respectively. Adjusted *P* values are shown for *LAPT5* and *HCLS1*. (D) Box plots comparing *LAPT5* and *HCLS1* mRNA expression across cancer types that were deposited in the TCGA. Normalized counts were obtained using RNA-Seq by Expectation Maximization (RSEM) software. The left and right ends of the boxes represent the 25th and 75th percentile values, respectively, and the segment in the middle is the median. The left and right extremes of the bars extend to the minimum and maximum values.

Figure 3A-B). TCGA revealed that the expression of *LAPT5* and *HCLS1* is specifically enriched in patients with lymphoid malignancies (diffuse large B-cell lymphoma) as compared with solid tumors (Figure 2D). These findings were further corroborated by analyzing 1037 tumor cell lines deposited in the Cancer Cell Line Encyclopedia, which showed a significant enrichment of both *LAPT5* and *HCLS1*

in hematopoietic-related tumors (supplemental Figure 4A-B). Taken together, these observations suggest that *LAPT5* and *HCLS1* are both relevant candidate genes for characterizing familial WM individuals and are also relevant to the tumor clone.

Sanger sequencing confirmed the *LAPT5*^{c.403t} variant in the original family, being present in 3/3 patients with familial WM and

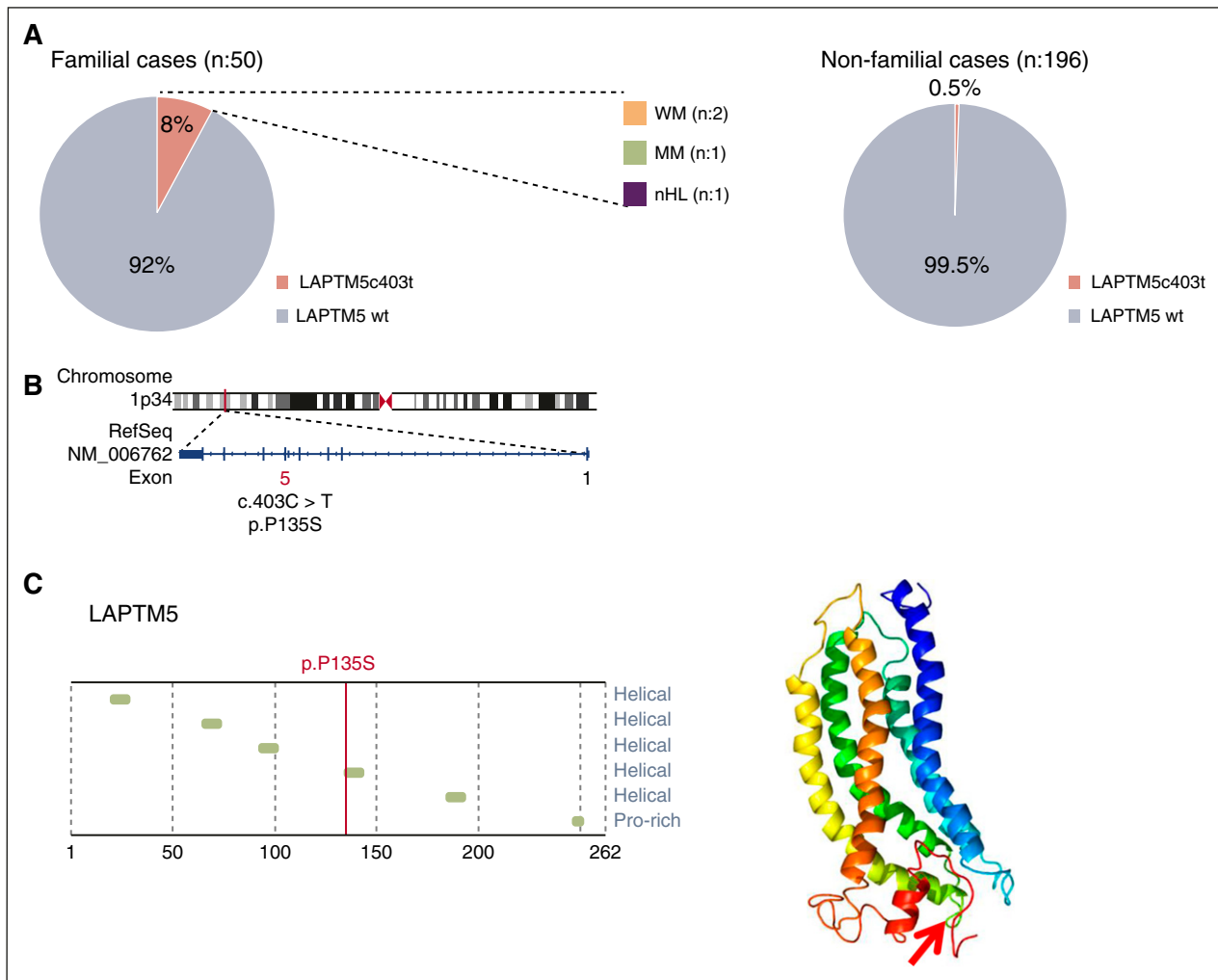


Figure 3. *LAPTMs*^{c403t/p135s} germ line variant characterizes familial WM cases. (A) Distribution of the variant among familial (n = 50) and nonfamilial (n = 196) cases. (B) Genomic structure: location of the *LAPTMs* locus is shown by a red line; the centromere is indicated by red triangles. Gene structure is also shown (National Center for Biotechnology Information Reference [RefSeq] number, NM_006762), with the blue box indicating exons. The variant is localized in exon 5. (C) Secondary structures for *LAPTMs* were predicted by Interpro (<http://www.ebi.ac.uk/interpro>). Mutation positions and amino acid changes are indicated in red. Three-dimensional reconstruction of the protein, as predicted by Phyre2: NH2-terminal and carboxyl-terminal domains are indicated in blue and red, respectively; P135S is indicated by a red arrow.

absent in the unaffected family member. We next performed Sanger sequencing in 246 independent, unrelated WM cases (50 probands were from familial cases, 196 of which were nonfamilial cases). Based on previously identified familial aggregation of WM cases and clustering of B-cell lymphoproliferative disorders among first-degree relatives of patients with WM,¹¹⁻¹⁶ we defined a familial case as a patient with first-degree relatives who were affected by a B-cell lymphoproliferative disorder, including WM, CLL, nHL, MM, IgM-MGUS, and IgG/IgA-MGUS; nonfamilial cases were individuals whose family members were either healthy or diagnosed with a solid tumor, but not a B-cell lymphoproliferative malignancy.

The *LAPTMs*^{c403t} variant was present in 4 of 50 (8%) of the familial WM cases, but was detected in only 1 of 196 (0.5%) of the nonfamilial cases, thus demonstrating a statistically significant difference in the presence of the *LAPTMs*^{c403t} variant in familial vs nonfamilial cases (supplemental Table 6; $P = .007$). Notably, the 4 familial members that presented with the *LAPTMs*^{c403t} germ line aberration had a family history of either WM (n = 2), MM (n = 1), or nHL (n = 1) (Figure 3A). The variant is located in exon 5 of the *LAPTMs* gene (Figure 3B). *LAPTMs* is a 29-kDa protein consisting of 5 hydrophobic transmembrane helical domains and is preferentially expressed in cells

of lymphoid and myeloid origin.³⁸ The variant is predicted as missense mutation that causes an amino-acid substitution from proline to serine at the fourth transmembrane helical domain (*LAPTMs*^{P135S}; Figure 3C).

Sanger sequencing also confirmed the *HCLSI*^{g496a} variant to be detectable in the original family, being present in 3/3 patients with familial WM and absent in the unaffected family member. The presence of the *HCLSI*^{g496a} variant was confirmed in 4 of 50 (8%) of independent familial WM cases (4/50) and in 1 of 196 (0.5%) of the nonfamilial cases, documenting a statistically significant difference in the numbers of patients carrying the *HCLSI*^{g496a} variant in familial vs nonfamilial cases (supplemental Table 6; $P = .007$). Notably, the 4 patients with familial WM that presented with the *HCLSI*^{g496a} germ line aberration had a family history of CLL (n = 2), WM (n = 1), or nHL (n = 1) (Figure 4A). The variant is located within exon 7 of the *HCLSI* gene (Figure 4B). *HCLSI* is a 79-kDa intracellular protein that consists of an Arp2/3 complex binding domain, 3.5 tandem repeats, and a coiled-coil region that binds to F-actin and a carboxyl-terminal SH3 domain.³⁹ The variant is predicted as a missense mutation that causes an amino-acid substitution from aspartic acid to asparagine at the third HS1 repeat (*HCLSI*^{D166N}; Figure 4C).

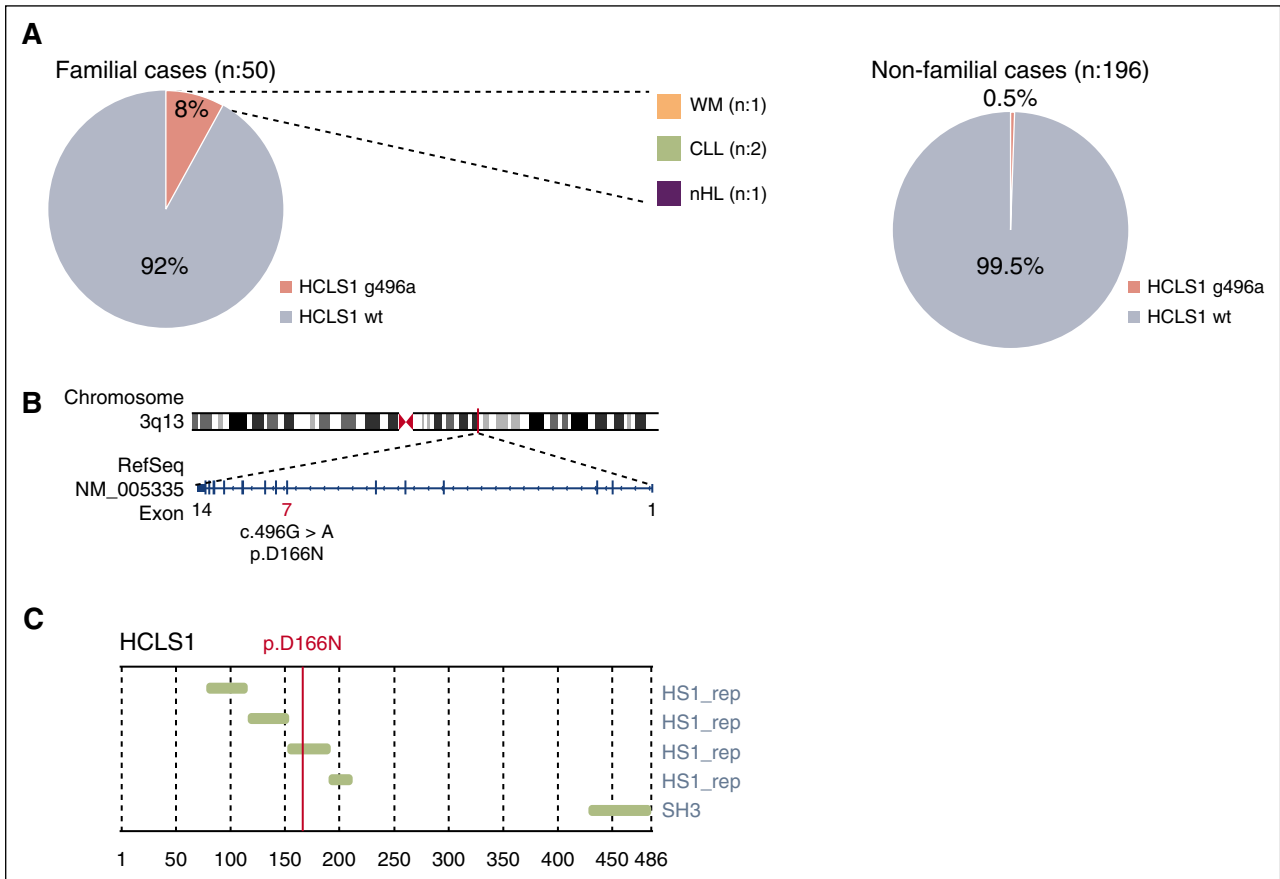


Figure 4. *HCLS1*^{g496a/D166N} germline variant characterizes familial WM cases. (A) Distribution of the variant among familial (n = 50) and nonfamilial (n = 196) cases. (B) Genomic structure: location of the *HCLS1* locus is shown by a red line; the centromere is indicated by red triangles. Gene structure is also shown (National Center for Biotechnology Information Reference [RefSeq] number, NM_005335). The exon is indicated by a blue box. (C) Secondary structures for *HCLS1* were predicted by Interpro (<http://www.ebi.ac.uk/interpro>). Mutation positions and amino acid changes are indicated in red.

LAPTM5 and HCLS1 impact on WM disease biology

To interrogate a possible involvement of *LAPTM5* and *HCLS1* in WM biology, we used the Human Experimental/Functional Mapper (HEFalMp) database to construct 2 modules for *LAPTM5* and *HCLS1*. The HEFalMp database provides a global gene–gene association map, predicted by integrating hundreds of publicly available genomic datasets,⁴⁰ and has led to the identification of the top 25 genes associated with *LAPTM5* or *HCLS1* (Figure 5A). Of note, *HCLS1* was part of the *LAPTM5* module, and vice versa, suggesting that both genes function through the same biological module. In addition, 52% of the identified genes were shared between the 2 modules.

Besides *LAPTM5* and *HCLS1*, the 2 other known WM-related genes, *MYD88* and *TNFAIP3*, were also present in this subnetwork: the *MYD88* L265P somatic mutation is present in 91% of patients with WM,⁶ and *TNFAIP3*, occurring in the context of the 6q deletion, is the most frequent cytogenetic event described in WM.⁴¹ Of note, although *MYD88* belonged to the *HCLS1* module, *TNFAIP3* appeared within the *LAPTM5* module.

We next interrogated the most significant pathways that were connected to each subnetwork by using a gene–pathway connectivity map that was generated by integrating a 70K microarray, Functional Annotation of the Mammalian Genome 5, and protein–protein interaction data, as previously reported.⁴² Statistically significant pathways were identified testing 184 Kyoto Encyclopedia of Genes and Genomes pathways and using a permutation test, with an adjusted *P* value of .05. The *LAPTM5* and *HCLS1* subnetworks shared

similar enriched pathways, including immune-related pathways, B-cell receptor-, Janus kinase/signal transducer and activator of transcription-, vascular endothelial growth factor receptor-, and chemokine-signaling pathways as well as cytokine–cytokine receptor interaction (supplemental Table 7).

We combined the *LAPTM5* and *HCLS1* modules into a single WM-associated module and investigated whether this module is disrupted in WM patients compared with HDs⁴³ using an independent gene expression profile dataset (GSE6691),⁴⁴ and demonstrated a statistically significant high gene-to-gene connectivity in HDs (Figure 5B-C). In contrast, gene-to-gene connectivity was significantly inferior in WM cells (Figure 5D), wherein the subnetwork was disrupted (Figure 5E). Using 10 000 random modules with the same size as background, we found that the change in mean connectivity was significantly different in the WM vs the HD modules (*P* < .0001, Figure 5F). Together, these findings suggest a conserved high interactivity between *LAPTM5* and *HCLS1* in normal B cells, whereas WM cells present a disrupted pattern of connectivity that likely impacts disease biology.

Discussion

Recent studies on the genomic landscape of WM have described recurrent somatic aberrations, including *MYD88*^{L265P}

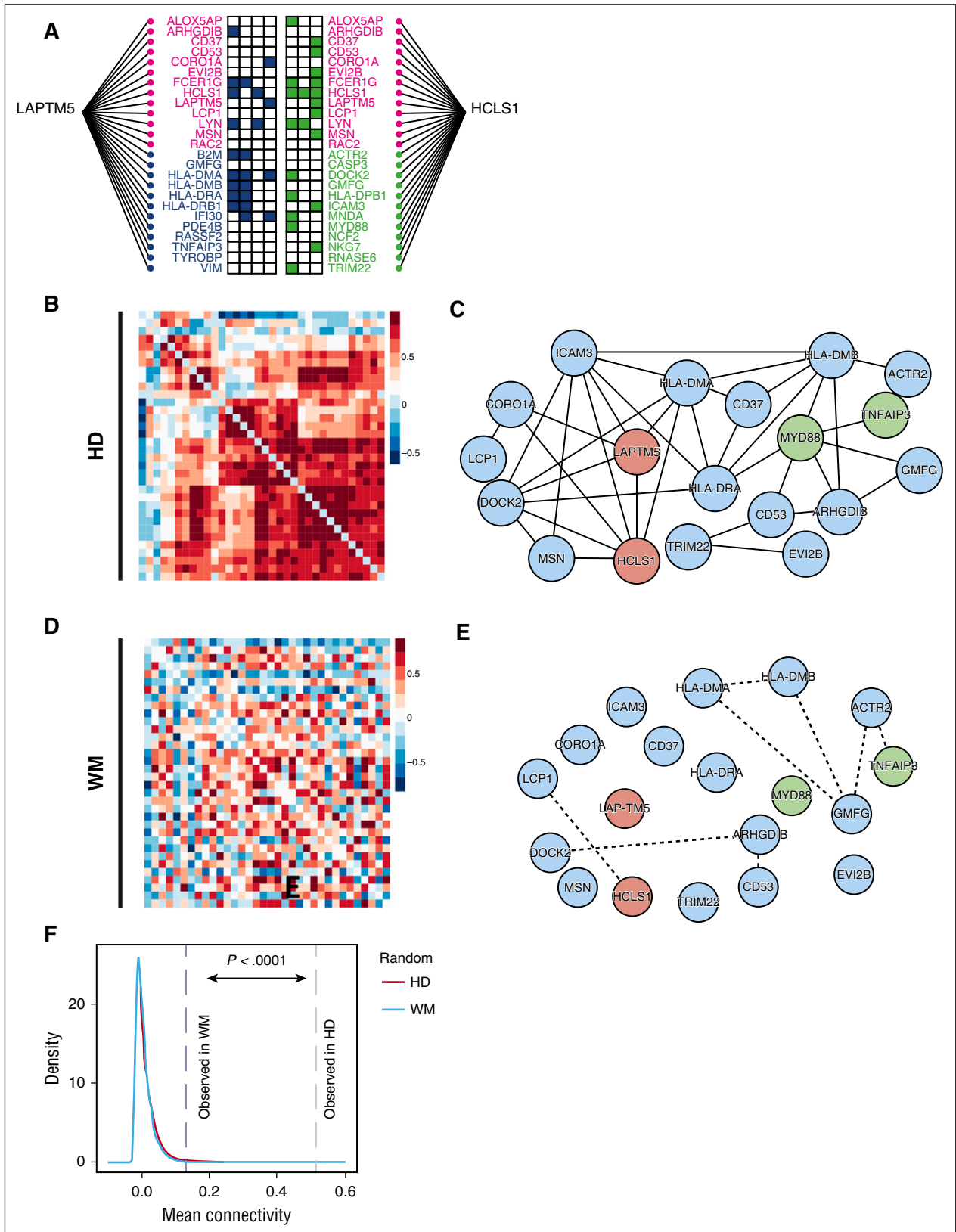


Figure 5. Involvement of *LAPTM5* and *HCLS1* subnetworks in WM. (A) *LAPTM5* and *HCLS1* modules were retrieved from the HEPaImP database as being among the top 25 WM-associated genes. Genes shared between the 2 modules are shown in pink; those unique to the *LAPTM5* module are in blue and *HCLS1*-unique genes are in green. (B) Gene-to-gene connectivities (quantified by Pearson correlation coefficients, PCC) in HDs for the *LAPTM5* and *HCLS1* combined module. (C) Combined *LAPTM5* and *HCLS1* subnetwork in HDs, with a stringent cutoff (PCC > 0.9). *LAPTM5* and *HCLS1* are highlighted in red, and 2 other known WM-associated genes are in blue. (D) Gene-to-gene connectivities in WM for the combined module. (E) Combined *LAPTM5* and *HCLS1* subnetwork in WM, with a less stringent cutoff (PCC > 0.8). (F) Ten thousand random subnetworks with the same size were sampled from WM and HD as background, and the distribution of mean connectivities was plotted as indicated; the observed change of mean connectivity in WM and HD was then compared with the background. *P* value is indicated.

and *CXCR4*^{S338X},^{6-8,45} whereas the germ line determinants of familial WM cases remain unexplored. We have shown that the *LAPTM5*^{c403t} variant may predispose to familial WM, and that in transformed WM cells there is a disrupted pattern of connectivity between *LAPTM5* and *MYD88*, leading to the hypothesis that these 2 genes may interplay in supporting the pathogenesis of this disease.

In summary, the identification of a germ line variant in genes that display oncogenic properties in B-cell lymphoproliferative disorders offers new insights into the molecular mechanisms of lymphoplasmacytic lymphoma pathogenesis, particularly in WM. The *LAPTM5* and *HCLSI* genes show relevant tissue-specific expression for WM and predicted functional relationship to WM phenotype by expression signature. The specific and recurrent *LAPTM5*^{c403t} and *HCLSI*^{g496a} variants observed in this study demonstrated segregation with disease in a striking familial WM pedigree and enrichment in other familial cases.

According to allele and genotype frequencies from the 1000 Genomes database, ~1 of 132 people (0.7%) in the general population carries either the *HCLSI*^{g496a} or the *LAPTM5*^{c403t} variant. It is therefore expected that roughly 1 in 17 000 (0.0058%) individuals in the general population carries both of these independently assorting variants. This latter number aligns with population estimates of WM prevalence. WM represents a rare phenotype that could plausibly require more predisposing genetic factors on average than associated and much more common B-cell lymphoproliferative disorders. This would be consistent with our finding of both the *LAPTM5*^{c403t} and *HCLSI*^{g496a} variants in the 3 individuals affected with WM in our primary study family for this report. Familial cases with less striking family history were enriched for 1 variant or the other, but none had both. One can envision a model by which 2 risk variants predispose to WM by a combinatorial mechanism. A polygenic mode also remains possible.

Previous studies have reported on *LAPTM5* overexpression in patients with B-cell lymphomas.⁴⁶ Our findings indicate *LAPTM5*^{c403t} as the most significantly predicted variant to be functionally related to the WM mRNA signature. Of note, *LAPTM5* has been shown to support nuclear factor κ B (NF- κ B) activation upon tumor necrosis factor- α stimulation,⁴⁷ and primary WM cells reportedly present with a constitutive activation of the NF- κ B pathway⁴⁸; we may therefore hypothesize that mutated *LAPTM5* may possibly contribute to NF- κ B

modulation in WM cells. Further studies would be needed to better characterize the relevance of *LAPTM5*^{c403t} in regulating canonical and noncanonical NF- κ B activity in WM cells at protein level and the potential relevance of the variant in regulating WM cell proliferation.

LAPTM5^{c403t} and *HCLSI*^{g496a} may represent predisposition alleles in patients with familial WM. Future studies will be needed to clarify the penetrance of specific alleles as well as possible combinatorial effects. Our findings suggest that the contribution of the *LAPTM5*^{c403t} and *HCLSI*^{g496a} variants to WM susceptibility should be further investigated.

Acknowledgments

We thank the International Waldenström's Macroglobulinemia Foundation, the Michelle and Steven Kirsch Laboratory for Waldenström's, and The Heje fellowship for Waldenström's. We also thank Dr Sonal Jhaveri for editing the manuscript.

Authorship

Contribution: A.M.R., A.S., J.S., and I.M.G. conceived and designed the experiments and analyzed the data; A.M.R. wrote the manuscript; J.S. and W.H. performed bioinformatics analysis; A.P.-G., A.S., M. Chiarini, Y.A., Y.M., and M.M. performed the experiments; M. Correll, S.M., S.G., E.M.V.A., and Y.K. analyzed the data; and J.J.C., S.P.T., L.I., G.R., J.R.B., M.R.I., M.L.F., I.R., E.H., and I.M.G. revised the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Irene M. Ghobrial, Dana-Farber Cancer Institute, Department of Medical Oncology, 450 Brookline Ave, Boston, MA 02115; e-mail: irene_ghobrial@dfci.harvard.edu; and Aldo M. Roccaro, Dana-Farber Cancer Institute, Department of Medical Oncology, 450 Brookline Ave, Boston, MA 02115; e-mail: aldo_roccaro@dfci.harvard.edu.

References

- Harris NL, Jaffe ES, Diebold J, et al. The World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues. Report of the Clinical Advisory Committee meeting, Airlie House, Virginia, November, 1997. *Ann Oncol*. 1999;10(12):1419-1432.
- Gertz MA. Waldenström macroglobulinemia: 2015 update on diagnosis, risk stratification, and management. *Am J Hematol*. 2015;90(4):346-354.
- Buske C, Leblond V. How to manage Waldenström's macroglobulinemia. *Leukemia*. 2013;27(4):762-772.
- Kyle RA, Dispenzieri A, Kumar S, Larson D, Therneau T, Rajkumar SV. IgM monoclonal gammopathy of undetermined significance (MGUS) and smoldering Waldenström's macroglobulinemia (SWM). *Clin Lymphoma Myeloma Leuk*. 2011;11(1):74-76.
- Kyle RA, Therneau TM, Rajkumar SV, et al. Long-term follow-up of IgM monoclonal gammopathy of undetermined significance. *Blood*. 2003;102(10):3759-3764.
- Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenström's macroglobulinemia. *N Engl J Med*. 2012;367(9):826-833.
- Hunter ZR, Xu L, Yang G, et al. The genomic landscape of Waldenström macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. *Blood*. 2014;123(11):1637-1646.
- Roccaro AM, Sacco A, Jimenez C, et al. C1013G/CXCR4 acts as a driver mutation of tumor progression and modulator of drug resistance in lymphoplasmacytic lymphoma. *Blood*. 2014;123(26):4120-4131.
- Cao Y, Hunter ZR, Liu X, et al. The WHIM-like CXCR4(S338X) somatic mutation activates AKT and ERK, and promotes resistance to ibrutinib and other agents used in the treatment of Waldenström's macroglobulinemia. *Leukemia*. 2015;29(1):169-176.
- Blattner WA, Garber JE, Mann DL, et al. Waldenström's macroglobulinemia and autoimmune disease in a family. *Ann Intern Med*. 1980;93(6):830-832.
- Kristinsson SY, Björkholm M, Goldin LR, McMaster ML, Turesson I, Landgren O. Risk of lymphoproliferative disorders among first-degree relatives of lymphoplasmacytic lymphoma/Waldenström macroglobulinemia patients: a population-based study in Sweden. *Blood*. 2008;112(8):3052-3056.
- Massari R, Fine JM, Metais R. Waldenström's macroglobulinemia observed in two brothers. *Nature*. 1962;196:176-178.
- McMaster ML, Csako G, Giambarresi TR, et al. Long-term evaluation of three multiple-case Waldenström macroglobulinemia families. *Clin Cancer Res*. 2007;13(17):5063-5069.
- Renier G, Ifrah N, Chevailler A, Saint-Andre JP, Boasson M, Hurez D. Four brothers with Waldenström's macroglobulinemia. *Cancer*. 1989;64(7):1554-1559.
- Royer RH, Koshiol J, Giambarresi TR, Vasquez LG, Pfeiffer RM, McMaster ML. Differential characteristics of Waldenström macroglobulinemia according to patterns of familial aggregation. *Blood*. 2010;115(22):4464-4471.
- Treon SP, Hunter ZR, Aggarwal A, et al. Characterization of familial Waldenström's macroglobulinemia. *Ann Oncol*. 2006;17(3):488-494.
- Crowther-Swanepoel D, Houlston RS. The molecular basis of familial chronic lymphocytic leukemia. *Haematologica*. 2009;94(5):606-609.

18. McMaster ML, Goldin LR, Bai Y, et al. Genomewide linkage screen for Waldenstrom macroglobulinemia susceptibility loci in high-risk families. *Am J Hum Genet.* 2006;79(4):695-701.
19. Broderick P, Chubb D, Johnson DC, et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet.* 2012;44(1):58-61.
20. Chubb D, Weinhold N, Broderick P, et al. Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nat Genet.* 2013;45(10):1221-1225.
21. Swaminathan B, Thorleifsson G, Jöud M, et al. Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. *Nat Commun.* 2015;6:7213.
22. Weinhold N, Johnson DC, Chubb D, et al. The CCND1 c.870G>A polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. *Nat Genet.* 2013;45(5):522-525.
23. Kuppusamy H, Ogmundsdottir HM, Baigorri E, et al. Inherited polymorphisms in hyaluronan synthase 1 predict risk of systemic B-cell malignancies but not of breast cancer. *PLoS One.* 2014;9(6):e100691.
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(4):1754-1760.
25. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum Mutat.* 2013;34(9):E2393-E2402.
26. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-498.
27. Cingolani P, Platts A, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012;6(2):80-92.
28. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* 2004;3:Article3.
29. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013; 6(269):pl1.
30. Raychaudhuri S, Plenge RM, Rossin EJ, et al; International Schizophrenia Consortium. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* 2009;5(6):e1000534.
31. Obayashi T, Okamura Y, Ito S, Tadaka S, Motoike IN, Kinoshita K. COXPRESdb: a database of comparative gene coexpression networks of eleven species for mammals. *Nucleic Acids Res.* 2013;41(Database issue):D1014-D1020.
32. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* 2009;4(3):363-371.
33. Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood.* 2011; 117(19):5019-5032.
34. Braggio E, Keats JJ, Leleu X, et al. Identification of copy number abnormalities and inactivating mutations in two negative regulators of nuclear factor-kappaB signaling pathways in Waldenstrom's macroglobulinemia. *Cancer Res.* 2009;69(8):3579-3588.
35. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4): 248-249.
36. Trynka G, Sandor C, Han B, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet.* 2013;45(2): 124-130.
37. Benita Y, Cao Z, Giallourakis C, Li C, Gardet A, Xavier RJ. Gene enrichment profiles reveal T-cell development, differentiation, and lineage-specific transcription factors including ZBTB25 as a novel NF-AT repressor. *Blood.* 2010;115(26): 5376-5384.
38. Scott LM, Mueller L, Collins SJ. E3, a hematopoietic-specific transcript directly regulated by the retinoic acid receptor alpha. *Blood.* 1996;88(7):2517-2530.
39. Takemoto Y, Sato M, Furuta M, Hashimoto Y. Distinct binding patterns of HS1 to the Src SH2 and SH3 domains reflect possible mechanisms of recruitment and activation of downstream molecules. *Int Immunol.* 1996;8(11):1699-1705.
40. Huttenhower C, Haley EM, Hibbs MA, et al. Exploring the human genome with functional maps. *Genome Res.* 2009;19(6): 1093-1106.
41. Braggio E, Philipsborn C, Novak A, Hodge L, Ansell S, Fonseca R. Molecular pathogenesis of Waldenstrom's macroglobulinemia. *Haematologica.* 2012;97(9):1281-1290.
42. Pers TH, Karjalainen JM, Chan Y, et al; Genetic Investigation of Anthropometric Traits (GIANT) Consortium. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun.* 2015;6:5890.
43. May RM, Okumura M, Hsu CJ, et al. Murine natural killer immunoreceptors use distinct proximal signaling complexes to direct cell function. *Blood.* 2013;121(16):3135-3146.
44. Gutiérrez NC, Ocio EM, de Las Rivas J, et al. Gene expression profiling of B lymphocytes and plasma cells from Waldenström's macroglobulinemia: comparison with expression patterns of the same cell counterparts from chronic lymphocytic leukemia, multiple myeloma and normal individuals. *Leukemia.* 2007;21(3): 541-549.
45. Cao Y, Hunter ZR, Liu X, et al. The WHIM-like CXCR4(S338X) somatic mutation activates AKT and ERK, and promotes resistance to ibrutinib and other agents used in the treatment of Waldenstrom's macroglobulinemia. *Leukemia.* 2015;29(1):169-176.
46. Seimiya M, O-Wang J, Bahar R, et al. Stage-specific expression of Clast6/E3/LAPTM5 during B cell differentiation: elevated expression in human B lymphomas. *Int J Oncol.* 2003;22(2): 301-304.
47. Glowacka WK, Alberts P, Ouchida R, Wang JY, Rotin D. LAPTM5 protein is a positive regulator of proinflammatory signaling pathways in macrophages. *J Biol Chem.* 2012;287(33): 27691-27702.
48. Leleu X, Eeckhoutte J, Jia X, et al. Targeting NF-kappaB in Waldenstrom macroglobulinemia. *Blood.* 2008;111(10):5068-5077.