Blood First Edition Paper, prepublished online February 22, 2016; DOI 10.1182/blood-2015-11-680199

Exome sequencing reveals recurrent germline variants in patients with familial Waldenstrom's Macroglobulinemia.

Aldo M. Roccaro,[1,4*] Antonio Sacco,[1*] Jiantao Shi,[2*] Marco Chiarini,[1,4] 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,[4] Giuseppe Rossi,[4] Jorge Castillo,[1] Steven P. Treon,[1] Matthew L. Freedman,[1] Eliezer M. Van Allen,[1] Winston Hide,[3] Elaine Hiller,[1] Irene Rainville,[1] Irene M. Ghobrial[1]

1) Dana-Farber Cancer Institute, Department of Medical Oncology, Harvard Medical School, Boston, 02215, MA, USA
2) Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, 02215, USA
3) Department of Biostatistics, Harvard School of Public Health, Boston, MA, 02215, USA
4) Spedali Civili di Brescia, Centro Ricerca Emato-oncologica AIL (CREA), Dept. of Hematology, 25123 Brescia, Italy

Corresponding Authors:
Irene M. Ghobrial, MD
Dana-Farber Cancer Institute
Dept. of Medical Oncology
450 Brookline Av., Boston, MA, 02115
Phone: +1.617.632.4198
Email: irene_ghobrial@dfci.harvard.edu

Aldo M. Roccaro, MD, PhD
Dana-Farber Cancer Institute
Dept. of Medical Oncology
450 Brookline Av., Boston, MA, 02115
Phone: +1.617.632.6733
Email: aldo_roccaro@dfci.harvard.edu

*AMR, AS, JS: Co-first authors: authors contributed equally to this work

Running title: Familial WM
Abstract: 162 words
Text: 3317 words
Figures: 5
Supplemental data: 4 supplemental figures; 7 supplemental tables

KEY POITNS

- ✓ *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.

ABSTRACT

Familial aggregation of Waldenström's 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 germline DNA obtained from four family members where coinheritance for WM was documented in three of them, and screened additional independent 246 cases, by using gene-specific mutation sequencing. Among the shared germline variants, *LAPTM5*^{c403t} and *HCLS1*^{g496a} represented the most recurrent ones, 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 non-familial 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.

INTRODUCTION

The evaluation of cancer occurrence within families is important for unraveling the molecular events that drive tumorigenesis. Waldenström's Macroglobulinemia (WM) represents a B-cell lymphoproliferative disorder, classified as a lymphoplasmacytic lymphoma, according to the WHO classification.[1] WM represents a rare B-cell malignancy that accounts for 1-2% of all hematologic neoplasms, with an incidence rate of 3-4 cases per million people, per year.[2,3] Evidence also exists that IgM 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 30-35% 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, GWAS studies have identified common variants associated with multiple myeloma.[19-22] Case control studies have similarly identified significant association with other lymphoproliferative disorders.[20,23] Whether relatively common germline variants may contribute specifically to familial WM cases, remains unexplored.

We therefore performed whole exome sequencing on germline DNA obtained from four members of a single family with documented coinheritance of WM (three affected; 1 unaffected) and applied bioinformatic tools to identify candidate germline variants likely to have a biological role in WM signaling pathways. We screened additional 246 independent, unrelated WM cases (50 probands from familial cases and 196 individuals from non-familial 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 non-familial cases and in <0.05 of a control population (1000 Genomes). These findings highlight potentially 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 three first-degree relatives that were affected with WM, and a healthy unaffected family member who was considered to be a control, and performed whole exome sequencing. A total of 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, while the remaining 196 were non-familial patients. Based on previous identification of familial aggregation of WM cases and clustering of B-cell lymphoproliferative disorders (including chronic lymphocytic leukemia-CLL, multiple myeloma-MM, non-Hodgkin lymphoma) 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, non-Hodgkin lymphoma, MM, IgM-MGUS and IgG/IgA-MGUS; non-familial 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 100x (Center for Cancer Computational Biology (CCCB), Dana-Farber Cancer Institute, Boston, MA). Reads were aligned to GRCh37 using BWA[24] with default parameters, and the resulting SAM files were converted to BAM files using Picard (http://picard.sourceforge.net/). Accession number: SRP053196.

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 around 51M, 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.[27] 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,[25] 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 single nucleotide variants (SNVs) were annotated by snpEff[26] and Oncotator (http://www.broadinstitute.org/oncotator/), and 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 non-synonymous

SNVs in the human genome.[27] Allele frequencies (AF) in 1000 Genomes, as well as PolyPhen-2 prediction, were used to quantify the deleteriousness of SNVs. Variants with AF >0.05 in 1000G 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 two-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 cbioportal (http://cbioportal.org),[29] which integrates data generated by the TCGA Research Network (http://cancergenome.nih.gov/) and by many other sources. Gene expression modules were assessed using an independent mRNA data set (GSE6691).

Disease gene prioritization

Given the selected potential familial WM-associated variants/genes, we used GRAIL[30] coupled with a global coexpression network COXPRESdb[31] to assess the functional relatedness between these genes and those that were differentially expressed between bone marrow-derived primary WM cells compared to their normal counterparts, using GSE12668.

Sanger Sequencing

The observed *LAPTM5*, *HCLS1* variants were validated by Sanger Sequencing. The Fisher exact test was used to assess the significance of variants observed in familial cases, compared to those in non-familial cases.

3D protein modeling

Three-dimensional *LAPTM5* protein reconstruction was obtained using the Phyre2 server, as described.[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 a FDR cutoff of 1%. Fisher exact test was used to assess the significance of variants observed in familial cases compared to those in non-familial cases.

RESULTS

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

We studied a kindred in which three members are affected by familial WM (Fig. 1A). The diagnosis of WM was confirmed in all cases by histology and immunohistochemistry,[33] and clinical features are summarized within Suppl. Table 1. Whole exome sequencing was performed on the three affected members and one of unaffected members. The criteria to identify candidate WM-associated variants were non-synonymous SNVs that were present only in the affected family members and absent in the unaffected member and an allele frequency <0.05 using the 1000 Genomes. This initial screen identified 132 candidate exonic, non-synonymous familial WM variants, mapping to 127 genes (Fig. 1B; Suppl. Table 2; Suppl. Fig. 1). We next performed a gene/variant prioritization in order to select the most significant WM relevant variants, by using Gene Relationships Across Implicated Loci (GRAIL),[30] which integrates text mining or co-expression databases for gene prioritization. GRAIL takes a group of seed genes to build a sub-network and test whether a query gene is functionally related to the seeds. Since 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 (HD); we found 393 genes that were significantly different between WM patients and healthy individuals, thus confirming the presence of a specific WM mRNA signature (false discovery rate –FDR-cutoff <1%), (Fig. 2A; Suppl. Table 3). We next assessed the functional relatedness between potential familial WM-associated genes and the observed WM mRNA signature, using the GRAIL algorithm [30] coupled with a global co-expression network COXPRESdb, as reported.[31] Using an FDR cutoff of 5%, 13 genes were predicted to be functionally related to the WM mRNA signature (Fig. 2B). In addition, the observed germline variants were annotated and interpreted by implementing Polymorphism Phenotyping v2

(PolyPhen-2), as described,[35] thus allowing a prioritization for the deleteriousness of single nucleotide variants (SNVs) shared between all three affected WM family members.

It has been reported that disease-related genes tend to be tissue specific;[36] we therefore assessed the tissue specificity of the selected genes for further prioritization. The Gene Enrichment Profiler (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 lymphoma:[1] we thus utilized 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 hematopoietic cell-specific Lyn substrate 1 (*HCLS1*) and lysosomal protein transmembrane 5 (*LAPTM5*) (Fig. 2C; Suppl. Table 4). Variants in these two genes co-segregated within the affected family members; being absent in the un-affected family member; and present in less than 1% of the control population, where allele frequency and heterozigote frequency were 0.38%/0.7% and 0.39%/0.7% for *LAPTM5* and *HCLS1*, respectively, using the 1000 Genomes (Suppl. Table 5). Both genes are highly B-cell tissue specific (Suppl. Fig. 2A-B-C). In contrast, the two control genes HRC and IL22RA1 (which are not functionally related to WM expression signature, but contain benign variants) are not B-cell specific (Suppl. Fig. 3A-B). The Cancer Genome Atlas (TCGA) revealed that the expression of *LAPTM5* and *HCLS1* is specifically enriched in patients with lymphoid malignancies (diffuse large B-cell lymphoma), as compared to solid tumors (Fig. 2D). These findings were further corroborated by analyzing 1037 tumor cell lines deposited in the Cancer Cell Line Encyclopedia (CCLE), which showed a significant enrichment of both *LAPTM5* and *HCLS1* in hematopoietic-related tumors (Suppl. Fig. 4A-B). Taken together, these observations suggest that *LAPTM5* and *HCLS1* are both relevant candidate genes for characterizing familial WM individuals, and are also relevant to the tumor clone.

Sanger sequencing confirmed the *LAPTM5*^{c403t} variant in the original family, being present in 3/3 patients with familial WM and absent in the unaffected family member. We next performed Sanger sequencing in 246 independent, unrelated WM cases (50 probands were from familial cases; 196 were non-familial 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, chronic lymphocytic leukemia (CLL), non-Hodgkin lymphoma (nHL), multiple myeloma (MM), IgM-MGUS and IgG/IgA-MGUS; non-familial cases were individuals whose family members were either healthy or diagnosed with a solid tumor, but not a B-cell lymphoproliferative malignancy.

The LAPTM5^{c403t} variant was present in four out of 50 (8%) of the familial WM cases, but was detected in only 1 out of 196 (0.5%) of the non-familial cases, thus demonstrating a statistically significant difference in the presence of the *LAPTM5*^{c403t} variant in familial versus non-familial cases (Suppl. Table 6; *P* 0.007). Notably, the 4 familial members that presented with the *LAPTM5*^{c403t} germline aberration, had a family history of either WM (n:2), MM (n:1), or nHL (n:1) (Fig. 3A). The variant is located in exon 5 of the *LAPTM5* gene (Fig. 3B). *LAPTM5* is a 29-kDA protein, consists of 5 hydrophobic trans-membrane 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 4th trans-membrane helical domain (*LAPTM5*^{P135S}; Fig. 3C).

Sanger sequencing also confirmed the *HCLS1*^{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 *HCLS1*^{g496a} variant was confirmed in four out of 50 (8%) of independent familial WM cases (4/50), and in 1 out of 196 (0.5%) of the non-familial cases, documenting a statistically significant difference in the numbers of patients carrying the *HCLS1*^{g496a} variant in familial versus non-familial cases (Suppl. Table 6; *P* 0.007). Notably, the 4 patients with familial WM that presented with the *HCLS1*^{g496a} germline aberration, had a family history of CLL (n: 2), WM (n: 1) or nHL (n: 1) (Fig. 4A). The variant is located within exon 7 of the *HCLS1* gene (Fig. 4B).

HCLS1 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 C-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 3rd HS1 repeat (*HCLS1*^{D166N}; Fig. 4C).

LAPTM5 and *HCLS1* impact on WM disease biology

To interrogate a possible involvement of *LAPTM5* and *HCLS1* in WM biology, we used the HEFalMp database to construct two 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 led to identification of the top 25 genes associated with *LAPTM5* or *HCLS1* (Fig. 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 two modules.

Besides *LAPTM5* and *HCLS1*, the two other known WM-related genes *MYD88* and *TNFAIP3*, were also present in this sub-network: 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, while *MYD88* belonged to the *HCLS1* module, *TNFAIP3* appeared within the *LAPTM5* module.

We next interrogated the most significant pathways that were connected to each sub-network by using a gene-pathway connectivity map, generated by integrating 70K microarray, FANTOM 5 and protein-protein interaction data, as previously reported.[42] Statistically significant pathways were identified testing 184 KEGG pathways and using permutation test, with adjusted p-value of 0.05. The *LAPTM5* and *HCLS1* sub-networks shared similar enriched pathways, including immune-related pathways, B-cell receptor-, JAK/STAT-, VEGF- and chemokine-signaling pathways, as well as cytokine-cytokine receptor interaction (Suppl. 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 to

HDs,[43] using an independent gene expression profile dataset (GSE6691),[44] and demonstrated a statistically significant high gene-to-gene connectivity in HDs (Fig. 5B-C); in contrast, gene-to-gene connectivity was significantly inferior in WM cells (Fig. 5D), wherein the sub-network was disrupted (Fig. 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- versus the HD-modules ($P<0.0001$, Fig. 5F). Together, these findings suggest a conserved high interactivity between *LAPTM5* and *HCLS1* in normal B cells, while WM cells present a disrupted pattern of connectivity, which likely impacts disease biology.

DISCUSSION

Recent studies on the genomic landscape of WM have described recurrent somatic aberrations, including *MYD88*^{L265P} and *CXCR4*^{S338X}, [6-8,45] while the germline 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 two genes may interplay in supporting the pathogenesis of this disease.

In summary, the identification of a germline 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 *HCLS1* 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 *HCLS1*^{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, about 1 out of 132 people (0.7%) in the general population carries either the *HCLS1*^{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 the *HCLS1*^{g496a} variants in the three individuals affected with WM in our primary study family for this report. Familial cases with less striking family history were enriched for one variant or the other, but none had both. One can envision a model by which two risk variants predispose to WM by a combinatorial mechanism. A polygenic mode also remains possible.

Previous studies have reported on *LAPTM5* over-expression 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 NF-κB activation upon TNF-α stimulation,[47] and primary WM cells reportedly present with a constitutive activation of the NF-κB pathway:[48] we may therefore hypothesize that mutated LAPTM5 may possibly contribute to NF-kB modulation in WM cells. Further studies would be needed to better characterize the relevance of *LAPTM5*^{c403t} in regulating canonical and non-canonical NF-kB activity in WM cells at protein level, and the potential relevance of the variant in regulating WM cell proliferation.

LAPTM5^{c403t} and *HCLS1*^{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 *HCLS1*^{g496a} variants to WM susceptibility should be further investigated.

ACKNOWLEDGEMENTS

International Waldenstrom's Macroglobulinemia Foundation (IWMF); Michelle and Steven Kirsch laboratory for Waldenstrom, The Heje fellowship for Waldenstrom. We thank Dr. Sonal Jhaveri for editing the manuscript.

AUTHORSHIP CONTRIBUTIONS

AMR, AS, JS, IMG: conceived and designed the experiments; analyzed the data

AMR: wrote the manuscript

JS, WH: performed bioinformatics analysis

APG, AS, MC, YA, YM, MM: performed the experiments

MC, SM, SG, EMVA, YK: analyzed the data

JC, SPT, JRB, MRI, MLF, RI, HE, IMG: revised the manuscript

CONFLICT OF INTEREST DISCLOSURE

Authors have no conflict of interest to disclose.

DATA AVAILABILITY

SRP053196

REFERENCES

1. 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.
2. Gertz MA. Waldenstrom macroglobulinemia: 2015 update on diagnosis, risk stratification, and management. *Am J Hematol*. 2015;90(4):346-354.
3. Buske C, Leblond V. How to manage Waldenstrom's macroglobulinemia. *Leukemia*. 2013;27(4):762-772.
4. Kyle RA, Dispenzieri A, Kumar S, Larson D, Therneau T, Rajkumar SV. IgM monoclonal gammopathy of undetermined significance (MGUS) and smoldering Waldenstrom's macroglobulinemia (SWM). *Clin Lymphoma Myeloma Leuk*. 2011;11(1):74-76.
5. 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.
6. Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia. *N Engl J Med*. 2012;367(9):826-833.
7. Hunter ZR, Xu L, Yang G, et al. The genomic landscape of Waldenstrom 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.
8. 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.
9. 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.
10. Blattner WA, Garber JE, Mann DL, et al. Waldenstrom's macroglobulinemia and autoimmune disease in a family. *Ann Intern Med*. 1980;93(6):830-832.
11. Kristinsson SY, Bjorkholm M, Goldin LR, McMaster ML, Turesson I, Landgren O. Risk of lymphoproliferative disorders among first-degree relatives of lymphoplasmacytic lymphoma/Waldenstrom macroglobulinemia patients: a population-based study in Sweden. *Blood*. 2008;112(8):3052-3056.
12. Massari R, Fine JM, Metais R. Waldenstrom's macroglobulinaemia observed in two brothers. *Nature*. 1962;196:176-178.
13. McMaster ML, Csako G, Giambarresi TR, et al. Long-term evaluation of three multiple-case Waldenstrom macroglobulinemia families. *Clin Cancer Res*. 2007;13(17):5063-5069.
14. Renier G, Ifrah N, Chevailler A, Saint-Andre JP, Boasson M, Hurez D. Four brothers with Waldenstrom's macroglobulinemia. *Cancer*. 1989;64(7):1554-1559.
15. Royer RH, Koshiol J, Giambarresi TR, Vasquez LG, Pfeiffer RM, McMaster ML. Differential characteristics of Waldenstrom macroglobulinemia according to patterns of familial aggregation. *Blood*. 2010;115(22):4464-4471.
16. Treon SP, Hunter ZR, Aggarwal A, et al. Characterization of familial Waldenstrom's macroglobulinemia. *Ann Oncol*. 2006;17(3):488-494.

17. 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, Joud 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. Roccaro AM, Sacco A, Thompson B, et al. MicroRNAs 15a and 16 regulate tumor proliferation in multiple myeloma. *Blood*. 2009;113(26):6669-6680.
25. 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.
26. Cingolani P, Platts A, Wang le 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.
27. 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):E2393-2402.
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. 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):D1014-1020.
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. 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. Gutierrez NC, Ocio EM, de Las Rivas J, et al. Gene expression profiling of B lymphocytes and plasma cells from Waldenstrom'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 somatic mutation activates AKT and ERK, and promotes resistance to ibrutinib and other agents used in the treatment of Waldenstrom's Macroglobulinemia. *Leukemia*. 2014.
46. Seimiya M, J OW, 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, Eeckhoute J, Jia X, et al. Targeting NF-kappaB in Waldenstrom macroglobulinemia. *Blood*. 2008;111(10):5068-5077.

FIGURE LEGENDS

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. For the family we studied, three members were identified as affected with WM (II:1, III:1, III:2). Arrow indicates the proband (first member diagnosed with WM). Asterisk indicates the sequenced samples. Age of diagnosis is indicated for the three affected WM members. The sequenced unaffected member was 75 years old 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.

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; non-significant 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 (GRCh37) as well as cDNA and protein changes are listed. Deleteriousness of a variant was predicted by Polyphen2 either as benign (B), damaging (D), or possibly damaging (P). **(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* value are shown for *LAPTM5* and *HCLS1*. **(D)** Box plots comparing *LAPTM5* and *HCLS1* mRNA expression across cancer types, deposited in the TCGA. Log2 scale of RSEM was used to quantify the expression

levels. 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 3. *LAPTM5*$^{c403t/P135S}$ germline variant characterizes familial WM cases.

(A) Distribution of the variant among familial (n=50) and non-familial (n=196) cases. **(B)** Genomic structure: location of the *LAPTM5* locus is shown by a red line; centromere is indicated by red triangles. Gene structure is also shown (National Center for Biotechnology Information Reference [RefSeq] number, NM_006762), with blue box indicating exons. The variant is localized in exon 5. **(C)** Secondary structures for *LAPTM5* 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: N- and C-terminal domains, are indicated in blue and red, respectively; P135S is indicated by a red arrow.

Figure 4. *HCLS1*$^{g496a/D166N}$ germline variant characterizes familial WM cases.

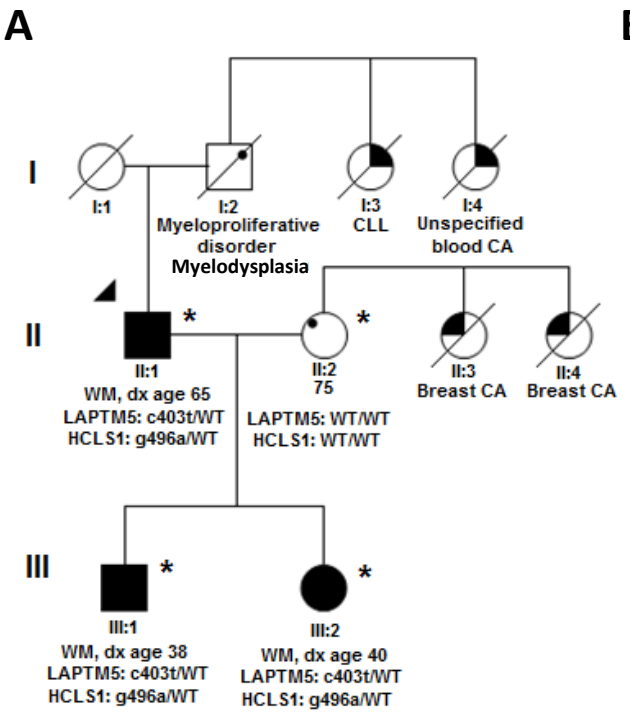
(A) Distribution of the variant among familial (n=50) and non-familial (n=196) cases. **(C)** Genomic structure: location of the *HCLS1* locus is shown by a red line; centromere is indicated by red triangles. Gene structure is also shown (National Center for Biotechnology Information Reference [RefSeq] number, NM_005335). Exon is indicated by a blue box. The variant localizes in exon 7. **(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.

Figure 5 Involvement of *LAPTM5* and *HCLS1* sub-networks in WM.

(A) *LAPTM5* and *HCLS1* modules were retrieved from the HEFalMp database, as being among the top 25 WM-associated genes. Genes shared between the two modules are shown in red; 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 healthy donors (HDs) for the *LAPTM5* and *HCLS1* combined module. **(C)** Combined *LAPTM5* and *HCLS1* sub-network in HDs, with a stringent cut-

off (PCC > 0.9). *LAPTM5* and *HCLS1* are highlighted in red, and two 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 HD, with a less stringent cut-off (PCC > 0.8). **(F)** Ten thousand random sub-networks 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 to the background. *P*-value is indicated.

FIGURE 1


A
I
II
III
I:1
I:2
Myeloproliferative disorder
Myelodysplasia
I:3
CLL
I:4
Unspecified blood CA
II:1
WM, dx age 65
LAPTM5: c403t/WT
HCLS1: g496a/WT
II:2
75
LAPTM5: WT/WT
HCLS1: WT/WT
II:3
Breast CA
II:4
Breast CA
III:1
WM, dx age 38
LAPTM5: c403t/WT
HCLS1: g496a/WT
III:2
WM, dx age 40
LAPTM5: c403t/WT
HCLS1: g496a/WT

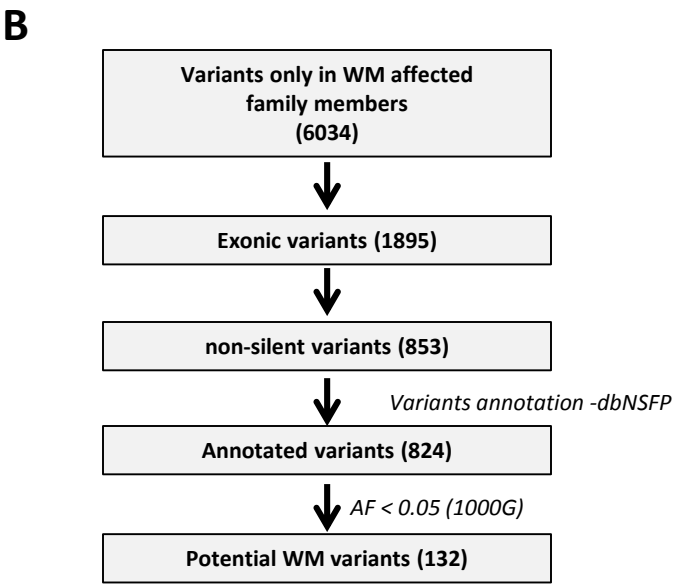
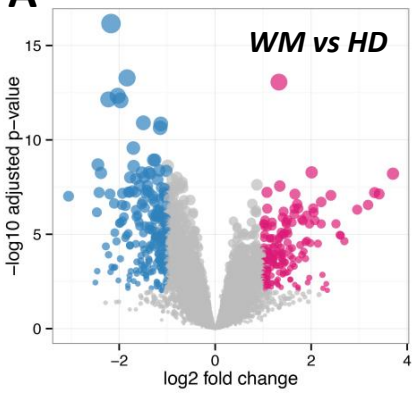

B
Variants only in WM affected family members (6034)
Exonic variants (1895)
non-silent variants (853)
Variants annotation -dbNSFP
Annotated variants (824)
AF < 0.05 (1000G)
Potential WM variants (132)

FIGURE 2

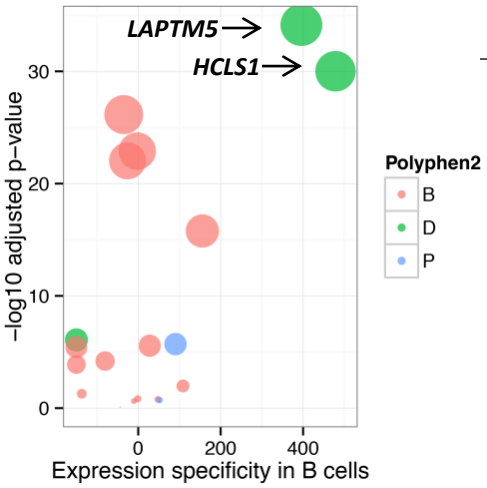
A



B

CHROM	POS	ref_allele	alt_allele	Hugo_Symbol	cDNA_Change	Protein_Change	Polyphen2	Adjusted p-value
1	31211894	G	A	LAPTM5	c.403C>T	p.P135S	D	7.35E-35
3	121356062	C	T	HCLS1	c.496G>A	p.D166N	D	9.86E-31
7	139661784	G	A	TBXAS1	c.886G>A	p.D296N	B	7.10E-27
22	43218397	T	C	ARFGAP3	c.691A>G	p.S231G	B	1.27E-23
12	8074055	T	C	SLC2A3	c.1445A>G	p.E482G	B	9.32E-23
1	159043196	C	T	AIM2	c.94G>A	p.E32K	B	1.65E-16
20	57599303	C	T	TUBB1	c.821C>T	p.T274M	D	8.34E-07
6	139495603	G	A	HECA	c.1394G>A	p.R465Q	P	1.98E-06
15	86122875	A	C	AKAP13	c.1576A>C	p.K526Q	B	2.74E-06
19	42091810	G	A	CEACAM21	c.428G>A	p.S143N	B	3.66E-06
16	56693058	G	C	MT1F	c.168G>C	p.K56N	B	6.42E-05
19	15569378	G	C	RASAL3	c.751C>G	p.L251V	B	1.29E-04
22	47095235	C	A	CERK	c.918G>T	p.L306F	B	1.06E-02

C



Symbol	adjusted P-value	Polyphen2	B cell specificity
LAPTM5	7.35E-35	D	395.88
HCLS1	9.86E-31	D	478.67

D

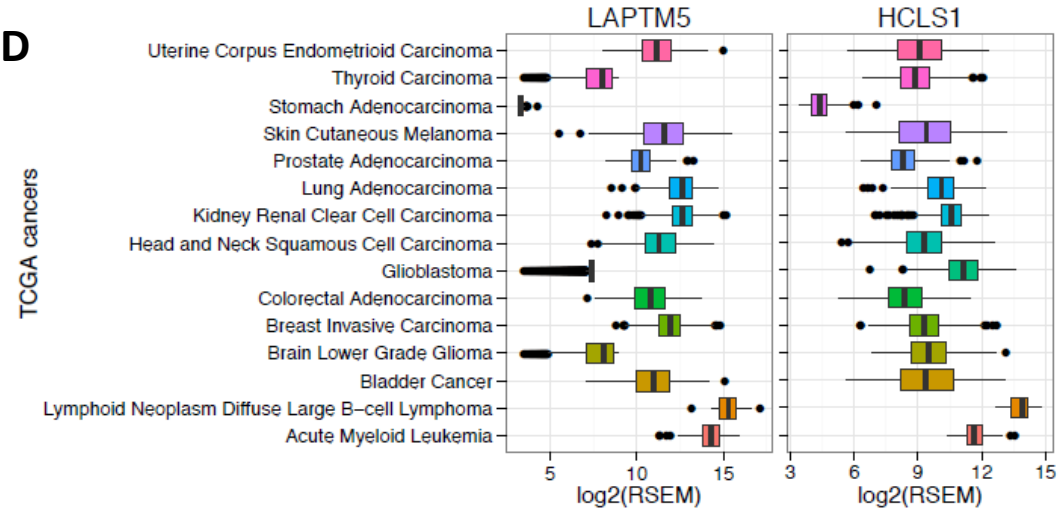
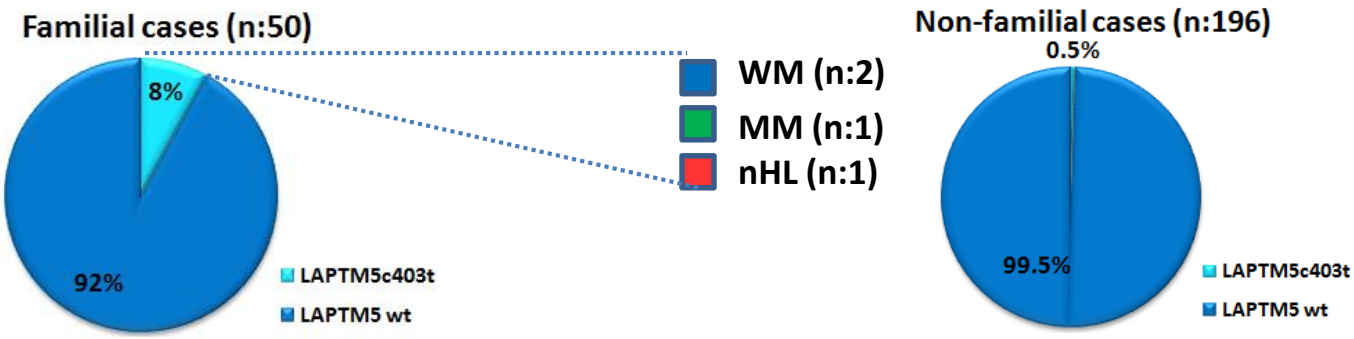
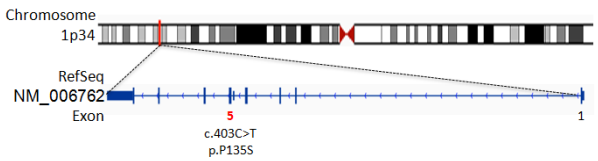


FIGURE 3

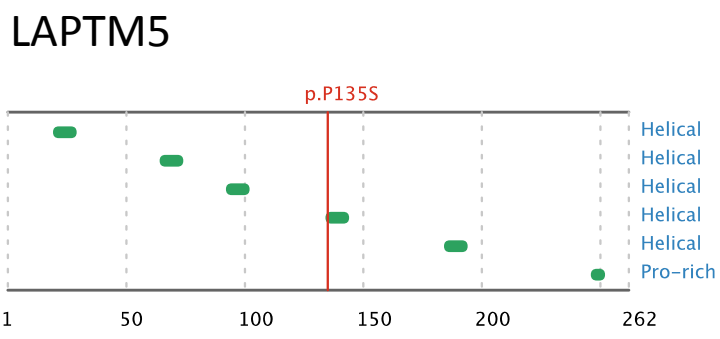
A


Familial cases (n:50)
8%
92%
LAPTM5c403t
LAPTM5 wt
WM (n:2)
MM (n:1)
nHL (n:1)
Non-familial cases (n:196)
0.5%
99.5%
LAPTM5c403t
LAPTM5 wt

B


Chromosome
1p34
RefSeq
NM_006762
Exon
5
1
c.403C>T
p.P135S

C


LAPTM5
p.P135S
Helical
Helical
Helical
Helical
Helical
Pro-rich
1
50
100
150
200
262

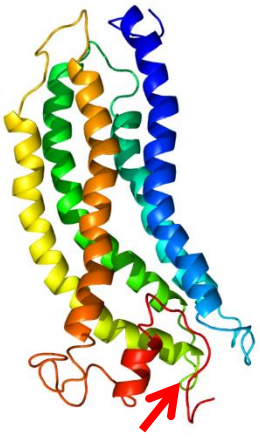
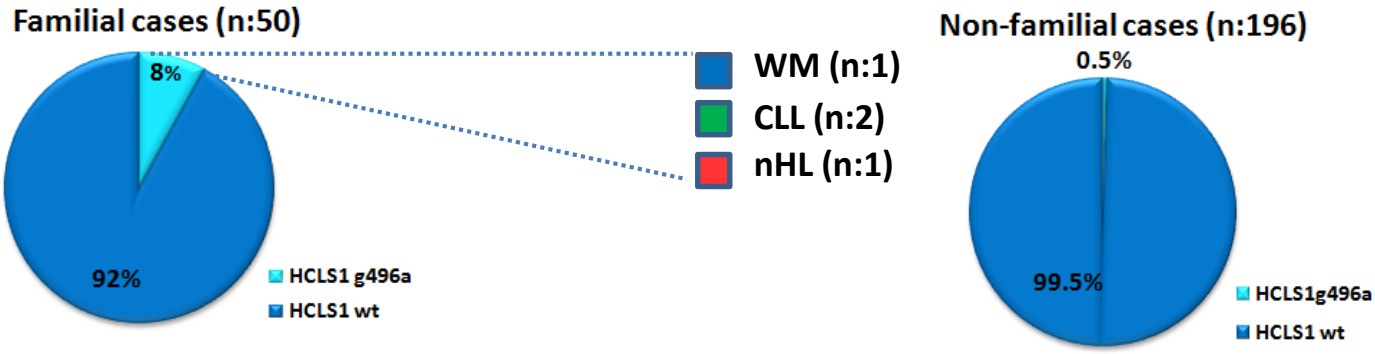
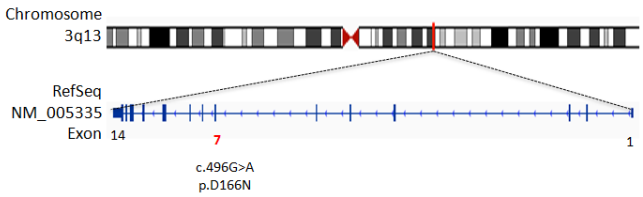


FIGURE 4

A


Familial cases (n:50)
8%
92%
HCLS1 g496a
HCLS1 wt
WM (n:1)
CLL (n:2)
nHL (n:1)
Non-familial cases (n:196)
0.5%
99.5%
HCLS1g496a
HCLS1 wt

B


Chromosome
3q13
RefSeq
NM_005335
Exon
14
7
1
c.496G>A
p.D166N

C

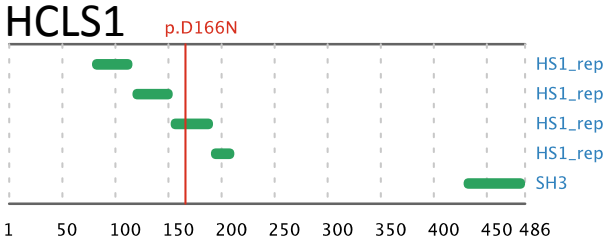
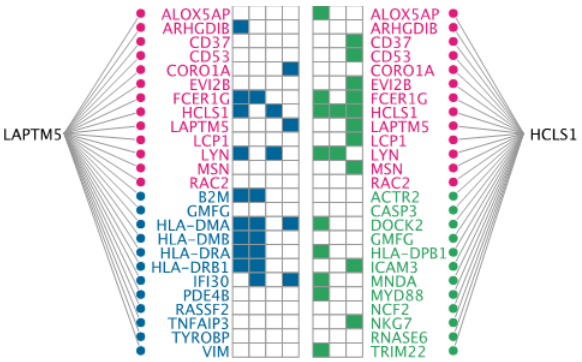
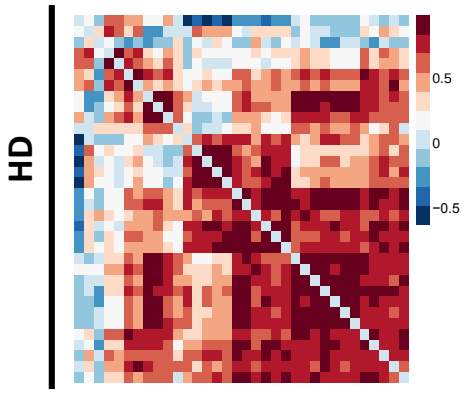

HCLS1
p.D166N
HS1_rep
HS1_rep
HS1_rep
HS1_rep
SH3
1
50
100
150
200
250
300
350
400
450
486

FIGURE 5

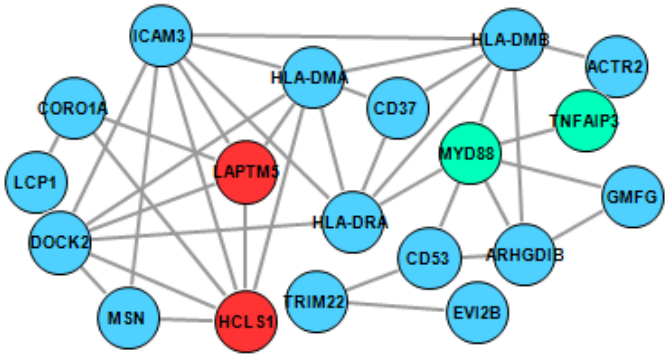
A



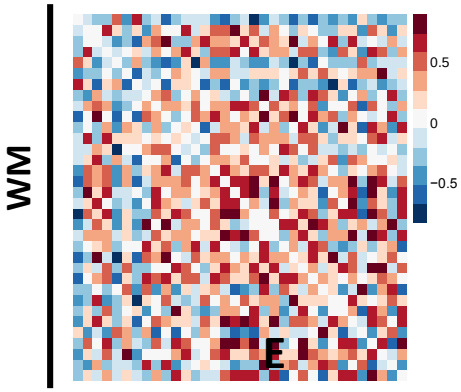
B



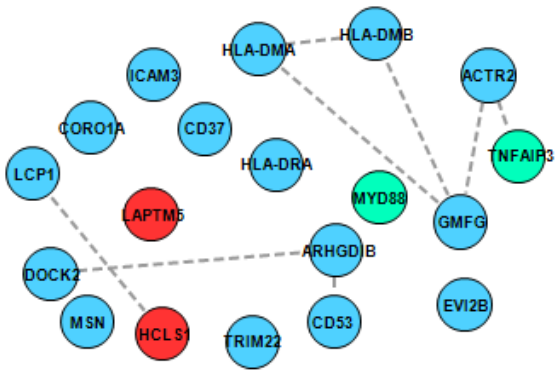
C



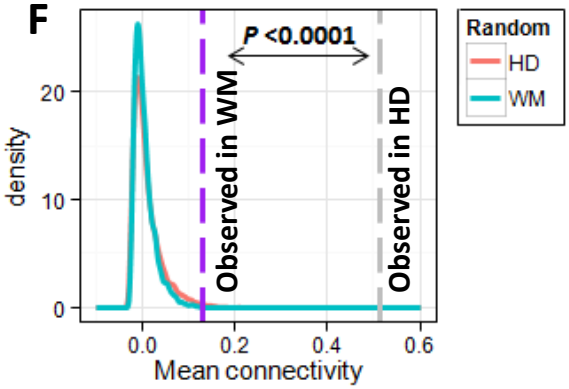
D



E



F



Prepublished online February 22, 2016;
doi:10.1182/blood-2015-11-680199

Exome sequencing reveals recurrent germline variants in patients with familial Waldenstrom's macroglobulinemia

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

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.