blood neoplasia

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
bloodneoplasia@hematology.org

The landscape of rare genetic variants in familial Waldenström macroglobulinemia

Tracking no: NEO-2023-000111R2

Alexander Pemov(DCEG,NCI,United States) Jung Kim(,NCI,United States) Wen Luo(,Leidos Biomedical, Inc.,United States) Jia Liu(,Leidos Biomedical Research Inc.,United States) Cole Graham(,NCI, DCEG,United States;,Tufts University Family Medicine Residency at Cambridge Health Alliance,United States) Kristine Jones(Cancer Genomics Research Laboratory,Leidos Biomedical, Inc.,United States) Delphine Demangel(,HCL,) Neal Freedman(,National Cancer Institute,United States) Charles Dumontet(Hematology,Hospices Civils de Lyon,France) bin zhu(,NIH,United States) Mary McMaster(Clinical Genetics Branch,National Cancer Institute,United States) Douglas Stewart(,NCI,United States)

Abstract:
Waldenström macroglobulinemia (WM) is a rare hematological malignancy. Risk for WM is elevated 20-fold among first-degree relatives of WM patients. However, the list of variants and genes that cause WM remains incomplete. In this study we analyzed exomes from 64 WM pedigrees for evi-dence of genetic susceptibility for this malignancy. We determined the frequency of pathogenic (P) or likely pathogenic (LP) variants among WM patients; performed variant- and gene-level association analyses with the set of 166 WM cases and 681 unaffected controls; and examined the segregation pattern of deleterious variants among affected members in each pedigree. We identified pathogenic variants in TREX1 and SAMHD1 (genes that function at the interface between innate immune re-sponse, genotoxic surveillance, and DNA repair) segregating in WM patients from two pedigrees. There were additional P/LP variants in cancer-predisposing genes, e.g., POT1, RECQL4, PTPN11, PMS2. In variant- and gene-level analyses, no associations were statistically significant after multiple testing correction. On a pathway level, we observed involvement of genes that play role in telomere maintenance (q-value=0.02), regulation of innate immune response (q-value=0.05) and DNA repair (q-value=0.08). Affected members of each pedigree shared multiple deleterious variants (median n=18) but the overlap between the families was modest. In summary, P/LP variants in highly pene-trant genes constitute a modest proportion of the deleterious variants, each pedigree is largely unique in its genetic architecture, and multiple genes are likely involved in the etiology of WM.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: A.P. designed the research, analyzed data, wrote and prepared manuscript for submission; J.K. designed the research and analyzed data; W.L designed the research and analyzed data; J.L. oversaw the bioinformatics analyses; C.G. designed the research and analyzed data; K.J. oversaw patients' samples processing and sequencing; D.D. oversaw sample collection and provided patients' samples; N.D.F. oversaw sample collection and provided control samples; C.D. oversaw sample collection and provided patients' samples; B.Z. designed the research, oversaw the data analyses and co-wrote the manuscript; M.L.M. designed the research, oversaw sample collection, provided patients' samples and co-wrote the manuscript; D.R.S. designed the research, oversaw the project and co-wrote the manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Genomic data are available through controlled access in dbGaP per the NIH genomic data sharing policy for the following studies: Waldenstroms macroglobulinemia genotyping (phs001284.v1.p1); CLL, Hodgkins, non-Hodgkins, Waldenstrom exome data (phs001219.v1.p1).

Clinical trial registration information (if any):

1 **The landscape of rare genetic variants in familial Waldenström macroglobulinemia**

2

3 **Running title:** Exome sequencing of 64 WM pedigrees

4

5 Alexander Pemov[1], Jung Kim[1], Wen Luo[2], Jia Liu[2], Cole Graham[1], Kristine Jones[2], Delphine DeMangel[3],

6 Neal D. Freedman[4], Charles Dumontet[3], Bin Zhu[5], Mary L. McMaster[1], and Douglas R. Stewart[1]

7

8 [1] Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute,

9 Bethesda, MD

10 [2] Frederick National Laboratory for Cancer Research, Division of Cancer Epidemiology and Genetics,

11 National Cancer Institute, Rockville, MD

12 [3] Hospices Civils de Lyon, University of Lyon, Lyon, France

13 [4] Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Insti-

14 tute, Bethesda, MD

15 [5] Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethes-

16 da, MD

17 **Corresponding author:** Douglas R. Stewart; 9609 Medical Center Drive, Rm. 6E450, Rockville, MD

18 20850; email: drstewart@mail.nih.gov; phone: 240-276-7238

19 **Main text word count**: 4,022

20 **Abstract word count**: 234

21 **Figure/table count:** 4

22 **Reference count:** 44

23

24

25

26

27 **KEY POINTS**

28

29 1) In familial WM, P/LP variants in highly penetrant genes constitute only a modest proportion of the dele-
30 terious variant load. 2) Each WM pedigree is largely unique in its genetic architecture; multiple genes and
31 pathways are likely involved in the etiology of WM.

32

33 **ABSTRACT**

34

35 Waldenström macroglobulinemia (WM) is a rare hematological malignancy. Risk for WM is elevated 20-
36 fold among first-degree relatives of WM patients. However, the list of variants and genes that cause WM
37 remains incomplete. In this study we analyzed exomes from 64 WM pedigrees for evidence of genetic sus-
38 ceptibility for this malignancy. We determined the frequency of pathogenic (P) or likely pathogenic (LP)
39 variants among WM patients; performed variant- and gene-level association analyses with the set of 166
40 WM cases and 681 unaffected controls; and examined the segregation pattern of deleterious variants
41 among affected members in each pedigree. We identified pathogenic variants in *TREX1* and *SAMHD1*
42 (genes that function at the interface between innate immune response, genotoxic surveillance, and DNA
43 repair) segregating in WM patients from two pedigrees. There were additional P/LP variants in cancer-
44 predisposing genes*, e.g., POT1, RECQL4, PTPN11, PMS2*. In variant- and gene-level analyses, no associa-
45 tions were statistically significant after multiple testing correction. On a pathway level, we observed in-
46 volvement of genes that play role in telomere maintenance (q-value=0.02), regulation of innate immune
47 response (q-value=0.05) and DNA repair (q-value=0.08). Affected members of each pedigree shared mul-
48 tiple deleterious variants (median n=18) but the overlap between the families was modest. In summary,
49 P/LP variants in highly penetrant genes constitute a modest proportion of the deleterious variants, each
50 pedigree is largely unique in its genetic architecture, and multiple genes are likely involved in the etiology
51 of WM.

52

53 **INTRODUCTION**

54

55 Waldenström macroglobulinemia (WM) is a rare hematological malignancy that belongs to the
56 spectrum of plasma cell disorders and is a subtype of lymphoplasmacytic lymphoma (LPL)[1-4]. Together,
57 WM and LPL account for approximately 2% of newly diagnosed non-Hodgkin lymphoma in the US[5]. WM
58 has an age-adjusted incidence of 0.36 per 100,000 in the US, and incidence increases markedly with age[6].
59 It is characterized by monoclonal immunoglobulin IgM in serum and abnormal cells that share characteris-
60 tics of lymphocytes and monoclonal plasma cells in the bone marrow. Lymph nodes, spleen and other or-
61 gans and tissues may be affected [7]. The disease is indolent in most patients, often remaining asymptomatic
62 for years[7,8]. However, despite recent advances in treatment with potential for long-term disease control,
63 WM remains incurable[9].

64 A genetic component for WM risk has long been suspected. Familial aggregation of WM has been
65 observed for more than 60 years[10-12]. Family history is the strongest risk factor in epidemiological stud-
66 ies[13], and a population-based registry study of first-degree relatives of WM/LPL patients documented sig-
67 nificantly elevated familial risk among for WM (20-fold) as well as for related B-cell malignancies and
68 monoclonal gammopathy of undetermined significance (MGUS)[14]. Early attempts to identify specific ge-
69 netic loci, including a linkage study and a candidate gene association approach conducted in high-risk fam-
70 ilies, were promising and notable in suggesting genetic heterogeneity. Linkage analysis of 11 high-risk
71 WM families found significant evidence of linkage on chromosomes 1q and 4q and suggestive evidence
72 for chromosomes 3 and 6 and provided the first conclusive evidence that IgM MGUS is part of this disease
73 spectrum[15]. Liang and co-authors' candidate gene association study in 165 unrelated familial cases with
74 WM or related B-cell tumor *vs.* 107 spouse controls reinforced the idea of WM genetic heterogeneity
75 based on identification of associations with multiple genes[16]. More recently, a genome-wide association
76 study (GWAS) performed in 530 unrelated WM/LPL cases and 4,362 controls of European ancestry[17]
77 identified two high-risk single nucleotide variants (SNV) at 6p25.3 and 14q32.13 that together explained
78 4% of the familial risk. Despite these early promising results, next-generation sequencing efforts to identi-

fy rare predisposing genetic variants have been limited. Roccaro et al. used exome sequencing to identify potential predisposition alleles in *LAPTM5* and *HCLS1* that segregated in three affected members of a multiplex WM family[18]. Follow-up screening of additional unrelated 246 WM cases identified significantly elevated frequency of these variants in familial cases compared to non-familial cases or unaffected controls. Another study identified a novel missense substitution in *FHL2* in identical twins, one of whom was affected with WM and the other with IgM MGUS, but not in their unaffected siblings[19]. The *FHL2* mRNA and protein expression levels were significantly lower in the peripheral blood cells of the WM patient compared to their healthy siblings, suggesting a role for the gene in the WM etiology. Somatic genome-sequencing studies revealed an *MYD88* c.794C>T (p.L265P) substitution affecting the NF-kappa-B pathway in 91% of WM patients[20], as well as inactivating variants in *ARID1A* (17% WM patients) and *CXCR4* (27% of WM patients)[21,22]. Multiple highly recurrent copy-number variants were also reported[21].

Etiologic heterogeneity for WM is also supported by epidemiological studies that have identified host and environmental WM risk factors, including personal history of autoimmune conditions (Sjögren's syndrome and systemic lupus erythematosus), adult weight, hay fever and infections, and occupational exposure and exposures to tobacco smoking, pesticides, wood dust and organic solvents, respectively[13, 23].

In this study, to better understand the genetic etiology of WM, we analyzed exomes of 64 WM pedigrees with 1-8 affected members in each for evidence of genetic susceptibility for this malignancy. Considering the strong familial clustering and the inability of the previous studies to identify highly penetrant variants associated with WM, we hypothesized that multiple low-penetrance variants specific to each family are likely to be involved in the predisposition to this malignancy. We determined the frequency of P/LP variants among affected members across all families. We performed variant- and gene-level association analyses in WM cases vs. unaffected controls. In addition, we examined the segregation pattern of deleterious variants among all affected members in each pedigree. To our knowledge, this is the largest set of WM families reported to date.

MATERIALS AND METHODS

(The full version of Materials and Methods can be found in Supplementary Methods)

Patients and sample collection. This study included patients diagnosed with WM, together with their consenting family members, enrolled in the US National Cancer Institute study of individuals and families at high risk of hematolymphoid cancers (NCT00039676), as previously described[15]. In addition, families containing at least two members with WM or MGUS were enrolled in a study of familial plasma cell dyscrasias (HCL/P 2007.460/3) conducted at the Hospices Civils de Lyon. Each protocol was approved by its respective Institutional Review Board (IRB), and all subjects provided written informed consent for sample collection and analysis. We included WM, MGUS and related B-cell tumors as cases for this analysis based on their known co-aggregation. For simplicity, in the text we refer to WM and related B-cell disorder patients as “WM cases”. Patients’ diagnoses and other relevant clinical information are shown in Supplementary Table 1. WM cases (n=173) included all known affected family members. Cancer-free controls (n=681) were selected from the Prostate, Lung, Colorectal and Ovarian (PLCO)[24] cancer screening trial samples. All pedigrees but one and all controls were of European descent.

Exome sequencing. Genomic DNA was extracted from blood using standard methods. DNA was captured with NimbleGen SeqCap EZ Exome Library and sequenced on the Illumina platform (HiSeq 2000/2500/4000 and NextSeq500 instruments).

Exome sequence data processing. The human reference genome and the “known gene” transcript annotation were downloaded from the UCSC database (http://genome.ucsc.edu/), version hg19 (corresponding to Genome Reference Consortium assembly GRCh37). Data processing was performed as described previously[25].

Data filtering and variant prioritization. For rare variant analyses, variants with population frequency >1% in gnomAD, v.2.1.1, and variants present in this study’s controls at a frequency above 10% were fil-

tered out. Variants with population frequency >1% and ≤5% in gnomAD, v.2.1.1, were analyzed in common variant analyses. To filter out putative somatic variants, only the variants with variant allele frequencies (VAF) ≥0.35 and ≤0.70 were considered in the analyses. Remaining variants were further prioritized: 1) variants classified in ClinVar as pathogenic (P) or likely pathogenic (LP); 2) loss-of-function (LOF) variants including frameshifting deletions and insertions, nonsense, start loss and canonical splice-sites; 3) missense variants with CADD_phred_score≥25, REVEL_score≥0.5 and MetaSVM_score=D (deleterious); 4) variants with deleterious splicing effects as determined by an *in silico* prediction tool spliceAI, delta_score≥0.5 for either acceptor or donor gain or loss; 5) non-frameshifting deletions/insertions and stop loss variants; and 6) all remaining variants of uncertain significance (VUS).

Statistical tests. Variant- and gene-level analyses were performed on rare coding variants. To address relatedness between the cases, SAIGE, and SAIGE-GENE+ methods (v1.0.4) were utilized for association analyses [26,27]. Age and sex were used as co-variates in the association analyses. Principal component analysis (PCA) was performed to cluster individuals with shared ancestry (Supplementary Figure 1). Multiple testing was adjusted by false discovery rate (FDR) computation in variant- and gene-based analyses (with the cutoff q-value≤0.05).

Frequency of P/LP variants among affected members across all families (variant frequency visualization with ONCOPRINT). Variants classified as P/LP in ClinVar were summarized in an oncoprint using associated clinical and genomic characteristics. The oncoprint was generated by using R library “Complex Heatmap.”

Variant segregation pattern analysis in pedigrees. Sixty-four pedigrees that included 1-8 affected individuals were analyzed. Prioritized variants present in all affected individuals within each pedigree were further considered.

156

157 **Candidate gene list compilation.** A list of genes that may be involved in the etiology of WM was com-

158 piled using results of previously published linkage and GWAS studies, somatic sequencing studies of WM,

159 and studies of germline variation in tumor predisposition syndromes[28,29]. Genes involved in lymphocyte

160 biology were included in the list as well (Supplementary Table 2).

161

162 **Pathway and ontological analysis of genes.** Pathway and ontological analyses of gene sets were per-

163 formed with Enrichr (Gene Ontology, Biological Process, 2021 database).

164

165 **RESULTS**

166

167 ***Overview.*** To identify variants and/or genes involved in the etiology of WM we examined the frequency of

168 pathogenic (P) or likely pathogenic (LP) variants in the WM cases across 64 pedigrees, determined the

169 segregation pattern of variants among WM affected individuals in each pedigree, and performed associa-

170 tion analyses with the exome sequencing data obtained from the germline DNA of 166 WM cases and 681

171 unrelated unaffected controls. The data sets and analyses performed in the study are schematically repre-

172 sented in Figure 1.

173 ***P/LP variants in WM cases.*** First, we examined the P/LP variant profile in WM cases (Figure 2, Table 1

174 and Supplementary Table 3A). We identified 46 P/LP variants in 44 genes in 55 WM cases from 35 pedi-

175 grees. Of these, 15 P/LP variants in 24 WM cases from 15 families resided in 14 genes with autosomal

176 dominant (AD), autosomal dominant/autosomal recessive (AD/AR) or X-linked (XL) mode of inheritance

177 (MOI) clinical traits (Tier 1 variants in Table 1 and Supplementary Table 3A). The remainder of the vari-

178 ants, all of which were heterozygous, were observed in AR genes (Tier 2). There were at most two P/LP

179 variants per affected individual. Pathway analysis of the 44 genes with P/LP variants using the Gene On-

180 tology (GO) database identified several statistically significant (q-value≤0.05) functional categories includ-

181 ing "Telomeric D-loop Disassembly" (*RECQL4, POT1*, q-value=2.0E-02), "Somatic Hypermutation Of

182 Immunoglobulin Genes" (*PMS2, SAMHD1*, q-value=2.0E-02), "Regulation Of Helicase Activity" (*POT1,*
183 *TP53*, q-value=2.0E-02), and suggestively significant (q-value≤0.10) categories, including "Regulation Of
184 Innate Immune Response" (*TREX1, PTPN11, SAMHD1*, q-value=5.1E-02) and "DNA Repair" (*RECQL4,*
185 *TREX1,PMS2, TP53,* q-value=8.0E-02; Supplementary Table 3B). Several genes (*PTPN11, TP53, POT1*
186 and *PMS2*) are associated with known autosomal dominant (AD) cancer-predisposition syndromes.
187 ***Segregation pattern of prioritized variants in 64 WM pedigrees.*** Next, we investigated whether any vari-
188 ants and genes were shared by all affected members in each of 64 WM families. The number of affected
189 members per family available for analysis ranged from one to eight. Thirty pedigrees included two affected
190 members, five pedigrees included a single affected individual, one pedigree had eight individuals and the
191 rest of the families included 3-5 individuals with WM or other B-cell neoplasms (Supplementary Figure 3).
192 In aggregation, there were 30,040 shared variants in 12,506 genes across all 64 families. To further priori-
193 tize segregating variants, we selected P/LP ClinVar variants, putative LOF, missense variants exceeding
194 threshold *in silico* scores, non-canonical splicing variants with spliceAI delta_score ≥0.5, non-
195 frameshifting deletions/insertions, stop-loss variants, and all remaining VUSs for subsequent analysis. This
196 prioritization resulted in 1,288 variants in 1,148 genes across 59 families (in five families there were no
197 shared variants after prioritization, Supplementary Table 4).
198 Next, we investigated whether there were genes harboring prioritized variants shared by multiple
199 pedigrees (Supplementary Table 5A). We found 130 such genes present in ≥2 families. Variants in
200 *ZC3H18* were shared by four families; variants in *GLMN, CP*, and *VWA2* were shared by three families
201 and the rest were shared by two families. Pathway analysis of these genes identified highly statistically
202 significant enrichment for Collagen Fibril Organization (q-value=1.1E-04). We also observed significant
203 association with Telomere Maintenance (q-value=2.9E-02) and DNA Repair genes (q-value=3.4E-02*)*
204 (Supplementary Table 5B).
205 To gain additional insight into biological relevance of the prioritized variants/genes segregating in
206 the families, we compiled a list of genes known for their involvement in WM, other LPDs, lymphocyte

207 biology and tumor predisposition disorders (Supplementary Table 2). We then performed a hypergeometric test for the significance of the overlap between the two gene-sets (Figure 3 and Supplementary Table 6). There were 90 genes in the overlap, which was statistically significant (p-value=0.006). Sixteen of these 90 genes, including *BLK, MLH1, HERC2* and *IKBKG*, were shared by multiple families.

211 In 59 out of 64 pedigrees that shared prioritized variants, we observed 21.8 variants per family on average (median=18, range=1-88). Most of the pedigrees (51/59, 86.4%) shared at least one variant or a variant-carrying gene with at least one other pedigree.

214 ***Association analysis in 166 WM cases vs. 681 controls, rare variants (population frequency ≤ 0.01).***

215 Next, we analyzed the association of single variants with the malignancy using the SAIGE statistical package. SAIGE methodology considers relatedness between samples, which facilitates inclusion of multiple related individuals from the same family, thus increasing the analyses' power. Prior to the analysis, seven samples that did not cluster closely with the rest of the samples on PCA were filtered out, thus leaving 166 WM cases, which were compared to 681 unrelated controls. There were no variants significant after multiple testing correction (Supplementary Table 7A). Among the variants with nominal p-values<0.001 we identified five that resided in biologically plausible genes including *PTPRK, ITGA1, PDGFB, FLT3LG* and *TOPBP1.* On the gene level SAIGE analysis (Supplementary Table 7B and Supplementary Figure 2), there were no significant associations after multiple testing correction. Thirty-nine genes were associated with WM at a nominal p-value<0.01. As in the variant-level analysis, *PDGFB* was also found among the genes that were nominally associated with WM (Table 2). Other biologically plausible genes included *EXO1* (DNA repair), *IGLL5* (B-cell receptor signaling) and *RBPJ* (ERBB signaling).

227 ***Association analyses in 166 WM cases vs. 681 controls, common variants (0.01 < population frequency ≤0.05).*** Association analysis of common variants with WM resulted in no significant associations after multiple testing correction (Supplementary Table 7C). A single SNV (rs9838238) in *DCBLD2* (discoidin, CUB and LCCL domain containing 2) was nearly significant after Bonferroni correction. The function of this gene is incompletely understood, and the consequences of this substitution appear to be modest as re-

232 flected by *in silico* scores, such as CADD (22.3), REVEL (0.177) and MetaSVM (tolerated) (Supplemen-
233 tary Table 7C).
234
235 **DISCUSSION**
236
237 In this study we exome-sequenced 64 WM pedigrees, most of which had at least three affected
238 members. To our knowledge, this is the largest set of WM families reported to date. We analyzed the se-
239 quenced data to identify genetic risk factors for WM. We examined the frequency of pathogenic or likely
240 pathogenic variants among affected members across all families, performed variant- and gene-level associ-
241 ation analyses in 166 WM cases vs. 681 unaffected controls, and investigated the segregation pattern of
242 deleterious variants in each pedigree.
243 The P/LP analysis in the families identified several biologically plausible genes. We observed a
244 pathogenic variant in *TREX1* in three WM patients from a single pedigree, and a pathogenic variant in
245 *SAMHD1* in two WM patients from a different family. These two genes are associated with Aicardi-
246 Goutières syndrome (AGS, OMIM##225750, 612952), a rare disorder affecting brain, immune system,
247 and skin[30]. TREX1 is a 3'-5' exonuclease that degrades both single- and double-stranded free cytoplasmic
248 DNA fragments, thus regulating the amount of interferon-stimulatory DNA present in the cell and sup-
249 pressing the senescence-associated secretory phenotype, a process known to turn a senescent cell into a
250 proinflammatory one with a potential for tumor initiation[31]. Another AGS-associated gene, *SAMHD1*, en-
251 codes a protein which functions at the interface between inflammation and DNA repair: it is involved in
252 innate immune response to viruses via regulation of dNTP pools, as well as in repairing DNA via stimula-
253 tion of the exonuclease activity of MRE11 at the sites of stalled replication forks[31]. Both proteins function
254 in cytosolic DNA-sensing pathway (https://www.kegg.jp/pathway/hsa04623), which intersects with the
255 larger NF-kappa-B signal transduction network including *MYD88*, the most frequently somatically mutated
256 gene in WM[20]. The pathogenic variants in *SAMHD1* (rs515726146) and *TREX1* (rs72556554) identified in
257 this study were heterozygous, while in AGS patients these SNVs were observed in a homozygous or com-

258 pound heterozygous states[32-34]. It is therefore unlikely that heterozygous rs515726146 or rs72556554
259 would cause symptoms of AGS, a predominantly AR disease[35]. However, the rs515726146 in *TREX1* was
260 reported in a heterozygous state as a pathogenic variant associated with systemic lupus erythematosus[36,37],
261 a clinically related disorder of an aberrant immune response and a well-known risk factor for WM. It is
262 noteworthy that two rare variants in the genes functioning in the same cellular process and associated with
263 the same rare genetic disease were identified in multiple patients from multiple WM pedigrees. These find-
264 ings warrant further investigation of possible roles of *TREX1* and *SAMHD1* in the etiology of WM.

265 In addition, we identified P/LP variants in the genes associated with telomere maintenance and
266 DNA repair, including *POT1* (protection of telomeres 1) and *RECQL4* (RecQ like helicase 4), in WM pa-
267 tients from two pedigrees. *POT1* tumor predisposition disorder (*POT1*-TPD) is autosomal dominant and is
268 associated with a lifetime increased risk of several solid malignancies as well as chronic lymphocytic leu-
269 kemia (CLL)[38]. Bi-allelic inactivating variants in *RECQL4* are associated with several AR disorders
270 (Rothmund-Thomson, Baller-Gerold, RAPADILINO) predisposing affected individuals to multiple malig-
271 nancies, including lymphomas (https://www.ncbi.nlm.nih.gov/books/NBK1237/).

272 *PTPN11* and *TP53* are well-known genetic disorder-causing genes involved in the etiology of AD
273 Noonan (OMIM#163950) and Li-Fraumeni (OMIM#151623) syndromes (LFS), respectively. Cancer risk
274 may be elevated in Noonan syndrome compared to the general population and is enriched for hematologic
275 malignancies[39]. Li-Fraumeni is a tumor-predisposing disorder and is associated with multiple solid tumors
276 as well as blood malignancies, including ALL[40]; however, *TP53* is one the most frequently mutated genes
277 that are associated with clonal hematopoiesis of indeterminate potential[41] and, therefore, the possibility of
278 the somatic origin of two *TP53* pathogenic variants in two WM patients cannot be excluded. Moreover, an
279 association of pathogenic *somatic* changes in *TP53* with progression and unfavorable prognosis in WM has
280 been reported[42]. Of note, prior to the analyses, all variants were filtered by their VAF (≥0.35 and ≤0.70)
281 and the variants in *PTPN11* (rs39751680, VAF=0.43) and *TP53* (rs730882005, VAF=0.62 and
282 rs28934576, VAF=0.69) were within this range.

283 Pathogenic variants in *PMS2* are associated with AD Lynch syndrome (OMIM#614337) and AR
284 constitutional mismatch repair deficiency syndrome (CMMRDS, OMIM#619101). Lynch syndrome is
285 typically associated with solid malignancies (colorectal, endometrium, ovary, stomach, small bowel, and
286 few others), whereas hematologic neoplasms have not been traditionally associated with this disorder[43]. On
287 the other hand, leukemia and lymphomas are frequent in CMMRDS patients[44]; however, a pathogenic
288 *PMS2* variant identified in one of the WM patients was heterozygous, and no other P/LP variants in this
289 gene were detected in that patient.

290 The P/LP variants in *TREX1, SAMHD1, RECQL4, PTPN11, TP53* (rs28934576) and *PMS2* are
291 listed in both ClinVar and The Human Gene Mutation Database, and the p.(Ser421*) nonsense substitution
292 in *POT1* and p.(Cys238Phe) missense variant in *TP53* (rs730882005) are listed in ClinVar only. In the lit-
293 erature, all of these variants were reported in patients with associated disorders. We also examined availa-
294 ble clinical records for this study's WM patients but did not identify specific symptoms suggestive of
295 AGS, Noonan, Li-Fraumeni, Lynch, CMMRDS, Rothmund-Thomson, Baller-Gerold or RAPADILINO
296 syndromes. However, it should be noted that in a substantial subset, the available clinical information was
297 limited to blood malignancies and clinical follow up for all patients was not feasible. In addition, a litera-
298 ture search did not demonstrate association of WM/LPL with these disorders.

299 In the segregation patterning analyses of the WM pedigrees, a comparison of the list of genes car-
300 rying prioritized segregating variants with the list of biologically plausible genes (Supplementary Table 2)
301 showed an overlap of 90 genes, 16 of which were found in multiple families. Most of these 16 genes' func-
302 tions are related to B-cell biology and immunoregulation. For instance, *BLK* (B-lymphoid tyrosine kinase
303 proto-oncogene) and *PRF1* (perforin 1) are highly specifically expressed in bone marrow, spleen, and
304 lymph nodes and play important roles in B-cell receptor signaling, B-cell development, and in lymphoid
305 malignancies. Finally, a pathway analysis of the set of 130 genes harboring prioritized variants that were
306 observed in ≥2 pedigrees demonstrated a statistically significant overlap (q-value≤0.05) with telomere
307 maintenance (*RTEL1, SLX4*) and DNA repair genes (*HERC2, MC1R, SLX4, IGHMBP2, ERCC5, ACTR8,*
308 *MLH1, FAN1).*

309 The etiology of WM initiation and progression is largely unknown. Most WM cases are associated
310 with a single somatic point mutation in *MYD88* and with substantial chromosomal instability[20-22]. In this
311 study, we did not observe the p.(Leu265Pro) MYD88 pathogenic variant in any of the WM cases, thus in-
312 directly confirming its somatic origin; however, in pathway analysis, we identified significant enrichment
313 in genes involved in telomere maintenance and DNA repair processes. A possible mechanism of WM de-
314 velopment may be dependent on an increased rate of accumulation of somatic hits or chromosomal rear-
315 rangements in hematopoietic cells due to the presence of constitutive defects in genes that control proper
316 DNA repair or chromosome maintenance. A modulating mutation or CNV in a driver gene (e.g., *MYD88,*
317 *ARID1A, CXCR4*) may confer a proliferative advantage to a single cell thus initiating a clonal tumor pro-
318 gression.

319 SAIGE association analysis identified a variant in *PDGFB* (rs143980537, p-value=1.44×10^{-5})
320 which was observed five times in 166 cases (five heterozygous WM patients in three pedigrees) but was
321 not found in any of 681 controls. This is a rare variant, it is classified as a variant of uncertain significance
322 (VUS) in ClinVar (VarID 2264365), has a highly conserved GERP++ score and resides in the
323 PDGF/VEGF domain; however, it has a mixture of high and low of *in silico* prediction scores (e.g.,
324 CADD=25.2, but REVEL=0.122). *PDGFB* encodes platelet derived growth factor B and is moderately
325 expressed in spleen and lymph nodes. It is a known oncogene associated with tumorigenesis in multiple
326 tissues and organs, but its rate of somatic mutations in hematopoietic and lymphoid tissues is relatively
327 low (0.22%) compared to other tissues
328 (https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=PDGFB#tissue). The gene-level SAIGE analysis also
329 identified *PDGFB* association with WM at the nominal p-value <0.01; however, after p-value correction
330 for multiple testing, neither the gene- nor the variant-level association was significant.

331 Our study has several limitations. Even though we exome-sequenced and analyzed one of the larg-
332 est sets of familial WM samples, including many multiplex families, the total number of cases (n=166) was
333 small for rare-variant association analyses. Even though the cases and controls were sequenced in the same
334 facility and on the same platform, the sequencing of the samples was done over a period of several years

335 and on different Illumina instruments, thus possibly introducing a batch effect into the data. In the associa-
336 tion analyses, we included samples from all affected family members and then took advantage of the SAI-
337 GE method which accounts for the relatedness of individuals in the sample set. To mitigate a possible se-
338 quencing batch effect, all samples were bioinformatically jointly processed from FASTQ files to a VCF
339 file as a single set. Another limitation of this study is that only the coding portion of the genome, which
340 constitutes less than 2% of the genetic material of the cell, was analyzed. Understanding the function of
341 regulatory regions of the genome and non-coding genes and their interaction with protein-coding loci will
342 be necessary to master the complete knowledge of how genomes operate; however, this task remains im-
343 mensely complex, labor intensive and costly and was outside the scope of this study.

344 In summary, we identified multiple deleterious rare variants and plausible candidate genes in WM
345 patients. In two pedigrees, we identified multiple WM patients with pathogenic variants in *TREX1* and
346 *SAMHD1*, the genes that function at the interface between innate immune response, genotoxic surveil-
347 lance, and DNA repair and are associated with the Aicardi-Goutières syndrome. There were additional
348 P/LP variants residing in genes associated with well-known cancer-predisposing disorders*, e.g., POT1,*
349 *RECQL4, PTPN11* and *PMS2*. On a pathway level, we observed statistically significant involvement of
350 genes that play role in telomere maintenance, DNA repair, and regulation of innate immune response. Af-
351 fected members of each pedigree shared multiple deleterious variants (median n=18) but the overlap be-
352 tween the pedigrees was modest. In association analyses, we observed several VUSs including
353 rs143980537 in *PDGFB* at the nominal p-value<0.001. This gene was also found to be associated with
354 WM in SAIGE_SKAT-O analysis at the nominal p-value<0.01. We conclude that multiple genes are likely
355 involved in the etiology of WM, each pedigree is largely unique in terms of its genetic risk architecture,
356 and highly penetrant P/LP variants account for only a small proportion of deleterious variant load in fami-
357 lies with WM. Larger studies are needed to identify a full catalogue of genes associated with elevated risk
358 for WM. Main challenges include the rarity of this disease and a likely oligo-/polygenic nature of the ge-
359 netic risk factors predisposing to WM. As a future research effort, methodological studies that examine the

incorporation of variants derived from WM GWAS (only one was published to date) into polygenic risk score models may become feasible.

ACKNOWLEDGEMENTS

This work used resources of the NIH High Performance Computing Biowulf cluster.

AUTHORSHIP CONTRIBUTIONS

A.P. designed the research, analyzed data, wrote and prepared manuscript for publication; J.K. designed the research and analyzed data; W.L designed the research and analyzed data; J.L. oversaw the bioinformatics analyses; C.G. designed the research and analyzed data; K.J. oversaw patients' samples processing and sequencing; D.D. oversaw sample collection and provided patients' samples; N.D.F. oversaw sample collection and provided control samples; C.D. oversaw sample collection and provided patients' samples; B.Z. designed the research, oversaw the data analyses and co-wrote the manuscript; M.L.M. designed the research, oversaw sample collection, provided patients' samples and co-wrote the manuscript; D.R.S. designed the research, oversaw the project and co-wrote the manuscript.

FUNDING

This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics of the National Cancer Institute, Bethesda, MD.

DISCLOSURE OF CONFLICTS OF INTEREST

All authors declare no conflicts of interest.

386

387 **REFERENCES**

388

389 1. Castillo JJ. Plasma Cell Disorders. *Prim Care*. 2016;43(4):677-691.

390 2. Castillo JJ, Olszewski AJ, Cronin AM, Hunter ZR, Treon SP. Survival trends in Waldenström
391 macroglobulinemia: an analysis of the Surveillance, Epidemiology and End Results database. *Blood*.
392 2014;123(25):3999-4000.

393 3. Castillo JJ, Olszewski AJ, Kanan S, Meid K, Hunter ZR, Treon SP. Overall survival and
394 competing risks of death in patients with Waldenström macroglobulinaemia: an analysis of the
395 Surveillance, Epidemiology and End Results database. *Br J Haematol*. 2015;169(1):81-89.

396 4. Sekhar J, Sanfilippo K, Zhang Q, Trinkaus K, Vij R, Morgensztern D. Waldenström
397 macroglobulinemia: a Surveillance, Epidemiology, and End Results database review from 1988 to 2005.
398 *Leuk Lymphoma*. 2012;53(8):1625-1626.

399 5. Swerdlow SH CE, Harris NL, et al., editors. . WHO Classification of Tumours of Haematopoietic
400 and Lymphoid Tissues. Revised 4th ed. Lyon: IARC; 2017.

401 6. McMaster ML. The epidemiology of Waldenström macroglobulinemia. *Semin Hematol*.
402 2023;60(2):65-72.

403 7. Gertz MA. Waldenstrom Macroglobulinemia: Tailoring Therapy for the Individual. *J Clin Oncol*.
404 2022;40(23):2600-2608.

405 8. Kyle RA, Treon SP, Alexanian R, et al. Prognostic markers and criteria to initiate therapy in
406 Waldenstrom's macroglobulinemia: consensus panel recommendations from the Second International
407 Workshop on Waldenstrom's Macroglobulinemia. *Semin Oncol*. 2003;30(2):116-120.

408 9. Dimopoulos MA, Kastritis E, Owen RG, et al. Treatment recommendations for patients with
409 Waldenström macroglobulinemia (WM) and related disorders: IWWM-7 consensus. *Blood*.
410 2014;124(9):1404-1411.

411 10. Treon SP, Hunter ZR, Aggarwal A, et al. Characterization of familial Waldenstrom's
412 macroglobulinemia. *Ann Oncol*. 2006;17(3):488-494.
413 11. Kristinsson SY, Goldin LR, Björkholm M, Koshiol J, Turesson I, Landgren O. Genetic and
414 immune-related factors in the pathogenesis of lymphoproliferative and plasma cell malignancies.
415 *Haematologica*. 2009;94(11):1581-1589.
416 12. McMaster ML. Familial Waldenström Macroglobulinemia: Families Informing Populations.
417 *Hematol Oncol Clin North Am*. 2018;32(5):787-809.
418 13. Vajdic CM, Landgren O, McMaster ML, et al. Medical history, lifestyle, family history, and
419 occupational risk factors for lymphoplasmacytic lymphoma/Waldenström's macroglobulinemia: the
420 InterLymph Non-Hodgkin Lymphoma Subtypes Project. *J Natl Cancer Inst Monogr*. 2014;2014(48):87-
421 97.
422 14. Kristinsson SY, Björkholm M, Goldin LR, McMaster ML, Turesson I, Landgren O. Risk of
423 lymphoproliferative disorders among first-degree relatives of lymphoplasmacytic
424 lymphoma/Waldenstrom macroglobulinemia patients: a population-based study in Sweden. *Blood*.
425 2008;112(8):3052-3056.
426 15. McMaster ML, Goldin LR, Bai Y, et al. Genomewide linkage screen for Waldenstrom
427 macroglobulinemia susceptibility loci in high-risk families. *Am J Hum Genet*. 2006;79(4):695-701.
428 16. Liang XS, Caporaso N, McMaster ML, et al. Common genetic variants in candidate genes and
429 risk of familial lymphoid malignancies. *Br J Haematol*. 2009;146(4):418-423.
430 17. McMaster ML, Berndt SI, Zhang J, et al. Two high-risk susceptibility loci at 6p25.3 and
431 14q32.13 for Waldenström macroglobulinemia. *Nat Commun*. 2018;9(1):4182.
432 18. Roccaro AM, Sacco A, Shi J, et al. Exome sequencing reveals recurrent germ line variants in
433 patients with familial Waldenström macroglobulinemia. *Blood*. 2016;127(21):2598-2606.
434 19. Wan Y, Cheng Y, Liu Y, Shen L, Hou J. Screening and identification of a novel FHL2 mutation
435 by whole exome sequencing in twins with familial Waldenström macroglobulinemia. *Cancer*.
436 2021;127(12):2039-2048.

437 20. Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenström's
438 macroglobulinemia. *N Engl J Med*. 2012;367(9):826-833.
439 21. Hunter ZR, Xu L, Yang G, et al. The genomic landscape of Waldenstrom macroglobulinemia is
440 characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic
441 deletions associated with B-cell lymphomagenesis. *Blood*. 2014;123(11):1637-1646.
442 22. Treon SP, Cao Y, Xu L, Yang G, Liu X, Hunter ZR. Somatic mutations in MYD88 and CXCR4
443 are determinants of clinical presentation and overall survival in Waldenstrom macroglobulinemia. *Blood*.
444 2014;123(18):2791-2796.
445 23. Royer RH, Koshiol J, Giambarresi TR, Vasquez LG, Pfeiffer RM, McMaster ML. Differential
446 characteristics of Waldenström macroglobulinemia according to patterns of familial aggregation. *Blood*.
447 2010;115(22):4464-4471.
448 24. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and
449 Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials*. 2000;21(6 Suppl):273s-309s.
450 25. Pemov A, Wegman-Ostrosky T, Kim J, et al. Identification of Genetic Risk Factors for Familial
451 Urinary Bladder Cancer: An Exome Sequencing Study. *JCO Precis Oncol*. 2021;5.
452 26. Zhou W, Zhao Z, Nielsen JB, et al. Scalable generalized linear mixed model for region-based
453 association tests in large biobanks and cohorts. *Nat Genet*. 2020;52(6):634-639.
454 27. Zhou W, Bi W, Zhao Z, et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based
455 rare variant association tests. *Nat Genet*. 2022;54(10):1466-1469.
456 28. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer
457 genome landscapes. *Science*. 2013;339(6127):1546-1558.
458 29. Rahman N. Realizing the promise of cancer predisposition genes. *Nature*. 2014;505(7483):302-
459 308.
460 30. Crow YJ, Chase DS, Lowenstein Schmidt J, et al. Characterization of human disease phenotypes
461 associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and
462 IFIH1. *Am J Med Genet A*. 2015;167a(2):296-312.

463 31. Coquel F, Neumayer C, Lin YL, Pasero P. SAMHD1 and the innate immune response to
464 cytosolic DNA during DNA replication. *Curr Opin Immunol*. 2019;56:24-30.
465 32. Crow YJ, Hayward BE, Parmar R, et al. Mutations in the gene encoding the 3'-5' DNA
466 exonuclease TREX1 cause Aicardi-Goutières syndrome at the AGS1 locus. *Nat Genet*. 2006;38(8):917-
467 920.
468 33. Ramesh V, Bernardi B, Stafa A, et al. Intracerebral large artery disease in Aicardi-Goutières
469 syndrome implicates SAMHD1 in vascular homeostasis. *Dev Med Child Neurol*. 2010;52(8):725-732.
470 34. Rice G, Patrick T, Parmar R, et al. Clinical and molecular phenotype of Aicardi-Goutieres
471 syndrome. *Am J Hum Genet*. 2007;81(4):713-725.
472 35. Crow YJ. Aicardi-Goutières Syndrome. In: Adam MP, Mirzaa GM, Pagon RA, Wallace SE, Bean
473 LJH, Gripp KW, Amemiya A, eds. GeneReviews(®). Seattle (WA): University of Washington, Seattle
474 Copyright © 1993-2023, University of Washington, Seattle. GeneReviews is a registered trademark of the
475 University of Washington, Seattle. All rights reserved.; 1993.
476 36. Lehtinen DA, Harvey S, Mulcahy MJ, Hollis T, Perrino FW. The TREX1 double-stranded DNA
477 degradation activity is defective in dominant mutations associated with autoimmune disease. *J Biol Chem*.
478 2008;283(46):31649-31656.
479 37. Orebaugh CD, Fye JM, Harvey S, Hollis T, Perrino FW. The TREX1 exonuclease R114H
480 mutation in Aicardi-Goutières syndrome and lupus reveals dimeric structure requirements for DNA
481 degradation activity. *J Biol Chem*. 2011;286(46):40246-40254.
482 38. Speedy HE, Kinnersley B, Chubb D, et al. Germ line mutations in shelterin complex genes are
483 associated with familial chronic lymphocytic leukemia. *Blood*. 2016;128(19):2319-2326.
484 39. Ney G, Gross A, Livinski A, Kratz CP, Stewart DR. Cancer incidence and surveillance strategies
485 in individuals with RASopathies. *Am J Med Genet C Semin Med Genet*. 2022;190(4):530-540.
486 40. Holmfeldt L, Wei L, Diaz-Flores E, et al. The genomic landscape of hypodiploid acute
487 lymphoblastic leukemia. *Nat Genet*. 2013;45(3):242-252.

488 41. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with
489 adverse outcomes. *N Engl J Med*. 2014;371(26):2488-2498.
490 42. Gustine JN, Tsakmaklis N, Demos MG, et al. TP53 mutations are associated with mutated
491 MYD88 and CXCR4, and confer an adverse outcome in Waldenström macroglobulinaemia. *Br J*
492 *Haematol*. 2019;184(2):242-245.
493 43. Bansidhar BJ. Extracolonic manifestations of lynch syndrome. *Clin Colon Rectal Surg*.
494 2012;25(2):103-110.
495 44. Wimmer K, Etzler J. Constitutional mismatch repair-deficiency syndrome: have we so far seen
496 only the tip of an iceberg? *Hum Genet*. 2008;124(2):105-122.
497
498 **FIGURE LEGENDS**
499
500 **Figure 1. Schematic representation of analyses and sample sets used in the study.** PCA – Principal
501 Component Analysis.
502
503 **Figure 2. P/LP variants identified in WM patients from 64 families.** Oncoprint representation of 46
504 P/LP variants residing in 44 genes identified in 55 WM patients. Variants are shown in rows and patients
505 are shown in columns. Clinical and demographic metadata are shown on the top of the table and the leg-
506 end is shown on the right. Individuals from the same family are denoted by the same symbols (*e.g.,* cir-
507 cles, squares, triangles, etc.) in the color-filled squares of the oncoprint.
508
509 **Figure 3. Overlap between a pre-compiled list of genes related to WM biology and the list of segre-**
510 **gating prioritized genes in WM families.**
511
512 **DATA AVAILABILITY**
513

514 Genomic data are available through controlled access in dbGaP per the NIH genomic data sharing policy

515 for the following studies: Waldenström macroglobulinemia genotyping (phs001284.v1.p1); CLL, Hodg-

516 kin's, non-Hodgkin's, Waldenström macroglobulinemia exome data (phs001219.v1.p1).

Table 1. Pathogenic and likely pathogenic variants in WM families. Variants are shown in GRCh37 genomic coordinates. AF_popmax - highest population allele frequency in gnomAD v.3. Tier 1 variants include variants in the genes associated with either AD or AD, AR or XL mode of inheritance (highlighted in bold face); tier 2 variants include variants found in AR genes.

Patient ID	Chr	Position	Variant ID	Ref	Var	Gene symbol	Mode of inheritance	Variant pathogenicity	ClinVar ID	AF_popmax	MetaSVM	REVEL	CADD phred	spliceAI	Tier
Fam20_Pt4	17	7577568	rs730882005	C	A	TP53	AD	Pathogenic	376574	.	damaging	0.958	31	tolerated	1
Fam26_Pt2	2	25469529	rs1234388246	G	GC	DNMT3A	AD	Pathogenic	985294	.	.	.	.	tolerated	1
Fam41_Pt1	9	139412682	rs1290954710	C	T	NOTCH1	AD	Likely pathogenic	445806	.	damaging	0.439	28.9	tolerated	1
Fam41_Pt2	9	139412682	rs1290954710	C	T	NOTCH1	AD	Likely pathogenic	445806	.	damaging	0.439	28.9	tolerated	1
Fam54_Pt2	17	7577120	rs28934576	C	T	TP53	AD	Pathogenic	12366	.	damaging	0.868	23.9	tolerated	1
Fam55_Pt1	17	48703927	rs767121010	G	A	CACNA1G	AD	Likely pathogenic	426883	.	tolerated	0.075	23.5	tolerated	1
Fam60_Pt1	12	112888193	rs397516801	A	G	PTPN11	AD	Pathogenic	44603	.	damaging	0.668	25.9	tolerated	1
Fam37_Pt1	20	35555631	rs515726146	A	AC	SAMHD1	AD, AR	Pathogenic	126413	.	.	.	.	tolerated	1
Fam37_Pt2	20	35555631	rs515726146	A	AC	SAMHD1	AD, AR	Pathogenic	126413	.	.	.	.	tolerated	1
Fam37_Pt3	9	111662096	rs111033171	A	G	ELP1	AD, AR	Pathogenic	6085	0.00006479	.	.	.	tolerated	1
Fam37_Pt4	9	111662096	rs111033171	A	G	ELP1	AD, AR	Pathogenic	6085	0.00006479	.	.	.	tolerated	1
Fam2_Pt1	9	111662096	rs111033171	A	G	ELP1	AD, AR	Pathogenic	6085	0.00006479	.	.	.	tolerated	1
Fam2_Pt2	9	111662096	rs111033171	A	G	ELP1	AD, AR	Pathogenic	6085	0.00006479	.	.	.	tolerated	1
Fam2_Pt4	9	111662096	rs111033171	A	G	ELP1	AD, AR	Pathogenic	6085	0.00006479	.	.	.	tolerated	1
Fam3_Pt1	3	48508395	rs72556554	G	A	TREX1	AD, AR	Pathogenic	.	0.0003	damaging	0.828	31	tolerated	1
Fam3_Pt2	3	48508395	rs72556554	G	A	TREX1	AD, AR	Pathogenic	4179	0.0003	damaging	0.828	31	tolerated	1
Fam3_Pt3	3	48508395	rs72556554	G	A	TREX1	AD, AR	Pathogenic	4179	0.0003	damaging	0.828	31	tolerated	1
Fam24_Pt1	17	40693092	rs104894592	C	T	NAGLU	AD, AR	Pathogenic	1562	0.0001	.	.	35	tolerated	1
Fam32_Pt3	7	6029587	rs587780064	C	A	PMS2	AD, AR	Pathogenic	127802	.	.	.	35	damaging	1
Fam44_Pt3	1	169500043	rs118203907	T	C	F5	AD, AR	Pathogenic	649	.	damaging	0.972	31	tolerated	1
Fam45_Pt1	8	55537568	rs760689800	C	T	RP1	AD, AR	Pathogenic	1065648	.	.	.	25.2	tolerated	1
Fam45_Pt2	8	55537568	rs760689800	C	T	RP1	AD, AR	Pathogenic	1065648	.	.	.	25.2	tolerated	1
Fam48_Pt1	7	124481134	.	G	T	POT1	AD, AR	Pathogenic	1042147	.	.	.	38	tolerated	1
Fam53_Pt1	X	153363118	rs179363901	G	A	MECP2	XL	Pathogenic	11845	.	tolerated	0.229	22.4	tolerated	1
Fam2_Pt1	12	49312533	rs1272967209	GTA	G	CCDC65	AR	Pathogenic	88685	.	.	.	.	tolerated	2
Fam7_Pt2	1	63868019	rs199682486	G	A	ALG6	AR	Pathogenic	95529	0.0008	.	.	.	damaging	2
Fam3_Pt3	2	31596738	rs148412639	C	G	XDH	AR	Likely pathogenic	505602	0.0008	.	.	34	damaging	2

Fam4_Pt2	1	17313566	rs765632065	AG	A	*ATP13A2*	AR	Pathogenic	465253	0.0012	.	.	.	tolerated	2
Fam6_Pt2	15	56736015	rs185005213	G	A	*MNS1*	AR	Pathogenic	973691	0.0001	.	.	37	tolerated	2
Fam6_Pt3	15	56736015	rs185005213	G	A	*MNS1*	AR	Pathogenic	973691	0.0001	.	.	37	tolerated	2
Fam10_Pt1	8	67986544	rs766020802	CAA	C	*CSPP1*	AR	Pathogenic	575640	0.0001	.	.	.	tolerated	2
Fam10_Pt3	8	67986544	rs766020802	CAA	C	*CSPP1*	AR	Pathogenic	575640	0.0001	.	.	.	tolerated	2
Fam11_Pt1	1	43221287	rs118203996	G	A	*P3H1*	AR	Pathogenic	1258	.	.	.	36	tolerated	2
Fam11_Pt2	1	43221287	rs118203996	G	A	*P3H1*	AR	Pathogenic	1258	.	.	.	36	tolerated	2
Fam14_Pt2	1	27121547	rs139073416	C	A	*PIGV*	AR	Pathogenic	1284	0.0001	damaging	0.83	25.1	tolerated	2
Fam14_Pt3	6	135754219	rs372659908	G	A	*AHI1*	AR	Pathogenic	217525	0.00006483	.	.	36	tolerated	2
Fam14_Pt6	11	88070745	rs587777533	A	C	*CTSC*	AR	Pathogenic	139655	0.0001	.	.	36	tolerated	2
Fam19_Pt1	2	26644264	rs142371860	C	T	*DRC1*	AR	Pathogenic	55840	0.0006	.	.	36	tolerated	2
Fam19_Pt2	2	26644264	rs142371860	C	T	*DRC1*	AR	Pathogenic	55840	0.0006	.	.	36	tolerated	2
Fam20_Pt1	11	68548130	rs80356779	G	A	*CPT1A*	AR	Pathogenic	65644	0.0012	damaging	0.785	24.8	tolerated	2
Fam22_Pt1	10	104590667	rs777638364	C	T	*CYP17A1*	AR	Pathogenic	1338524	.	damaging	0.908	31	tolerated	2
Fam24_Pt1	19	44012925	rs761661864	A	C	*ETHE1*	AR	Likely pathogenic	577939	.	.	.	35	damaging	2
Fam24_Pt2	19	44012925	rs761661864	A	C	*ETHE1*	AR	Likely pathogenic	577939	.	.	.	35	damaging	2
Fam27_Pt1	8	68030481	rs766633448	A	C	*CSPP1*	AR	Likely pathogenic	1464866	.	.	.	32	damaging	2
Fam27_Pt1	6	74348214	rs727504156	TG	T	*SLC17A5*	AR	Pathogenic	167693	0.0001	.	.	.	tolerated	2
Fam27_Pt2	17	18058028	rs184435771	G	A	*MYO15A*	AR	Pathogenic	228276	0.0002	damaging	0.89	32	tolerated	2
Fam27_Pt2	6	74348214	rs727504156	TG	T	*SLC17A5*	AR	Pathogenic	167693	0.0001	.	.	.	tolerated	2
Fam29_Pt1	16	88876475	rs745594160	T	TA	*APRT*	AR	Pathogenic	203396	0.0001	.	.	.	tolerated	2
Fam34_Pt2	17	8076835	rs201558321	T	C	*SNORD118*	AR	Pathogenic	929280	.	.	.	.	tolerated	2
Fam35_Pt1	3	52409343	rs762545991	T	TG	*DNAH1*	AR	Pathogenic	478376	.	.	.	.	tolerated	2
Fam35_Pt1	1	12026314	.	AAGG	A	*PLOD1*	AR	Likely pathogenic	14367	.	.	.	.	tolerated	2
Fam38_Pt1	8	145740366	rs386833845	CA	C	*RECQL4*	AR	Pathogenic	6066	0.0012	.	.	.	tolerated	2
Fam38_Pt2	8	145740366	rs386833845	CA	C	*RECQL4*	AR	Pathogenic	6066	0.0012	.	.	.	tolerated	2
Fam40_Pt2	11	66637890	rs200030109	C	A	*PC*	AR	Likely pathogenic	203916	0.0012	damaging	0.79	25.5	tolerated	2
Fam42_Pt1	2	69627594	rs374514431	C	A	*NFU1*	AR	Pathogenic	30700	0.0012	damaging	0.972	29.5	tolerated	2
Fam43_Pt1	9	133946976	.	AG	A	*LAMC3*	AR	Likely pathogenic	504359	.	.	.	.	tolerated	2
Fam43_Pt2	7	50544323	rs201951824	C	T	*DDC*	AR	Pathogenic	202181	0.0012	tolerated	0.381	31	tolerated	2
Fam43_Pt2	3	46747402	rs375470385	G	A	*TMIE*	AR	Likely pathogenic	504683	0.00006479	.	.	.	damaging	2
Fam43_Pt3	3	46747402	rs375470385	G	A	*TMIE*	AR	Likely pathogenic	504683	0.00006479	.	.	.	damaging	2
Fam43_Pt3	6	151751289	rs144972972	T	C	*RMND1*	AR	Pathogenic	225255	0.0002	tolerated	0.803	25.2	tolerated	2
Fam57_Pt2	2	211460259	.	G	C	*CPS1*	AR	Likely pathogenic	449391	.	damaging	0.959	27.4	tolerated	2
Fam60_Pt1	14	76201609	rs541400148	C	T	*TTLL5*	AR	Pathogenic	1072044	0.0012	.	.	33	tolerated	2
Fam60_Pt2	14	76201609	rs541400148	C	T	*TTLL5*	AR	Pathogenic	1072044	0.0012	.	.	33	tolerated	2
Fam61_Pt1	3	49137706	rs751537797; rs149346696	G	A	*QARS1*	AR	Pathogenic	941594	.	.	.	34	tolerated	2

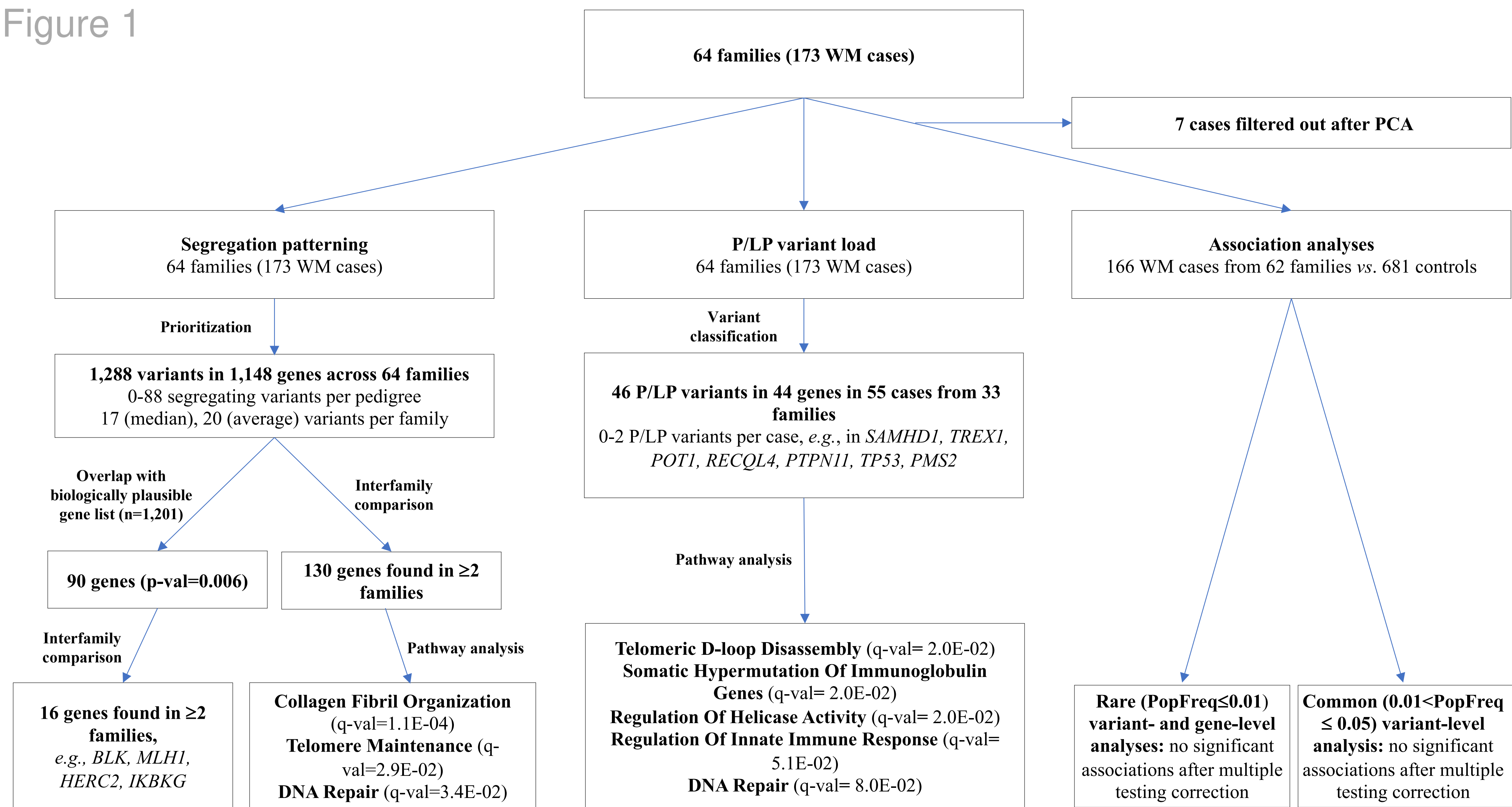

Figure1

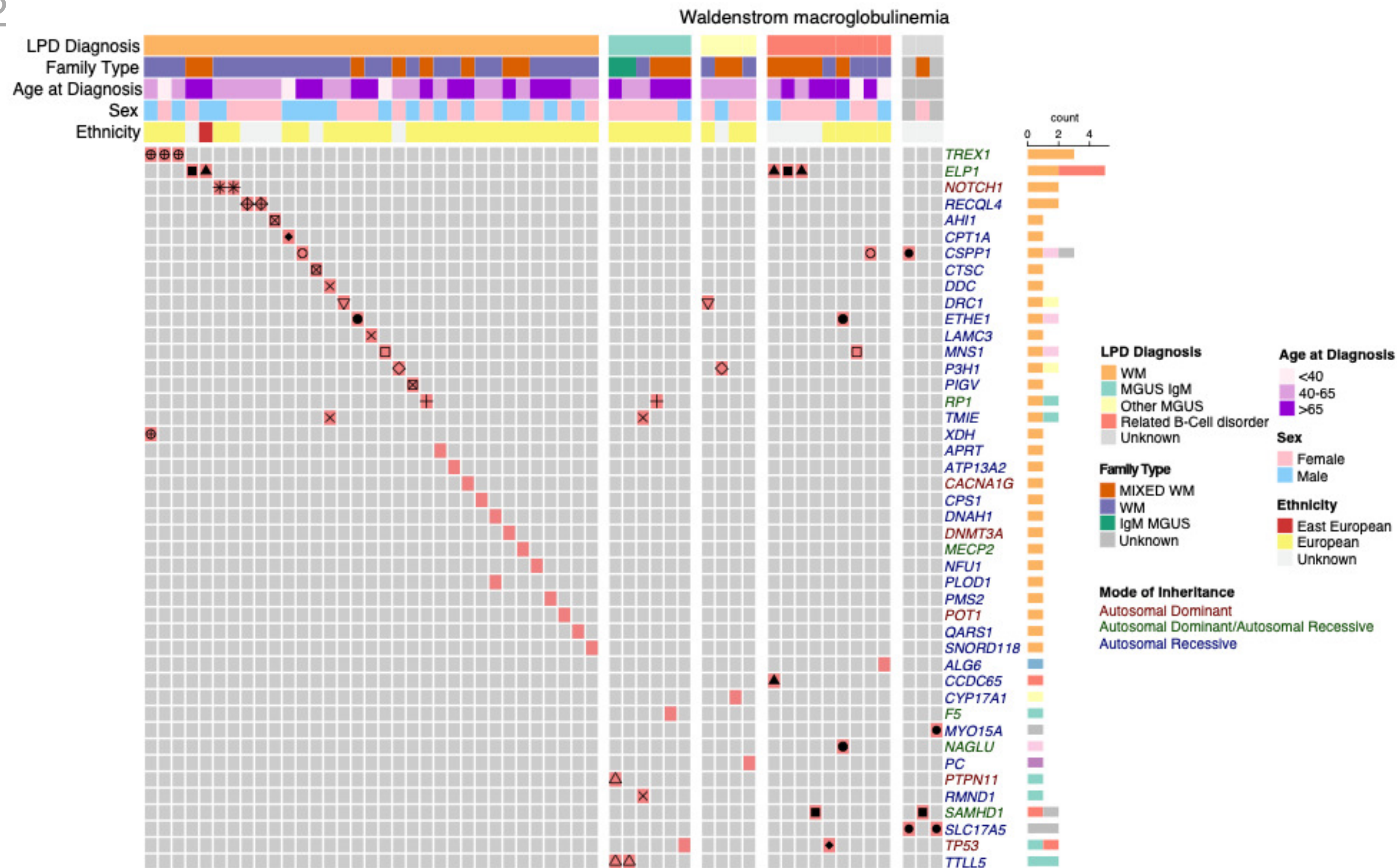

Figure 2

Figure 3

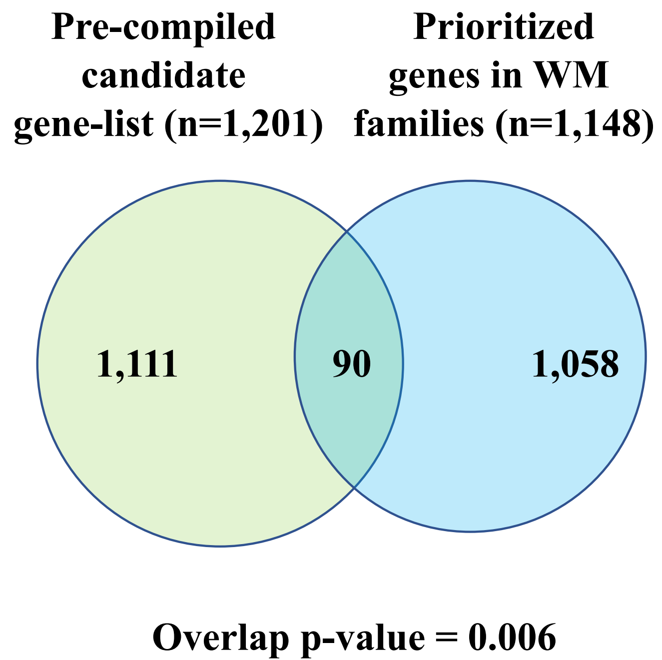

Figure 3