Table 2. Gene mutations identified by WES (PNH1–10) and targeted deep NGS (PNH15–60)

Patient	No. of mutations	*PIGA*	Other gene mutations	Ancestral mutated gene	Subclonal gene
PNH1	3	G68E, intron 5 splice	*NTNG1* (P24S)	*NTNG1* (P24S)	*PIGA*
PNH2	4	L83fs, S127X	*MAGEC1* (C747Y), *BRPF1*(N797S)	*MAGEC1* (C747Y)	*PIGA*
PNH3	7	L76fs	*TMC1* (E80G), *WDR96* (T1115I), *NRXN3* (Y9C), *CCR9* (M188T), *ALDH1B1* (K81X), *CPD*(P472S)	*TMC1* (E80G), *WDR96* (T1115I), *NRXN3* (Y9C)	*CCR9* (M188T), *ALDH1B1* (K81X), *CPD* (P472S)
PNH4	2	Microdeletion Xp22.2	*CELSR1* (E2404D)	–	–
PNH5	2	Intron 4 splice	*KDM3B* (L125I)	*PIGA*	*KDM3B*
PNH6	3	L243fs	*STAC3* (F97V), *TET2* (E1250X)	*STAC3* (F97V), *TET2* (E1250X)	*PIGA*
PNH7	3	T192fs, R233fs	*SLC20A1* (I671L)	*SLC20A1* (I671L)	*PIGA*
PNH8	4	Q18X	*MUC7* (R358X), *RBP3* (G560S), *C11orf34* (F15L)	*PIGA*	*MUC7* (R358X), *RBP3* (G560S), *C11orf34* (F15L)
PNH9	3	E395fs, R119fs	*MAN1A2* (splice)	*MAN1A2* (splice)	*PIGA*
PNH10	4	T192fs	*PEX14* (Y290C), *SYNE2* (K5198N), *FBN1* (T1340A)	*PEX14* (Y290C), *SYNE2* (K5198N)	*PIGA*, *FBN1* (T1340A)
PNH15	1	–	*SUZ12* (intron 2 splice)	–	–
PNH16	2	–	*TET2* (S1556fs, A671fs)	–	–
PNH17	1	Exon 3 splice	–	–	–
PNH19	1	S130fs	–	–	–
PNH22	2	V150fs	*BCOR* (Q1606X)	–	–
PNH24	1	I53fs	[B]	–	–
PNH25	1	V150fs	–	–	–
PNH26	1	Exon 3 splice	–	–	–
PNH27	2	I269fs	*ASXL1* (M1345L)	–	–
PNH31	1	H154R	–	–	–
PNH34	1	Q252X	–	–	–
PNH37	1	S368fs	[B]	–	–
PNH38	2	M326fs	*DHX29* (K498N)	–	–
PNH39	1	V193fs	[B]	–	–
PNH40	1	V67fs	–	–	–
PNH41	1	I417fs	–	–	–
PNH42	1	H128R	[B]	–	–
PNH43	1	T71fs	[B]	–	–
PNH44	1	–	*MECOM* (P18S)	–	–
PNH45	5	L171R[A], exon 3 splice[A], G146fs[A]	*BRCC3* (A53fs)[A], *RIT1* (Q212X)[A]	–	–
PNH47	2	G263R	*MECOM* (K613fs)	–	–
PNH49	2	Microdeletion Xp22.2	*JAK2V617F*	–	–
PNH50	2	Microdeletion Xp22.2	*JAK2V617F*	–	–
PNH51	2	Intron 4 splice	*U2AF1* (Q157P)	–	–
PNH52	2	Q100fs, intron 4 splice	–	–	–
PNH53	1	Intron 4 splice	–	–	–
PNH54	2	Intron 4 splice	*KDM6A* (K151R)	–	–
PNH57	2	Intron 4 splice	*ETV6* (A52V)	–	–
PNH59	2	Exon 3 splice[A]	*RIT1* (V58G)[A]	–	–
PNH60	1	–	*TET2* (L1514fs)	–	–

[A]Not confirmed by Sanger sequencing; [B]DNA quality insufficient for targeted deep sequencing; –, not detected or undetermined.

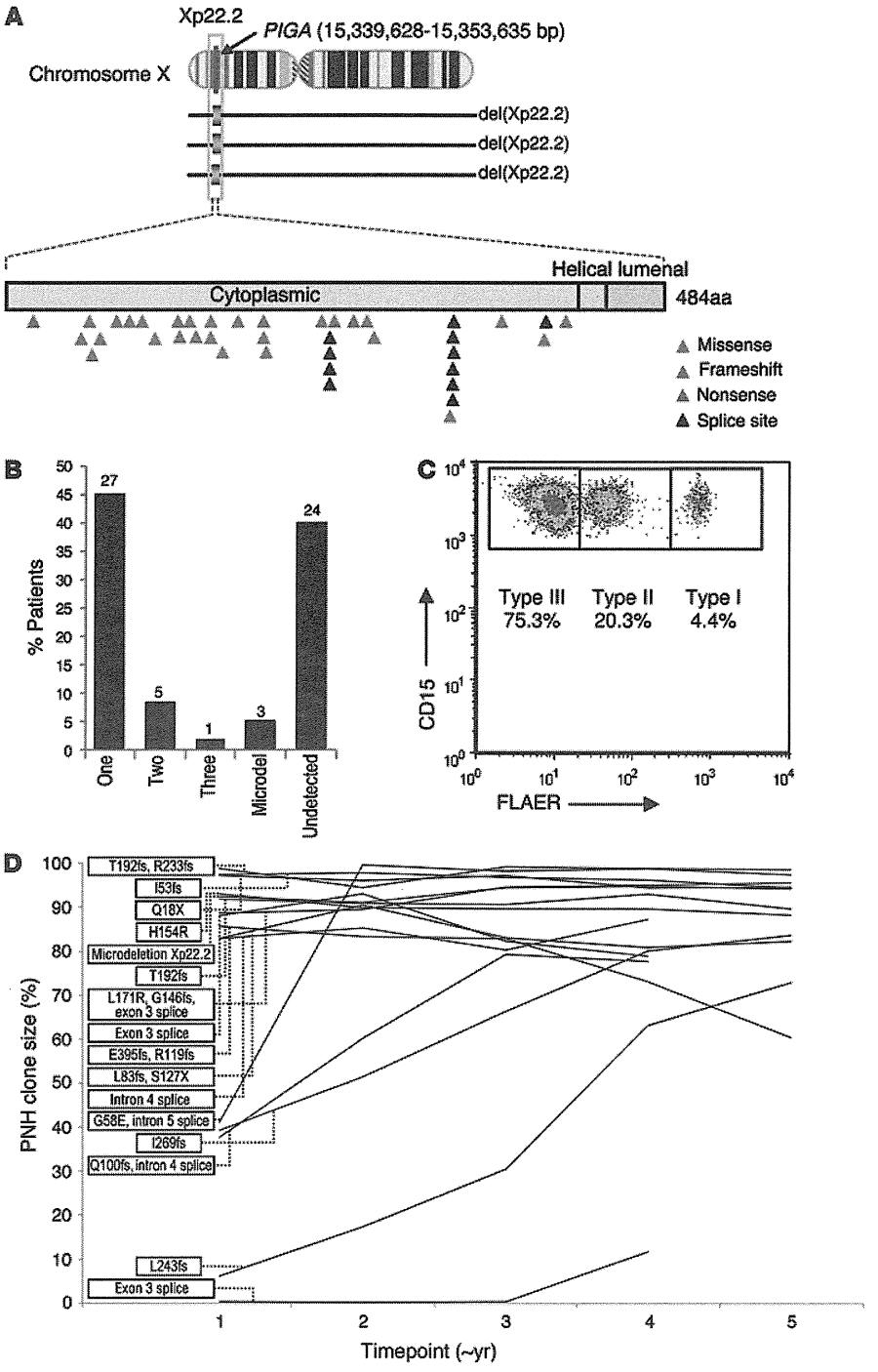

Figure 2. *PIGA* mutations and longitudinal analysis. (**A**) Distribution of *PIGA* missense, frameshift, nonsense, splice site, and microdeletions. (**B**) Proportion and absolute number of patients with 1, 2, or 3 mutations as well as microdeletions and those patients with the PNH phenotype in which a *PIGA* mutation was not detected. (**C**) Representative flow cytometric plot quantifying the number of type I (normal), type II (intermediate GPI anchor loss), and type III (complete GPI anchor loss) cells. (**D**) Longitudinal flow cytometry quantifying the wbc PNH clone size in patients with approximately 4 years or more of follow-up (n = 16). Each time point corresponds to approximately 1 year.

As we continued our analysis of the data, it became evident that additional mutations further increase the complexity of clonal architecture in PNH. In an illustrative case (PNH8), we identified 1 *PIGA* mutation and an additional 3 somatic mutations (*C11orf34*, *RBP3*, *MUC7*; Figure 3A). Again, based on targeted deep NGS and single-colony sequencing (Figure 3B), we deduced that the *PIGA* mutation constituted an initial ancestral event followed by subclonal defects in 3 other genes (Figure 3C). The diversity of clonal architecture in PNH became apparent when more cases were analyzed. In PNH2, another case with 2 *PIGA* mutations, additional somatic events including *MAGEC1* and *BRPF1* mutations, were found and independently confirmed. Comparison of VAFs obtained through targeted deep NGS revealed that the corresponding clonal size for the *MAGEC1* mutation was at least comparable to, if not larger than, that of the largest *PIGA* mutation, suggesting that the initial event may not have been a *PIGA* mutation, but indeed a *MAGEC1* mutation (Supplemental Figure 1). In PNH9, we found that the initial event that may be either a malignant or a passenger mutation, again was not a *PIGA* mutation, but a novel mutation in *MAN1A2*, which was followed by the appearance of 2 *PIGA* mutations, thus creating 2 independent clones, both of which carried the original *MAN1A2* mutation. In a somewhat similar case (PNH10, Supplemental Figure 2), somatic *SYNE2* and *PEX14* gene mutations were the initial events, followed by a *PIGA* frameshift mutation and an additional *FBN1* mutation. Perhaps the most complex case, PNH3, demonstrated 7 somatic gene mutations that were all confined to the PNH fraction; single-colony sequencing confirmed that the *PIGA* mutation occurred after 2 other clonal mutations had already been acquired (*TMC1* and *WDR96*) and, again, was not the initial event (Supplemental Figure 3).

In addition to the detection of concomitant mutations within the PNH clone, we found somatic events that predated the *PIGA* mutations and were present in both PNH and non-PNH myeloid fractions but not in germline DNA derived from T cells. These mutations included *TET2*, *SUZ12*, and *JAK2*. In PNH6, we observed concomitant *TET2* and *STAC3* mutations. However, mutant fractions were larger than those of the *PIGA* mutant, and the *TET2* mutation was also present in the non-PNH fraction, which, in turn, was negative for *STAC3* and *PIGA* mutations (Supplemental Figure 4). These results indicate that PNH, in this case,

Table 3. Clinical characteristics of PNH cases according to the presence of additional somatic mutations

Characteristic	PNH with additional mutations (n = 24)	PNH without additional mutations (n = 26)	P value
Age at diagnosis, yr Mean (range)	37 (16–71)	36 (5–77)	0.77
Age at sequencing, yr Mean (range)	45 (20–75)	40 (10–77)	0.48
wbc PNH clone size, % Mean (range)	68.9 (0.8–98.9)	49.8 (1.3–95)	0.04[A]
Absolute neutrophil count, /mm^3 Mean (range)	2,845 (850–8,130)	3,259 (1,050–9,530)	0.56
Hemoglobin, g/dl Mean (range)	10.8 (8.6–16)	11 (7–17)	0.97
Platelets, 109/l Median(range)	95.3 (14–255)	117 (6–467)	0.37
Hemolysis, n (%)	12 (52)	14 (54)	1
Thrombosis, n (%)	5 (25)	6 (16)	0.4866
Cytogenetics			
Normal, n (%)	14 (58)	7 (26)	1
Abnormal, n (%)	1 (4)	1 (4)	1
NA	9 (38)	18 (69)	

[A]Statistical significance at $\alpha = 0.05$.

evolved as a subclone after a clonal *TET2* mutation was acquired. However, in the bone marrow of another case (PNH3), we identified dysplastic changes along with trisomy 8 in 20 of 20 metaphases. FISH analysis resolved the origin of the trisomy 8, which was present only in the non-PNH fraction (data not shown). These results suggest that PNH in patient 3 evolved independently of the acquisition of trisomy 8.

Overall, our clonal analysis of 9 WES cases suggested that *PIGA* mutations were often acquired at a later stage (6 of 9 cases, Table 2), and mutations in other genes were the initial clonal events. Of note is that in a number of cases, the clonal composition showed significant overlap, preventing precise recapitulation of clonal hierarchy. In summary, our experiments indicate that the cohort can be stratified into 4 different scenarios (Figure 4): *PIGA* as the initial ancestral event accompanied by secondary mutations (Figure 4B, patients 5 and 8); *PIGA* as an event secondary to other mutations (Figure 4C, patients 1, 2, 6, 7, 9, and 10); *PIGA* as the lone mutation (Figure 4D, patients 17, 19, 25, 26, 31, 34, 40, 41, 52, and 53); and a *PIGA* mutation coexisting with other mutations responsible for the development of an MDS clone (Figure 4E, patient 3).

Clinical correlations. The discovery of additional somatic mutations in PNH prompted further investigation into the potential clinical impact these mutations may have. Analysis of various relevant clinical parameters with regard to the presence (n = 24) or absence (n = 26) of an additional mutation largely failed to find significant differences (Table 3). However, the presence of an additional mutation beyond *PIGA* was associated with a larger wbc PNH clone size as assessed by flow cytometry (P = 0.04; Table 3). Furthermore, based on the instructive discovery that PNH45 harbored 3 *PIGA* mutations (triclonal PNH) and a relatively large PNH type II wbc population (Figure 2C), we hypothesized that the presence of more than 1 *PIGA* mutation leads to the presence of both type II and type III PNH cells. Indeed, patients with more than 1 *PIGA* mutation had a higher incidence of a type II wbc population of greater than 3% than those with only 1 mutation (4 of 5 vs. 4 of 12, respectively; P = 0.048, data not shown).

Discussion

One of the discoveries of cancer NGS projects was the realization of genetic heterogeneity due to a combination of somatic mutations and the complexity of clonal architecture reflective of the sequential acquisition of genetic defects (11, 17–21). While general principles of clonal dynamics have been derived from the study of leukemia, they have not been explored in nonmalignant diseases such as PNH. The results of our study lead to a conceptual understanding of this disease: PNH is subject to clonal dynamics and selection forces similar to those observed in hematopoietic neoplasms. Consequently, in PNH, additional clonal and subclonal mutations corroborate with the pathognomonic *PIGA* mutation responsible for the prevailing cell phenotype. An important conclusion of this new concept is that the nature and composition of additional clonal somatic mutations may modify the behavior of the *PIGA* clone and have the potential to explain various distinct clinical courses seen in individual PNH patients. Thus, the analogy to MDS is apparent with multiple mutations that arise within the clonal population and undergo clonal selection along with *PIGA* mutations. Furthermore, the mutational events discovered in our research are not unique to PNH, but overlap significantly with the spectrum of mutations seen in typical MDS. Many of these mutations, or other mutations in the same genes, have been identified as key drivers of clonal evolution in MDS or cancer, including *U2AF1* (22, 23), *TET2* (2, 24), *MAGEC1* (25), *BRPF1* (26), *NRXN3* (27), *KDM3B* (28), *SLC20A1* (29), *MUC7* (30), *PEX14* (31), *FBN1* (32), *SUZ12* (33), *ASXL1* (34), *BCOR* (35), *DHX29* (36), *MECOM* (37), *RIT1* (38), and *JAK2* V617F (39). In some ways, PNH appears genetically similar to, but pathogenetically distinct from, MDS. While additional mutations have been described in PNH (6), our results illustrate for the first time in a large PNH cohort that genetic clonal selection and evolution may be operative not only in malignant conditions, but also in otherwise benign hematologic diseases.

Our stringent bioanalytic approach has led to effective exclusion of germline variants based on comparatively similar high VAFs in both experimental and control fractions. If variant reads that derived from clonal mutations were present in the CD59$^+$ cell population due to contamination, their VAF was low and was correspondingly high in the PNH$^+$ population. As such, this would be incompatible with germline alterations, which would be expected to oscillate around 50% of reads in both fractions. With this in mind, all of the mutational events we report in this article are exclusively somatic. While we found that some mutations had similar VAFs in both fractions, they were clearly somatic, since the VAF was much lower than would be expected in heterozygous germline variants. In such situations, we verified the absence of the mutation in the CD3$^+$ fraction and concluded that the *PIGA* mutation arose in the context of a preexisting clone. Future prospective studies would benefit from the use of nonhematopoietic

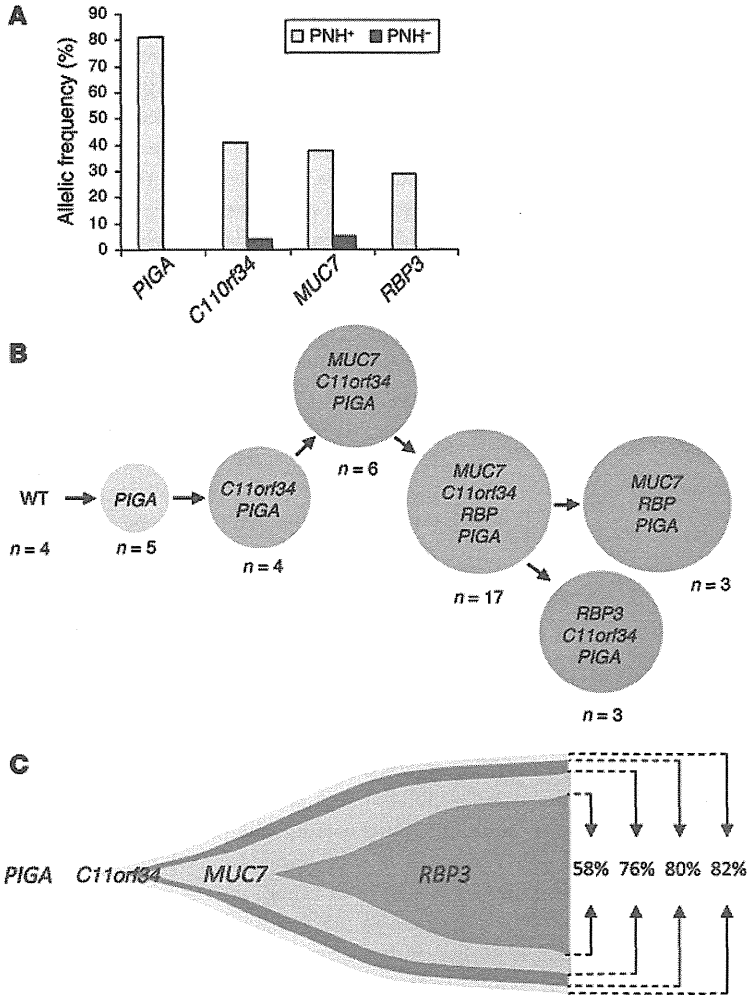

Figure 3. *PIGA* mutation as the initial ancestral event in PNH8. (**A**) A *PIGA* mutation together with 3 other somatic mutations (*C11orf34*, *RBP3*, and *MUC7*) were detected via WES at various allelic frequencies. All mutations were confined to the PNH fraction. Of note is that, in this case, the *PIGA* mutation was hemizygous, rendering the clone size equivalent to the allelic frequency. (**B**) Single-colony Sanger sequencing further confirmed the results obtained from deep sequencing and suggested that a *PIGA* mutation was the initial event. *n* = the number of colonies with the indicated mutations observed. Colonies that were not reproducible in independent experiments are not shown. (**C**) Proposed model of clonal architecture in this case based on VAFs and colony sequencing results.

tissue to clarify the mutational spectrum in PNH, as it is possible that some pre-*PIGA* mutations were excluded due to their presence in the entire hematopoietic compartment.

Our study, while showing that intrinsic somatic factors may contribute to clonal expansion, is also consistent with the immune selection theory of the evolution of PNH. Multiple independent clones characterized by *PIGA* mutations illustrate that a growth advantage may promote selection of several privileged clones, which in the process of disease may be further enabled by subsequent somatic events or by primordial passenger events in the ancestral stem cell affected by *PIGA* mutations. Phenotypically, this intraclonal diversity may not be easily distinguishable, although we have identified a significant relationship between the presence of both type II and type III PNH cells in patients with more than 1 *PIGA* mutation, supporting previous research (40). Nevertheless, our results suggest that deep targeted NGS of *PIGA* may have ancillary diagnostic potential in PNH, including quantitating the clonal size and composition of aberrant cells, yet flow cytometry remains the most effective diagnostic method, as we failed to detect *PIGA* mutations in 3 of 12 cases by WES, 1 of 10 by targeted deep sequencing, and 21 of 36 by Sanger sequencing. Other genes in the biosynthetic pathway, including all other PIG genes such as *PIGT* (41), were examined in WES cases, but no lesions were found, raising the possibility that intronic mutations may lead to the loss of GPI anchors on the cell surface. Furthermore, the high failure rate of Sanger sequencing could be due to a number of reasons including small clone size, intronic mutations, or micro/large deletions only detectable by SNP array.

At the onset of our study, we postulated several theoretical possibilities as to the clonal architecture of PNH: (a) *PIGA* gene defects are the initial ancestral events, and other mutations are acquired subsequently, similar to the situation in MDS; (b) other initial somatic events are followed by *PIGA* mutations constrained within and completely overlapping the PNH clone, suggesting that the initial non-*PIGA* events are passenger mutations; (c) secondary *PIGA* mutations arise in the context of other clonogenic mutations that are present in both *PIGA* mutant and wild-type cells; and (d) both myelodysplastic and PNH clones independently coexist. While *PIGA* mutations appear to be the initial event in many cases, we identified concomitant somatic mutations in a large proportion of PNH patients. In such cases, subsequent subclonal events can occur and are similar to those present in more aggressive hematologic malignancies. In cases in which the preceding clonal mutation was found but was limited to the PNH fraction, one could suggest that a passenger mutation was clonally "fixed" by the subsequent *PIGA* event. However, in some cases, permissive leukemogenic effects appeared to be instigated by mutations in *JAK2*, *TET2*, or *STAC3*, suggesting that these primary events arose, conveying an initial growth advantage, with a *PIGA* mutation as a subclone conveying an additional growth advantage. In particular, the presence of 2 cases with "myeloproliferative" *JAK2* mutations suggests that the propensity toward clonal proliferation may be further modified by the presence of *PIGA* defects and that *JAK2* mutations in concert with *PIGA* mutations lead to a markedly different phenotype than that of a myeloproliferative neoplasm (MPN). In addition, there were a significant number of cases in which *PIGA* was the lone mutation, although WES data suggest that this may be less frequent when entire exomes are analyzed.

Regardless, in many cases, the subclonal or clonal occurrence of associated mutations resembles the typical architecture of MDS (10). We and others have previously described additional somatic events in PNH patients demonstrating that even *PIGA*

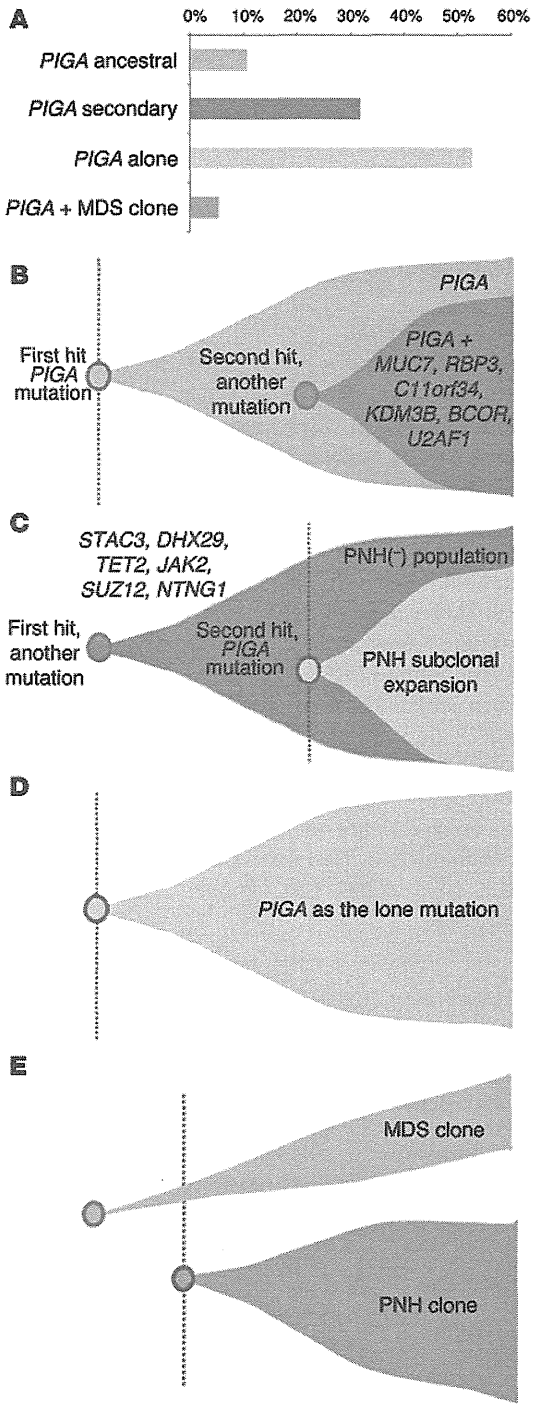

Figure 4. Mutational summary and clonal architecture scenarios. (A) Cohort frequency of 4 different scenarios for the clonal architecture encountered in PNH: ancestral *PIGA* mutation (**B**), secondary *PIGA* mutation (**C**), *PIGA* mutation as an isolated genetic event (**D**), and *PIGA* mutation leading to a PNH clone that independently coexists with an MDS clone (**E**).

mutation–negative cases were clonal (6–8). The inability to detect significant recurrent additional mutations may be due to the relatively small cohort and surprising diversity discovered in this study. Nevertheless, our results suggest that an obligatory secondary event does not occur in PNH, yet PNH clonal expansions appear to be aided by the presence of additional mutations. Depending on the permissivity of additional genetic events, the *PIGA* mutant clone may expand quickly, smolder, or disappear. This heterogeneous scenario would also be consistent with a relative lack of correlation between immunosuppression and the size of the clone, which may be driven under some circumstances by additional subclonal events. Our results suggest that future investigations into PNH will offer the opportunity to examine the implications of hematopoiesis sustained primarily by 1 or 2 mutated stem cells, with many cases demonstrating an accumulation of additional mutations that may be due to an increase in the self-renewal burden placed on stem cells as a consequence of the disease.

Methods

Patients. Bone marrow aspirates and/or blood samples were collected from 60 patients with PNH at the Cleveland Clinic. Diagnosis was confirmed and assigned according to the guidelines for the diagnosis and management of PNH (42).

Cytogenetics and SNP arrays. Cytogenetic analysis was performed according to standard banding techniques based on 20 metaphases, if available. Technical details regarding sample processing for SNP array (SNP-A) assays were previously described (43, 44). Affymetrix 250K and 6.0 kits were used. A stringent algorithm was applied for the identification of SNP-A lesions. Patients with SNP-A lesions concordant with metaphase cytogenetics or typical lesions known to be recurrent required no further analysis. Changes reported in our internal or publicly available (Database of Genomic Variants; http://projects.tcag.ca/variation) copy number variation (CNV) databases were considered nonsomatic and excluded. Results were analyzed using CNAG version 3.0 (20) or Genotyping Console (Affymetrix). All other lesions were confirmed as somatic or germline by analysis of CD3-sorted cells (45).

WES. WES was performed as previously reported (46). Briefly, tumor DNA was extracted from patients' bone marrow or peripheral blood mononuclear cells. DNA was obtained from paired CD59-positive and -negative cells. Whole-exome capture was accomplished based on liquid-phase hybridization (SureSelect; Agilent Technologies) according to the manufacturer's protocol. The SureSelect Human All Exon 50 Mb kit was used for 12 cases. The captured targets were subjected to massive sequencing using Illumina HiSeq 2000 according to the manufacturer's instructions. The raw sequence data were processed through the in-house pipeline constructed for whole-exome analysis of paired cancer genomes at the Human Genome Center, Institute of Medical Science, The University of Tokyo, and are summarized in a previous report (46). Data processing was divided into 2 steps: (a) generation of a BAM file (http://samtools.sourceforge.net/) for paired normal and tumor samples for each case, and (b) detection of somatic SNVs and indels by comparing normal and tumor BAM files. Alignment of sequencing reads on hg19 was visualized using Integrative Genomics Viewer (IGV) software (http://www.broadinstitute.org/igv/) (47).

Targeted deep sequencing and Sanger sequencing. We applied targeted exon sequencing to the validation of WES results as previously described (46, 48). Briefly, the number of reads containing SNVs and indels in both tumor and reference samples was determined using SAM tools, and the null hypothesis of equal allele frequencies in PNH and non-PNH samples was tested using the 2-tailed Fisher's exact test. A variant was adopted as a candidate somatic mutation if it had a *P* value

of less than 0.01. First, amplicons of each observation were subjected to Sanger sequencing by standard techniques on an ABI 3730xl DNA analyzer (Applied Biosystems). All mutations were confirmed by bidirectional sequencing and scored as pathogenic if not present in nonclonal, paired CD59-derived DNA. Then, deep NGS was applied to additional validation for variant allelic frequency and small clone detections. The amplicon libraries were generated according to an Illumina pair-end library protocol and subjected to deep sequencing on an Illumina MiSeq sequencer according to the standard protocol. For *PIGA* (exons 2–6) and *KDM3B* (exons 1–24), direct genomic sequencing was performed as previously described, together with deep sequencing, so that even very small mutated events would not be missed.

Single-cell colony culture. Colony formation assays in semi-solid cultures were used to obtain colonies for single-colony sequencing. A total of 10^5 bone marrow cells from PNH patients were plated in 1 ml of methylcellulose supplemented with G-SCF, GM-SCF, and EPO cytokines (STEMCELL Technologies) as well as FBS in a 35-mm culture plate at 37°C with 5% CO_2. After 7 to 10 days of culturing, individual colonies were removed, and DNA was extracted for sequencing.

Cell sorting. Following lysis of rbc, immunomagnetic selection of cells was performed using anti-CD59 PE (Life Technologies) followed by anti-PE microbeads (Miltenyi Biotec). Samples were separated using LS Columns (Miltenyi Biotec), and purity was verified in each fraction by flow cytometry on a Beckman Coulter FC500.

Flow cytometry. Whole blood was stained with antibodies against GlyA, CD15, CD24, CD55, CD59 (Beckman Coulter), CD59 (Life Technologies), and FLAER (Alexa 488 Proaerolysin Variant; Cedarlane Scientific) in various combinations to quantitate the number of GPI-deficient rbc and wbc in each patient. Samples were run on a Beckman Coulter FC500 or XL-MCL.

Accession codes. WES results have been deposited in the Sequence Read Archive (SRA) public database (PRJNA254174).

Statistics. JMP Pro 10 (SAS Institute Inc.) was used for statistical analysis. Either a Wilcoxon or Fisher's exact test was performed to evaluate statistically significant differences between groups, with an α of 0.05.

Study approval. Research was conducted according to protocols approved by the IRB of the Cleveland Clinic and in accordance with Declaration of Helsinki principles. Written informed consent was received from all patients prior to enrollment in the study.

Acknowledgments

This work was supported by grants from the NIH (RO1HL-082983, U54 RR019391, and K24 HL-077522, to J.P. Maciejewski); the Aplastic Anemia & MDS International Foundation (to J.P. Maciejewski and H. Makishima); the Robert Duggan Charitable Fund (to J.P. Maciejewski); Scott Hamilton CARES (to H. Makishima); the Ministry of Health, Labor and Welfare of Japan and KAKENHI (grants-in-aid 23249052, 22134006, and 21790907, to S. Ogawa); the Project for Development of Innovative Research on Cancer Therapeutics (P-DIRECT) (to S. Ogawa); and the Japan Society for the Promotion of Science (JSPS) through the Funding Program for World-Leading Innovative R&D on Science and Technology, initiated by the Council for Science and Technology Policy (CSTP) (to S. Ogawa). W. Shen was partly supported by the Jiangsu Health International Exchange Program (China).

Address correspondence to: Jaroslaw P. Maciejewski, Taussig Cancer Institute/R40, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. Phone: 216.445.5962; E-mail: maciejj@ccf.org.

1. Brodsky RA. Paroxysmal nocturnal hemoglobinuria: stem cells and clonality. *Hematology Am Soc Hematol Educ Program.* 2008;111–115.
2. Takeda J, et al. Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria. *Cell.* 1993;73(4):703–711.
3. Young NS, Calado RT, Scheinberg P. Current concepts in the pathophysiology and treatment of aplastic anemia. *Blood.* 2006;108(8):2509–2519.
4. Maciejewski JP, Sloand EM, Sato T, Anderson S, Young NS. Impaired hematopoiesis in paroxysmal nocturnal hemoglobinuria/aplastic anemia is not associated with a selective proliferative defect in the glycosylphosphatidylinositol-anchored protein-deficient clone. *Blood.* 1997;89(4):1173–1181.
5. Araten DJ, Nafa K, Pakdeesuwan K, Luzzatto L. Clonal populations of hematopoietic cells with paroxysmal nocturnal hemoglobinuria genotype and phenotype are present in normal individuals. *Proc Natl Acad Sci U S A.* 1999;96(9):5209–5214.
6. Inoue N, et al. Molecular basis of clonal expansion of hematopoiesis in 2 patients with paroxysmal nocturnal hemoglobinuria (PNH). *Blood.* 2006;108(13):4232–4236.
7. O'Keefe CL, et al. Deletions of Xp22.2 including PIG-A locus lead to paroxysmal nocturnal hemoglobinuria. *Leukemia.* 2011;25(2):379–382.
8. Mortazavi Y, Tooze JA, Gordon-Smith EC, Rutherford TR. N-RAS gene mutation in patients with aplastic anemia and aplastic anemia/ paroxysmal nocturnal hemoglobinuria during evolution to clonal disease. *Blood.* 2000;95(2):646–650.
9. Sugimori C, et al. Paroxysmal nocturnal hemoglobinuria and concurrent JAK2(V617F) mutation. *Blood Cancer J.* 2012;2(3):e63.
10. Walter MJ, et al. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med.* 2012;366(12):1090–1098.
11. Ley TJ, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature.* 2008;456(7218):66–72.
12. Bessler M, Mason P, Hillmen P, Luzzatto L. omatic mutations and cellular selection in paroxysmal nocturnal haemoglobinuria. *Lancet.* 1994;343(8903):951–953.
13. Mazelin L, et al. Netrin-1 controls colorectal tumorigenesis by regulating apoptosis. *Nature.* 2004;431(7004):80–84.
14. Arakawa H. Netrin-1 and its receptors in tumorigenesis. *Nat Rev Cancer.* 2004;4(12):978–987.
15. van Gils JM, et al. The neuroimmune guidance cue netrin-1 promotes atherosclerosis by inhibiting the emigration of macrophages from plaques. *Nat Immunol.* 2012;13(2):136–143.
16. Cavard C, et al. Gene expression profiling provides insights into the pathways involved in solid pseudopapillary neoplasm of the pancreas. *J Pathol.* 2009;218(2):201–209.
17. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature.* 2013;501(7467):338–345.
18. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature.* 2013;501(7467):328–337.
19. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature.* 2008;455(7216):1061–1068.
20. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368(22):2059–2074.
21. Stransky N, et al. The mutational landscape of head and neck squamous cell carcinoma. *Science.* 2011;333(6046):1157–1160.
22. Graubert TA, et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet.* 2011;44(1):53–57.
23. Makishima H, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood.* 2012;119(14):3203–3210.
24. Ko M, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature.* 2010;468(7325):839–843.
25. Koh SS, et al. Differential gene expression profiling of primary cutaneous melanoma and sentinel lymph node metastases. *Mod Pathol.*

2012;25(6):828–837.
26. Shima H, et al. Bromodomain-PHD finger protein 1 is critical for leukemogenesis associated with MOZ-TIF2 fusion. *Int J Hematol.* 2014; 99(1):21–31.
27. Sun HT, Cheng SX, Tu Y, Li XH, Zhang S. FoxQ1 promotes glioma cells proliferation and migration by regulating NRXN3 expression. *PLoS One.* 2013;8(1):e55693.
28. Kim JY, et al. KDM3B is the H3K9 demethylase involved in transcriptional activation of lmo2 in leukemia. *Mol Cell Biol.* 2012;32(14):2917–2933.
29. Liu L, Sanchez-Bonilla M, Crouthamel M, Giachelli C, Keel S. Mice lacking the sodium-dependent phosphate import protein, PiT1 (SLC20A1), have a severe defect in terminal erythroid differentiation and early B cell development. *Exp Hematol.* 2013;41(5):432–443.
30. Byrd JC, Bresalier RS. Mucins and mucin binding proteins in colorectal cancer. *Cancer Metastasis Rev.* 2004;23(1–2):77–99.
31. Cheung KJ, et al. High resolution analysis of follicular lymphoma genomes reveals somatic recurrent sites of copy-neutral loss of heterozygosity and copy number alterations that target single genes. *Genes Chromosomes Cancer.* 2010;49(8):669–681.
32. Summers KM, et al. Experimental and bioinformatic characterisation of the promoter region of the Marfan syndrome gene, FBN1. *Genomics.* 2009;94(4):233–240.
33. Liu C, et al. SUZ12 is involved in progression of non-small cell lung cancer by promoting cell proliferation metastasis. *Tumour Biol.* 2014;35(6):6073–6082.
34. Abdel-Wahab O, Dey A. The ASXL-BAP1 axis: new factors in myelopoiesis, cancer and epigenetics. *Leukemia.* 2013;27(1):10–15.
35. Panagopoulos I, et al. Fusion of the ZC3H7B and BCOR genes in endometrial stromal sarcomas carrying an X;22-translocation. *Genes Chromosomes Cancer.* 2013;52(7):610–618.
36. Parsyan A, et al. The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. *Proc Natl Acad Sci U S A.* 2009;106(52):22217–22222.
37. Roy S, et al. BCR-ABL1 tyrosine kinase sustained MECOM expression in chronic myeloid leukaemia. *Br J Haematol.* 2012;157(4):446–456.
38. Gomez-Segui I, et al. Novel recurrent mutations in the RAS-like GTP-binding gene RIT1 in myeloid malignancies. *Leukemia.* 2013; 27(9):1943–1946.
39. Kralovics R, et al. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med.* 2005;352(17):1779–1790.
40. Rollinson S, Richards S, Norfolk D, Bibi K, Morgan G, Hillmen P. Both paroxysmal nocturnal hemoglobinuria (PNH) type II cells and PNH type III cells can arise from different point mutations involving the same codon of the PIG-A gene. *Blood.* 1997;89(8):3069–3071.
41. Krawitz PM, et al. A case of paroxysmal nocturnal hemoglobinuria caused by a germline mutation and a somatic mutation in PIGT. *Blood.* 2013;122(7):1312–1315.
42. Parker C, et al. Diagnosis and management of paroxysmal nocturnal hemoglobinuria. *Blood.* 2005;106(12):3699–3709.
43. Maciejewski JP, Tiu RV, O'Keefe C. Application of array-based whole genome scanning technologies as a cytogenetic tool in haematological malignancies. *Br J Haematol.* 2009;146(5):479–488.
44. Gondek LP, et al. Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. *Blood.* 2008;111(3):1534–1542.
45. Tiu RV, et al. New lesions detected by single nucleotide polymorphism array-based chromosomal analysis have important clinical impact in acute myeloid leukemia. *J Clin Oncol.* 2009;27(31):5219–5226.
46. Yoshida K, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 2011;478(7367):64–69.
47. Robinson JT, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29(1):24–26.
48. Makishima H, et al. Somatic SETBP1 mutations in myeloid malignancies. *Nat Genet.* 2013;45(8):942–946.

Leukemia (2015) 29, 126–136

www.nature.com/leu

ORIGINAL ARTICLE

PRPF8 defects cause missplicing in myeloid malignancies

A Kurtovic-Kozaric[1,10], B Przychodzen[1], J Singh[2], MM Konarska[3], MJ Clemente[1], ZK Otrock[1], M Nakashima[4], ED Hsi[4], K Yoshida[5], Y Shiraishi[6], K Chiba[6], H Tanaka[7], S Miyano[6,7], S Ogawa[5,8], J Boultwood[9], H Makishima[1], JP Maciejewski[1] and RA Padgett[2]

Mutations of spliceosome components are common in myeloid neoplasms. One of the affected genes, *PRPF8,* encodes the most evolutionarily conserved spliceosomal protein. We identified either recurrent somatic *PRPF8* mutations or hemizygous deletions in 15/447 and 24/450 cases, respectively. Fifty percent of *PRPF8* mutant and del(17p) cases were found in AML and conveyed poor prognosis. *PRPF8* defects correlated with increased myeloblasts and ring sideroblasts in cases without *SF3B1* mutations. Knockdown of *PRPF8* in K562 and CD34+ primary bone marrow cells increased proliferative capacity. Whole-RNA deep sequencing of primary cells from patients with *PRPF8* abnormalities demonstrated consistent missplicing defects. In yeast models, homologous mutations introduced into *Prp8* abrogated a block experimentally produced in the second step of the RNA splicing process, suggesting that the mutants have defects in proof-reading functions. In sum, the exploration of clinical and functional consequences suggests that *PRPF8* is a novel leukemogenic gene in myeloid neoplasms with a distinct phenotype likely manifested through aberrant splicing.

Leukemia (2015) **29,** 126–136; doi:10.1038/leu.2014.144

INTRODUCTION

The application of next generation sequencing (NGS) has enabled the discovery of novel somatic mutations and their genetic significance in myeloid neoplasms including myelodysplastic syndrome (MDS).[1–6] Various cell functions are affected by these mutations including epigenetic regulation of gene expression at various levels (*DNMT3A, TET2, EZH2*),[7–11] as well as signal transduction (*SETBP1, RAS* pathway).[12–14] Furthermore, NGS projects have revealed a novel class of leukemogenic mutations affecting spliceosomal genes. Such mutations are present in more than 50% of MDS patients and can be associated with particular phenotypes.[3,4,15–17]

The spliceosome is an intricate multi-component RNA–protein complex composed of small nuclear ribonucleoprotein particles (snRNPs) that catalyzes the splicing reaction—the excision of introns and joining of the remaining exon sequences to form a functional messenger RNA. Splicing proceeds through two consecutive steps: (i) the branch site initiates a nucleophilic attack on the 5′ splice site, producing a lariat intermediate and a cleaved 5′ exon; ii) the 5′ exon then attacks the 3′ splice site, yielding spliced mRNA and lariat intron products.[18,19]

The *PRPF8* gene encodes the largest and most evolutionarily conserved protein of the spliceosome, with 61% amino-acid identity between yeast *Prp8* and the human protein.[20] As a key part of the catalytic core of the spliceosome, it forms interactions with all substrates, including the 5′ splice site, branch point, and 3′ splice site in the pre-mRNA, as well as with snRNAs and the excised intron.[21,22] Research indicates that it is essential for vast majority of pre-mRNA splicing and is required in all tissues.[23,24] In higher eukaryotes, PRPF8 is responsible for processing the majority of intron-containing transcripts, including alternatively spliced mRNAs.[18] Research indicates that it is essential for pre-mRNA splicing and is required in all tissues.

Germline mutations clustered in the C terminus of human *PRPF8* lead to type 13 autosomal dominant retinitis pigmentosa (RP13).[25,26]

Recurrent somatic mutations of other components of the spliceosome, such as *U2AF1*, *SRSF2* and *SF3B1*, are common in myeloid neoplasms and are associated with specific phenotypes. For example, ~65% of refractory anemia with ringed sideroblast (RARS) cases have mutations in *SF3B1*, a component of the U2 snRNP.[4] Ringed sideroblasts (RS) are characterized by aberrant accumulation of non-heme iron within mitochondria, but the precise mechanism of RS formation in MDS or *SF3B1* mutant cells remains unknown.

During the investigation of spliceosomal mutations in myeloid neoplasms, we have identified novel, recurrent, somatic mutations in *PRPF8*.[16] However, clinical correlations, prevalence and the functional relevance of *PRPF8* mutations in MDS and AML have not been systematically evaluated. Here, we show that *PRPF8* mutations and deletions correlate with the RS phenotype and have a significant role in the second step of splicing. In addition, we demonstrate that *PRPF8* mutations lead to widespread aberrant splicing of numerous gene transcripts, most notably those known to be involved in the hematopoietic pathway and iron metabolism in mitochondria.

MATERIALS AND METHODS

Patients

Informed consent was obtained according to protocols approved by the institutional review board of Cleveland Clinic. Diagnoses were assigned according to WHO 2008 criteria.[27]

Validation of mutations by Sanger and deep sequencing

All exons of *PRPF8* were amplified and underwent direct genomic sequencing by the ABI 3730xl DNA analyzer (Applied Biosystems, Foster

[1]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland, OH, USA; [2]Department of Molecular Genetics, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA; [3]Rockefeller University, New York, NY, USA; [4]Department of Pathology, Cleveland Clinic, Cleveland, OH, USA; [5]Cancer Genomics Project, Graduate School of Medicine, Tokyo, Japan; [6]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; [7]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; [8]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan and [9]LLR Molecular Haematology Unit, NDCLS, RDM, John Radcliffe Hospital, Oxford, UK. Correspondence: Dr H Makishima, Department of Translational Hematology/Oncology Research, Taussig Cancer Center, Taussig Cancer Institute, 9500 Euclid Ave, R40, Taussig Cancer Institute/R40, 9500 Euclid Avenue, Cleveland, OH 44195, USA.
E-mail: makishh@ccf.org
[10]Current address: Clinical Center of the University of Sarajevo, Sarajevo, Bosnia and Herzegovina.
Received 19 December 2013; revised 7 April 2014; accepted 21 April 2014; accepted article preview online 30 April 2014; advance online publication, 30 May 2014

City, CA, USA) and deep sequencing by an IlluminaMiSeq sequencer as previously described.[13,28,29] All mutations were detected by bidirectional sequencing and scored as somatic if not present in paired CD3-positive cell-derived DNA.

RNA deep sequencing

For detecting exon inclusion/exclusion ratios of targeted genes, we amplified the region of interest by reverse transcription and PCR using primers spanning the adjacent exons. Sequencing libraries were generated and subjected to deep sequencing as previously described.[13]

Single-nucleotide polymorphism array (SNP-A)

SNP-A analysis was performed using Affymetrix 250K and 6.0 Kit (Affymetrix, Santa Clara, CA, USA) according to the standard protocols, followed by copy-number analysis using CNAG (v3.0) (http://www.genome.umin.jp/) or Genotyping Console (Affymetrix).

Inhibitor and semi-solid culture assay

Cells were grown in IMDM media (Gibco, Life Technologies, Grand Island, NY, USA) supplemented with a cocktail of cytokines (100 ng of TPO, SCF, IL3, IL6 and FLT3L; PeproTech, Rocky Hill, NJ, USA) and plated in 48-well plates with or without the spliceosome inhibitor meayamycin (CAS# 933474-21-1).[15] For human cell colony formation (semi-solid culture) assays, a total of 10^5 cells (K562 or normal bone marrow) were plated in triplicate in 1 ml of methylcellulose medium supplemented with G-SCF, GM-SCF and EPO cytokines (MethoCult, StemCell Technologies, Vancouver, British Columbia, Canada) in 35-mm culture plates at 37 °C with 5% CO_2.[15]

Lentiviral infection

Lentiviral shRNA constructs targeting *PRPF8* were obtained from Sigma-Aldrich (St Louis, MO, USA). Viral packaging cell supernatants containing either the control virus vector (without insert) or shRNA vectors were used to infect 5×10^5 K562 cells, unfractionated normal bone marrow cells or CD34+ cells isolated from normal bone marrow using magnetic beads (Miltenyi Biotech, Auburn, CA, USA). Cells were cultured for 2 days in the presence of puromycin (2 μg/μl) to enrich for virus-infected cells following which total RNA was extracted for measurement of mRNA levels.

Quantitative RT-PCR

Total RNA was extracted from bone marrow mononuclear cells and cell lines using the Nucleospin RNA II Kit (Macherey-Nagel, Bethlehem, PA, USA). cDNA was synthesized from 100 ng total RNA using random hexamer primers (Invitrogen, Carlsbad, CA, USA). Gene expression levels were measured using real-time PCR with the ABI PRISM 7500 Fast Sequence Detection System (Applied Biosystems). Primers and probes were purchased from Life Technologies (*PRPF8*: Hs00197615_m1 and Hs01556852_m1; *GAPDH*: Hs99999905_m1). The expression level of target genes was measured in duplicate and normalized to the *GAPDH* mRNA as previously described.[30]

Yeast strains and copper assays

The *S. cerevisiae* strain yJU75, which is deleted for chromosomal *Prp8*, was transformed with plasmids containing either the yeast wild-type *Prp8* gene or the yeast *Prp8* gene carrying each single point mutation corresponding to the human MDS-associated mutations. To monitor splicing efficiency, we used the ACT1-CUP1 reporter gene comprised of the yeast actin (ACT1) intron inserted upstream of the copper resistance gene, CUP1 open-reading frame. Growth of the cells in the presence of copper provides a measure of splicing efficiency of the ACT1 intron. To test the function of *Prp8* mutants, the ACT1 intron was modified by various 5′ splice site, branch site or 3′ splice site mutations as described[31] or prepared by overlapping PCR and *in vivo* gap-repair cloning.[32] For the copper resistance assays, cultures were grown to mid-log phase in –Leu medium, diluted to $A_{600} = 0.2$, and equal volumes were dropped onto –Leu plates containing 0–2.0 mM $CuSO_4$.[32] Plates were scored and photographed after 3 days at 30 °C.

Caspase assay

The Caspase-glo 9 assay was used to measure caspase-9 activity (Promega, Madison, WI, USA) according to the manufacturer's recommended protocol. Luciferase activity generated is proportional to the amount of caspase activity present.

Splice site analyses

For analysis of 5′ and 3′ splice site sequences, 30 nucleotides spanning the exon/intron junctions were extracted from the UCSC genome browser (http://www.genome.ucsc.edu/). Splice site strength was scored using MaxEntScan.[33] The control set consisted of a randomly selected set of 1000 exons from exons generated by the TCGA data set used in the analyses. Sequence logos were generated using the WebLogo online application (http://weblogo.berkeley.edu).

Publicly available databases and analytical tools

The February 2009 human reference sequence (GRCh37) produced by the Genome Reference Consortium was used as the reference genome (UCSC genome browser; http://genome.ucsc.edu/cgi-bin/hgGateway). Expression array data were extracted from Oncomine (https://www.oncomine.org/). Somatic mutation data were obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC) database (http://www.sanger.ac.uk/genetics/CGP/cosmic/). Cytogenetic and mutational status was obtained for AML patients ($n = 197$) from The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/).

RESULTS

Identification of *PRPF8* mutations and deletions

On the basis of two index cases of patients with *PRPF8* mutations previously published,[13,16] we screened for lesions affecting this gene in a larger cohort ($N = 447$) of patients with MDS and related conditions (Supplementary Table 1). We found 15 heterozygous mutations in *PRPF8* gene: 13 mutations were somatic missense and 2 somatic nonsense mutations. In total, we characterized 15 samples from 13 patients (2 patients had serial samples before and after AML progression; Figure 1a, Supplementary Table 1). The most common mutation was in D1598, found in four different patient samples. Two additional mutations from TCGA primary acute myeloid leukemia (pAML) database have been added to the analysis (A687P and G1750E, Figure 1a). The positions and identity of *PRPF8* mutations identified in myeloid neoplasms are given in Figure 1a along with the distribution of mutations found in other types of tumors (Figure 1a). The somatic nature of *PRPF8* mutations found in our patient cohort was confirmed using CD3+ fraction DNA (Figure 1b).

PRPF8 mutations are generally missense and scattered throughout the gene (Figures 1a and d). The high level of conservation between the yeast *Prp8* and human *PRPF8* proteins allows for precise mapping of homologous mutations (Figure 1d). In addition, 24 cases in our cohort and 12 cases in the TCGA pAML data set contained deletions of one copy of the *PRPF8* locus (Figure 1a) and exhibited *PRPF8* haploinsufficiency (Figure 1c and Supplementary Figure 1A, Supplementary Table 2, Supplementary Table 3). A further analysis of a cohort of 447 cases of MDS and related conditions showed that 20% contain mutations in various spliceosomal protein genes, including *SF3B1*, *SRSF2*, *U2AF1*, *ZRSR2* and *LUC7L2* (Supplementary Figure 1B). Interestingly, in general, the various spliceosomal factor mutations were mutually exclusive (Figures 2a and b). Other mutations coinciding with *PRPF8* such as *TET2*, *CBL* and *TP53* were identified (Figure 2a).

PRPF8 defects were most frequently identified in pAML and sAML (primary and secondary acute myeloid leukemia; Figure 2c), suggesting an association with more aggressive disease phenotypes as opposed to low-risk MDS (Figure 2c; $P < 0.01$, Supplementary Table 3). Serial sample sequencing in two exemplary cases showed that *PRPF8* mutations were present at disease onset as assessed by comparison of variant allelic frequencies obtained for mutant and wild-type alleles at sequential time points (Supplementary Figures 2A and 2B).

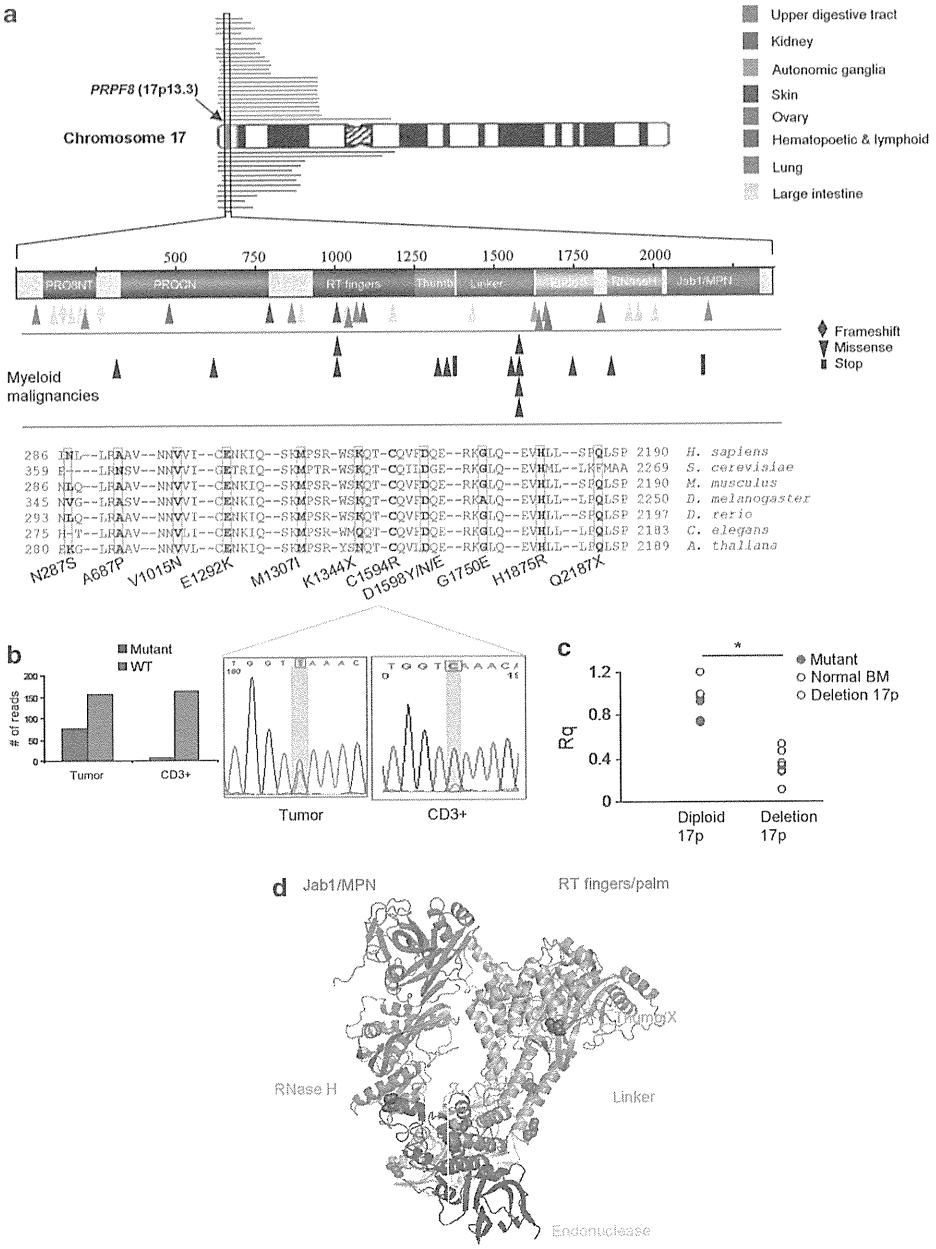

Figure 1. *PRPF8* mutations and deletions in myeloid malignancies. (**a**) The *PRPF8* gene is located on 17p13.3 (red vertical box indicated by an arrow). In this study, 24 patients with 17p deletions were analyzed (green bars). TCGA database contained 12 patients with similar deletions of the 17p region (blue bars). Also shown is the domain structure of the *PRPF8* protein including several conserved domains: Reverse Transcriptase (RT fingers/palm), Thumb/X, Endonuclease and RNase H domains. Tumor-associated missense mutations forming a variety of tumor types are scattered throughout the protein-coding region. Myeloid malignancy missense mutations (shown in black) are located in regions that are conserved from yeast to humans as shown by the sequences of a representative sample of organisms (mutated residues shown in bold, the mutations are listed below the figure). (**b**) The somatic nature of the mutations was confirmed using either targeted deep sequencing (right) or Sanger sequencing (left) in tumor and CD3 + cells. (**c**) Expression levels of *PRPF8* using RT-PCR (Rq) normalized to normal bone marrow values (NBM). Mutant *PRPF8* (blue) and NBM samples (yellow) are in the diploid 17p column (left) and deletion 17p samples ($n = 6$) are in right column (white, $P < .001$). (**d**) Disease-associated mutations (black) are mapped onto the yeast *Prp8* protein structure.

Phenotypic/genotypic associations of *PRPF8* mutations

In 60% of cases (54% of mutants and 65% of deletions), an increased presence of RS was noted by Prussian blue staining (Figures 2a and d; Supplementary Figure 1C and Supplementary Table 2). The RS phenotype without *PRPF8* mutation or deletion was also found in 24 additional cases out of 447, 16 of which harbored *SF3B1* mutations.

In most of the *PRPF8* mutant and deleted samples, pseudo Pelger-Huet anomaly (PHA), characterized by bilobed or hypolobular nuclei were noted (Figure 2d; Supplementary Tables 3, 4). As *PRPF8* is located on 17p13.3 and *TP53* is on 17p13.1, some of the del(17p) cases also involve the *TP53* locus (Figure 1a). However, increased RS and PHA were also found in cases without *TP53* deletion.

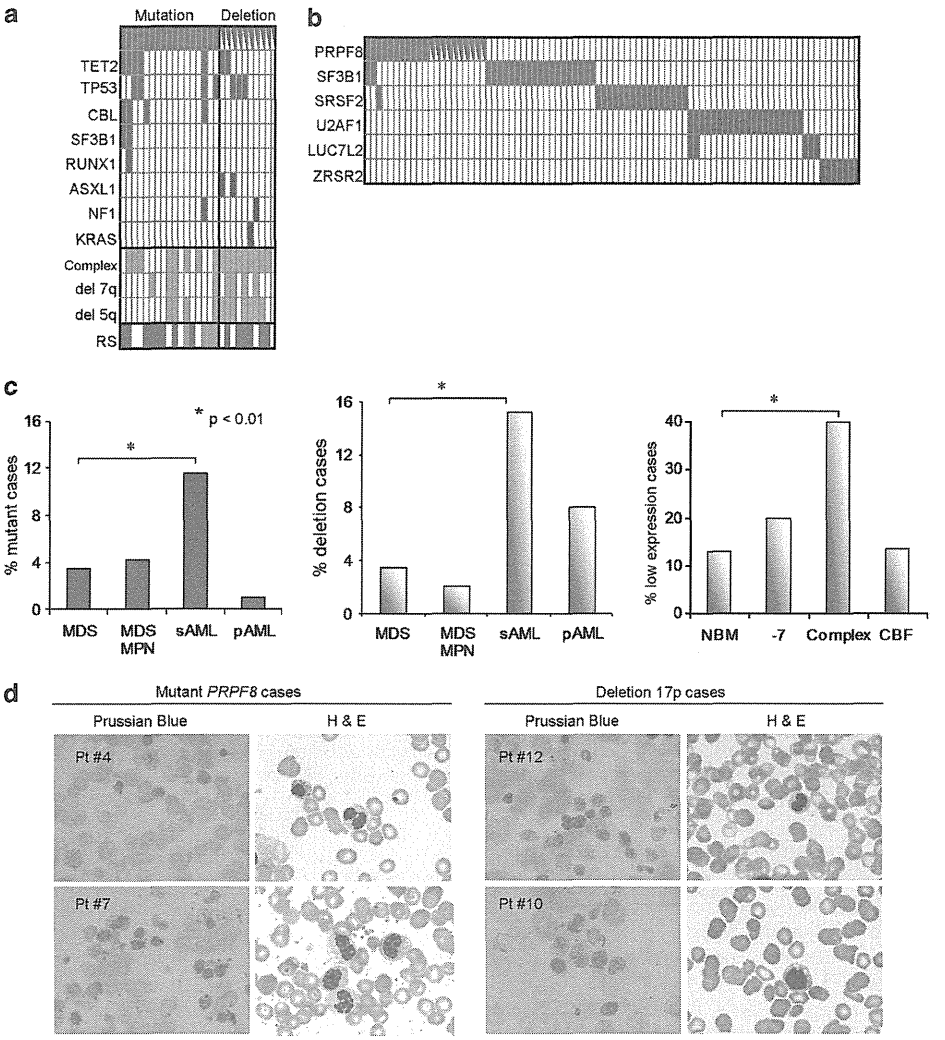

Figure 2. Morphological characteristics of *PRPF8* mutants and deletions. (**a**) Concomitant mutational spectrum of *PRPF8* mutation and deletion cases. Included are other clinical features of the screened cohort (cytogenetics: complex (yes or no); deletion 7q; deletion 5q) and the presence of RS (red), no increase in RS (white) and unknown (grey). Deletion cases are indicated with half-blue rectangles. (**b**) The cases with *PRPF8* mutations and deletions are not concomitant with other spliceosomal mutations except for two cases of *SF3B1* and one case of *SRSF2* mutations. Deletion cases are indicated with half-blue rectangles. (**c**) Etiology of *PRPF8* mutation (left), deletion (center) and low expression cases (right, Oncomine database), all which are significantly associated with aggressive disease course (secondary AML or complex karyotype). The pAML cases with mutations or deletions are taken from the TCGA. For del(17p) deletion cases, 450 cases were analyzed by SNP-A. (**d**) Two exemplary mutant *PRPF8* cases and two deletion cases confirm a RS phenotype with Prussian blue staining (left panels, blue staining surrounding nuclei). Quantification of RS in each analyzed case is given in Supplementary Table 3. H&E staining (right panels) shows pseudo Pelger-Huet anomaly (bi- or hypo-lobular nuclei), characteristic for 17p deletion cases.

The gene expression changes associated with RS phenotype with defective *PRPF8* function are shown in Supplementary Figure 3. When we analyzed the 35 samples from TCGA database for which *PRPF8* mutational status was available (mutation and low expression as compared with controls with normal PRPF8 expression, $P<0.01$), 20% of the 200 most differentially expressed genes were associated with mitochondrial function. In addition, when we analyzed the overlap between differentially expressed genes in RS cases with and without *PRPF8* mutations, several genes involved in the electron transport chain complex were found (Supplementary Figure 3).

Decreased expression of *PRPF8* leads to increased cellular proliferation

To evaluate the effect of reduced *PRPF8* expression on cellular proliferation, we knocked down *PRPF8* mRNA levels using two shRNA lentivirus constructs (Supplementary Figure 4). Decreased expression levels of *PRPF8* were associated with increased cellular proliferation (Figure 3a) and increased clonogenicity (Figure 3b). Control and knockdown cells showed indistinguishable activity in a caspase 9 assay, indicating no difference in apoptosis (Figure 3c). When *PRPF8* mutant and deletion cells were treated with meayamycin, a potent pre-mRNA splicing inhibitor that targets the splicing factor 3b complex,[15] growth of *PRPF8* defective cells was more susceptible to inhibition than normal bone marrow cells (Figure 3d). The AML cell line KG1 that is monosomic for chromosome 17 is also more sensitive to meayamycin than normal bone marrow cells (Figure 3d).

Functional consequences of homologous mutations

An important question arises as to whether the *PRPF8* mutations confer an altered function or they are loss of function alleles, and

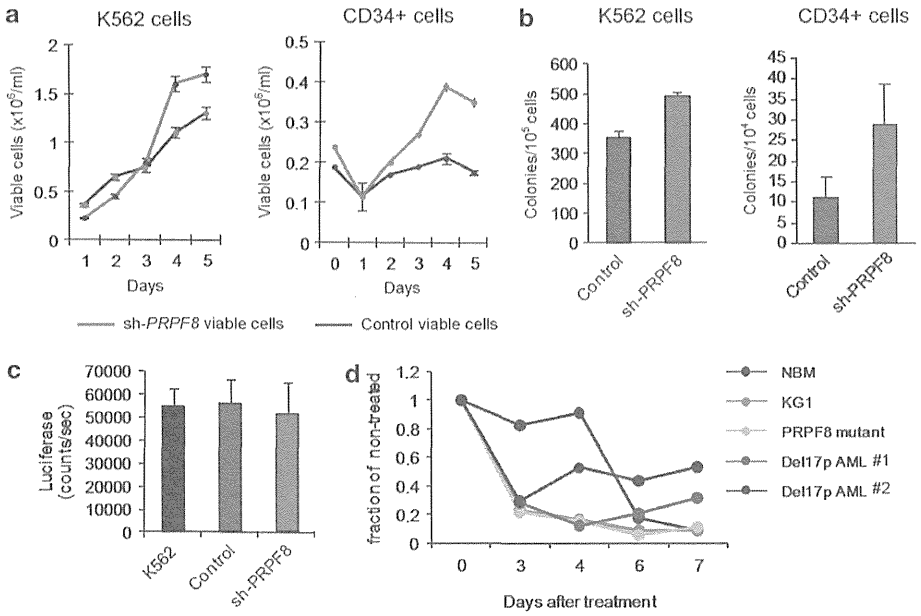

Figure 3. Cell proliferation and viability following *PRPF8* knockdown. (**a**) Proliferation of cells with reduced *PRPF8* levels. Growth curve axis (number of viable cells in 10^6/ml) is shown on left y-axis and days on x-axis. In both types of cells, knockdown of *PRPF8* leads to increased viable cell counts. Error bars are calculated as 1 s.e. (**b**) Colony formation of K562 and CD34+ cells with *PRPF8* knockdown showing increased colony counts. Cells, 10^4 K452 and 10^5 CD34+, were plated in semi-solid culture supplemented with cytokines and counted after 4 and 10 days, respectively. (**c**) Caspase 9 activity assay shows that there is no increase in apoptotic activity in *PRPF8* knockdown cells. (**d**) Effect of treatment with the pre-mRNA splicing inhibitor meayamycin on patient cells with deletion 17p (two primary samples, Del17p AML #1 and #2) and mutant *PRPF8* patient cells as well as normal bone marrow cells (NBM) and the KG1 cell line.

thus merely recapitulate the haploinsufficiency of the deletion cases. As the *PRPF8* protein is exceptionally conserved between yeast and humans (Figure 1a), we created homologous mutations in yeast *Prp8* to determine whether the human disease-associated *PRPF8* mutations cause splicing defects. Nine different yeast mutations, N760P (hA687P), V1088N (hV1015N), E1364K (hE1292K), M1379I (hM1307I), D1670Y/N (hD1598Y/N), G1822E (hG1750E), H1947R (h1875R), corresponding to the human somatic mutations were constructed and introduced into cells in which the chromosomal *Prp8* gene was deleted, thus making the mutant *Prp8* the only source of activity. In all cases, cell growth at all tested temperatures was indistinguishable from wild type, suggesting that these mutations do not grossly affect function. To detect more subtle defects, we used a modification of the classical yeast splicing suppressor assay (Figures 4a and b). In brief, we used a reporter plasmid that contains the ACT1–CUP1 reporter gene in which the ACT1 exon 1 and intron are inserted into the CUP1 gene. If splicing function is normal, the ACT1 intron is spliced out allowing the production of active CUP1 protein, which confers a level of copper resistance to the cells that is proportional to the level of pre-mRNA splicing. If splicing is aberrant because of mutations in the ACT1 intron splice sites or *Prp8* mutations, copper resistance is lowered. Previous studies have shown that certain splice site mutations in the ACT1 intron can be suppressed by second site mutations in splicing factors, particularly *Prp8*.[32,34]

We therefore tested the *Prp8* mutants for their ability to suppress splicing defects and thus restore copper resistance in strains containing reporter genes with several different intron mutations known to inhibit splicing in *Prp8* wild-type strains. These mutations were located at the 5′ splice site (for example, U2A and G5A of the GUAUG consensus), the branch site (for example, A>G at the branch point), the 3′ splice site (for example, GAG/ instead of CAG/) or an intron containing a suboptimally short (8 vs 43 nt in wild type) distance between the branch site and the 3′ splice site (Figure 4b).

Previous work has shown that different intron mutations block splicing at different steps in the reaction (Figure 4a).[32,34] On the basis of this paradigm, we tested the disease-associated *Prp8* mutants with several types of intron mutations. Splicing of introns with 5′ splice site mutations (defective in the first step of splicing) was not rescued by disease-associated *Prp8* alleles. However, we did observe improvement of splicing for intron mutants defective in the second step, in particular for the short branch site-3′SS distance mutant, as indicated by the improved growth in the presence of copper (Figure 4c). In this experiment, the previously described *Prp8* alleles, *Prp8*-p.R1753K (known to remove first step blocks), and *Prp8*-p.V1870N and p.W1575R (known to remove second step blocks) were used as controls. As expected, splicing of the short branch to 3′ splice site intron mutant (a prototypical 2nd step defect mutant) is not improved by the first step *Prp8*-p.R1753K suppressor allele, but only by the second step *Prp8*-p.V1870N and p.W1575R suppressor alleles (Figure 4c). Among the disease-associated *Prp8* mutations, changes at five different positions (p.V1088, p.E1364, p.M1379, p.D1670 and p.H1947) all improved splicing of the short branch-3′SS intron mutant in the yeast model (Figure 4c). At position p.V1088, three different substitutions (to N, D, and I) had a similar effect (not shown), suggesting that transition to the second step may be favored by any mutation at this position. By contrast, p.D1670Y and p.D1670N, but not p.D1670H, exhibited the suppressor effect (Figure 4c). Only two of the tested disease-associated *Prp8* mutants, p.N769P and p.G1822E, showed little or no suppression in the yeast assay. These results suggest that the *PRPF8* mutations in MDS may alter the internal dynamics of the spliceosome leading to changes in splicing patterns.

PRPF8 mutations and low *PRPF8* expression affect splicing patterns in human cells

As disease-associated *PRPF8* mutations showed splicing alterations when tested in yeast, we hypothesized that they might have an

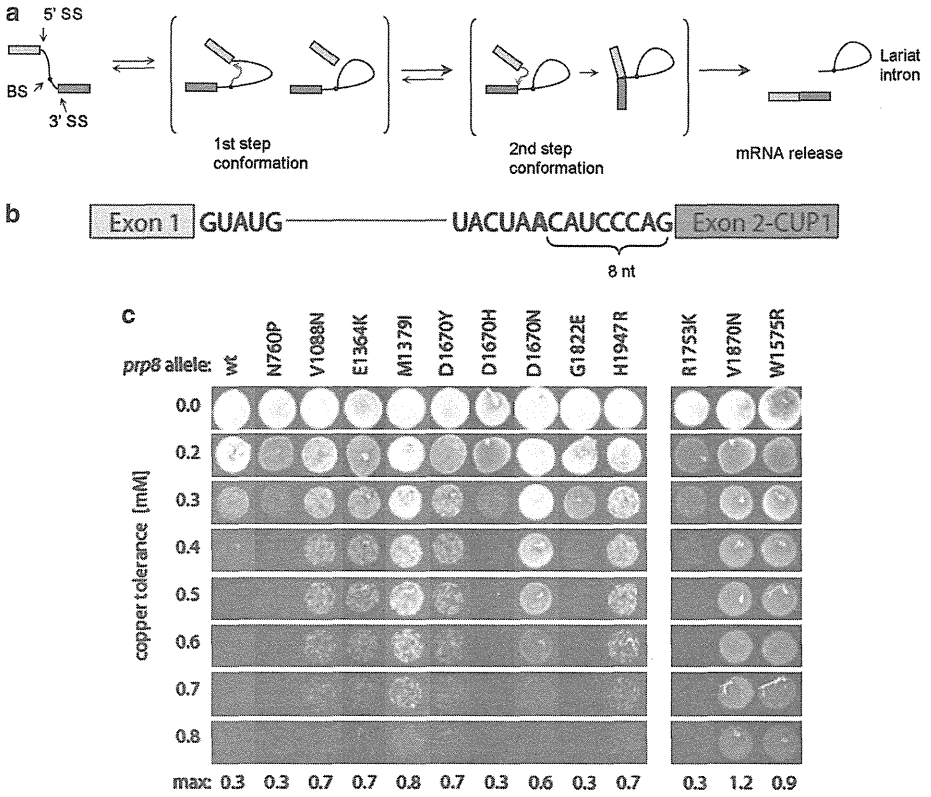

Figure 4. Yeast *Saccharomyces cerevisiae in vivo Prp8* splicing assay. (**a**) Schematic representation of the two-step splicing pathway (SS, splice site; BS, branch site). In brief, two consecutive trans-esterifications involve three sites of the intron: in the first step, the BS attacks the 5′ SS, producing a lariat intermediate and the cleaved 5′ exon. In the second step, the 5′ exon attacks the 3′ SS, resulting in spliced mRNA and the lariat intron. *Prp8* is required for both chemical steps and has been shown to regulate the conformational changes in the spliceosome that accompany the process. (**b**) Diagram of the *ACT1-CUP1* reporter gene. The *CUP1* gene allows growth in copper-containing media and is interrupted with an intron (originally from *ACT1* gene) such that increased splicing results in increased resistance to copper. The intron tested here has been mutated so that the BS adenosine (bold) is placed eight nucleotides away from exon 2, instead of 43 nucleotides as found in the wild-type *ACT1* intron. (**c**) Yeast growth on different copper concentrations (y-axis). Different columns represent wt and disease-associated mutant alleles in the left panel, and control mutant alleles in the right panel. Amino-acid designations are from yeast *Prp8*. Control p.R1753K is a first step suppressor mutant, while p.V1870N and p.W1575R are second step suppressor mutants. Mutations p.V1088D/I/N (hV1015), p.E1364K (hE1292), p.M1379I (hM1307I), p.D1670Y/H/N (hD1598Y) and p.H1947R (hH1875R) show a second-step suppressor phenotype, while p.N760P (hA687P) and p.G1822E (hG1750E) show a wild-type phenotype.

effect on alternative splicing in human AML samples. Using deep RNA sequencing data in cases of pAML from TGCA, alternative exon usage patterns were analyzed with SpliceTrap software (Supplementary Table 4) in *PRPF8* mutants ($n=2$, A687P and G1750E), del(17p)/low-expression cases ($n=17$) and wild-type *PRPF8* cases ($n=14$). *PRPF8* defect-specific alternative splicing patterns were categorized into two groups: those genes with increased exon exclusion (lower exon usage) and genes with increased exon inclusion (higher exon usage; Figure 5a). Globally, we observed that hundreds of exons are affected in both groups and in both types of *PRPF8* defects (Figures 5b and c). *PRPF8* low expressers displayed modestly increased exon exclusion as compared with controls, while, in mutant cases, more robust results were found, in that approximately twice as many exons are affected and far more of these are altered to a greater extent than seen in *PRPF8* low expressers (Figure 5b). These results indicate that *PRPF8* mutations (and likely haploinsufficient expression) result in widespread changes in alternative splicing.

We also compared the overlaps of the sets of exons excluded or included in mutant and low expression samples. If the *PRPF8* mutations resulted simply in a loss of function, the effects on splicing should be similar in mutant and deletion cases. Although there was significant overlap (>50%) between mutant and low expresser cases in the spectrum of more skipped and more retained exons (Figure 5c), there were far more exons affected by *PRPF8* mutations, and these were less likely to be altered in low expression samples. In addition, the magnitude of exon splicing alterations was higher for the mutant cases than the low expression cases (Figure 5b). These results support the hypothesis that the *PRPF8* mutant proteins have a neomorphic functional phenotype rather than simply a defective function.

The yeast *Prp8* mutant phenotype suggested that mutations might affect the ability of the spliceosome to reject poor splice sites. To determine whether *PRPF8* mutants and low expressers are more permissive for poor sequences surrounding the 3′ and 5′ splice sites of the alternatively spliced exons, we selected the 20 exons with the greatest changes from control samples (either exon exclusion or inclusion) and calculated the frequencies of nucleotides at each position within the consensus 5′ and 3′ splice site sequences (Figure 5d). In addition, we computed the average splice site strength as measured by MaxEnt scores[33] for each splice site of each set. The results show that there are significant differences in splice site strength adjacent to affected exons (Figure 5e). The top section of Figure 5d shows that the scores for splice sites adjacent to a random control set of exons are in the narrow range of 8.0–8.5. The center section shows the splice sites adjacent to exons affected by *PRPF8* mutations. In these cases, the splice sites adjacent to the more included exons are quite weak

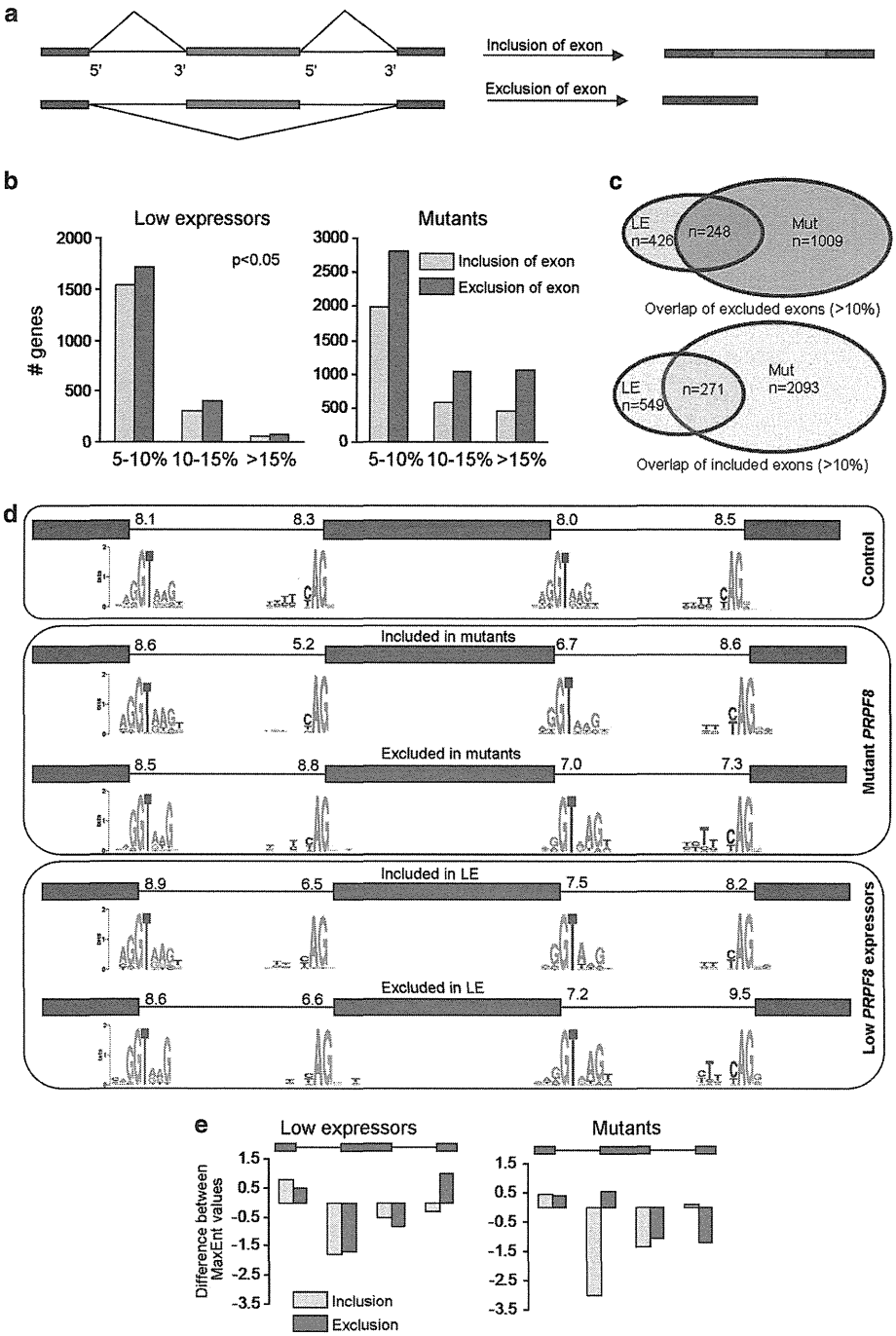

Figure 5. Aberrant alternative splicing in *PRPF8* mutants and deletions. (**a**) Schematic representation of alternative splicing (inclusion or exclusion) of the middle exon. (**b**) Number of genes whose splicing is affected in the RNA sequencing analysis of low expressers (left) and mutants (right). The x-axis shows the percentage difference of low expressers or mutants to control 'wild type' expressers of *PRPF8* of either exon inclusion or exclusion, binned as 5–10%, 10–15% and >15% absolute change. The number of genes that show alternative exon exclusion is higher in all categories, in low expressers as well as in mutants. Overall, the number of genes with either inclusion or exclusion is higher in mutants. (**c**) The overlap of affected exons between mutants (M) and low expressers (LE) in either included (top) or excluded exons (bottom) using a threshold of >10% absolute change. (**d**) The splice site sequences of the 20 most affected exons in each category are compared. The average strength of each set of splice sites was determined using the MaxEntScan program, represented by the number above each splice site. The first box represents control, the middle box mutant *PRPF8* and the lower box the low expressers. The graphical representation of nucleotide frequencies, calculated by Weblogo, at each splice site is shown for each case. (**e**) Using the data shown in **d**, the deviation of the MaxEnt scores from the control values for each set of exons is plotted.

compared with the flanking splice sites. Thus, these exons would be expected to be poorly included under normal conditions. The fact that they are activated by the *PRPF8* mutations suggests that splice site recognition has been altered and, specifically, that poor 3′ splice sites adjacent to the regulated exons are most affected (Figure 5e). In contrast, the 3′ splice sites adjacent to exons that

are skipped more frequently in *PRPF8* mutants are stronger than average (8.8 compared with 8.3) while the downstream 3′ splice sites are relatively weak. These differences again suggest that mutant *PRPF8* is affecting splice site strength or efficiency. The changes seen in *PRPF8* low expression samples are less marked and more consistent between more and less included exons supplying further evidence that the disease-associated *PRPF8* mutants are functionally different from loss of function mutations.

These results are consistent with the phenotype of the homologous yeast mutants in which a suboptimal 3′ splice site was activated by the mutations (Figure 4). We conclude that the *PRPF8* mutants facilitate splicing of the upstream 5′ splice sites to weak 3′ splice sites preceding the alternative exon. As a result, the subsequent splicing of the downstream pair of splice sites would be forced, leading to enhanced inclusion of alternative exons. Similarly, in cases where the alternative exon is more frequently skipped in the mutant samples, the downstream 3′ splice site, which can be thought of as competing for the first 5′ splice site with the alternative 3′ splice site, is weaker than average. Again, the mutant *PRPF8* protein could be promoting splicing to the weaker 3′ splice site, in this case leading to enhanced exon exclusion.

Validation of aberrant splicing in candidate target genes

To validate the above splicing analysis in primary samples, we first tested the twenty most misspliced genes using RT-PCR gel band intensity and selected five genes for further analysis (*GATA1, NDUFAP6, SLC25A19, SFXN2* and *RPS24*). We asked whether similar splicing patterns are seen in *PRPF8* mutant and deletion cases and in K562 cells in which *PRPF8* expression was knocked down using shRNA. Normal bone marrow cells were used as controls. Alternative exon usage was measured in two ways after RT-PCR was performed on cellular RNA samples using primers located in the two exons flanking the alternative exon (Figure 6a): (1) the percentages of exon included to exon excluded products were determined by measurement of gel band intensity; and (2) targeted deep RNA sequencing (Figures 6b and c). The results were also compared with the ratios obtained from the TCGA data set.

Overall, the results show that samples from *PRPF8* knock down cells, low expression cells ($n = 5$) and mutant patient cells ($n = 3$) all had similar exon splicing ratios to those obtained from the TCGA data set (Figure 6c). For the *SFXN2* gene, exon inclusion percentages were similar in mutant and low expressers. In the case of the *NDUFAF6* gene, all samples showed similar changes in exon exclusion. A different pattern is seen in the *RPS24* gene where knockdown of *PRPF8* in K562 cells showed increased inclusion of the exon, while the low expressers and mutants showed preferential exon exclusion. In concordance with a recent report,[35] we also found that *GATA1*, a transcription factor that has a critical role in hematopoiesis, is misspliced in human *PRPF8* mutants, low expressers and knockdown cells (Figure 6b).

DISCUSSION

Spliceosomal protein genes have been recently identified as an important class of genes affected by somatic mutations in MDS.[3,4,15–17,36] Here we present novel evidence that the spliceosomal gene *PRPF8* has a direct role in MDS. *PRPF8* mutations and haploinsufficiency lead to distinct oncogenic and phenotypic features including the presence of RS and the PHA phenotype. The association of PHA with del(17p) has been reported previously.[37] RS are observed in both *SF3B1* and *PRPF8* mutations. However, *SF3B1* mutations are associated with a mild disease phenotype, while *PRPF8* defects are associated with a more aggressive phenotype. Furthermore, knockdowns of these two factors have different effects on cells. Unlike *SF3B1*,[4,15] *PRPF8* knockdown leads to increased cellular proliferation and increased clonogenicity. Nevertheless, the association of *SF3B1* and *PRPF8* mutations with the RS phenotype suggests that there may be common genes or pathways affected by both *SF3B1* and *PRPF8*.

PRPF8 is located on chromosome 17p13.3, close to the *TP53* locus (17p13.1). While precise mapping of 17p deletions did show two minimally deleted regions centered on *TP53* and *PRPF8*, both genes were affected by a deletion in multiple cases (data not shown). Frequent cases were also identified with deletion or mutation of *PRPF8* concomitant with *TP53* mutations. Mechanistically, loss of function mutations in zebrafish *PRPF8* have been shown to result in missplicing of *TP53*,[35] while *PRPF8* knockdown in mammalian cells led to increased compensatory transcriptional activity of *TP53*,[38] suggesting another possibility of synergetic interaction between *PRPF8* defects and *TP53*.

While *PRPF8* has not been previously implicated in the regulation of mitochondrial genes, our analysis of patients with RS and *PRPF8* defects demonstrated changes in the expression profiles of several genes involved in electron transport chain complexes. Some of the most misspliced genes seen in our RNA sequencing analysis are involved in mitochondrial function, for example, *NDUFAF6, SFXN2* and *SLC25A19. NDUFAF6* has a crucial role in the assembly of complex I (NADH-ubiquinone oxidoreductase) of the mitochondrial respiratory chain.[39,40] Sideroflexin-2, *SFXN2*, is a putative mitochondrial iron transporter.[41,42] *SLC25A19* is a mitochondrial thiamine pyrophosphate transporter.[43,44] *RPS24*, another highly misspliced gene that we analyzed in primary patient samples, is mutated in Diamond-Blackfan anemia.[45,46] A further connection between *PRPF8* and hematopoiesis is seen in zebrafish where a *PRPF8* mutation resulted in accumulation of many aberrantly spliced transcripts along with the impairment of myeloid differentiation.[35] This study reported the missplicing of *GATA1* in zebrafish. In agreement with this observation, we found that human *GATA1* is aberrantly spliced in association with *in vivo* and *in vitro* lesions in *PRPF8*. All of these results support the notion that *PRPF8* mutations or haploinsufficiency affects the alternative splicing of genes involved in mitochondrial metabolism and hematopoietic differentiation.

We found large-scale alterations in alternative splicing patterns in samples from patients with *PRPF8* mutations and deletions. Several studies in yeast have shown that *Prp8* mutations can improve splicing of suboptimal splice sites by modulating spliceosome conformation.[34,47–52] In agreement with this notion, our study demonstrated that mutations in yeast *Prp8* alter the protein's propensity to allow splicing of certain suboptimal 3′ splice sites normally blocked in the second step. Alternative splice sites can also be activated at a late stage of spliceosome assembly after mutant *PRPF8* has joined the complex.[48] In human cells, we can detect enhanced activation of suboptimal alternative splice sites when *PRPF8* is mutated or haploinsufficient. These findings suggest that the deregulated selection of splice sites might be one of the significant pathophysiologic mechanisms in *PRPF8* defects.

We have argued above that (1) *PRPF8* mutations and haploinsufficiency produce similar pathophysiologies in myeloid neoplasms, yet (2) *PRPF8* mutations produce a molecular phenotype distinct from haploinsufficiency, suggesting that the mutations have a neomorphic character. To reconcile these findings, we first note that there is significant overlap between the gene sets whose alternative splicing is affected by mutations and haploinsufficiency. This suggests that one or more downstream genes involved in neoplastic progression are similarly affected by both defects. Second, the mutational phenotype may also encompass some degree of reduced *PRPF8* function in that the proposed defects in proof-reading might also lead to increased lifetimes of non-functional spliceosomal complexes. This would sequester much of the mutant *PRPF8* and partially phenocopy haploinsufficiency.

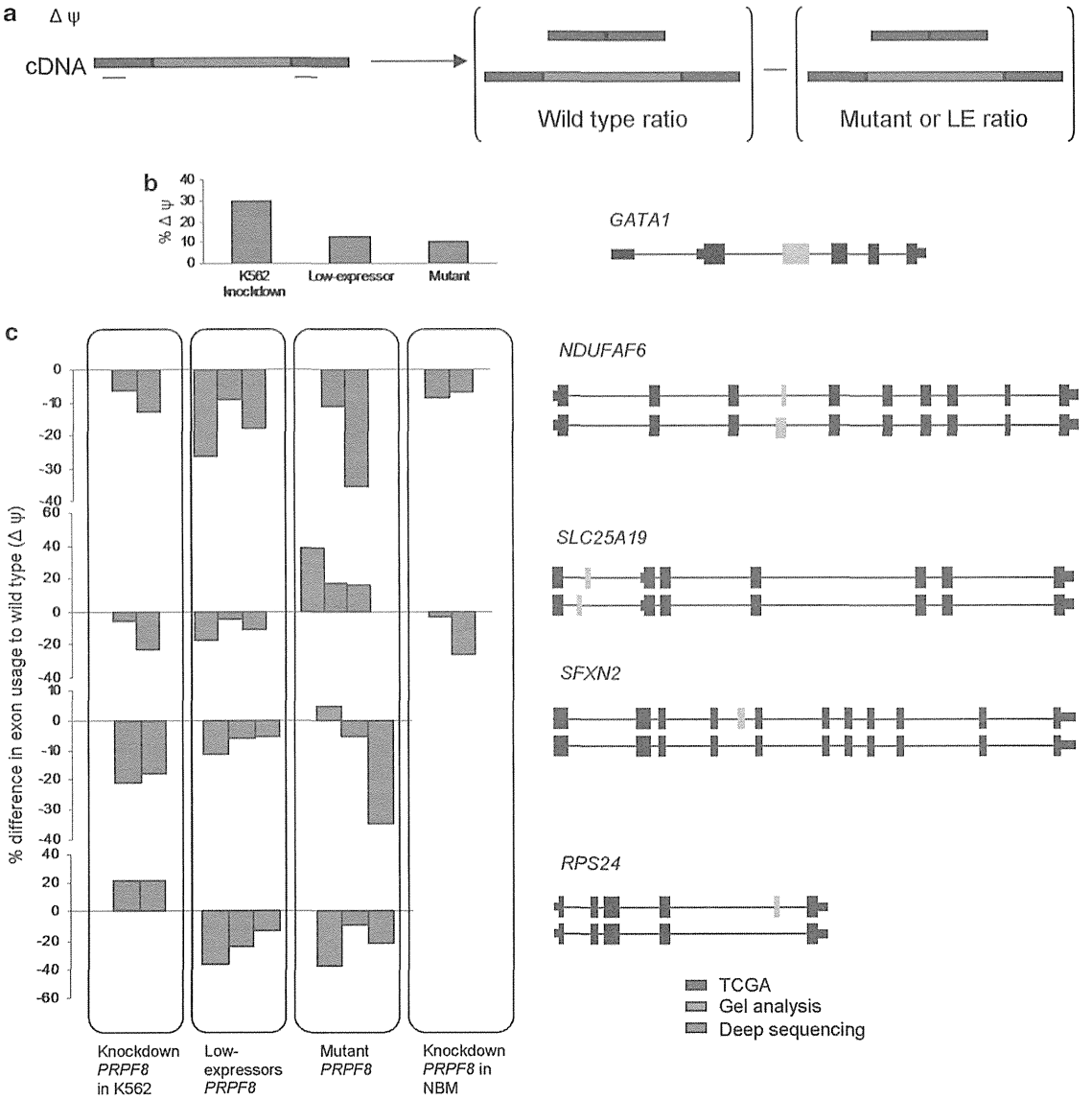

Figure 6. Confirmation of alternatively spliced genes. (**a**) Gene-specific primers were designed to flank the alternative exon. The size of each PCR product was between 100 and 200 bp. DeltaPsi value is calculated as the difference in the percent of intron retention of the wild type as compared with the percent of intron retention in cells with aberrant *PRPF8*. (**b**) *GATA1* exon2 inclusion as determined by gel analysis of RT-PCR of RNA from K562 cells with shRNA knockdown of *PRPF8*, patient cells with a *PRPF8* mutation, and patient cells with a 17p deletion. The results are presented as the change in the percent of intron retention compared with control K562 cells or normal bone marrow cells, respectively. (**c**) Four genes were selected for validation of RNA-sequencing results. *RPS24*, *SLC25A19*, *SFXN2* and *NDUFAF6* were analyzed for changes in alternative splicing using data from TCGA database, deep sequencing or gel analysis of RT-PCR products. The schematic representation of each gene is given on the right, with the two alternative isoforms given below each other. The alternative exon in each case is colored yellow. Exon inclusion values for *PRPF8* knockdown, low-expresser or mutant (p.A687P and p.G1750E) samples were subtracted from the corresponding wild-type values for each group (that is, positive values represent increased inclusion and negative values represent increased exon exclusion). For the deep sequencing analysis, at least three primary samples were used for each category.

In conclusion, the core spliceosomal protein *PRPF8* is frequently mutated or deleted in myeloid neoplasms and joins other spliceosomal factors as potential driver genes. We show that *PRPF8* lesions lead to neomorphic spliceosomal activity and increased cellular proliferation resulting in a distinct phenotype of aggressive myeloid malignancies with increased RS. Our alternative splicing analyses demonstrate that these effects are likely mediated through missplicing of genes involved in iron accumulation in mitochondria and abnormal hematopoiesis. Our functional studies further suggest that *PRPF8* mutations confer a distinct gene expression phenotype that only partially overlaps that of the haploinsufficient state.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health grants R01 HL082983 (JPM), RO1 GM093074 (RAP), and RO1 GM049044 (MMK), Scott Hamilton CARES Initiative

(HM), Aplastic Anemia & MDS International Foundation (HM), Fulbright Scholar Fellowship, US Department of State (AK). We thank Naoko Hosono, Kathryn Guinta and Brittney Dienes for experimental and technical assistance.

AUTHOR CONTRIBUTIONS

AK, JPM, RAP, MMK and HM designed research, performed research, collected data, performed statistical analysis and wrote the manuscript; BP and MC designed research, performed statistical analysis, contributed analytical tools, interpreted data and wrote the manuscript. JS, MMK, MN, ZKO, EDH, KY, SO and JB collected and interpreted data.

REFERENCES

1 Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008; **456**: 66–72.
2 Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013; **368**: 2059–2074.
3 Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 2011; **478**: 64–69.
4 Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D *et al.* Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med* 2011; **365**: 1384–1395.
5 Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G *et al.* Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 2013; **28**: 241–247.
6 Walter MJ, Shen D, Shao J, Ding L, White BS, Kandoth C *et al.* Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia* 2013; **27**: 1275–1282.
7 Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* 2010; **363**: 2424–2433.
8 Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A *et al.* Mutation in TET2 in myeloid cancers. *N Engl J Med* 2009; **360**: 2289–2301.
9 Ernst T, Chase AJ, Score J, Hidalgo-Curtis CE, Bryant C, Jones AV *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet* 2010; **42**: 722–726.
10 Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 2009; **361**: 1058–1066.
11 Makishima H, Jankowska AM, Tiu RV, Szpurka H, Sugimoto Y, Hu Z *et al.* Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies. *Leukemia* 2010; **24**: 1799–1804.
12 Damm F, Itzykson R, Kosmider O, Droin N, Renneville A, Chesnais V *et al.* SETBP1 mutations in 658 patients with myelodysplastic syndromes, chronic myelomonocytic leukemia and secondary acute myeloid leukemias. *Leukemia* 2013; **27**: 1401–1403.
13 Makishima H, Yoshida K, Nguyen N, Przychodzen B, Sanada M, Okuno Y *et al.* Somatic SETBP1 mutations in myeloid malignancies. *Nat Genet* 2013; **45**: 942–946.
14 Gomez-Segui I, Makishima H, Jerez A, Yoshida K, Przychodzen B, Miyano S *et al.* Novel recurrent mutations in the RAS-like GTP-binding gene RIT1 in myeloid malignancies. *Leukemia* 2013; **27**: 1943–1946.
15 Visconte V, Makishima H, Jankowska A, Szpurka H, Traina F, Jerez A *et al.* SF3B1, a splicing factor is frequently mutated in refractory anemia with ring sideroblasts. *Leukemia* 2012; **26**: 542–545.
16 Makishima H, Visconte V, Sakaguchi H, Jankowska AM, Abu Kar S, Jerez A *et al.* Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood* 2012; **119**: 3203–3210.
17 Damm F, Kosmider O, Gelsi-Boyer V, Renneville A, Carbuccia N, Hidalgo-Curtis C *et al.* Mutations affecting mRNA splicing define distinct clinical phenotypes and correlate with patient outcome in myelodysplastic syndromes. *Blood* 2012; **119**: 3211–3218.
18 Hoskins AA, Moore MJ. The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem Sci* 2012; **37**: 179–188.
19 Maciejewski JP, Padgett RA. Defects in spliceosomal machinery: a new pathway of leukaemogenesis. *Br J Haematol* 2012; **158**: 165–173.
20 Grainger RJ, Beggs JD. *Prp8* protein: at the heart of the spliceosome. *RNA* 2005; **11**: 533–557.
21 Galej WP, Oubridge C, Newman AJ, Nagai K. Crystal structure of *Prp8* reveals active site cavity of the spliceosome. *Nature* 2013; **493**: 638–643.
22 Konforti BB, Konarska MM. U4/U5/U6 snRNP recognizes the 5′ splice site in the absence of U2 snRNP. *Genes Dev* 1994; **8**: 1962–1973.
23 Luo HR, Moreau GA, Levin N, Moore MJ. The human *Prp8* protein is a component of both U2- and U12-dependent spliceosomes. *RNA* 1999; **5**: 893–908.
24 Schellenberg MJ, Wu T, Ritchie DB, Fica S, Staley JP, Atta KA *et al.* A conformational switch in *PRP8* mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nat Struct Mol Biol* 2013; **20**: 728–734.
25 McKie AB, McHale JC, Keen TJ, Tarttelin EE, Goliath R, van Lith-Verhoeven JJ *et al.* Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum Mol Genet* 2001; **10**: 1555–1562.
26 Towns KV, Kipioti A, Long V, McKibbin M, Maubaret C, Vaclavik V *et al.* Prognosis for splicing factor PRPF8 retinitis pigmentosa, novel mutations and correlation between human and yeast phenotypes. *Hum Mutat* 2010; **31**: E1361–E1376.
27 Shaffer LG, Tommerup N. *ISCN 2009. An International System for Human Cytogenetics Nomenclature*. Karger: Basel, 2009.
28 Przychodzen B, Jerez A, Guinta K, Sekeres MA, Padgett R, Maciejewski JP *et al.* Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood* 2013; **122**: 999–1006.
29 Makishima H, Jankowska AM, McDevitt MA, O'Keefe C, Dujardin S, Cazzolli H *et al.* CBL, CBLB, TET2, ASXL1, and IDH1/2 mutations and additional chromosomal aberrations constitute molecular events in chronic myelogenous leukemia. *Blood* 2011; **117**: e198–e206.
30 Khan SN, Jankowska AM, Mahfouz R, Dunbar AJ, Sugimoto Y, Hosono N *et al.* Multiple mechanisms deregulate EZH2 and histone H3 lysine 27 epigenetic changes in myeloid malignancies. *Leukemia* 2013; **27**: 1301–1309.
31 Lesser CF, Guthrie C. Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics* 1993; **133**: 851–863.
32 Query CC, Konarska MM. Suppression of multiple substrate mutations by spliceosomal *Prp8* alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol Cell* 2004; **14**: 343–354.
33 Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; **11**: 377–394.
34 Liu L, Query CC, Konarska MM. Opposing classes of *Prp8* alleles modulate the transition between the catalytic steps of pre-mRNA splicing. *Nat Struct Mol Biol* 2007; **14**: 519–526.
35 Keightley MC, Crowhurst MO, Layton JE, Beilharz T, Markmiller S, Varma S *et al.* In vivo mutation of pre-mRNA processing factor 8 (Prpf8) affects transcript splicing, cell survival and myeloid differentiation. *FEBS Lett* 2013; **587**: 2150–2157.
36 Visconte V, Makishima H, Maciejewski JP, Tiu RV. Emerging roles of the spliceosomal machinery in myelodysplastic syndromes and other hematological disorders. *Leukemia* 2012; **26**: 2447–2454.
37 Soenen V, Preudhomme C, Roumier C, Daudignon A, Lai JL, Fenaux P. 17p Deletion in acute myeloid leukemia and myelodysplastic syndrome. Analysis of breakpoints and deleted segments by fluorescence in situ. *Blood* 1998; **91**: 1008–1015.
38 Allende-Vega N, Dayal S, Agarwala U, Sparks A, Bourdon JC, Saville MK. p53 is activated in response to disruption of the pre-mRNA splicing machinery. *Oncogene* 2012; **32**: 1–14.
39 Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 2008; **134**: 112–123.
40 McKenzie M, Tucker EJ, Compton AG, Lazarou M, George C, Thorburn DR *et al.* Mutations in the gene encoding C8orf38 block complex I assembly by inhibiting production of the mitochondria-encoded subunit ND1. *J Mol Biol* 2011; **414**: 413–426.
41 Yoshikumi Y, Mashima H, Ueda N, Ohno H, Suzuki J, Tanaka S *et al.* Roles of CTPL/Sfxn3 and Sfxn family members in pancreatic islet. *J Cell Biochem* 2005; **95**: 1157–1168.
42 Ye X, Xu J, Cheng C, Yin G, Zeng L, Ji C *et al.* Isolation and characterization of a novel human putative anemia-related gene homologous to mouse sideroflexin. *Biochem Genet* 2003; **41**: 119–125.
43 Lindhurst MJ, Fiermonte G, Song S, Struys E, De Leonardis F, Schwartzberg PL *et al.* Knockout of Slc25a19 causes mitochondrial thiamine pyrophosphate depletion, embryonic lethality, CNS malformations, and anemia. *Proc Natl Acad Sci USA* 2006; **103**: 15927–15932.
44 Spiegel R, Shaag A, Edvardson S, Mandel H, Stepensky P, Shalev SA *et al.* SLC25A19 mutation as a cause of neuropathy and bilateral striatal necrosis. *Ann Neurol* 2009; **66**: 419–424.
45 Badhai J, Frojmark AS, E JD, Schuster J, Dahl N. Ribosomal protein S19 and S24 insufficiency cause distinct cell cycle defects in Diamond-Blackfan anemia. *Biochim Biophys Acta* 2009; **1792**: 1036–1042.
46 Boria I, Quarello P, Avondo F, Garelli E, Aspesi A, Carando A *et al.* A new database for ribosomal protein genes which are mutated in Diamond-Blackfan Anemia. *Hum Mutat* 2008; **29**: E263–E270.
47 Collins CA, Guthrie C. Allele-specific genetic interactions between *Prp8* and RNA active site residues suggest a function for *Prp8* at the catalytic core of the spliceosome. *Genes Dev* 1999; **13**: 1970–1982.
48 House AE, Lynch KW. Regulation of alternative splicing: more than just the ABCs. *J Biol Chem* 2008; **283**: 1217–1221.

49 Umen JG, Guthrie C. A novel role for a U5 snRNP protein in 3′ splice site selection. *Genes Dev* 1995; **9**: 855–868.
50 Badhai J, Frojmark AS, Razzaghian HR, Davey E, Schuster J, Dahl N. Posttranscriptional down-regulation of small ribosomal subunit proteins correlates with reduction of 18S rRNA in RPS19 deficiency. *FEBS Lett* 2009; **583**: 2049–2053.
51 Yang K, Zhang L, Xu T, Heroux A, Zhao R. Crystal structure of the beta-finger domain of *Prp8* reveals analogy to ribosomal proteins. *Proc Natl Acad Sci USA* 2008; **105**: 13817–13822.
52 Hahn D, Kudla G, Tollervey D, Beggs JD. Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. *Genes Dev* 2012; **26**: 2408–2421.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)

Cancer Science
Japanese Cancer Association JCA
Open Access

High-resolution genomic copy number profiling of primary intraocular lymphoma by single nucleotide polymorphism microarrays

Ludan Wang,[1] Aiko Sato-Otsubo,[2,3] Sunao Sugita,[4,5] Hiroshi Takase,[4] Manabu Mochizuki,[4] Yoshihiko Usui,[6] Hiroshi Goto,[6] Takatoshi Koyama,[7] Hiroki Akiyama,[1] Osamu Miura,[1] Seishi Ogawa[2,3] and Ayako Arai[1]

[1]Department of Hematology, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo; [2]Cancer Genomics Project, University of Tokyo, Tokyo; [3]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto; [4]Department of Ophthalmology and Visual Science, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo; [5]Laboratory for Retinal Regeneration, Riken Center for Developmental Biology, Kobe; [6]Department of Ophthalmology, Tokyo Medical University, Tokyo; [7]Department of Laboratory Molecular Genetics of Hematology, Graduate School of Health Care Sciences, Tokyo Medical and Dental University, Tokyo, Japan

Key words

Central nervous system, IL-10, intraocular lymphoma, single nucleotide polymorphism, vitreous body

Correspondence

Ayako Arai, Department of Hematology, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo, Japan.
Tel: +81-3-3813-6111; Fax: +81-3-5803-0131;
E-mail: ara.hema@tmd.ac.jp

Funding information

Ministry of Education, Culture, Sports, Science, and Technology of Japan (23591375). Japan Leukemia Research Fund.

Received September 17, 2013; Revised February 18, 2014; Accepted February 22, 2014

Cancer Sci **105** (2014) 592–599

doi: 10.1111/cas.12388

Primary intraocular lymphoma (PIOL) is a rare lymphoma. Because of difficulties in obtaining tissue samples, little is known about the disease's genetic features. In order to clarify these features, we carried out single nucleotide polymorphism array karyotyping of IOL using genomic DNA extracted from vitreous fluid. We analyzed 33 samples of IOLs consisting of 16 PIOLs, 12 IOLs with a central nervous system (CNS) lesion at diagnosis (IOCNSL), and five secondary IOLs following systemic lymphoma. All were B-cell type. We identified recurrent copy number (CN) gain regions in PIOLs, most frequently on chromosome 1q followed by 18q and 19q. Chromosome 6q was the most frequent loss region. Although these CN gain regions of PIOL were in common with those of IOCNSL, loss of 6q22.33 containing *PTPRK* and 9p21.3 containing *CDKN2A* were more frequently deleted in IOCNSL. Large CN loss in 6q was detected in three of four PIOL patients who had early CNS development and short survival periods, whereas long-term survivors did not have such deletions. There was a correlation between gain of the *IL-10* gene located on 1q and intravitreal interleukin-10 concentration, which was higher in IOL than in benign uveitis. The results suggest that IOCNSL is a highly malignant form of PIOL that infiltrates into the CNS at an early stage. They also indicate that genetic differences between PIOL and primary CNS lymphoma need to be clarified.

Primary intraocular lymphoma (PIOL) is a lymphoma whose lesion is located in the eye ball, including the vitreous body, the retina, the choroid, the iris, and the ciliary muscle. It is a rare form of non-Hodgkin's lymphoma, accounting for less than 1% of all non-Hodgkin's lymphoma cases[1] and of all intraocular neoplasms.[2] The majority of PIOL are vitreoretinal lymphoma whose lesions are found exclusively in the retina and the vitreous body of the eye, resulting in the primary symptoms of blurred vision and decreased visual acuity.[3] Ophthalmoscopic findings are a cloudy vitreous body and/or subretinal proliferative lesions due to tumor cell infiltration.[3,4] Most PIOLs are B-cell lymphomas; T-cell lymphomas may occur, but are extremely rare.[3]

The diagnosis of IOL is quite difficult due to several reasons. First, it is hard to obtain sufficient biopsy material from the lesions. Retinal biopsy may cause visual field defects, and enucleation inflicts irreversible damage to visual acuity. Therefore, vitrectomy with vitreous sampling is used to obtain the specimen. Diagnosis is usually made by cytology for samples of vitreous fluid, although the detection rate is low (20–44.5%) because of degradation of the infiltrating cells or lack of cells in the samples.[5,6] Polymerase chain reaction analysis using DNA extracted from vitreous fluid has been shown to have a much higher detection rate and has become the most persuasive procedure for the diagnosis of IOL.[7] Using this method, we detected IgH rearrangements in 21 of 22 patients with IOL (95.5%).[5] Diagnosis can be supported by the interleukin-10 (IL-10) : IL-6 cytokine ratio in the vitreous fluid.[6,8,9] We also showed that 86% patients with an IL-10 : IL-6 ratio of >1.0 had IOL, whereas the majority of patients with benign uveitis had a ratio of <1.0.[5] Using these methods, the accurate rate of IOL diagnosis has improved dramatically, and consequently, the number of PIOL patients has been increasing.

Many issues, however, relating to PIOL remain unresolved. According to previous reports, primary vitreoretinal lymphomas, which are the major types of PIOLs, are reported to be diffuse large B-cell lymphomas (DLBCL).[10–12] In the clinical situation, however, the histological types of PIOLs are often undetermined because the material for diagnosis is usually vitreous fluid. Furthermore, the genetic features of PIOL have not yet been elucidated. In order to solve them, we carried out

Table 1. Clinical information of IOL patients

No	Age	Gender	PS	Extra-ocular lesions at diagnosis of IOL				Initial treatment	Clinical course (time after diagnosis)	Survival period (after diagnosis)
				Brain involvement	CSF infiltration	Systemic lesions	BM infiltration			
PIOL-1	83	M	0	(–)	(–)	(–)	(–)	VT, ivMTX	CNS development (4 Ms), Dead	6 Ms
PIOL-2	71	M	0	(–)	(–)	(–)	(–)	VT, ORT	CNS development (11 Ms), Alive	>17 Ms
PIOL-3	77	F	0	(–)	(–)	(–)	(–)	VT, WBRT	LFU	Unknown
PIOL-4	57	F	0	(–)	(–)	(–)	(–)	VT, sMTX, WBRT	CNS development (30 Ms), Dead	48 Ms
PIOL-5	80	M	0	(–)	(–)	(–)	ND	VT, enucleation	LFU	Unknown
PIOL-6	70	F	0	(–)	(–)	(–)	ND	VT, ivMTX	Alive	>36 Ms
PIOL-7	65	F	1	(–)	ND	(–)	ND	VT, itMTX	CNS development (ND), Dead	26 Ms
PIOL-8	82	F	0	(–)	ND	(–)	ND	VT, ivMTX	Alive	>52 Ms
PIOL-9	85	F	0	(–)	ND	(–)	ND	VT, ivMTX	LFU	Unknown
PIOL-10	68	F	0	(–)	ND	(–)	(–)	VT	CNS development (6 Ms), Dead	8 Ms
PIOL-11	74	F	0	(–)	(–)	(–)	ND	VT	CNS development (40 Ms), Alive	>53 Ms
PIOL-12	82	M	1	(–)	ND	(–)	ND	VT, ivMTX	CNS development (4 Ms), Dead	18 Ms
PIOL-13	51	F	0	(–)	ND	(–)	ND	VT	Alive	>53 Ms
PIOL-14	55	M	0	(–)	(–)	(–)	(–)	VT, ivMTX	CNS development (3 Ms), Dead	12Ms
PIOL-15	79	F	1	(–)	ND	(–)	ND	VT, ORT	LFU	Unknown
PIOL-16	57	M	0	(–)	ND	(–)	ND	VT, ORT	Alive	>70 Ms
IOCNSL-1	70	F	1	(+)	(–)	(–)	(–)	VT, enucleation, sMTX	Alive in CR	>27 Ms
IOCNSL-2	45	F	0	(+)	ND	(–)	ND	VT, WBRT	LFU	Unknown
IOCNSL-3	52	M	0	(+)	ND	(–)	(–)	VT, WBRT, sCT	LFU	Unknown
IOCNSL-4	47	M	0	(+)	(–)	(–)	(–)	VT, WBRT, sCT	Alive	>36 Ms
IOCNSL-5	47	M	0	(+)	(–)	(–)	(–)	VT, WBRT, sCT	Alive	>36 Ms
IOCNSL-6	66	M	3	(+)	ND	(–)	ND	VT, ivMTX	LFU	Unknown
IOCNSL-7	52	M	4	(+)	ND	(–)	ND	VT, ivMTX	PD, Dead	29 Ms
IOCNSL-8	54	M	3	(+)	ND	(–)	(–)	VT, ivMTX, sCT	PD, Dead	8 Ms
IOCNSL-9	61	M	1	(+)	ND	(–)	ND	VT, WBRT, sMTX	LFU	Unknown
IOCNSL-10	66	M	0	(+)	(–)	(–)	(–)	VT, ivMTX, sCT	Alive	>69 Ms
IOCNSL-11	70	M	1	(+)	(–)	(–)	(–)	VT, ivMTX	PD, Dead	14 Ms
IOCNSL-12	88	F	2	(+)	(–)	(–)	(–)	VT, ORT, sMT	LFU	Unknown
SIOL-1	75	M	0	(–)	ND	(–)	ND	VT	LFU	Unknown
SIOL-2	73	M	0	(–)	(–)	(–)	(–)	VT, sMTX	LFU	Unknown
SIOL-3	45	M	0	(–)	(–)	(–)	(–)	VT	Alive	>70 Ms
SIOL-4	69	M	0	(–)	(–)	(–)	(–)	VT, sMTX	CNS development (34 Ms), Alive	>99 Ms
SIOL-5	68	F	ND	(–)	ND	(–)	ND	ND	LFU	Unknown

+, Present; –, not present; BM, bone marrow; CNS, central nervous system; CR, complete response; CSF, cerebrospinal fluid; F, female; IOCNSL, IOL with central nervous system lesion at diagnosis; itMTX, intrathechal methotrexate injection; ivMTX, intravitreal methotrexate injection; LFU, lost to follow-up; M, male; Ms, months; ND, not described; ORT, ocular radiation therapy; PD, progressive disease; PIOS, primary IOS; PS, performance status, according to Eastern Cooperative Oncology Group criteria; sCT, systemic chemotherapy; SIOL, secondary IOL; sMTX, systemic methotrexate injection; VT, vitrectomy; WBRT, whole brain radiation therapy.