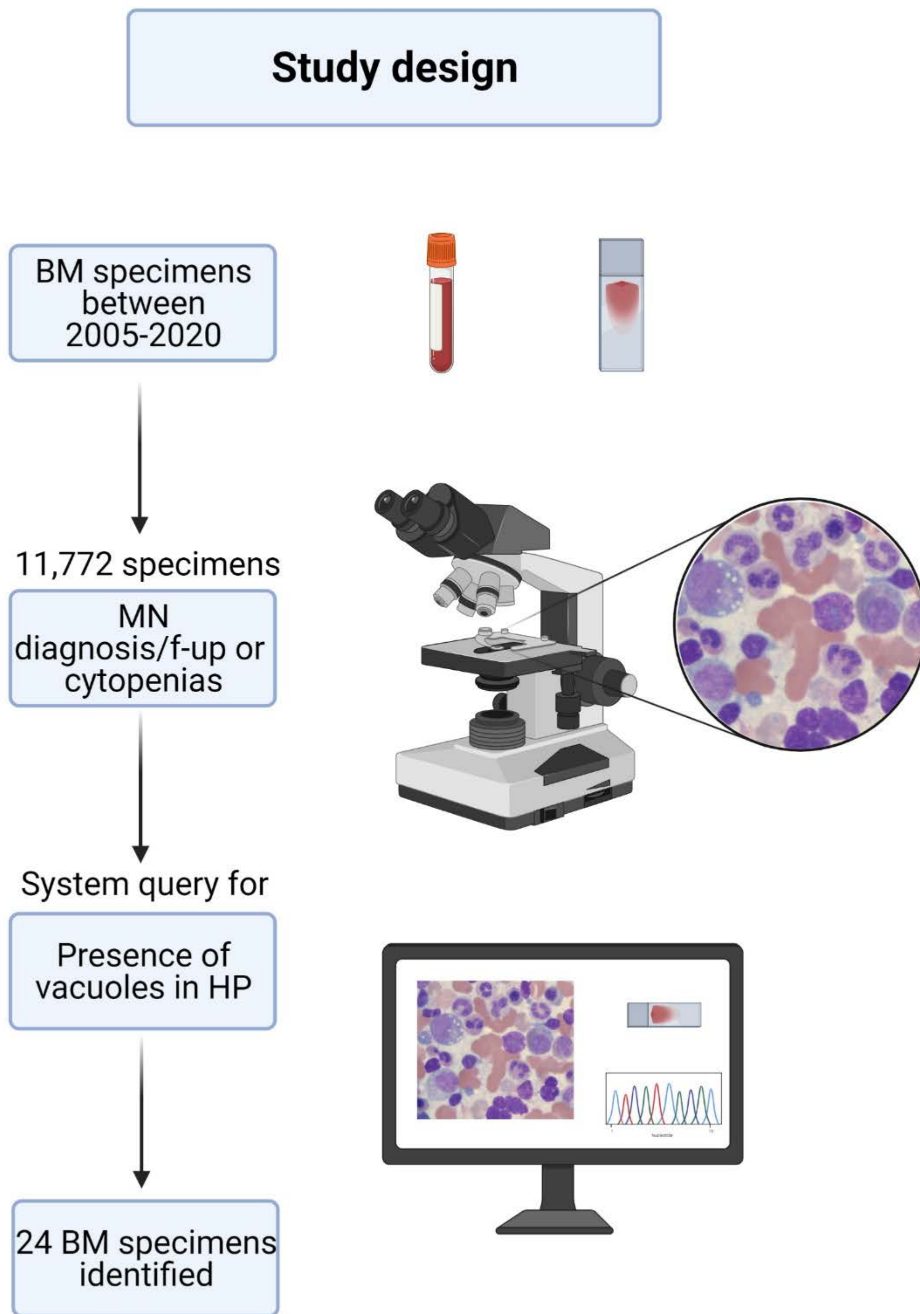
Supplementary Appendix

METHODS

Patients. Clinical data were collected from medical charts of patients enrolled at The Cleveland Clinic Foundation between 2005 and 2020 in accordance to the protocols and written consent approved by appropriate IRBs and the Declaration of Helsinki. A total of 11,772 bone marrow (BM) specimens from patients undergoing evaluation for diagnosis and/or follow-up of a myeloid neoplasia or unexplained cytopenias were selected for the purpose of this study. After launching a system query for presence of vacuoles in hematopoietic precursors reported in the patients' chart, we identified a total of 24 cases. These patients were studied for clinical and laboratory characteristics and also subjected to search for frozen cell specimens, paraffin blocks and/or available DNA specimens for DNA sequencing studies (see appropriate section below) (**Supplemental Figure 1**).

Supplementary Figure 1. Study conceptualization and design. BM: bone marrow; MN: myeloid neoplasms; HP: hematopoietic precursors; f-up: follow-up



DNA Sequencing studies. *Deep targeted sequencing:* For the detection of somatic mutations of the myeloid genes (**Table S1**), DNA was extracted from whole blood or mononuclear cells and subjected to targeted sequencing. Mutations were annotated using Annovar and their somatic status was called using an in-house bio-analytic pipeline (1-3), by comparison with sequences of healthy subjects and databases such as dbSNP 138, 1000 Genomes Project, NHLBI Exome Sequencing Project 6500, Exome Aggregation Consortium (ExAC) and genome aggregation database (gnomAD). *Whole exome sequencing* (*WES*)*:* WES was performed as previously described. (4) Paired tumor and germline DNAs were used. Whole-exome capture was accomplished through hybridization of sonicated genomic DNA to a bait cDNA library synthesized on magnetic beads (SureSelect Human All Exon 50 Mb or V4 kit, Agilent Technologies). Captured targets were subjected to massively parallel sequencing using a HiSeq 2000 (Illumina) according to the standard protocol for 100-bp paired-end reads. Sequencing reads were aligned to the human genome 19 and a GATK pipeline was used to extract variants/ polymorphisms and to remove sequencing errors. (4) *Sanger sequencing:* DNA was extracted from whole blood (Nuclei Lysis Solution, Promega) and/or paraffin blocks (QIAamp DNA FFPE Tissue Kit, Qiagen). Twenty ng was used as source for amplification through polymerase chain reaction (PCR) using primers (forward primer 5'-CCCCTCTTTGCTGTAAAATG-3'; reverse primer 5'-CTCATGGCCCAACACATACC-3') covering exonic regions containing methionine 41 of *UBA1* gene. Primers were adopted by https://genome.ucsc.edu/. from The PCR amplification consisted of an initial denaturation for 3 min at 94°C, followed by 35 cycles of 30 sec denaturation at 94°C, annealing for 30 sec at 60°C, and extension for 1 min at 72°C. The final extension was carried out for 7 min at 72°C. Direct sequencing was performed with 8 μL of the purified PCR product and 4 μL of forward primer and validated using reverse primer (Eurofins Genomics LLC Chromatograms traces were viewed and analyzed using FinchTV (Geospiza, Inc).

References

1) Makishima H, Yoshizato T, Yoshida K, et al. Dynamics of clonal evolution in myelodysplastic syndromes. *Nat Genet.* 2017; 49: 204-12.

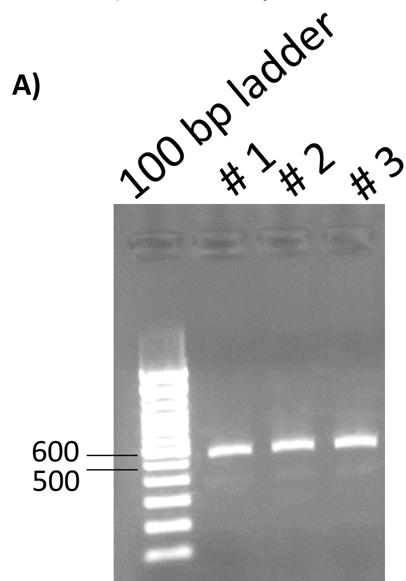
2) Hirsch CM, Nazha A, Kneen K, et al. Consequences of mutant TET2 on clonality and subclonal hierarchy. *Leukemia*. 2018; 32:1751-1761.

3) Nagata Y, Makishima H, Kerr CM, et al. Invariant patterns of clonal succession determine specific clinical features of myelodysplastic syndromes. *Nat Commun*. 2019; 10:5386.

4) Makishima, H, Yoshida K, Nguyen, N et al. Somatic SETBP1 mutations in myeloid malignancies. *Nat. Genet*. 2013; 45, 942–946.

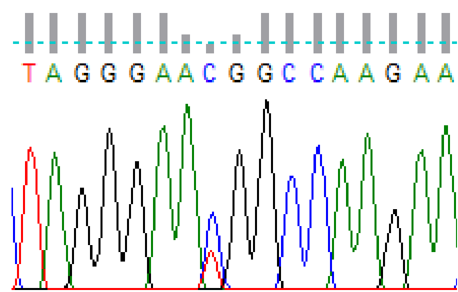
Supplementary Figure 2. Sanger sequencing results. A) Gel Electrophoresis (1.5% agarose) of coding region containing methionine 41. Three representative cases (#1 and 2: *UBA1* mutant; #3: *UBA1* wild type); B) Chromatograms showing forward sequences of two patients with *UBA1* variant (c.122 T>C, p.Met41Thr) and one wild type patient.

A)

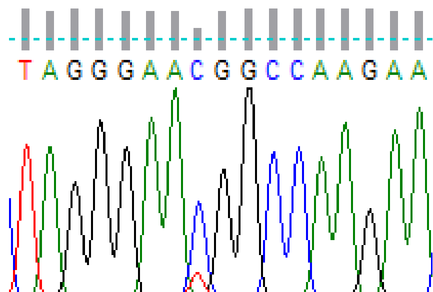


B)

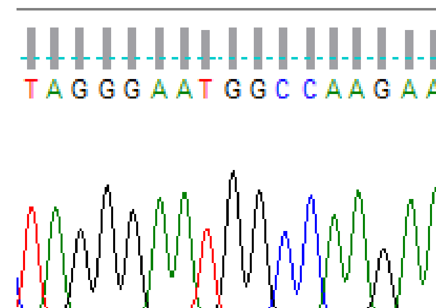
c.122 T>C; p.Met41Thr



c.122 T>C; p.Met41Thr

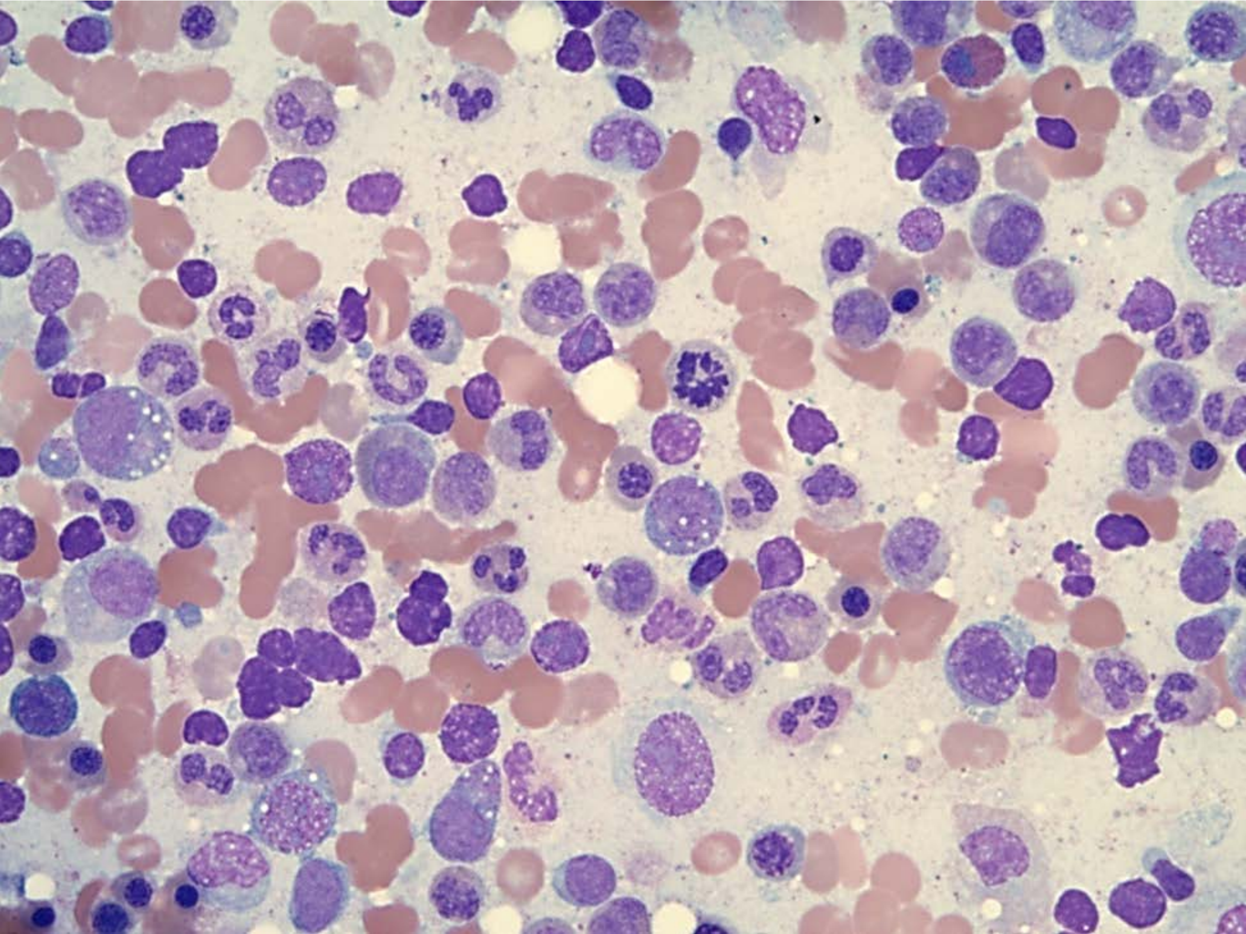


Wild type

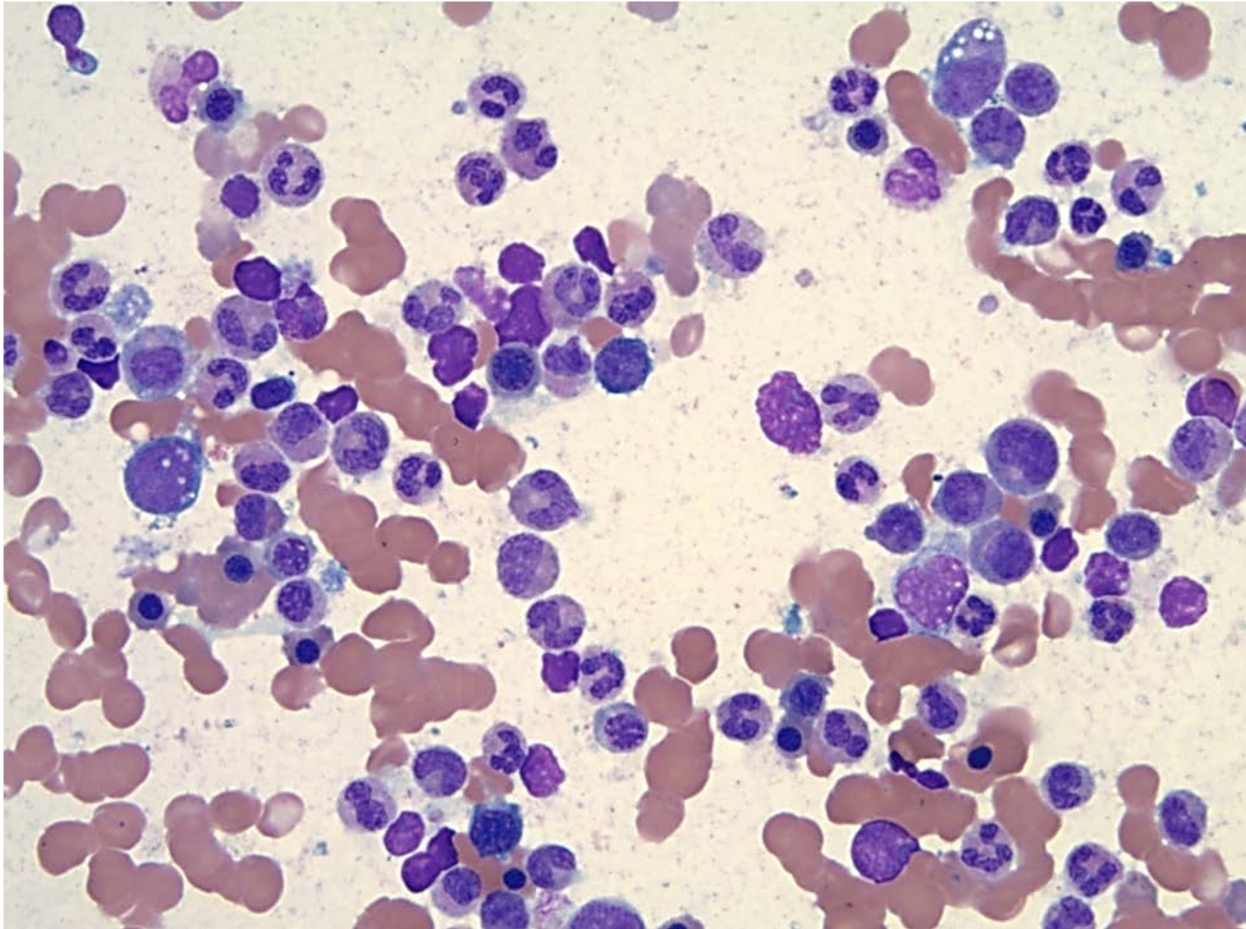


Supplementary Figure 3. Bone marrow aspiration pictures of 3 exemplificative patients showing a vision at a glance of the vacuolization phenomenon (500x Wright stain).

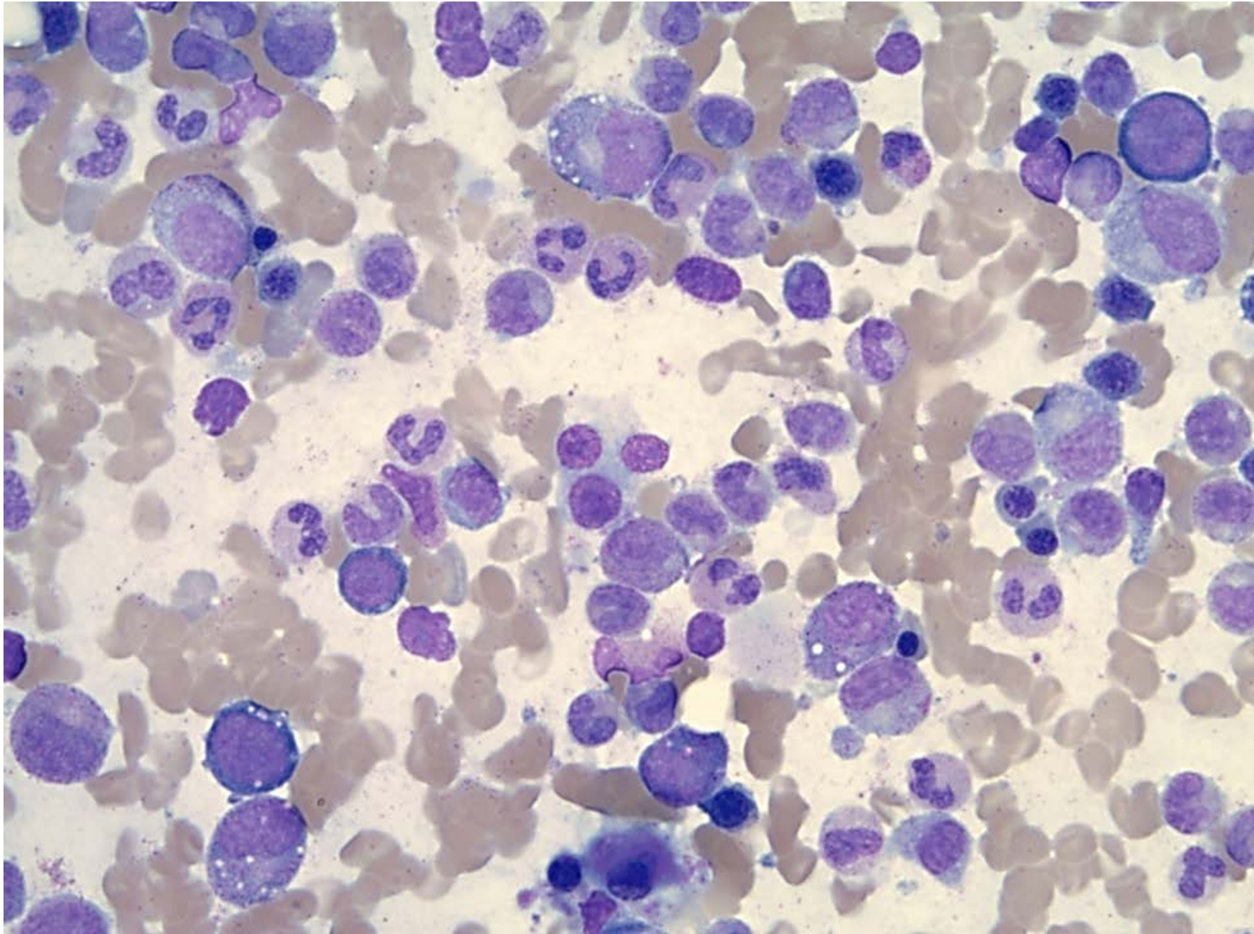
A



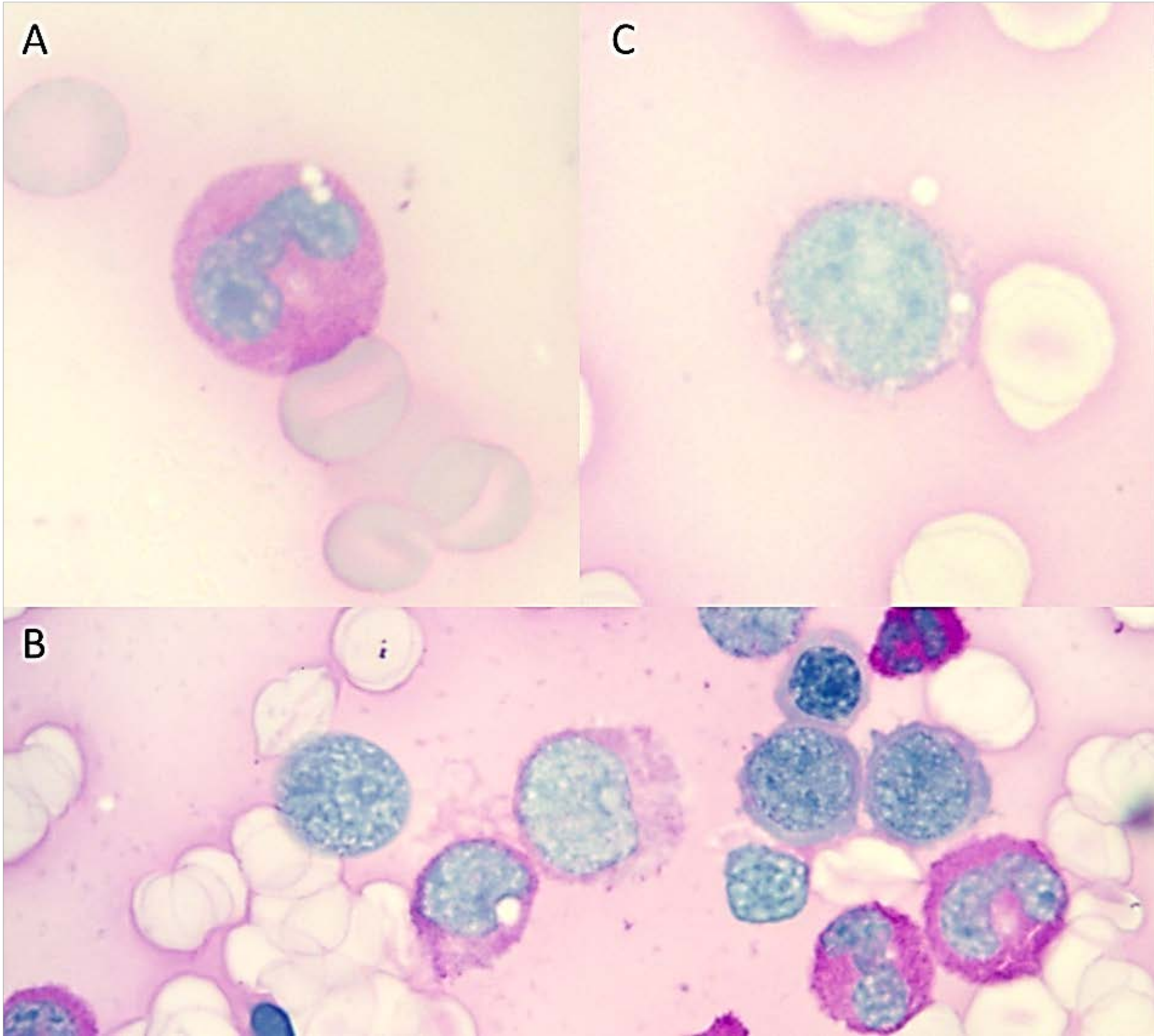
B



C



Supplementary Figure 4. Periodic acid–Schiff (PAS) staining on bone marrow smears of 2 *UBA1* positive (A-B) and 1 *UBA1* negative patients (C) illustrating PAS negative vacuoles.



Supplementary Table 1. List of 63 genes included in the targeted DNA sequencing panel of The Cleveland Clinic

Gene list										
ABL1	*CBL*	*DNMT3A*	*FLT3*	*IKZF1*	*KRAS*	*NOTCH1*	*PRPF8*	*SETBP1*	*STAG2*	*U2AF1*
ASXL1	*CDKN2A*	*EED*	*GATA1*	*JAK2*	*LUC7L2*	*NPM1*	*PTEN*	*SF3B1*	*STAT3*	*WT1*
BCOR	*CEBPA*	*ETKN1*	*GATA2*	*JAK3*	*MECOM*	*NRAS*	*PTPN11*	*SH2B3*	*STAT5B*	*ZRSR2*
BCORL1	*CSF3R*	*ETV6*	*GNAS*	*KDM6A*	*MPL*	*PHF6*	*RAD21*	*SMC1A*	*SUZ12*	
BRAF	*CUX1*	*EZH2*	*IDH1*	*KIT*	*MYD88*	*PIGA*	*RIT1*	*SMC3*	*TET2*	
CALR	*DDX41*	*FBXW7*	*IDH2*	*KMT2A*	*NF1*	*PPM1D*	*RUNX1*	*SRSF2*	*TP53*	

Supplementary Table 2. Somatic mutations.

UPN	Genetic alterations (VAF %)[£]	Protein Change
1	DNMT3A (26%), UBA1[¥]	p.R882H, p.M41T
2	DNMT3A (4%)	p.W601*
3	NA	NA
4	TP53 (34%)	p.R248W
5	ASXL1 (52%), U2AF1 (9%)	p.A1312V, p.Q157R
6	UBA1[¥]	p.M41T
7	ASXL1 (36%), IDH2 (19%), SF3B1 (21%), TET2 (4%), TP53 (23%)	p.L775*, p.R140Q, p.R625C,p.L1312Rfs*3, p.R196*
8	WT	-
9	JAK2 (36%)	p.V617F
10	21q21.1-q22.3 LOH[$]	NA
11	NA	NA
12	WT	-
13	NA	NA
14	CUX1 (73%)	p.Q808L
15	NA	NA
16	NA	NA
17	NA	NA
18	BCOR (46%)	p.V679I
19	IDH1 (43%), RUNX1 (39%)	p.R132C, p.A56fs
20	NA	NA
21	DNMT3A (34%)	p.S618L
22	DDX41 (45%), DNMT3A (14%), PHF6 (15%)	p.R525H, p.C394R, p.C20fs
23	DNMT3A (44%), JAK2 (8%), TET2 (46%), TET2 (43%)	p.536_536del; p.V617F; p.Q530X, p.H762fs
24	NA	NA

[£]Somatic variants are reported at the time of collection of bone marrow specimens.

[¥]*UBA1* mutations were detected by Sanger sequencing.

[$]This patient had a copy neutral loss of heterozygosity (LOH) involving most of the long arm of chromosome 21 (21q21.1-q22.3) in a small but significant percentage of the cell population. This region would encompass *RUNX1* gene and the lesion is classified as likely pathogenic.

Supplementary Table 3. List of possible disorders with vacuolization of hematopoietic precursors

Myelodysplastic syndromes
Acute myeloid leukemia
Alcoholic abuse
Pearson disease
Hemophagocytic lymphohistiocytosis
Copper deficiency
Protein-losing enteropathies
Malnourishment
Zinc-toxicity
Bariatric or upper gastrointestinal surgery
VEXAS syndrome