GUIDE TO SUPPLEMENTARY INFORMATION

FOR

Age-related cancer mutations associated with clonal hematopoietic expansion

Mingchao Xie[1,2]*, Charles Lu[1]*, Jiayin Wang[1,2]*, Michael D. McLellan[1], Kimberly J. Johnson[3], Michael C. Wendl[1,4,5], Joshua F. McMichael[1], Heather K. Schmidt[1], Venkata Yellapantula[1,2], Christopher A. Miller[1], Bradley A. Ozenberger[1,2], John S. Welch[2,6], Daniel C. Link[2,6], Matthew J. Walter[2,6], Elaine R. Mardis[1,2,4,6], John F. Dipersio[2,6], Feng Chen[2,6], Richard K. Wilson[1,2,4,6], Timothy J. Ley[1,2,4,6], Li Ding[1,2,4,6#]

[1]The Genome Institute, [2]Department of Medicine, [4]Department of Genetics, [5]Department of Mathematics, [6]Siteman Cancer Center, Washington University in St. Louis, MO 63108, USA
[3]Brown School Master of Public Health Program, Washington University in St. Louis, St. Louis, MO 63130, USA

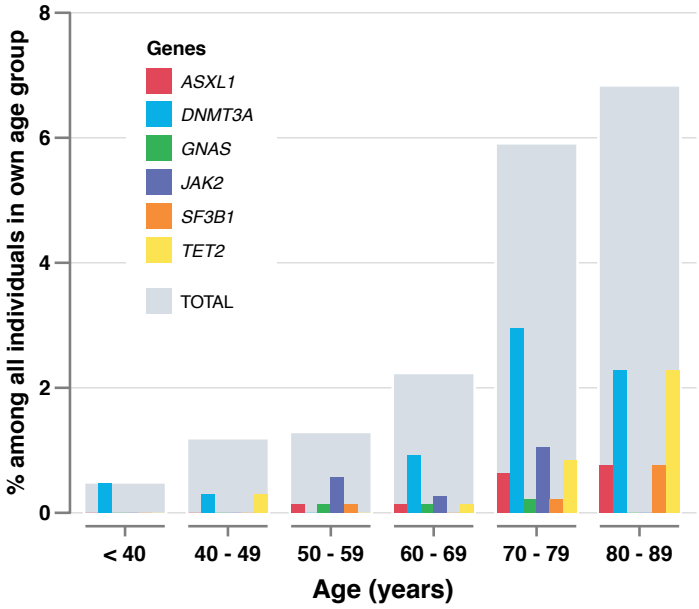
*These authors contributed equally to this work.

Corresponding Author:
Li Ding, Ph.D.
The Genome Institute
Division of Oncology, Department of Medicine
Washington University School of Medicine
St. Louis, MO 63108
Email: lding@genome.wustl.edu

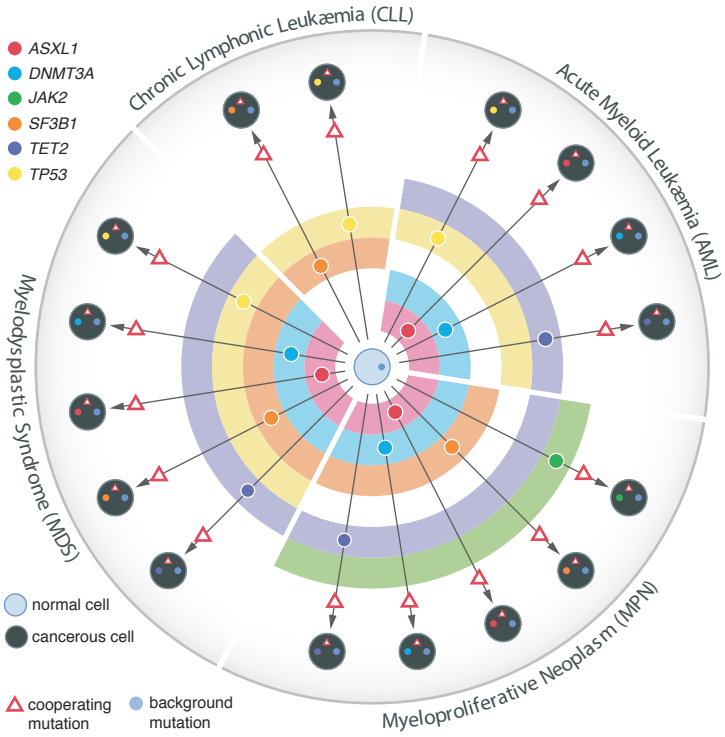
Journal: Nature Medicine

Article Title:	Age-related cancer mutations associated with clonal hematopoietic expansion
Corresponding Author:	Dr. Li Ding

Supplementary Item & Number (add rows as necessary)	**Title or Caption**
Supplementary Figure 1	Distribution of blood-specific mutations in *DNMT3A*, *TET2*, *JAK2*, *ASXL1*, *SF3B1*, *GNAS*, and all 31 genes across different age groups. The 91 sites include 77 detected by our processing pipeline and 14 low VAF sites (2 to 10%) identified by read count-based analysis. The total includes all blood-specific mutations in 556 cancer associated genes identified in each age group.
Supplementary Figure 2	Distinct and common connections among normal blood samples, MPN, MDS, CLL, and AML cases. A combination of precursor, initiating mutations in the normal blood samples may rarely collaborate with subsequent, progression mutations to develop MPN, MDS, CLL, and/or AML.
Supplementary Table 1a	Sample IDs for the 2,728 TCGA cases included in this study.
Supplementary Table 1b	Samples included in the study and their clinical characteristics.
Supplementary Table 1c	The distribution of germline variants across 2,728 TCGA samples. TCGA Ovarian counts were collected from the previous report.
Supplementary Table 2	Somatic mutations in 2,241 TCGA tumor samples included in the study. Somatic mutation data are unavailable for a subset of samples.
Supplementary Table 3a	Somatic mutations in 3,355 TCGA tumor samples from 12 cancer types used for identifying recurrent mutations.
Supplementary Table 3b	Recurrent somatic mutations from 12 TCGA cancer types used for hotspot analysis.
Supplementary Table 4	556 cancer-associated genes used in this study.
Supplementary Table 5a	77 blood-specific events detected in 2,728 cases using our standard discovery pipeline.
Supplementary Table 5b	Low-level blood-specific events detected in *DNMT3A*, *JAK2*, *SF3B1*, *GNAS*, and *IDH2* in TCGA samples.
Supplementary Table 5c	Deep-sequencing based validation of low-level blood-specific events detected in *DNMT3A*, *JAK2*, and *SF3B1* in TCGA samples.
Supplementary Table 6	Truncation and hotspot variants in four prominent genes (*DNMT3A*, *TET2*, *JAK2*, and *ASXL1*) involved in HSPC clonal expansion in 6,503 ESP samples.
Supplementary Table 7	Rare truncation variants and known hotspot variants detected in *DNMT3A*, *TET2*, *ASXL1*, *GNAS*, *JAK2*, *SF3B1*, *IDH1*, and *IDH2* in 557 WHISP samples.
Supplementary Table 8	Exome capture sequencing coverage for 11 TCGA cancer types analyzed.



Supplementary Figure 1 Distribution of blood-specific mutations in *DNMT3A*, *TET2*, *JAK2*, *ASXL1*, *SF3B1*, *GNAS*, and all 31 genes across different age groups. The 91 sites include 77 detected by our processing pipeline and 14 low VAF sites (2 to 10%) identified by read count-based analysis. The total includes all blood-specific mutations in 556 cancer associated genes identified in each age group.



Supplementary Figure 2 Distinct and common connections among normal blood samples, MPN, MDS, CLL, and AML cases. A combination of precursor, initiating mutations in the normal blood samples may rarely collaborate with subsequent, progression mutations to develop MPN, MDS, CLL, and/or AML.