Genomics of therapy-related myeloid neoplasms

Teodora Kuzmanovic,[1] Bhumika J. Patel,[1] Srinivasa R. Sanikommu,[1] Yasunobu Nagata,[1] Hassan Awada,[1] Cassandra M. Kerr,[1] Bartlomiej P. Przychodzen,[1] Babal K. Jha,[1] Devendra Hiwase,[2] Deepak Singhal,[2] Anjali S. Advani,[3] Aziz Nazha,[3] Aaron T. Gerds,[3] Hetty E. Carraway,[3] Mikkael A. Sekeres,[3] Sudipto Mukherjee,[3] Jaroslaw P. Maciejewski[1,3,#] and Tomas Radivoyevitch[1,4,#]*

#These authors are co-senior authors

[1]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA; [2]Department of Haematology, Royal Adelaide Hospital, Port Road, Adelaide, Australia; [3]Leukemia Program, Department of Hematologic Oncology and Blood Disorders, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH and [4]Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

Correspondence: JAROSLAW P. MACIEJEWSKI - maciejj@ccf.org

doi:10.3324/haematol.2019.219352

Supplementary Appendix

Content

Myelodysplastic/Myeloproliferative Neoplasm, unclassifiable; RARS-T, MDS/MPN with ring sideroblasts and thrombocytosis

Figure S8. Categorization of AML cases by 2016 WHO criteria. Categories include: Undifferentiated; inv(16), AML with inversion 16; inv(3), AML with inversion 3; NPM1+, AML with mutated *NPM1*; RUNX1+, AML with mutated *RUNX1*; MDS-related changes, AML with myelodysplasia-related changes; t(6;9), AML with translocation (6;9); t(8;21), AML with translocation (8;21); t(9;11); AML with translocation (9;11); NOS, AML not otherwise specified; APL, APL with *PML-RARA*.

Figure S9. Overall Survival of *EZH2* and *TP53*-mutated tMN patients. (A) Effects of disease group, DNA change, and domain mutated in *EZH2*-mutated patients. **(B)** Effects of disease group, DNA change, domain mutated, clonal hierarchy, and number of mutations for *TP53*-mutated patients.

Figure S10. Relationship between pMN and CHIP. Frequencies of ancestral mutations in pMN (top, dark blue) versus frequencies of mutations found in CHIP (bottom, light blue).

Figure S11. Relationship between sMN and CHIP. Frequencies of ancestral mutations in sMN (top, dark green) versus frequencies of mutations found in CHIP (bottom, light green).

Figure S12. Estimation of Mutation Derivation. In order to be CHIP-derived the frequency of a gene mutated in CHIP> frequency as an ancestral event in Myeloid Neoplasms. In order to be *de novo*-derived, the gene mutated was not found in CHIP, but was found in Myeloid Neoplasm. To be classified as Ambiguous, the gene mutated is found in both CHIP and Myeloid Malignancies. tMN are noted in red, sMN in green, and pMN in blue.

Figure S13. Schematic of origin of CHIP mutations in tMN. (A) In cases where CHIP was not present at initiation of therapy for primary malignancy. **(B)** In cases where CHIP was present prior to initiation of therapy for primary malignancy. Tx, cytotoxic therapy (chemotherapy and/or radiation); HSC, hematopoietic stem cell.

1. Supplementary Materials and Methods

1.1 Multi-amplicon deep sequencing

Multi-amplicon deep sequencing (TruSeq; Illumina) was performed for 70 gene targets, according to the manufacturer's instructions (Illumina) (**Supplementary Figure 1; Supplementary Table 1**). TruSeq custom amplicon generation protocol was applied to customized probe sets to amplify target exons and whole exons of target genes. The sequencing libraries were generated according to an Illumina pair-end library protocol and subjected to deep sequencing on MiSeq (Illumina) sequencers according to the standard protocol.[1-4] Subsequent validation and confirmatory sequencing are described as below (**Supplementary Figure 1**).

1.2 Confirmatory Sanger sequencing and deep sequencing

Exons of selected genes were amplified and underwent direct genomic sequencing by standard techniques on the ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, CA), as previously described.[5,6] All mutations were detected by bidirectional sequencing and scored as pathogenic if not present in non-clonal paired GL DNA. When the marginal volume of mutant clone size was not confirmed by Sanger sequencing, cloning and sequencing individual colonies (TOPO TA cloning, Invitrogen, Carlsbad, CA) was performed for validation. For detecting allelic frequency of mutations or SNPs, we applied deep sequencing to targeted exons as previously described. [7] Briefly, we analyzed for possible or validated mutations amplicons of around 250 bps, targeting the locus with each specific primer pair. The sequencing libraries were generated according to an Illumina pair-end library protocol and subjected to deep sequencing on HiSeq2000 or MiSeq sequencers according to the standard protocol (Illumina).[7, 8]

1.3 Detection of molecular defects by SNP-A

Single nucleotide polymorphism (SNP)-array karyotyping for confirming metaphase cytogenetics and detecting copy-number normal loss of heterozygosity was performed as previously described.[8, 9] Briefly, Affymetrix 250K and 6.0 SNP-arrays were used to evaluate copy number and loss of heterozygosity. Using our internal and publicly available databases, the screening algorithm validated each lesion as somatic [5, 6]. Non somatic lesions were excluded for further analysis. [10] Affected genomic positions in each lesion were visualized and extracted by CNAG (v3.0) or the Genotyping Console (Affymetrix) software.

1.4 Adjustment of variant allele frequency

Variant allelic frequencies (VAFs) of mutations were adjusted according to the zygosity and copy number confirmed by SNP-array. VAF of homozygous mutations as well as mutations of the genes located on chromosome X in the male cases were reduced to the half value of raw data. Hemizygous mutation VAFs were adjusted based on the formula as "Adjusted VAF = a/1+a (a = raw VAF value)." These adjustments of VAF were not required for heterozygous mutations. The adjusted VAF value of each mutation was categorized into large and small size dichotomized by mean VAF of all the identified mutations (**Supplementary Table 2**).

1.5 Distinction of founder and subclonal mutations

For distinction between ancestral and secondary mutations present in each case, we used the following criteria: 1.) In serial analyses, mutations appearing in progression but not present initially were deemed subclonal; 2.) In each case, VAFs of significant mutations adjusted by copy number variations and zygosity were compared and the largest clone was deemed founder in that case.[11,12] Cases without conclusive founder mutation (co-dominant) were excluded from further analyses (**Supplementary Table 3**). For distinction between ancestral or sub-clonal genes on the population level, we applied a ranking approach wherein, for each mutation, the proportion of cases in which that mutation was ancestral was calculated and the values compared to select the most likely ancestral events.

2. Supplementary Tables

Table S1. Genes tested in Next Generation Sequencing

Gene Name				
APC	DDX54	IDH1	NOTCH1	SF3B1
ASXL1	DHX29	IDH2	NPM1	SIMC1
BCOR	DNMT1	IRF4	NRAS	SMC3
BCORL1	DNMT3A	JAK2	NSD1	SRSF2
BTRC	EED	JAK3	OGT	STAG2
CALR	ERBB4	KDM6A	PHF6	STAT3
CBL	ETV6	KIT	PRPF40B	SUZ12
CCDC42B	EZH2	KRAS	PRPF8	TET1
CDH23	FLT3	LUC7L2	PTCH1	TET2
CEBPA	GATA2	MECOM	PTPN11	TP53
CFTR	GLI1	MED12	RAD21	U2AF1
CSF1R	GLI2	MLL	RNF25	U2AF2
CUX1	GNB1	MPL	RUNX1	WT1
DDX41	GPR98	NF1	SETBP1	ZRSR2

Table S2. Adjustment of Variant Allele Frequency

Mutation type	Adjustment performed
Homozygous	$VAF_{adjusted}=0.5(VAF_{raw})$
Found on X chromosome in males	$VAF_{adjusted}=0.5(VAF_{raw})$
Hemizygous	$VAF_{adjusted}= VAF_{raw}/1+VAF_{raw}$
Heterozygous	$VAF_{adjusted}=VAF_{raw}$

Table S3. Definitions of Ancestral versus Secondary Mutations

Mutation type	Criteria for definition
Ancestral (Dominant)	In serial sampling, appear as largest adjusted VAF in first sampling Largest adjusted VAF
Co-dominant	Adjusted VAF within 5% of Dominant ($VAF_{adjusted\ co\text{-}dominant} \leq VAF_{adjusted\ dominant}$ -5%)
Secondary (Subclonal)	In serial sampling, appear at second sampling but not first sampling >5% difference between Dominant ($VAF_{adjusted\ subclonal} < VAF_{adjusted\ dominant} - 5\%$)

Table S4. Clinical Characteristics of Sequenced Cases

	pMN (*n*=683)	sMN (*n*=65)	tMN (*n*=145)
Demographics			
Median age (years) at primary malignancy diagnosis (range)	64 (9-88)	60 (21-79)	60
Median age (years) at MN diagnosis (range)	64 (9-88)	71.5 (47-92)	68 † ‡
Median latency (years)	NA	8	7
Sex- Female: Male (%)	272:411 (40%:60%)	17: 48 (26%: 74%)	78:67 (54%:46%) * † ‡
Presentation			
MN presentation as advanced disease	328 (48%)	33 (51%)	75 (52%)
MN presentation as non-advanced disease	355 (52%)	32 (49%)	70 (48%)
Cytogenetics			
Normal	311 (46%)	32 (49%)	53 (37%) † ‡
Complex	111 (16%)	5 (8%)	35 (24%) * † ‡
del5	94 (14%)	4 (6%)	22 (15%)
del7/-7	84 (12%)	7 (11%)	36 (25%) * † ‡
del17	22 (3%)	1 (2%)	9 (6%)
del20	63 (9%)	3 (5%)	12 (8%)
trisomy 8	62 (9%)	3 (5%)	17 (12%)
delY	24 (4%)	2 (3%)	3 (2%)
Family history of cancer			
1st degree	280 (41%)	32 (49%)	63 (43%)
2nd degree	79 (12%)	4 (6%)	23 (16%)
1st and 2nd degree	70 (10%)	8 (12%)	22 (15%)
Total family history	429 (63%)	44 (68%)	108 (74%) † ‡
1st degree- hematologic	57 (8%)	7 (11%)	17 (12%)
2nd degree- hematologic	28 (4%)	1 (2%)	4 (3%)
1st and 2nd degree-hematologic	8 (1%)	2 (3%)	2 (1%)
Total family history-hematologic	93 (13%)	10 (16%)	23 (16%)
Top 3 primary malignancies			
Primary malignancy 1	NA	Prostate, 17 (26%)	Breast, 50 (34%)*
Primary malignancy 2	NA	Colorectal, 11 (17%)*	NHL, 28 (19%)*
Primary malignancy 3	NA	Breast, 10 (15%)*	Prostate, 24 (17%)

Table S5. Chemotherapy types used for treatment of a primary malignancy

	Total tMN	Ctx only	Ctx+Rtx	Rtx only
n total	266	87	90	89
Ctx treatment type unknown	59	28	31	NA
Ctx treatment type known	118	59	59	NA
Ctx type- Alkylating	94 (80%)	47 (80%)	47 (80%)	NA
Ctx type- Topoisomerase II	31 (26%)	19 (32%)	12 (20%)	NA
Ctx type- Other	111 (91%)	60 (97%)	51 (85%)	NA

Table S6. Mutations enriched in sMN

Gene	sMN *vs.* pMN	*p*	sMN *vs.* tMN	*p*	tMN *vs.* pMN	*p*
SRSF2	1.9967	0.0439	2.1667	0.0776	0.9216	0.88418
PHF6	4.8596	0.005	2.3559	0.1959	2.06027	0.2286
CUX1	3.0787	0.0417	2.333	0.2905	1.3194	0.7822
IDH1	3.0787	0.0417	12	0.0115	0.26	0.2243
IDH2	1.3396	0.581	3.944	0.0619	0.3396	0.0644
SF3B1	1.0915	0	3.3793	0.0508	0.323	0.0144
ZRSR2	2.2934	0.1123	3.5847	0.072	0.6398	0.41916
PTPN11	1.8817	0.2828	4.6885	0.0753	0.4013	0.2871

Table S7. Ages and latencies (in years) of CHIP and *de novo* tMN patients

	CHIP-derived	*De novo*- derived	*p* average	*p* median
Average age at primary malignancy diagnosis (median age)	**62 (63)**	**53 (53)**	**0.0175**	**0.0238**
Average age at Myeloid Neoplasm Diagnosis (median age)	**71 (70)**	**63 (64)**	**0.019**	**0.025**
Average latency from primary malignancy to myeloid malignancy (median latency)	**9 (7)**	**11 (7)**	**0.739**	**0.506**

3. Supplementary Figures

Figure S1. Bioanalytical Algorithm

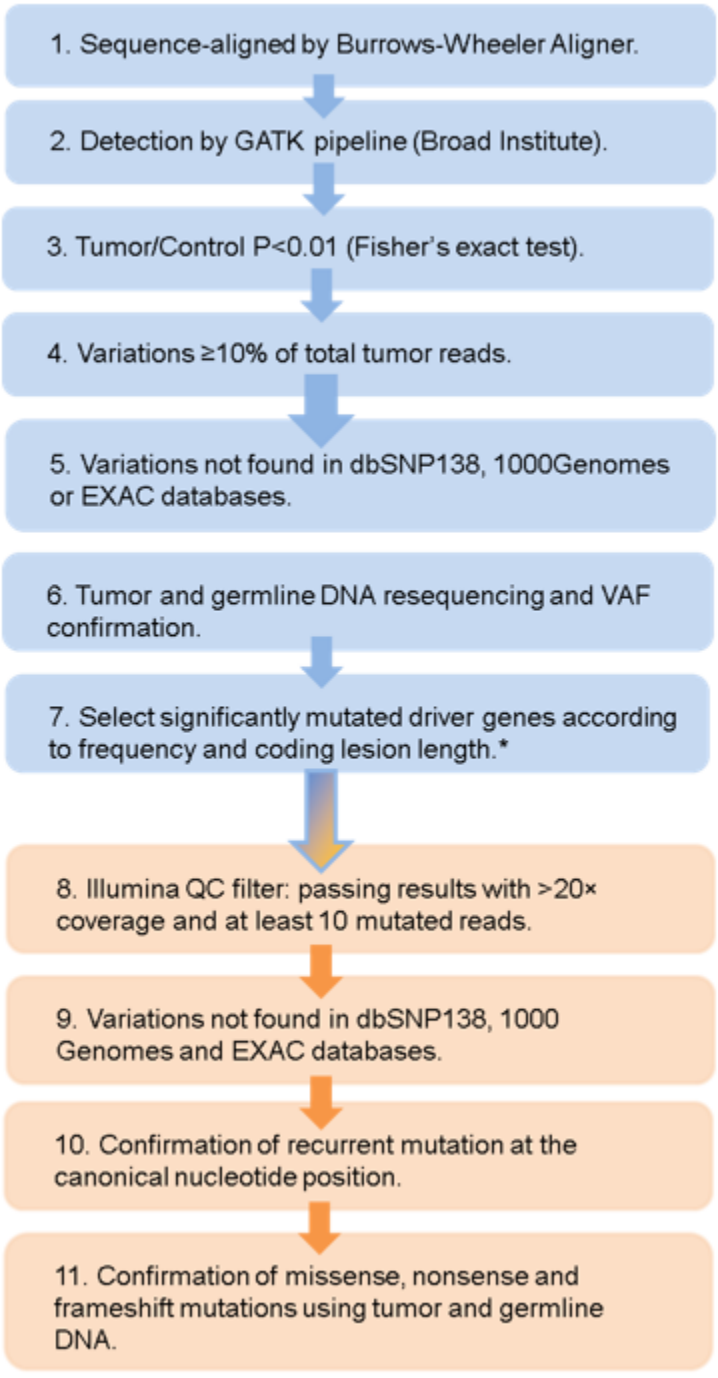


Figure S2. Common Mutations tested in pMN

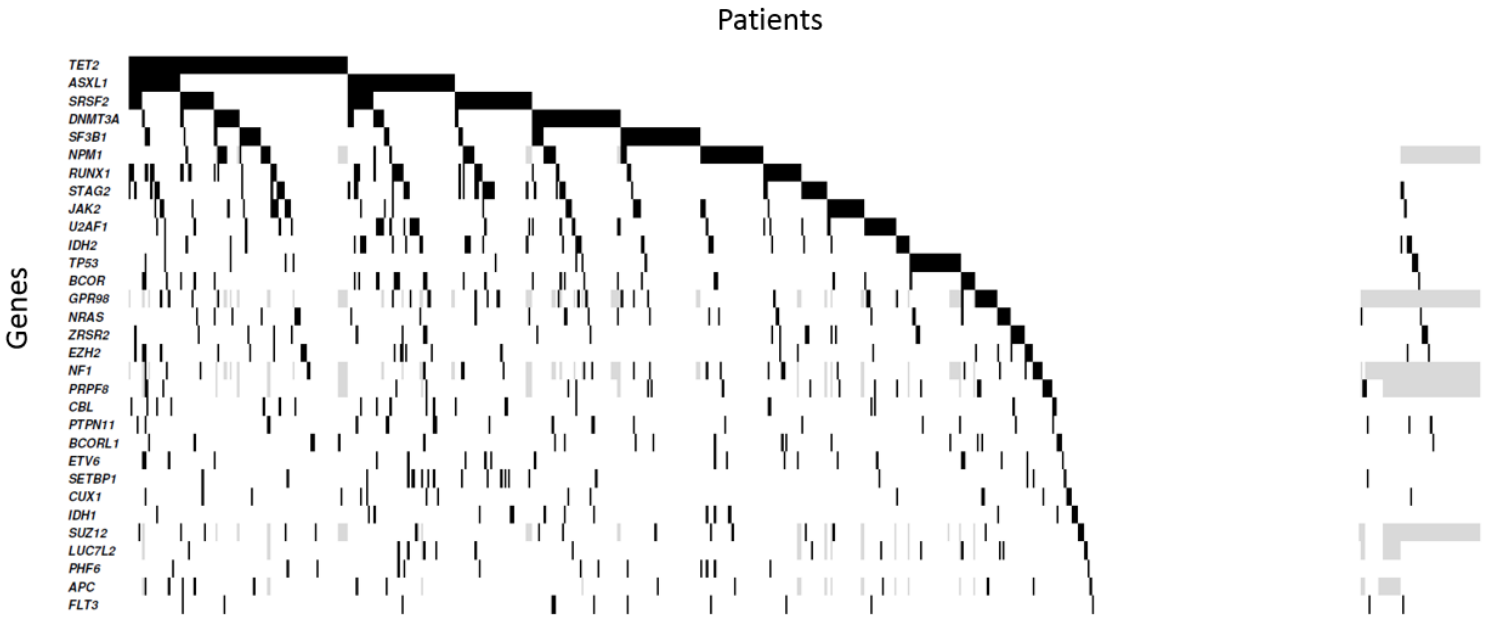


Figure S3. Common Mutations tested in sMN

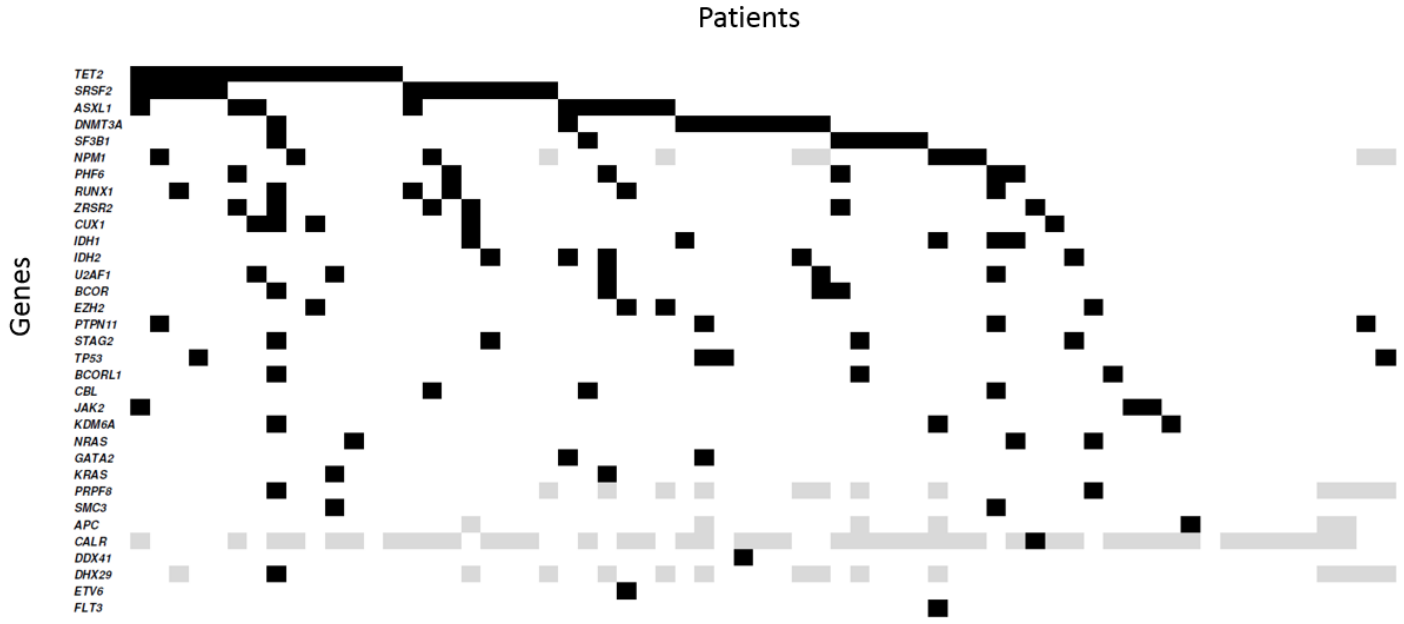


Figure S4. Common Mutations tested in tMN

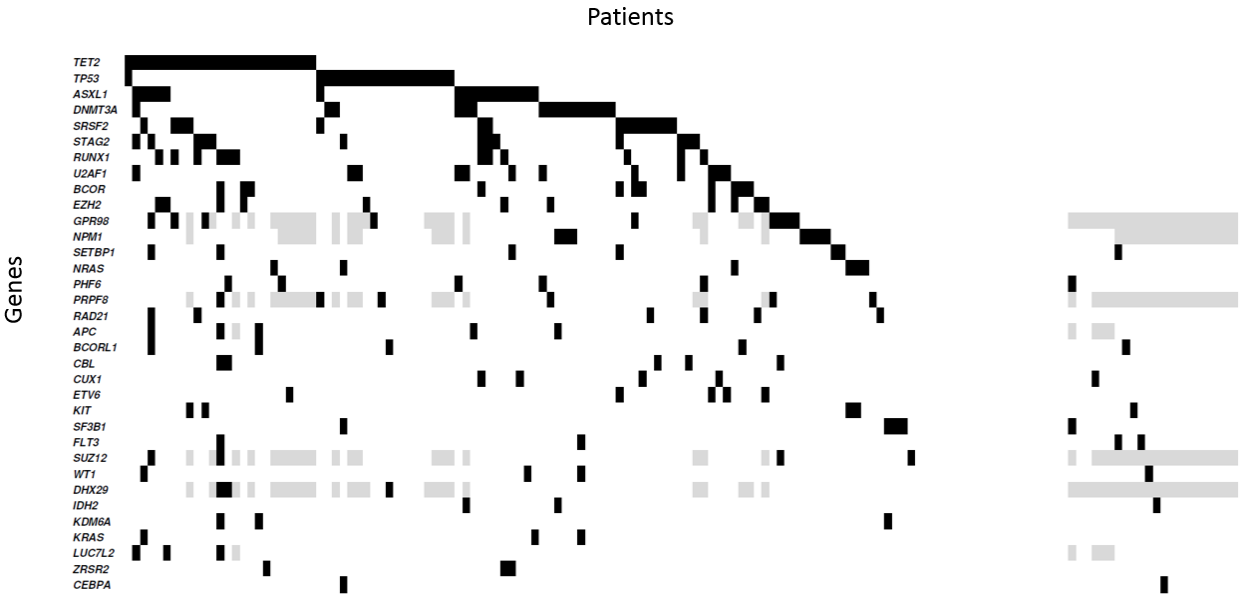


Figure S5. Disease Categorization of pMN, sMN, and tMN

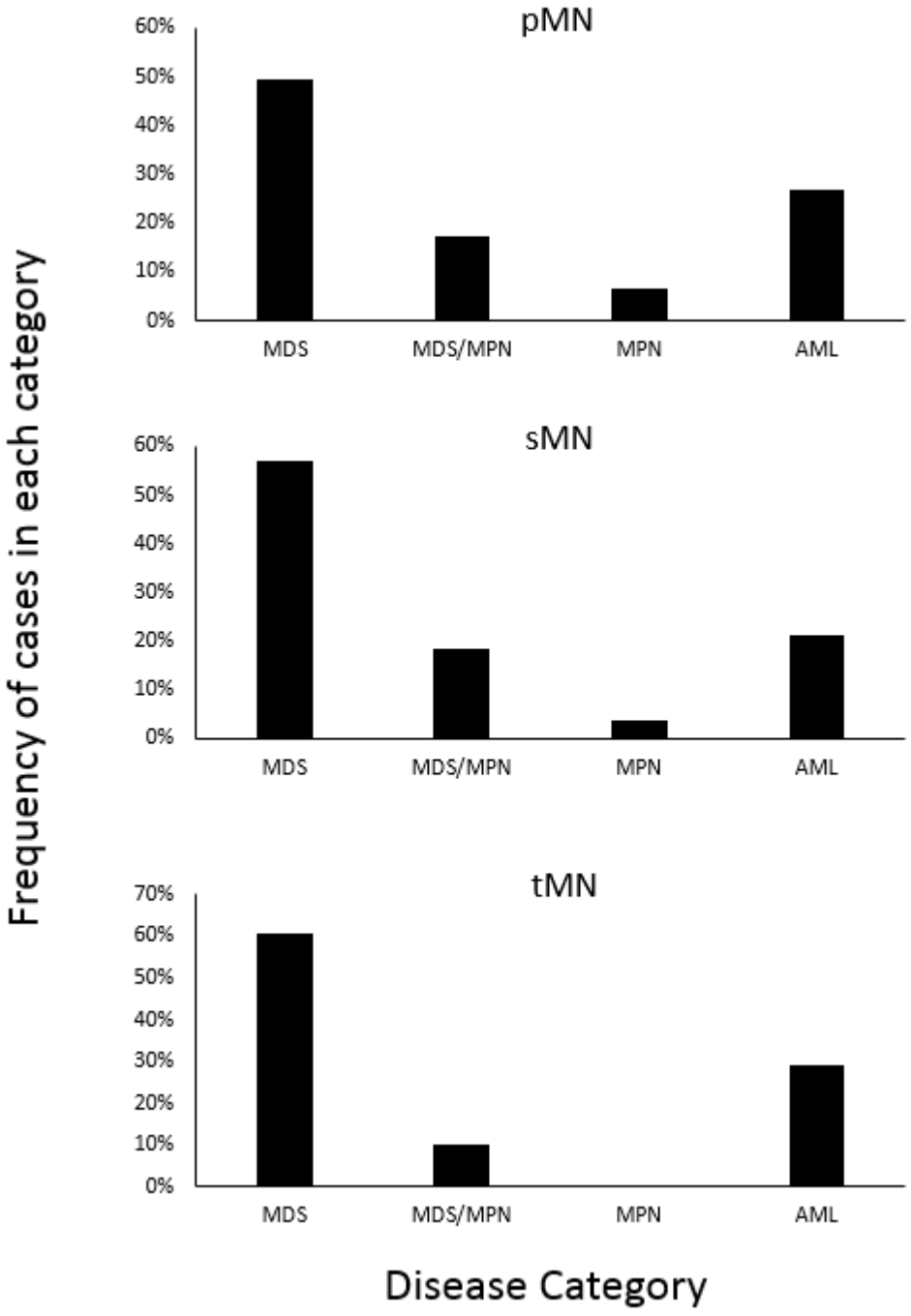


Figure S6. Categorization of MDS cases by 2016 WHO criteria

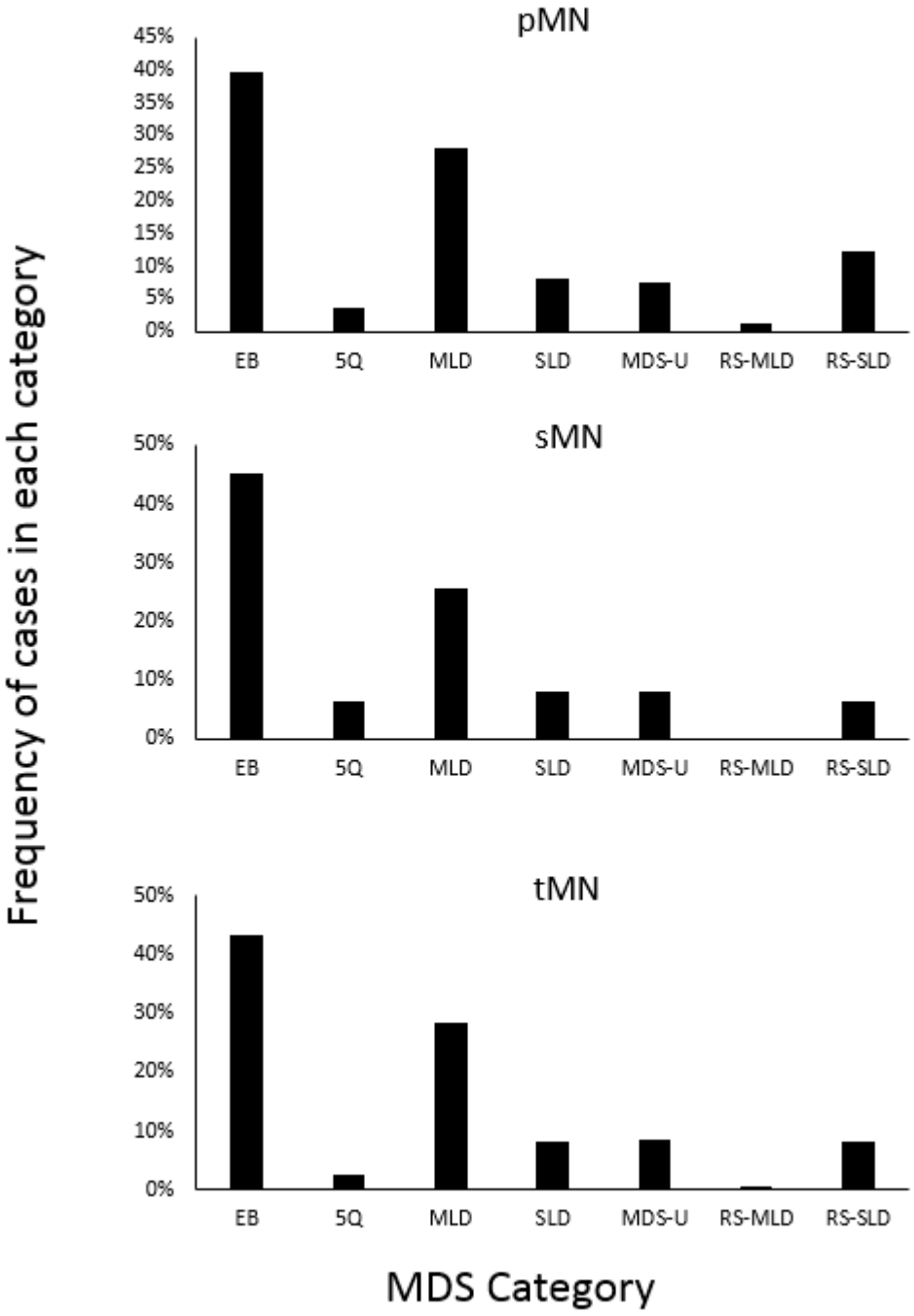


Figure S7. Classification of MDS/MPN cases by 2016 WHO criteria

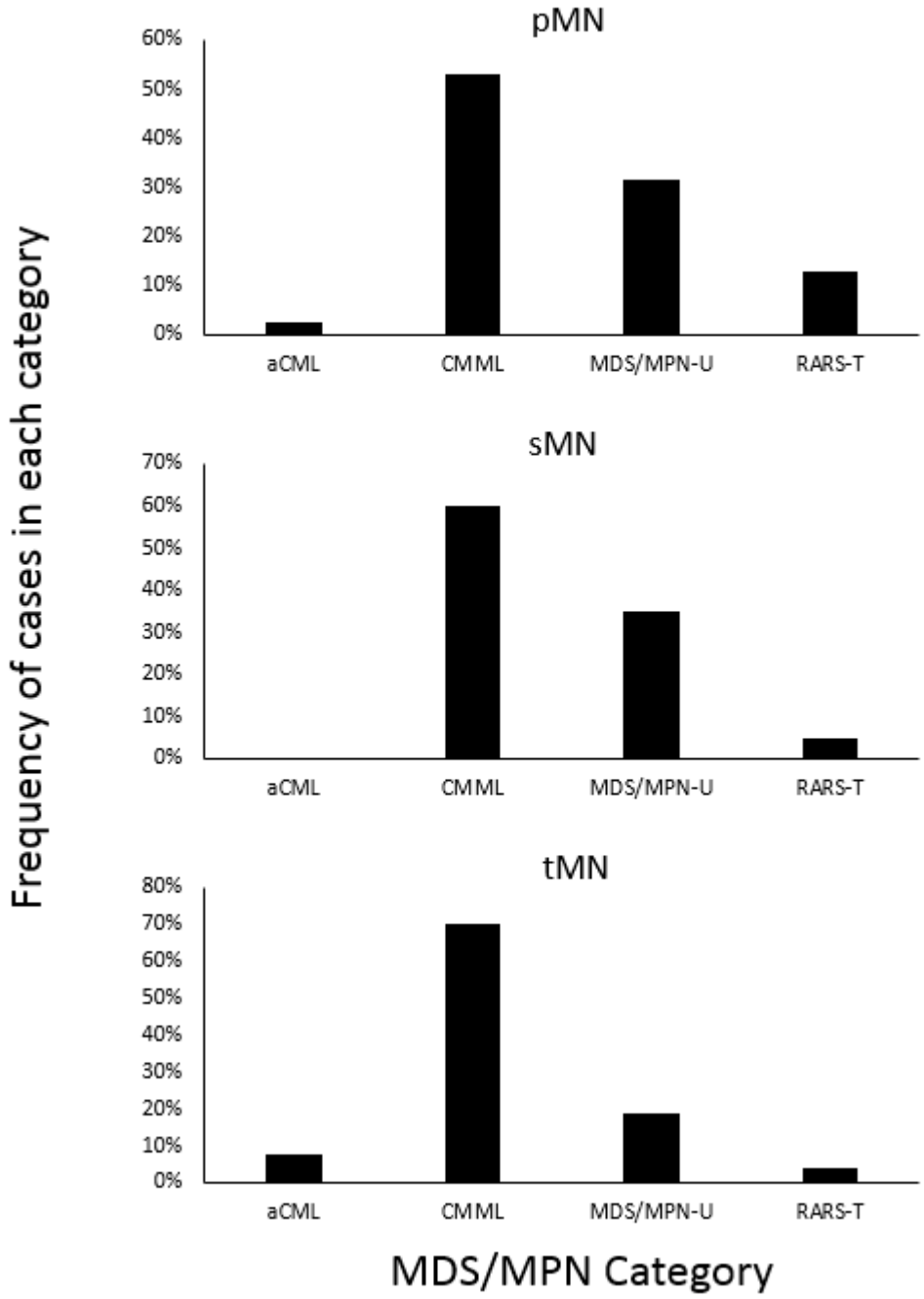


Figure S8. Classification of AML cases by 2016 WHO criteria

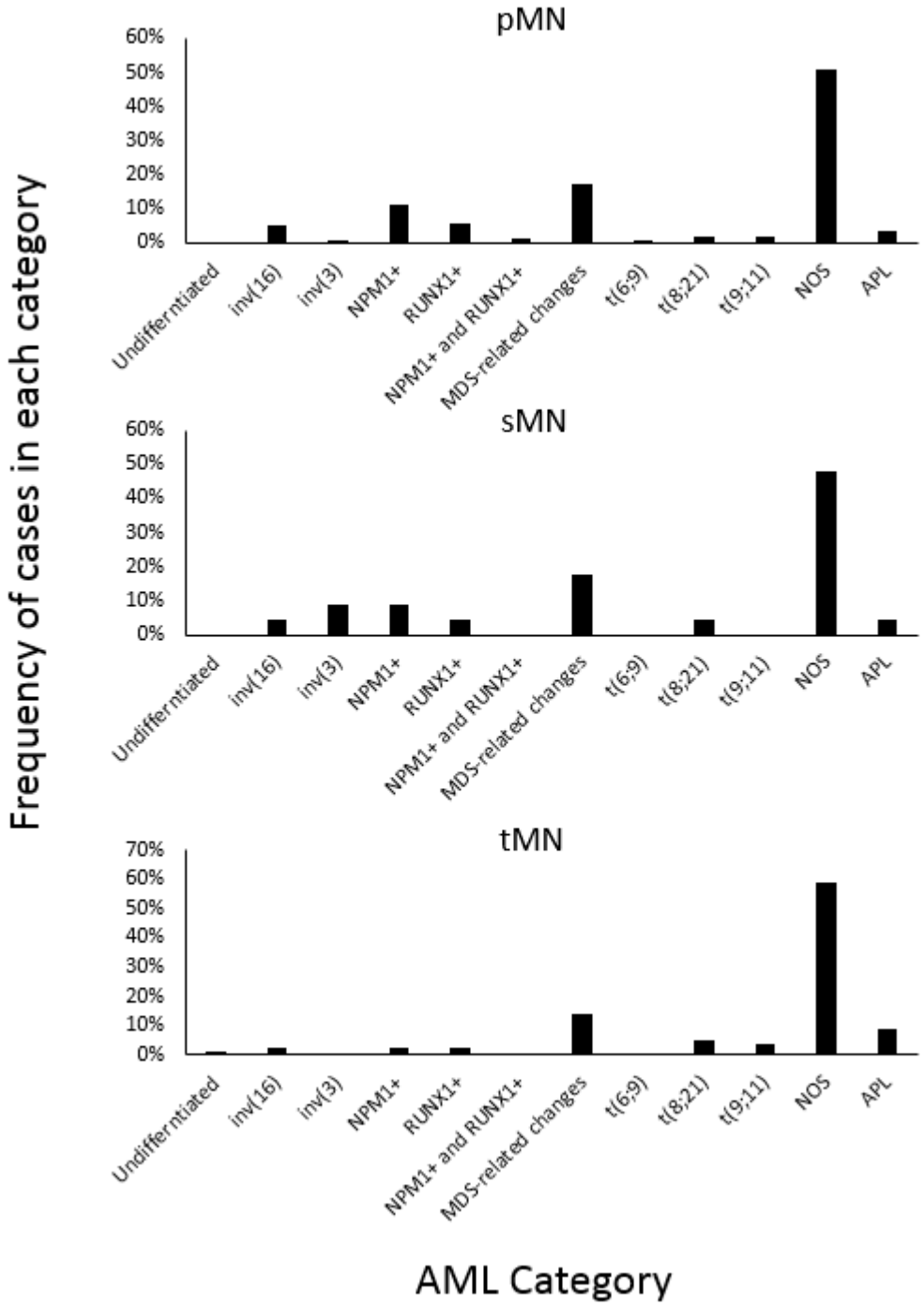
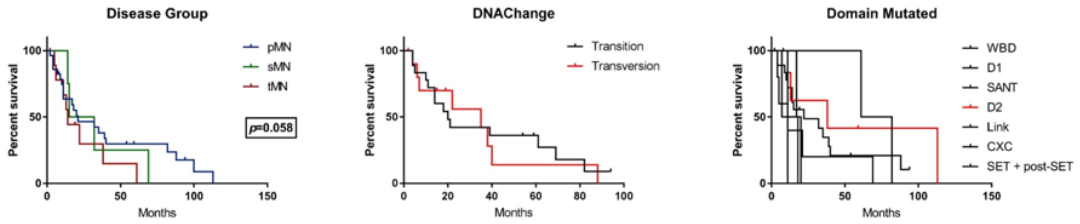


Figure S9. Effect of mutation or MN status on overall survival of *EZH2* and *TP53*-mutated tMN patients

A. *EZH2*



B. *TP53*

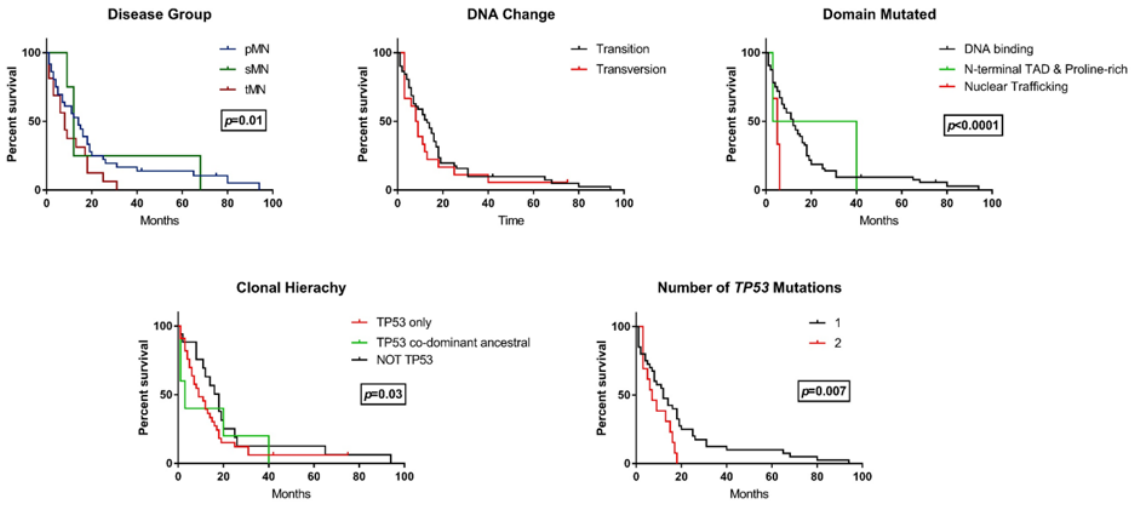


Figure S10. Relationship between pMN and CHIP

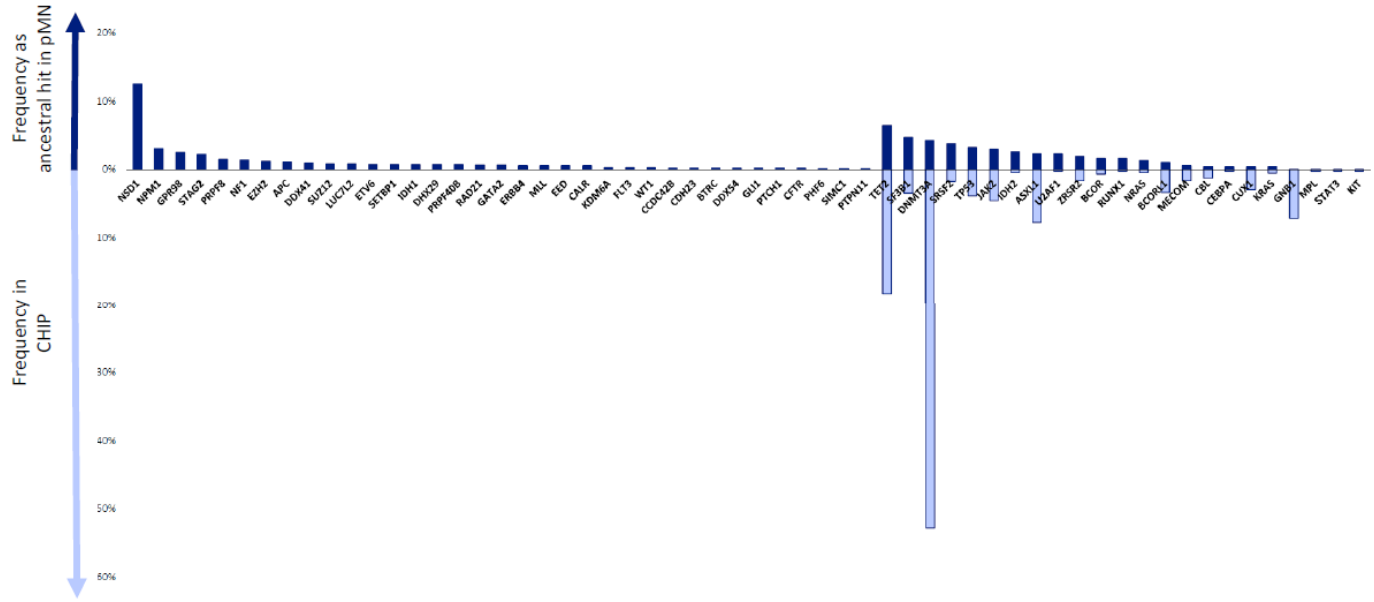


Figure S11. Relationship between sMN and CHIP

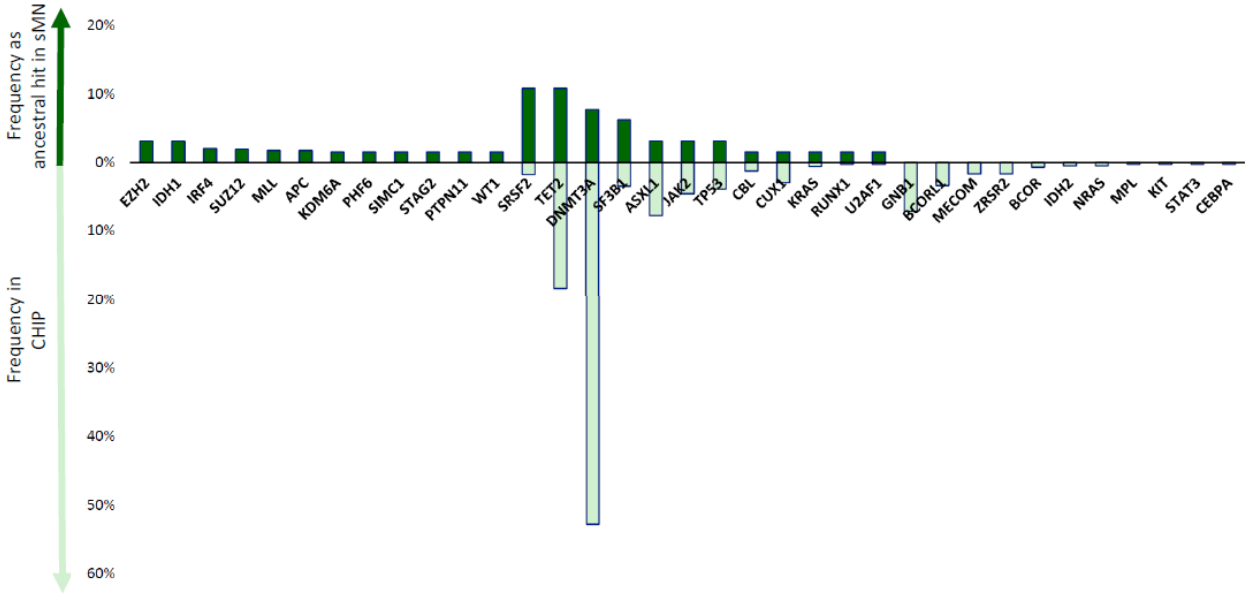


Figure S12. Estimation of Mutation Derivation

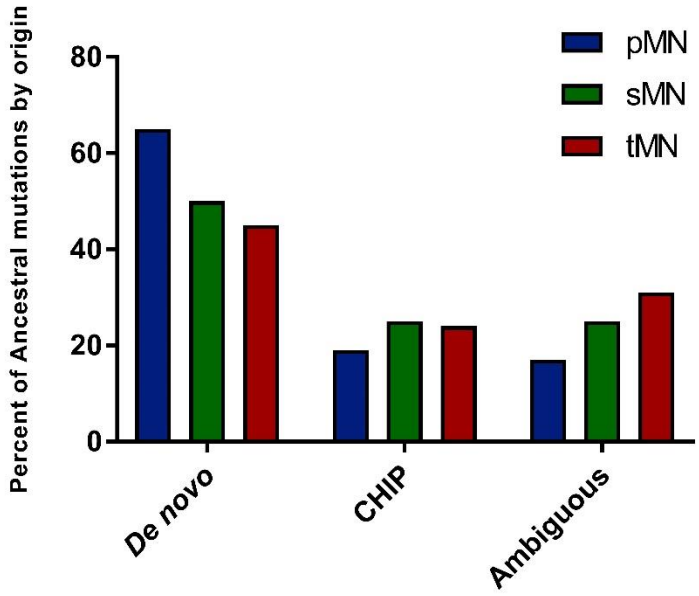
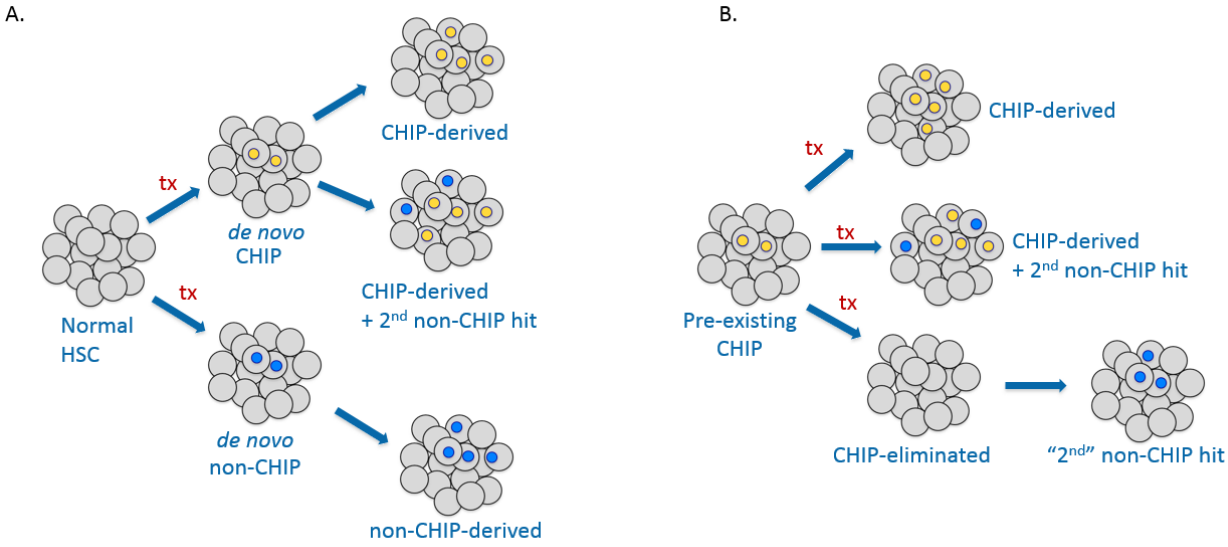


Figure S13. Schematic of origin of CHIP mutations in tMN



Supplementary References

1. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.
2. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595.
3. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.
4. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498.
5. Tiu RV, Gondek L, O'Keefe CL, Huh J, Sekeres MA, Elson P, 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;27:5219-26.
6. Huh J, Tiu R, Gondek LP, O'Keefe CL, Jasek M, Makishima H, et al. Characterization of chromosome arm 20q abnormalities in myeloid malignancies using genome-wide single nucleotide polymorphism array analysis. Genes Chromosomes Cancer 2010;49:390-9.
7. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 2011;478:64-9.
8. 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;119:3203-10.
9. Makishima H, Jankowska AM, McDevitt MA, et al. CBL, CBLB, TET2, ASXL1, and IDH1/2 mutations and additional chromosomal aberrations constitute molecular events in chronic myelogenous leukemia. Blood 2011;117:e198-206.
10. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
11. Hirsch C, Nazha A, Kneen K, Abazeed M, Meggendorfer M, et. al Consequences of mutant TET2 on clonality and subclonal hierarchy. Leukemia 2018; 32: 1751-1761
12. Patel BJ, Przychodzen B, Thota S, et al. Genomic determinants of chronic myelomonocytic leukemia. *Leukemia*. 2017;31(12):2815-2823.