

Supplementary Material
Risk stratification for younger and older patients with Acute Myeloid Leukemia
through transcriptomics, clinical data and machine learning

Raïssa Silva, Cédric Riedel, Jerome Reboul, Benoit Guibert, Florence Ruffle,
Nicolas Gilbert, Anthony Boureux, Thérèse Commes

Supplementary Table 1. Importance by k-mer

K-mer	Importance	Gene	Type
CTTTCTTGTTGGTAATCAGCTGAGCCTTGCA	0.6574002	GSTA4	ref
CCTCCGACAGTCTCCAGCAGTGGCACCCCA	0.41347378	MICALL2	ref
CTGGCCAATGCCAGGGATCTTTAGGTCAGCA	0.40869108	CD109	ref
AAATGCGTAAGAGAACAACTTCCTAAGCAC	0.40762982	ADRB2	NA
AACATCTTAAAAACAAAAGCTGACAACCAAGT	0.34261805	CD109	alt
TTCAGCAATGCAACCTTCTAATACTTGAGCA	0.31804562	CARMIL1	ref
AAAAATGAGAAAAGGTTAAAAAGACAGAATA	0.31195727	unannotated_1	NA
TTATCCCAGCCCACAGCACATGGTACAAACC	0.31044275	PTK2	ref
AACCATTGTTTATATGGCAAATGTCATCAAC	0.2972135	ADRB2	NA
CATGTACACCAGCAGCCTGGCCACCCTCACC	0.28174144	KCTD1	NA
AAAGGGACATCCGGCCACGCCAGGCCAGGGC	0.25231528	BAHCC1	ref
CACCAACACTACGTGACAAGTCTCAAGTTCA	0.24100202	BRF1	ref
AAAAAAAAAAAAAAAAATGAATAAGACAGTAGTC	0.22449535	TMED7-TICAM2	ref
AAAAGTTTCAGAGCCACTGGTCTACAGCATA	0.20582753	IL13RA1	NA
CCCAACCCAAAGAGACTTCTTCCATCTCTAA	0.19903265	MBOAT2	ref
CAAGCTGTTCTATCCATTTGATTCTTGCCAG	0.18329492	unannotated_1	NA
ATCGCGAAGCATCTGAACGAACGAGGAAGCC	0.17911354	BRF1	ref
ACTGTGAGGGGGAGTGCTCCTTCCCACTGGA	0.17721322	BMP8B	ref
AGAGACTGGCTCCAGGTCTGACTCAGTCCAC	0.17222512	CLIC2	ref
AGGGTGGGCCAGTAACCTCAGCAGGAGACGA	0.16615881	EFCC1	ref
ATGCTGCCACCTGAGCGGTTAAGCAGCTCCC	0.15589766	ANKS6	alt
CCCCTTGGCAAACTGATCAGTTTTAAACA	0.13830331	KCTD1	ref
ACAACCTAAGGATGGCATAGATGCCAGTGGAC	0.13776498	TRIM46	ref
CCGGCTAATTTTTGTATATTTAGTAGAGAC	0.13698669	H2BC10	NA
CTGTTAACCAAGATTCTCCATTTAAATGC	0.13467756	IQCK	ref
CCCTGGCCTGGCGTGGCCGGATGTCCCTTTC	0.13349365	BAHCC1	ref
AAAAATGAGTAAGTGTAGGTCATGTTGCCAT	0.13244861	PCAT18	ref
CGATTAACCCAAGTCAATAGAAGCAGGCGTA	0.12645255	unmapped_2	NA
AAGAGGATTTGAAATCGTGGGCTCCTGTGGG	0.11816129	KCTD1	ref
CCACTTCAGAGGGAGGAAACAGGCCCGGAGG	0.115323946	SLC16A1-AS1	ref
CTGCCCTTATTCACACCTTCCTTCCTATAGC	0.10388955	GDF11	ref
AGTCAGCTTTGCTGATGGGATTTAGGGCCAG	0.094916396	CMTM4	ref
CGTTCTGTGGGCAAATTTAAAGTTGAACTAA	0.091522425	LARP1B	alt

CTCACAGCAGTCGTGCCCCGGTCACTTCCCC	0.090130255	GAS6	ref
AGGGACATCCGGCCACGCCAGGCCAGGGCCA	0.086395524	BAHCC1	ref
CTAGCGGTGATGAAAAGCGGGAATGTGCTCC	0.085734725	LAMA5	alt
AAGGTGCGGGTTCTCCTGGGCGCGCACTCCC	0.084651895	CFD	ref
ACACTGGGCAAACGTGGTAAATCTTGGCCTC	0.08415953	CD109	ref
ACTTGAAAGATGCTCTCTACGATGTCATCAC	0.07976251	INPP4B	alt
CCGCTCTGCGGTGCTGAGGCTGCGGCCGCCG	0.079055645	SCRN1	ref
ACAGGAAGCTGTTGGAAGGCGAGGAGAGCAG	0.075669065	VIM	ref
AAATTAACTTTACTTAAAGCCTAGCAAGTCA	0.07410463	LPO	ref
TGGGTGCCAGGTGGCTGGGCGGGGCTGCCCC	0.07222159	HOMER3-AS1	ref
AGCAATCCACCAGCATGAGGTGAACCCTGGA	0.07192942	SNHG14	ref
ATCTCCCAAGACCAAGATTGTCCCTCGTGCC	0.07077692	SNHG14	ref
GAAATATTGGAATGTTCTGCTCAGATTGGC	0.07062769	PCAT18	ref
AAGCTTTGCCTCCTGGGTTACGCCATTCTC	0.070265464	CLEC40	NA
ACAGCCAGACCCCCCATCAGTGGAACAAGGGG	0.069603965	APOC4-APOC2	ref
TGCTTCTTACCTGCCTTGCTCTTAACTGACA	0.068664126	ANKS6	ref
CCAAGTTTCTTACCATGGGGATGTGGTCCA	0.06817956	PRXL2A	ref
AAGCGGGAATGTGCTCCGGCTGGAGGTGGAC	0.06605618	LAMA5	ref
TGTGTTCACTCACCAGGTACCTGCAGAAGGC	0.06598201	PEAR1	ref
ACTGCAACCTCTGTCTCTCAGGTTCAAGTGA	0.065595776	CD109	NA
ACGTTGCCCAGCGGACTGCTTGAGAAGTGC	0.064303696	RPS6KA1	ref
AGCCCTGGGGAAGATTTTGCTGAGCACGCTG	0.064099886	EFCC1	ref
CCGTCCGGGAAAAAAGGCGGGAGAAGCCCCGG	0.060177624	unmapped_1	NA
AAAACTCCATGAATCCTCTGTATGGGGAGT	0.05918478	unannotated_1	NA
TGACTTACATTGCAGTGACTAGATTATTCAC	0.058622096	PLXNC1	ref
AAGTCGTAGAAGCTGATCCAGCGGGTAAAAA	0.05803569	ANKS6	ref
GCTGCTCTAGACCCAGGAGCTTCTCTAGACC	0.05617371	ST6GALNAC4	ref
AAACCGCTCAACTACATGGAAAATGAACAAC	0.05557132	unannotated_2	NA
GGCAGAGGCTGAAGGAGATGCCCCCGGGCCC	0.052201163	CGREF1	NA
AATGGATAAATTCCTCGACATATACACCCTC	0.051751036	unannotated_1	NA
AAATCTTTATCTTTTTTGCTAATGAGAGTAA	0.05063609	GALNT12	ref
TAGAAAGCCTTCTCAAGAAGAAAGTCTCAAA	0.05034343	DR1	NA
AGAAAAGTATTGAGAATACTTCTATCACCTC	0.048316404	SNHG14	ref
AAAAAAAAAAAAATACAAAAATTAGCCAAGCAT	0.047399834	unmapped_3	NA
ACAGGTGAGGTTCCCCAGGAAGTAAACTCTG	0.0465446	IL13RA1	NA
AGGCGAGCAACTGGGGTTGGGCGGCAGCAGG	0.045928236	ANKS6	ref
ACCTCCAACCTAGCCCTACTGCCACCCAC	0.044179983	CD109	ref
AATGAAATAGTGCTGTTTGAGACTGTCCAAA	0.043540265	KCTD1	ref
CAGAGAGGAAGCAGCTCCGGGACAAGCTCCA	0.042981487	GALNT12	ref
CCGGAGGCAGTTTCTTCTGGCCTCACCCACC	0.042177584	ANKS6	ref

AAACAGAATTCCTCCGATACTGTTTCCATTG	0.041635726	CRHBP	ref
AACCTGGATGGCGGAGGTTGCAGTGAGCCGA	0.041516256	NA	NA
AAACACAAGGCCTAACTTTCCCTAATACTG	0.04011723	EXOSC3	ref
GATTTGAAGATATGGCAAAGCGGACAAGGC	0.04002869	HMGB1	alt
AAACCAGGATTGTCTGATTCCAACATCTCTG	0.03955902	PCAT18	ref
TAAATTCTCTCAGCCTTACACTCACAGAAGC	0.039424986	ANKS6	ref
ATCAAGAAAGAGATCGGAAGAGCACACGTCT	0.03822584	CALR	NA
TGAAAATGTACAAGTTGTAGGAGACTAAACA	0.03627722	DGKH	ref
AAGACAAGGTGGGAGGATCACTTGAGGCCAG	0.035173643	DNM1	NA
CTGAAACATTGAAGCACAAAGTCTCTCTCCA	0.033989966	KCTD1	ref
ATTCGCTGAACTGGGACGAGTCTACTCCTCC	0.033664405	KCTD1	ref
TCATCCATGTCCCTGCAAAGACATAAACTC	0.033451058	unannotated_1	NA
TTGTAAATGAGTTTCTGATGCCTTGAAAGAC	0.0325762	OPTN	ref
AGGTGATTGATACTCCTGATGCGAGTCATAC	0.032402795	unannotated_3	NA
CCCGTTCTCCAGGTATCTGCGGAGCCACTGC	0.032320283	HLA-H	alt
AACATATGTGTGCATGCATCTTTATAGTAGA	0.032124177	unannotated_1	NA
ACCTCTTCATCTCGCTGTAGAACACTGAGAA	0.031790778	SYTL4	ref
CTTCAGAAAATTCTGTAAAGACAAAGTAACA	0.030777127	MBOAT2	ref
AACAATTTGCCCTCTCCCTGCAGGCTGATTC	0.029595088	MMP19	ref
AAATTTTCATATGGAATAAAAAAGAGCCCAC	0.029195921	unannotated_1	NA
ATGCGCTATGTGAAAAATTACTCTGAATTGG	0.02624577	DDX42	ref
CCCTCCTGGGGAGGCTGACGGGTGTGGCCAC	0.023207586	TPPP	ref
AAAAACACAAAAATTTAGCCGGCTGTGGTGG	0.022889378	HOMER3-AS1	ref
AGGGGCCAGCCGTGGGTTCAGGGTGGGAGCC	0.022880722	VAC14	ref
ACATCCCTGATACCCAGAAGCTCCTCCCTA	0.022830663	TXNIP	alt
AAGATACTCTGGCAGTGGCGGGGATGGTGGC	0.021292085	SETBP1	ref
ACATAGCGCATATGTACAGGTAAGAAATGAA	0.0207838	DDX42	ref
AATTCGGCTGTGAATCCATCTGGCCCTGGAC	0.020057777	unannotated_1	NA
AAAAAAAAAAAAAGAGAAAAAGAAAAAGTGAC	0.019975632	CFD	ref
AACTGTAAGGAGACTGTGTCCTCACCCTCA	0.019662598	KCTD1	ref
CCTGCTGGGCTCCTACATTGAGCAGCCCGCC	0.019518739	LAPTM5	ref
AAAAAAAAAAAAAAAAAATAGCTGGGCATC	0.019173747	unmapped_4	NA
AAGCTTGGGCCTGGTCCCTGCCCTCACAGAT	0.018992797	SLC6A9	alt
TGCCGACCATTAGCAGTCGTACCCTCATCAC	0.018285051	PHF21A	ref
TTCCGATCTGTGCCAGGTCCTCTCCTAGGAC	0.018012764	CANX	alt
AACAGAGATAACATTGCAGAATCTTGAGTT	0.017984064	LINC00676	NA
AACTACTAAGTCCACTTCATAAATAAGTGTA	0.017704766	DUSP10	alt
ACCATTGCACTCCAGCCCGGGTGACAGTGCG	0.01767736	TMCO3	NA
TTTTCTAAGTATACAATCATGTAATCTGCAA	0.017542565	unannotated_1	NA
AATTTGGTTTAAAGTTTACCTAGTGACTGAC	0.01748757	IQCK	ref

CCCCCTCTATGCAGTGGTAAAGGAGCACATC	0.017340831	PRXL2A	NA
AAGTCTGGTACATTAATATTATATCCAGGGG	0.017183727	DTWD2	ref
ATATTTTGACAGCTCCTCCCCTCCCCTGCCC	0.017178835	ADAMTS10	ref
CCCCACGCCATCACACCTGCCACCAGACCA	0.017124208	SERPINB6	ref
AGAAGAATATCCATGACTACCTGAGATGGCC	0.016742615	ADRB2	NA
AAAACCTCCATGAATCCTCTGTATGGGGAGTT	0.016613971	unannotated_1	NA
AAACACCCGTACGTTACCTCTTAAGTAGTTG	0.016500216	unannotated_1	NA
GAGATCGGGCACACGTGAGGCCAGGGCACAC	0.016441066	ADAMTS10	ref
CACATTCGGCATGAACCATGACGGCGTGGGA	0.016423956	ADAMTS10	alt
AAAATAATTCTCTATTATTTTATTACCAAG	0.016384877	GAS6	ref
CCCGACCCTGCACTGCTGCCGTCCGCCCGCC	0.016148146	SMAD3	ref
ACTGTAGACATTGTCCTAGATGTTGGTCACA	0.015929727	TRIM46	ref
CCTAGGGGCTGGGGGCCACTCTTGCGGCAAG	0.015793737	AKT3	ref
ACACAATAAAACACTAATGAGTCACCCATAA	0.015568112	MEF2C	ref
TTTGATCTGTTGACGCTGTGCATTGTTTAC	0.015061573	ANXA1	alt
ACACAACCGCCACGGGCTTCCGAAGCGCCCG	0.015031795	BCOR	ref
CTTTTCTTTCTCTCCTTTCTTCTTTGCC	0.015017967	CD109	ref
CATAGTACATAGAAAATAAGGAAAGGGGGG	0.01492348	SERPINB6	ref
ACGGTGAAACCCCGTCTCTGCTGAAGATACA	0.014918538	ATP9A	alt
AACTTCAGGCTGACCAGAACTTTGCAACCTA	0.014820531	OSBPL5	ref
ATAGATTACATTTATTGATTTGCATATGTTA	0.014586778	unannotated_1	NA
ACTCCCTGTGCGAGCCGGAGCCCTCCTAGCG	0.014405567	CFD	ref
AGCTGATAATTTGTTAGCTGGAACCTATTCA	0.0142535735	SCRN1	ref
CCTCCAGCGCCGCTTTCTCCCCTCCAGCCAC	0.014230713	PLPPR3	ref
AACCCAGACCTTCTCTTTACCCCTTGCCATT	0.013970674	SHANK3	ref
ACAAATTTTCTCCCATTCTGTAGGTTGCCTG	0.013448307	unannotated_4	NA
AAAAAGTCCAGGGCCAGATGGATTCACAGCC	0.012917136	unannotated_5	NA
ACTCAGGAGGCTAGGGCAGGAGAATGGCGTG	0.011109659	XR_001738227.1	NA
ATATATGCAAGTGAATCAACTTGCCAGATAA	0.011046859	SPART	ref
AAAAAATCTATGCAGCTGGGTGCAGTGGCT	0.00909234	STAR	ref
AAAAAAAAAAAAAATTTTTTTTTTCTGGT	0.008661332	RARS2	alt
AGCAGCTCCCATGCAGGGGCAGGAAGACAAG	0.008544628	BAHCC1	ref
AAAAGTTCTTGTTTTGTTGAATTTATAGTA	0.007375727	GNL3	ref
ACCTGCCTGCCTTAGTCTCCCAAAGTGCTGG	0.006720087	CRCP	NA
AATCTCTGCTCACTGCAACCTCCACCTTCTG	0.006540393	LINC00869	NA
AACCCAGTCCCATGCCAGACCCTTGCAGTAG	0.005587928	MMP19	ref
AAAAAAAAAAACAAAAAAAAAACCCCTCCTT	0.0051020076	FOXO1	ref
ACGTGACAAGTCTCAAGTTCAGCCGCCTGGG	0.00496057	BRF1	ref
AAAGTTCTGGTCAGCCTGAAGTTCACAGTTC	0.0044764616	OSBPL5	ref
AAAAATTAAGTTATTAACATCTTAAAATTAT	0.003865281	PROM1	ref

AAAGGATGGCTGTGAGGTAGACGTCGTGGAC	0.0038279193	FANK1	alt
AAAGCAAGTCCTGAGTGACCTAAAAAGAGAC	0.0035685285	CRYBG3	NA
AAGACAAGTCTGGTTAACATGGCAAAATCCC	0.0033761482	ADAMTS2	ref
AGACTGTAAGATTTTCAGTTTATCCAATAATA	0.0032213798	DUSP10	ref

Supplementary Table 2. List of genes by group

Gene	Age group	Number of k-mers
ADAMTS10	older	3
ADAMTS2	older	1
ADRB2	older	3
AKT3	older	1
ANXA1	older	1
APOC4-APOC2	older	1
BAHCC1	older	4
BCOR	older	1
BRF1	older	3
CMTM4	older	1
CRCP	older	1
CRHBP	older	1
CRYBG3	older	1
DNM1	older	1
DTWD2	older	1
DUSP10	older	2
EFCC1	older	2
FANK1	older	1
FOXO1	older	1
GAS6	older	2
GDF11	older	1
GNL3	older	1
H2BC10	older	1
HLA-H	older	1
IL13RA1	older	2
IQCK	older	2
LAPTM5	older	1
LARP1B	older	1
LINC00676	older	1
LINC00869	older	1
MBOAT2	older	2
MEF2C	older	1
MMP19	older	2
OSBPL5	older	2
PEAR1	older	1
PHF21A	older	1

PLPPR3	older	1
PROM1	older	1
PRXL2A	older	2
SCRN1	older	2
SERPINB6	older	2
SETBP1	older	1
SHANK3	older	1
SLC6A9	older	1
SMAD3	older	1
SPART	older	1
STAR	older	1
SYTL4	older	1
TMED7-TICAM2	older	1
TRIM46	older	2
unannotated_3	older	1
unannotated_4	older	1
unannotated_5	older	1
unmapped_2	older	1
unmapped_3	older	1
unmapped_4	older	1
VAC14	older	1
VIM	older	1
XR_001738227.1	older	1
ANKS6	younger	6
ATP9A	younger	1
BMP8B	younger	1
CALR	younger	1
CANX	younger	1
CARMIL1	younger	1
CGREF1	younger	1
CLEC40	younger	1
CLIC2	younger	1
DDX42	younger	2
DGKH	younger	1
DR1	younger	1
EXOSC3	younger	1
GALNT12	younger	2

GSTA4	younger	1
HMGB1	younger	1
INPP4B	younger	1
LAMA5	younger	2
LPO	younger	1
MICALL2	younger	1
OPTN	younger	1
PLXNC1	younger	1
PTK2	younger	1
RARS2	younger	1
RPS6KA1	younger	1
SLC16A1-AS1	younger	1
SNHG14	younger	3
ST6GALNAC4	younger	1
TMCO3	younger	1
TPPP	younger	1
TXNIP	younger	1
unannotated_2	younger	1
unmapped_1	younger	1
CD109	younger/older	6
CFD	younger/older	3
HOMER3-AS1	younger/older	2
KCTD1	younger/older	7
PCAT18	younger/older	3
unannotated_1	younger/older	12

Supplementary Table 3. ToppCluster for younger patients

Category	ID	Title (or Source)	☐	Cluster 1_logP	Cluster 1_GeneSet
GO: Molecular Function				pValues	Gene List
GO: Molecular Function	GO:0005178	integrin binding	☐	☐ 3.8411	CALR HMGB1 LAMA5 PTK2
GO: Biological Process				pValues	Gene List
GO: Biological Process	GO:0034446	substrate adhesion-dependent cell spreading	☐	☐ 10.0000	CALR CARMIL1 LAMA5 MICALL2 PTK2
Pathway				pValues	Gene List
Pathway	MM15592	REACTOME CALNEXIN CALRETICULIN CYCLE	☐	☐ 4.0843	CALR CANX
Coexpression				pValues	Gene List
Coexpression	M5777	GSE9006 HEALTHY VS TYPE 1 DIABETES PBMC 1MONTH POST DX UP	☐	☐ 5.1747	CANX DR1 GALNT12 HMGB1 TXNIP

Supplementary Table 3. ToppCluster for older patients

Category	ID	Title (or Source)	Cluster 1_logP	Cluster 1_GeneSet
GO: Biological Process			pValues	Gene List
GO: Biological Process	GO:0099601	regulation of neurotransmitter receptor activity	☐ 5.9301	ADRB2 CRHBP MEF2C SHANK3 SLC6A9
GO: Biological Process	GO:0070167	regulation of biomineral tissue development	☐ 4.7105	ADRB2 BCOR GAS6 MEF2C SMAD3
Pathway			pValues	Gene List
Pathway	MM15656	REACTOME FOXO MEDIATED TRANSCRIPTION	☐ 4.4795	AKT3 FOXO1 SMAD3
Transcription Factor Binding Site			pValues	Gene List
Transcription Factor Binding Site	V\$AP1Fj Q2	V\$AP1Fj Q2	☐ 10.0000	AKT3 DNM1 LAPTMS MMP19 SCRNI SLC6A9 VIM
Transcription Factor Binding Site	V\$AP1 Q4	V\$AP1 Q4	☐ 5.2222	AKT3 DNM1 LAPTMS MMP19 SCRNI VIM
Transcription Factor Binding Site	M40788	NFXL1 TARGET GENES	☐ 4.0572	DNM1 GDF11 LARP1B MEF2C PEAR1 PHF21A
Transcription Factor Binding Site	V\$AP1 Q2	V\$AP1 Q2	☐ 4.0560	DNM1 LAPTMS MMP19 SCRNI VIM
Coexpression			pValues	Gene List
Coexpression	M15107	JAATINEN HEMATOPOIETIC STEM CELL UP	☐ 10.0000	AKT3 CD109 CRHBP PROM1 PRXL2A SCRNI SHANK3 SPART
Coexpression	M10382	IIZUKA LIVER CANCER EARLY RECURRENCE	☐ 5.8199	LAPTMS MEF2C VIM
Coexpression	M45039	CARRILLOREIXACH MRS3 VS LOWER RISK HEPATOBLASTOMA DN	☐ 5.7972	AKT3 ANXA1 CRYBG3 MEF2C SETBP1 VIM
Coexpression	MM1072	SERVITJA ISLET HNF1A TARGETS UP	☐ 5.5545	ANXA1 DNM1 HLA-H LAPTMS SMAD3 VIM
Coexpression	MM16606	ZHANG UTERUS CO SECRETORY STROMAL3 NPPC HIGH CELL	☐ 5.1274	ADAMTS2 ANXA1 GAS6 GNL3 MMP19 SERPINB6 VIM
Coexpression	M19231	EPPERT HSC R	☐ 5.1080	CD109 CRHBP FOXO1 H2BC10 PROM1
Coexpression	M2356	KOINUMA TARGETS OF SMAD2 OR SMAD3	☐ 4.9781	CD109 DUSP10 FANK1 KCTD1 MBOAT2 PRXL2A SCRNI SETBP1 SMAD3 VIM
Coexpression Atlas			pValues	Gene List
Coexpression Atlas	GSM399454 500	Stem Cells, SC.STSL.BM, CD117+ IL7R- CD150- CD48- AA4.1+ CD43+, Bone marrow, avg-3	☐ 10.0000	ANXA1 BAHCC1 EFCC1 KCTD1 MBOAT2 MEF2C PEAR1 PLPPR3 PRXL2A SHANK3

Additional Information 1. The k-mers selected in region unannotated_1 and genome position for each k-mers can be seen in table A1. Moreover, the mean of k-mers expressed for unannotated_1 region in younger and older patients by ELN2017 classification is present in Figure AI 1. The expression relative to NPM1 mutation status is present in Figure AI 2.

Table AI 1. K-mers selected in region unannotated_1

K-mer	Genome position
AAAACTCCATGAATCCTCTGTATGGGGAGT	chr2:89,174,923-89,174,953
AATGGATAAATTCCTCGACATATACACCCCTC	chr2:89,166,009-89,166,039
TCATCCATGTCCCTGCAAAAGACATAAACTC	chr2:89,164,248-89,164,278
AACATATGTGTGCATGCATCTTTATAGTAGA	chr2:89,164,403-89,164,433
AAATTTTCATATGGAACATAAAAAAGAGCCCAC	chr2:89,164,761-89,164,791
AATTCGGCTGTGAATCCATCTGGCCCTGGAC	chr2:89,165,883-89,165,913
TTTTCTAAGTATACAATCATGTAATCTGCAA	chr2:89,165,197-89,165,227
AAACTCCATGAATCCTCTGTATGGGGAGTT	chr2:89,174,922-89,174,952
AAACACCCGTACGTTACCTCTTAACTAGTTG	chr2:90,012,322-90,012,352
ATAGATTACATTTATTGATTTGCATATGTTA	chr2:89,165,573-89,165,603
AAAAATGAGAAAAGGTTAAAAAGACAGAATA	chr2:90,014,403-90,014,433
CAAGCTGTTCCCTATCCATTTGATTCTTGCA	chr2:89,166,833-89,166,863

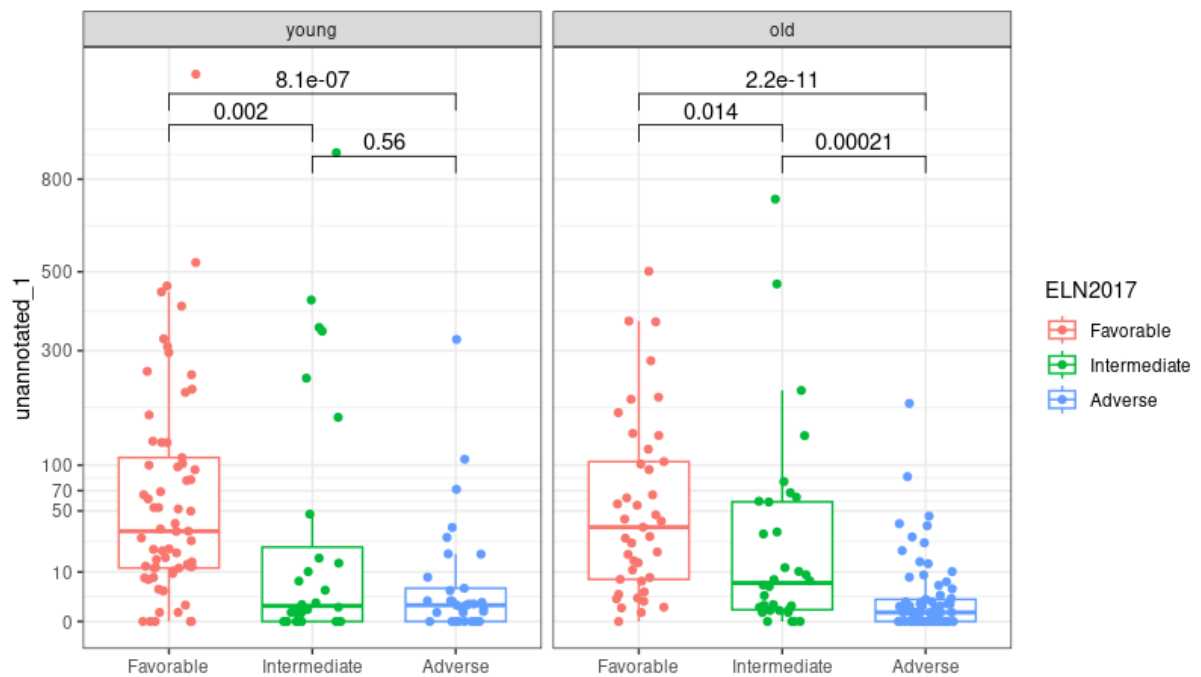


Figure AI 1. Expression of mean of k-mers from unannotated_1 in younger and older patients by ELN2017 classification.

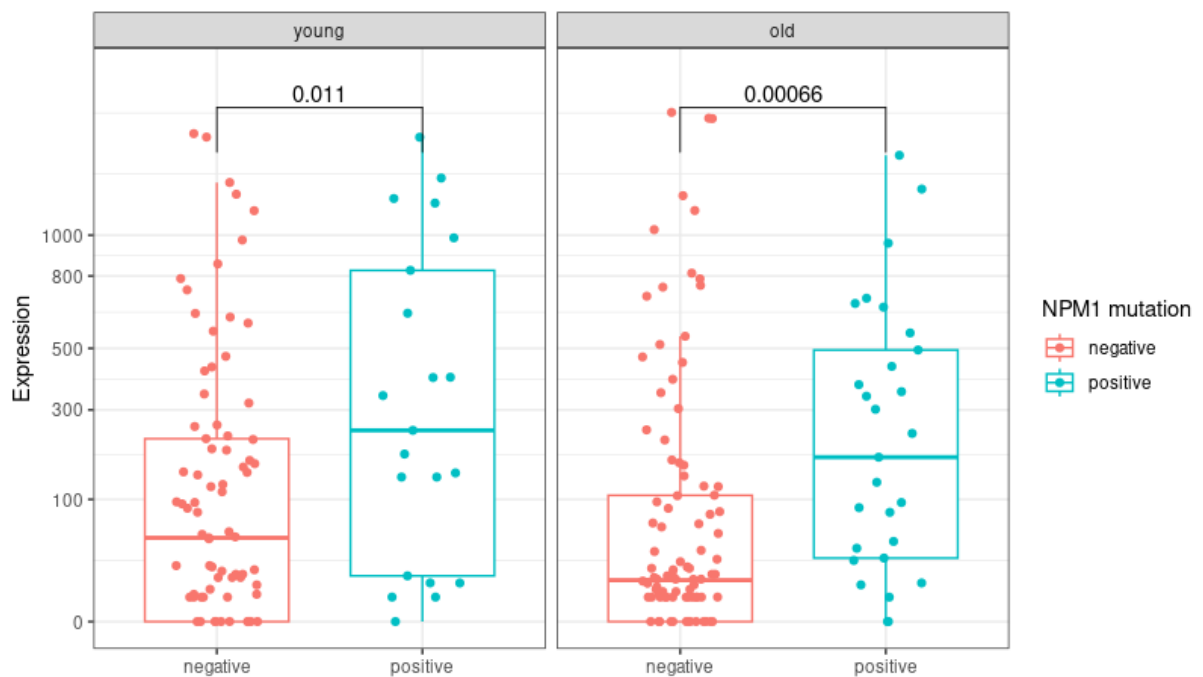


Figure AI 2. Expression of mean of k-mers from unannotated_1 relative to NPM1 mutation status

Additional information 2. To try to understand the differences between patients from different age groups but with the same risk stratification, we carried out additional tests. We selected all patients classified as Favorable risk to predict whether the patient is young or old. We applied the same pipeline described in the paper, we built a k-mer counting matrix with the Favorable risk patients from the Beat-AML cohort, applied the feature selection step, trained the models, and tested in the Beat-AML2.0 cohort. The performance to predict the age group can be seen in Figure AI 3 by ROC curve and AUC metrics.

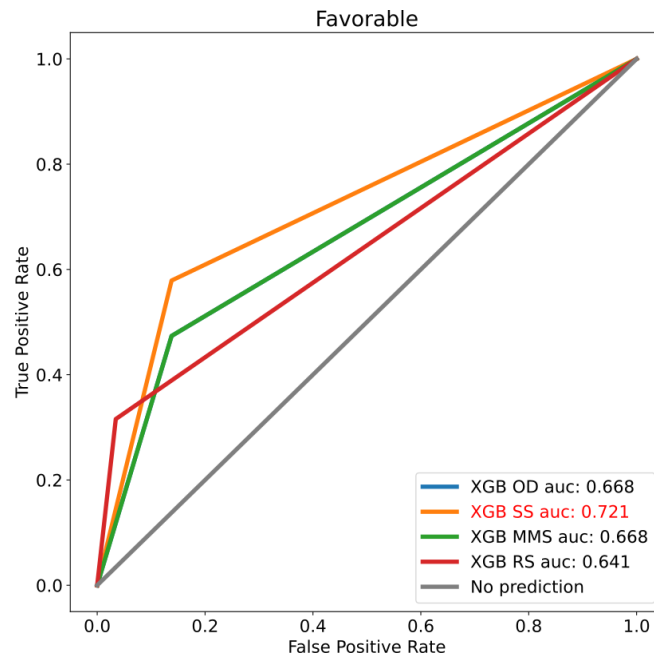


Figure AI 3. ROC curve and AUC in predicting age group with Favorable risk.

Comparing the XGB models with original data (OD), StandardScaler (SS), MinMaxScaler (MMS), and RobustScaler (RS), the best performance was XGB model with SS normalization, achieving 72.1% of AUC. Then, we used the XGB SS model with SHAP to select the most important k-mers, allowing us to identify 63 different annotations. We identified that the CD226 gene was the best annotation to predict younger and older patients in the Favorable risk group. 51 out of 65 younger patients had CD226 expressed (Figure AI 4). We used the same process to identify the difference between younger and older patients in the Adverse risk group, however, our methodology was not able to distinguish the patients.

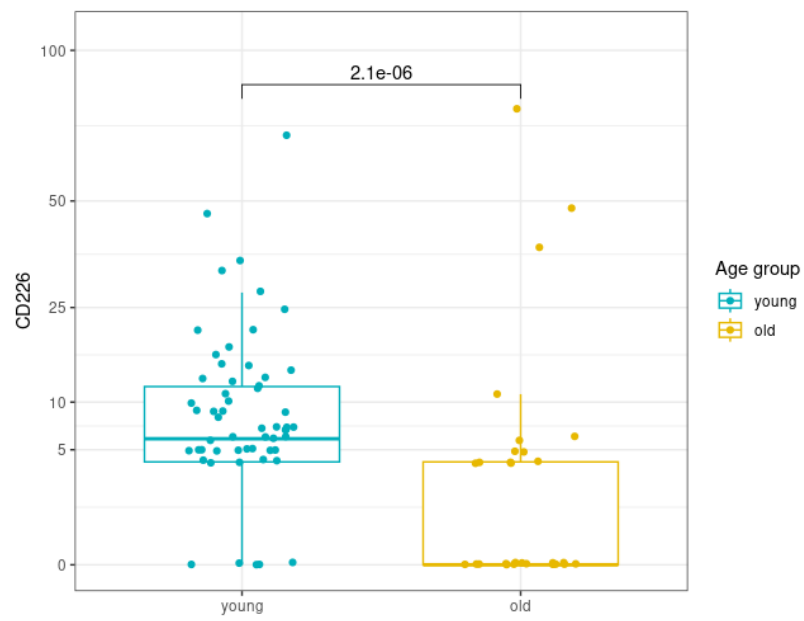


Figure AI 4. Expression of CD226 gene (k-mer) in younger and older patients with Favorable risk.