

## Supplementary Material

### Acute myeloid leukemia risk stratification in younger and older patients through transcriptomic machine learning models

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

#### Additional information 1

The pipeline to generate ML models and predict RNA-seq data can be seen in Figure A11. From an RNA-seq cohort, we generate a training k-mer count table using Kmtricks. Then, we apply a feature selection step to select k-mers relevant to distinguish the prognosis and we generate a new training k-mer count table with the selected k-mers. This new table is used to train the ML models. For a new patient, we extract the count of the selected k-mers from the RNA-seq data using “Back To Sequences” and we generate a test k-mer count table. This table is used by the trained model to predict the prognosis.

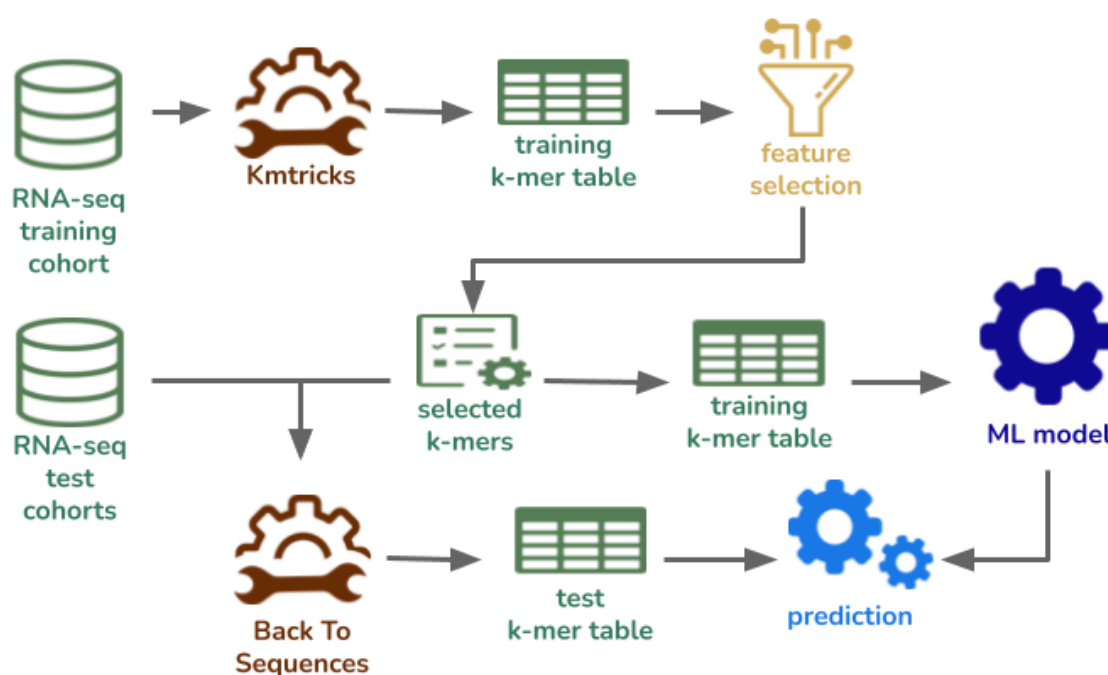


Figure A11. Pipeline to training and prediction for RNA-seq data.

## Additional information 2

As an initial test, we analyzed the prediction if the patient has a favorable or adverse risk in younger and older patients using a k-mer approach with ML, we performed initial tests with 70% of the Beat-AML cohort to train and 30% to validate. The metrics to evaluate the initial models can be seen in Table AI2.

Table AI2. Predicting favorable and adverse risk with initial models. Performance of models in younger and older patients.

Younger - Beat-AML2.0 and Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.63	0.53	0.81	0.25	0.07
KNN	0.84	0.75	1	0.5	0.64
LR	0.87	0.84	0.92	0.75	0.69
NN	0.89	0.86	0.96	0.75	0.75
RF	0.84	0.75	1	0.5	0.64
XGB	0.79	0.78	0.81	0.75	0.54
Older - Beat-AML2.0 and Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.79	0.77	0.69	0.86	0.56
KNN	0.82	0.83	0.85	0.81	0.64
LR	0.88	0.89	0.92	0.86	0.76
NN	0.85	0.88	1	0.76	0.74
RF	0.85	0.87	0.92	0.81	0.71
XGB	0.94	0.95	1	0.9	0.89

The best performance in younger patients was achieved by the Neural Network (NN) model achieving an accuracy of 89%. In older patients, eXtreme Gradient Boosting (XGB) achieved the best performance in all the metrics, with an accuracy of 94%.

### Additional information 3

We searched for the difference in peripheral blood samples, Natural killer (NK) cells, Fibroblasts, and Endothelial cells showed a difference between favorable and adverse in younger patients. In older, only Endothelial cells showed to be different between the risks.

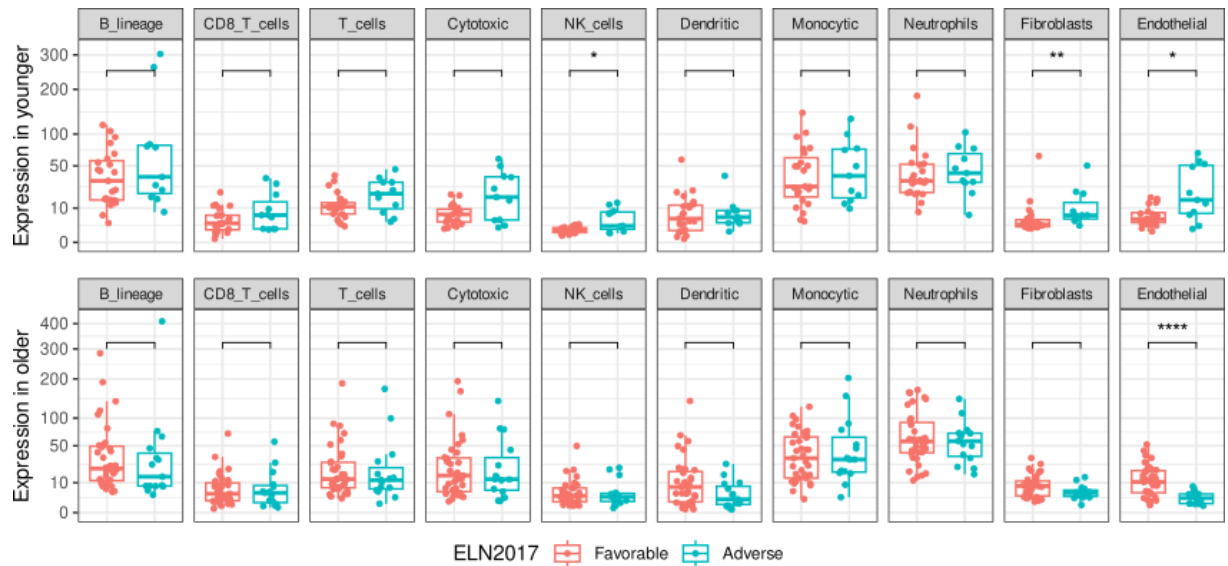


Figure AI3. Expression of immune and stromal cells in peripheral blood (PB) samples and comparison of favorable and adverse risk.

**Table S1. Metrics Method A**

Younger - Beat-AML2.0 and Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.59	0.62	0.4	0.85	0.27
KNN	0.83	0.8	0.97	0.63	0.66
LR	0.92	0.92	0.92	0.91	0.83
NN	0.85	0.82	0.97	0.67	0.69
RF	0.83	0.8	0.95	0.65	0.65
XGB	0.86	0.86	0.89	0.83	0.72
Younger - Beat-AML2.0					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.72	0.73	0.66	0.8	0.46
KNN	0.83	0.82	1	0.64	0.7
LR	0.94	0.94	1	0.88	0.89
NN	0.85	0.84	1	0.68	0.73
RF	0.83	0.82	0.97	0.68	0.68
XGB	0.85	0.85	0.86	0.84	0.7
Younger - Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.46	0.55	0.19	0.9	0.13
KNN	0.82	0.78	0.94	0.62	0.62
LR	0.89	0.91	0.86	0.95	0.79
NN	0.84	0.81	0.94	0.67	0.66
RF	0.82	0.78	0.94	0.62	0.62
XGB	0.88	0.86	0.92	0.81	0.73
Older - Beat-AML2.0 and Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.73	0.71	0.65	0.78	0.43
KNN	0.77	0.74	0.65	0.84	0.5
LR	0.86	0.85	0.81	0.9	0.71
NN	0.84	0.84	0.84	0.84	0.67
RF	0.78	0.75	0.65	0.86	0.52
XGB	0.91	0.92	0.94	0.9	0.82
Older - Beat-AML2.0					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.72	0.73	0.66	0.8	0.46
KNN	0.83	0.82	1	0.64	0.7
LR	0.94	0.94	1	0.88	0.89
NN	0.85	0.84	1	0.68	0.73
RF	0.83	0.82	0.97	0.68	0.68
XGB	0.85	0.85	0.86	0.84	0.7
Older - Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.61	0.62	0.42	0.82	0.25
KNN	0.74	0.75	0.5	1	0.57
LR	0.87	0.88	0.75	1	0.77
NN	0.78	0.78	0.83	0.73	0.56
RF	0.52	0.53	0.33	0.73	0.07
XGB	0.87	0.86	1	0.73	0.76

**Table S2. Metrics Method B**

Younger - Beat-AML2.0 and Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.57	0.57	0.55	0.59	0.14
KNN	0.84	0.82	0.94	0.7	0.67
LR	0.88	0.87	0.92	0.83	0.76
NN	0.86	0.84	0.92	0.76	0.7
RF	0.88	0.87	0.95	0.78	0.76
XGB	0.81	0.81	0.8	0.83	0.62
Younger - Beat-AML2.0					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.52	0.51	0.62	0.4	0.02
KNN	0.85	0.84	0.97	0.72	0.72
LR	0.87	0.87	0.93	0.8	0.74
NN	0.81	0.8	0.97	0.64	0.65
RF	0.87	0.86	1	0.72	0.76
XGB	0.81	0.82	0.79	0.84	0.63
Younger - Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.61	0.65	0.5	0.81	0.31
KNN	0.82	0.79	0.92	0.67	0.62
LR	0.89	0.89	0.92	0.86	0.77
NN	0.89	0.9	0.89	0.9	0.78
RF	0.89	0.89	0.92	0.86	0.77
XGB	0.81	0.81	0.81	0.81	0.6
Older - Beat-AML2.0 and Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.72	0.7	0.61	0.78	0.4
KNN	0.78	0.77	0.74	0.8	0.54
LR	0.89	0.87	0.81	0.94	0.76
NN	0.83	0.8	0.71	0.9	0.63
RF	0.73	0.7	0.58	0.82	0.41
XGB	0.84	0.83	0.77	0.88	0.66
Older - Beat-AML2.0					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.83	0.83	0.84	0.82	0.64
KNN	0.79	0.82	0.89	0.74	0.6
LR	0.88	0.86	0.79	0.92	0.72
NN	0.9	0.88	0.84	0.92	0.77
RF	0.83	0.8	0.74	0.87	0.61
XGB	0.91	0.9	0.84	0.95	0.8
Older - Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.43	0.44	0.25	0.64	-0.12
KNN	0.74	0.75	0.5	1	0.57
LR	0.91	0.92	0.83	1	0.84
NN	0.65	0.66	0.5	0.82	0.33
RF	0.48	0.48	0.33	0.64	-0.03
XGB	0.65	0.65	0.67	0.64	0.3

**Table S3. Metrics Method C**

Younger - Beat-AML2.0 and Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.78	0.78	0.82	0.74	0.55
KNN	0.69	0.65	0.91	0.39	0.36
LR	0.77	0.75	0.89	0.61	0.53
NN	0.71	0.66	0.98	0.33	0.44
RF	0.84	0.81	0.97	0.65	0.68
XGB	0.81	0.8	0.85	0.76	0.61
Younger - Beat-AML2.0					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.78	0.78	0.79	0.76	0.55
KNN	0.69	0.68	0.79	0.56	0.36
LR	0.78	0.77	0.86	0.68	0.55
NN	0.7	0.68	0.97	0.4	0.45
RF	0.87	0.86	1	0.72	0.76
XGB	0.8	0.8	0.76	0.84	0.6
Younger - Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.79	0.77	0.83	0.71	0.55
KNN	0.7	0.6	1	0.19	0.36
LR	0.77	0.72	0.92	0.52	0.49
NN	0.72	0.62	1	0.24	0.41
RF	0.81	0.76	0.94	0.57	0.58
XGB	0.82	0.79	0.92	0.67	0.62
Older - Beat-AML2.0 and Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.67	0.6	0.32	0.88	0.25
KNN	0.62	0.5	0	1	0
LR	0.72	0.67	0.45	0.88	0.37
NN	0.69	0.65	0.48	0.82	0.32
RF	0.68	0.58	0.16	1	0.33
XGB	0.65	0.55	0.1	1	0.25
Older - Beat-AML2.0					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.74	0.67	0.47	0.87	0.38
KNN	0.67	0.5	0	1	0
LR	0.74	0.66	0.42	0.9	0.37
NN	0.74	0.69	0.53	0.85	0.39
RF	0.76	0.63	0.26	1	0.44
XGB	0.72	0.58	0.16	1	0.33
Older - Leucegene					
Model	Accuracy	AUC	Sensitivity	Specificity	MCC
DT	0.48	0.5	0.08	0.91	-0.01
KNN	0.48	0.5	0	1	0
LR	0.65	0.66	0.5	0.82	0.33
NN	0.57	0.57	0.42	0.73	0.15
RF	0.48	0.5	0	1	0
XGB	0.48	0.5	0	1	0

**Table S4. Inversed models**

<b>Younger kmers in older - Beat-AML2.0 and Leucegene</b>					
<b>Model</b>	<b>Accuracy</b>	<b>AUC</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>MCC</b>
DT	0.58	0.54	0.35	0.72	0.08
KNN	0.78	0.78	0.77	0.78	0.54
LR	0.8	0.79	0.74	0.84	0.58
NN	0.74	0.73	0.71	0.76	0.46
RF	0.81	0.83	0.87	0.78	0.63
XGB	0.79	0.76	0.65	0.88	0.55
<b>Older kmers in younger - Beat-AML2.0 and Leucegene</b>					
<b>Model</b>	<b>Accuracy</b>	<b>AUC</b>	<b>Sensitivity</b>	<b>Specificity</b>	<b>MCC</b>
DT	0.49	0.51	0.38	0.63	0.02
KNN	0.65	0.67	0.55	0.78	0.34
LR	0.7	0.71	0.66	0.76	0.42
NN	0.65	0.66	0.62	0.7	0.31
RF	0.41	0.45	0.23	0.67	-0.11
XGB	0.63	0.67	0.46	0.87	0.35

**Table S5. Genes identified for younger patients**

<b>Gene</b>	<b>Selected k-mers</b>
ABR	15
ACVR2B	12
AMOT	13
ANKS6	434
ATP1A3	14
B4GAT1-DT	25
BAIAP2L2	14
BCAS4	28
BMI1	21
BMP8B	58
CD44	30
CD96	47
CEP192	18
CEP72	13
CLCN5	85
CLDND2	12
CLEC11A	61
CLEC4OP	24
CLEC9A	18
COL8A2	20
COMMD3-BMI1	23
CPNE8	12
CRACR2A	11
CROCC2	22
CRTC3	20
CYP2U1	12
CYP2U1-AS1	12
DGKH	38
DNMT3B	12
EHMT1	13
EPN1	17
F2R	12
FAM86B3P	28
FARP2	31
FBN1	20
GAPDH	11
GLCCI1	27
GPR84-AS1	50
HIVEP3	12
IGKV2-23	11
IGKV2-24	12
INPP5B	21
KHDC1	13
KIAA1549	37
KLRG1	12
LAMA5	275
LINC02811	12
LINC02891	32
LINGO3	14
LPO	261



MALAT1	12
MAMDC2	14
MAMDC2-AS1	12
MGST1	31
MPO	5502
MRPL42	17
NEIL1	39
PAQR6	13
PFKM	18
PI4KA	12
PIF1	12
PLXNA1	54
PPIE	58
PPP1R9B	11
QRFP	35
RACK1	13
RBFOX2	91
RPS6KA1	12
SAMD11	39
SCAMP5	14
SH3PXD2B	42
SLC16A1-AS1	52
SLC25A23	52
SLC29A2	25
SNORD116-4	15
SPATS2L	11
STK25	21
STXBP5	29
TALAM1	12
TBC1D29P	44
TBCD	13
TIAM1	48
TMEM213	22
TMEM273	37
TPM1	35
TPPP	13
TRIP10	19
TRO	16
TSPAN14-AS1	35
TUBG2	16
TXNIP	85
UICLM	22
VSTM1	17
XKR6	14
XYLB	14
ZG16B	71
ZHX3	30
ZNF263	12
ZNF532	91

**Table S6. Genes identified for older patients**

<b>Gene</b>	<b>Selected k-mers</b>
AAK1	13
ABCB1	55
ABCG1	71
ABLIM1	20
ADCY3	28
AGBL2	17
AMPD3	22
ANPEP	12
APOL3	53
ARAP2	28
ARHGEF10	11
ARHGEF17	12
ARMCX5-GPRASP2	14
ATP10A	218
ATP8B3	35
ATP8B4	23
BAALC	190
BAALC-AS1	198
BAALC-AS2	78
BACH1-IT1	12
BCL2L14	23
BIN1	36
BMAL1	16
BRF1	23
C16orf46	35
C16orf74	17
C2orf88	67
C8orf88	22
CA5B	56
CAB39L	18
CALN1	20
CASK	11
CAT	148
CATSPER2	16
CCDC136	15
CCDC24	32
CD34	115
CD74	11
CDC42BPA	62
CELSR1	36
CIITA	40
CNGB3	33
COBLL1	48
COL24A1	53

COL6A2	61
CRELD2	18
CRHBP	40
CRNDE	140
CRYBG3	11
CTTN	20
CUL4A	12
CYFIP2	17
DAZAP2	14
DGAT1	11
DGKE	11
DIPK1B	19
DLGAP3	29
DMXL2	206
DTWD2	31
DUSP1	17
EFNB1	11
EGFL7	44
EIF1B-AS1	14
ELOVL7	33
EPHA1-AS1	20
ERG	13
ESAM	24
F2RL1	186
FAM124B	24
FAM30A	14
FANK1	12
FBLN5	13
FBXO22	16
FGFBP2	22
FILNC1	11
FIS1	12
FLJ12825	13
FOXO1	17
GATA3	35
GBP4	18
GBP5	15
GGTA1	12
GNAI1	29
GNG11	15
GNL3	11
GPR34	11
GPRASP1	14
GPRC5B	34
GRAP2	13

GYPC	302
H1-0	14
H2BC4	11
HAL	19
HIBCH	67
HILPDA-AS1	36
HLA-B	20
HLA-DPA1	28
HLA-DPB1	33
HLA-DRA	27
HOMER3	14
HPGD	14
HSD17B4	23
IFI16	37
IFI44	77
IFT140	11
IGF2BP3	15
IL12RB2	52
IL6ST	19
INHBA	17
IPO9-AS1	162
IQCG	54
IQCK	64
IQSEC2	11
ISG20	49
JUN	20
KANK1	17
KBTBD11-OT1	11
KCNA3	20
KIAA2026	13
LAPTM5	43
LDOC1	76
LGMN	62
LINC00174	14
LINC00467	11
LINC01237	12
LINC02147	26
LINC02573	15
LINC02767	35
LMLN	72
LRP6	30
LTB	14
MAN1A1	50
MAP6D1	16
MCOLN2	14

MEF2C-AS2	11
MIAT	71
MIR130AHG	20
MITF	20
MLANA	13
MLYCD	12
MMP19	144
MMP28	28
MN1	510
MON1A	14
MREG	63
MRPL33	19
MTAP	64
MX1	80
MYRIP	14
NAV1	163
NECAB1	17
NEDD4L	78
NEK11	12
NHSL1	39
NOXRED1	16
NPR3	88
NREP	21
NRG4	24
OCEL1	11
OIP5	25
OIP5-AS1	26
OTULINL	23
PAK1	38
PAWR	74
PBX3	53
PEAR1	30
PHKA2-AS1	11
PIWIL4	95
PIWIL4-AS1	95
PPHLN1	22
PPP1R26	11
PREP	20
PREPL	17
PRICKLE1	49
PRKCH	132
PROM1	100
PRR5	24
PRR5L	55
PSMB8	15

PSMB9	37
PTGFRN	50
PXDC1	16
QPRT	13
RAB8A	34
RABGAP1L	17
RAMP1	18
RBKS	18
RBPMS-AS1	13
RDX	197
RFX8	20
RIPOR3	187
RNF130	21
RNF217	111
ROBO4	117
RORA	20
RORA-AS1	20
RTN1	12
SAMD15	11
SCARF1	70
SCD	27
SCN1B	14
SCRN1	93
SERPINB6	125
SHANK3	510
SIAE	17
SLC2A9	29
SLC37A3	86
SLC39A11	13
SLC39A8	14
SLC41A3	31
SLC6A9	32
SMAGP	14
SNX33	11
SPAG16	65
SPRY1	12
ST3GAL3	54
STARD4-AS1	17
STOM	71
SUSD3	30
SV2A	32
SYN3	73
SYNE2	15
TAL1	38
TAP1	37

TBCK	15
TBXAS1	17
TC2N	24
TGFBR3	20
TICAM2-AS1	30
TIMP3	73
TMC4	23
TMC6	16
TMC8	12
TMED7-TICAM2	30
TMEM204	11
TP53I11	37
TRH	41
TRPS1	264
TSPAN18	37
TSPAN7	30
USP54	15
VSIG2	12
YEATS2	23
YEATS2-AS1	23
YPEL4	20
ZBP1	79
ZC3H12C	195
ZNF277-AS1	14
ZNF626	13
ZNF662	185

**Table S7. Summary of Cox model in younger patients**

Call:  
coxph(formula = formule\_y, data = training\_y)

n= 131, number of events= 46

	coef	exp(coef)	se(coef)	z	Pr(> z )	
age	0.02954	1.02999	0.01367	2.161	0.03072	*
MP01	-0.94206	0.38983	0.65894	-1.430	0.15282	
CLCN51	0.34664	1.41431	0.38033	0.911	0.36207	
TPM11	0.44778	1.56484	0.36518	1.226	0.22013	
CLEC40P1	-0.72477	0.48444	1.14734	-0.632	0.52759	
TIAM11	-0.02798	0.97240	0.41319	-0.068	0.94600	
GLCCI11	1.20000	3.32011	0.42733	2.808	0.00498	**
SLC29A21	1.49507	4.45964	0.39911	3.746	0.00018	***
RACK11	-0.92005	0.39850	0.44670	-2.060	0.03943	*
NEIL11	-0.12939	0.87863	0.41653	-0.311	0.75608	
LING031	0.76455	2.14803	0.37248	2.053	0.04011	*
IGKV2_241	-0.46193	0.63007	0.53574	-0.862	0.38856	
BMI11	0.38799	1.47401	0.34729	1.117	0.26391	

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
age	1.0300	0.9709	1.00275	1.0580
MP01	0.3898	2.5652	0.10715	1.4183
CLCN51	1.4143	0.7071	0.67113	2.9805
TPM11	1.5648	0.6390	0.76494	3.2012
CLEC40P1	0.4844	2.0643	0.05112	4.5905
TIAM11	0.9724	1.0284	0.43266	2.1855
GLCCI11	3.3201	0.3012	1.43685	7.6718
SLC29A21	4.4596	0.2242	2.03974	9.7505
RACK11	0.3985	2.5094	0.16603	0.9564
NEIL11	0.8786	1.1381	0.38838	1.9877
LING031	2.1480	0.4655	1.03511	4.4575
IGKV2_241	0.6301	1.5871	0.22048	1.8006
BMI11	1.4740	0.6784	0.74626	2.9115

Concordance= 0.848 (se = 0.03 )  
Likelihood ratio test= 78.96 on 13 df, p=2e-11  
Wald test = 53.81 on 13 df, p=7e-07  
Score (logrank) test = 78.68 on 13 df, p=2e-11



**Table S8. Schoenfeld residuals for Cox model in younger patients**

	chisq	df	p
age	0.04959	1	0.824
MPO	2.90832	1	0.088
CLCN5	0.64535	1	0.422
TPM1	0.59275	1	0.441
CLEC40P	1.67329	1	0.196
TIAM1	0.77515	1	0.379
GLCCI1	0.68569	1	0.408
SLC29A2	0.00189	1	0.965
RACK1	0.00289	1	0.957
NEIL1	0.93715	1	0.333
LING03	3.61748	1	0.057
IGKV2_24	0.04301	1	0.836
BMI1	0.55816	1	0.455
GLOBAL	14.77814	13	0.321

**Table S9. Summary of Cox model in older patients**

```
Call:
coxph(formula = formule_o, data = training_o)

n= 147, number of events= 97

      coef exp(coef) se(coef)      z Pr(>|z|)
age 0.06466  1.06680  0.01411  4.583 4.58e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
age      1.067      0.9374      1.038      1.097

Concordance= 0.629 (se = 0.033 )
Likelihood ratio test= 20.15 on 1 df,  p=7e-06
Wald test               = 21.01 on 1 df,  p=5e-06
Score (logrank) test = 21.75 on 1 df,  p=3e-06
```

**Table S10. Schoenfeld residuals for Cox model in older patients**

	chisq	df	p
age	0.744	1	0.39
GLOBAL	0.744	1	0.39

**Table S11. Gene ontology analysis for younger group**

source - term name	term id	adjusted p-value	genes
GO:MF enzyme regulator activity	GO:0030234	< 0.01	ABR,ACVR2B,BMI1,FARP2,GAPDH,PIF1,PPP1R9B,RACK1,RPS6KA1,SH3PXD2B,STXBP5,TBCD,TIAM1,TXNIP,ALDH1A1,ANXA1,ARHGEF1,CCNA1,CCNK,CD109,DMPK,DMWD,DOCK1,DOCK4,DOCK9,GAS6,KRTCAP2,PREX1,RASAL2,RASGRP1,SH2D3A,SH2D3C,SH3BP5,SOCS2,SORL1
GO:MF guanyl-nucleotide exchange factor activity	GO:0005085	< 0.01	ABR,FARP2,TIAM1,ARHGEF1,DOCK1,DOCK4,DOCK9,PREX1,RASGRP1,SH2D3A,SH2D3C,SH3BP5
GO:MF receptor tyrosine kinase binding	GO:0030971	0.01	RACK1,TIAM1,ARTN,CPNE3,DOCK4,GAS6,PLCG1
GO:BP positive regulation of locomotion	GO:0040017	< 0.01	AMOT,F2R,MALAT1,RACK1,TALAM1,TIAM1,AKT3,ANXA1,ARTN,AZU1,CALR,CARMIL1,CPNE3,DOCK1,DOCK4,GAS6,MYLK,PECAM1,PLCG1,PPM1F,PREX1,PTK2,RREB1,SRPX2,VTN
GO:BP response to stimulus	GO:0050896	< 0.01	ABR,ACVR2B,AMOT,ATP1A3,BMI1,BMP8B,CD44,CD96,CEP192,CLEC11A,CPNE8,CRACR2A,CRTC3,CYP2U1,DGKH,EHMT1,EPN1,F2R,FARP2,FBN1,GAPDH,IGKV2-24,INPP5B,KLRG1,LAMA5,LPO,MGST1,MPO,NEIL1,PI4KA,PIF1,PLXNA1,PPP1R9B,QRFP,RACK1,RBFOX2,RPS6KA1,SCAMP5,SLC25A23,STK25,TIAM1,TPM1,TRIP10,TXNIP,VSTM1,ABCB6,ADGRG1,ADGRL1,AKT3,ALDH1A1,ANXA1,AOC1,ARHGEF1,ARTN,ASH2L,AZU1,BACH1,C1R,C1RL,CACNA2D4,CALCRL,CALR,CCNK,CD109,CFD,CLIC2,CPNE3,DMPK,DNAAF4,DOCK1,DOCK4,DOCK9,DTX3,ECHDC3,ELF1,EVL,EXT2,GAS6,GRAMD1B,GSTA4,HLA-E,HTR1F,IFI44L,IFITM2,INPP4B,ITPKA,ITPR3,LAG3,LPAR4,LPIN2,LSP1,MAD2L2,MEF2C,MLLT3,MYLK,NCOA7,NDFIP1,NRIP1,OPTN,OXR1,PECAM1,PFKFB2,PLCG1,PLIN2,PLXNC1,PPM1F,PPP3CC,PREX1,PRKD3,PRXL2A,PTK2,PTP4A3,RAB39B,RASAL2,RASGRP1,RBPMS,RREB1,RUNX3,SARM1,SH2D3A,SH2D3C,SH3BP5,SLC24A5,SLC41A1,SLC44A2,SNRPN,SOCS2,SORBS3,SORL1,SREBF1,SYNE1,SYTL4,TJP2,TRPM2,TSPAN14,VIM,VTN
GO:BP metal ion transport	GO:0030001	< 0.01	ATP1A3,COMMD3-BMI1,CRACR2A,F2R,SLC25A23,ABCB6,CACNA2D4,CALCRL,CLIC2,DMPK,FHL1,GAS6,ITPR3,KLHL3,MYLK,NDFIP1,PLCG1,PPP3CC,SFXN1,SLC24A5,SLC38A1,SLC41A1,TMCO3,TMEM163,TRPM2,TSPAN13
GO:BP cell communication	GO:0007154	0.02	ABR,ACVR2B,AMOT,ATP1A3,BMI1,BMP8B,CD44,CLEC11A,CRACR2A,CRTC3,DGKH,EPN1,F2R,FARP2,FBN1,GAPDH,INPP5B,KLRG1,LAMA5,PFKM,PI4KA,PLXNA1,PPP1R9B,QRFP,RACK1,RBFOX2,RPS6KA1,STK25,STXBP5,TIAM1,TRIP10,TXNIP,VSTM1,ADGRG1,ADGRL1,AKT3,ANXA1,AOC1,ARHGEF1,ARTN,AZU1,CALCRL,CALR,CCNK,CD109,CLIC2,CPNE3,DLG3,DMPK,DNAAF4,DOCK1,DOCK4,DOCK9,DTX3,ELF1,EVL,GAS6,HTR1F,IFITM2,INPP4B,ITPKA,ITPR3,LAG3,LPAR4,LSP1,MAD2L2,MEF2C,MLLT3,NDFIP1,OPTN,PECAM1,PFKFB2,PLCG1,PLIN2,PLXNC1,PPM1F,PPP3CC,PREX1,PRKD3,PTK2,PTP4A3,RAB39B,RASAL2,RASGRP1,RBPMS,SARM1,SH2D3A,SH2D3C,SH3BP5,SLC38A1,SLC44A2,SOCS2,SORBS3,SORL1,SREBF1,SYNE1,SYTL4,TRPM2,TSPAN14,VIM,VTN
GO:BP intracellular signaling cassette	GO:0141124	0.03	ABR,AMOT,CD44,CRACR2A,F2R,FARP2,GAPDH,PPP1R9B,RACK1,STK25,TIAM1,ADGRG1,ADGRL1,ARHGEF1,AZU1,CALR,CLIC2,DMPK,DOCK1,DOCK4,DOCK9,GAS6,LPAR4,MEF2C,NDFIP1,OPTN,PLCG1,PPP3CC,PREX1,PTK2,PTP4A3,RAB39B,RASGRP1,RBPMS,SH2D3A,SH2D3C,SLC44A2,SORBS3,TRPM2,VIM
GO:BP positive regulation of stress fiber assembly	GO:0051496	0.04	AMOT,TPM1,CARMIL1,EVL,PPM1F,SORBS3
GO:CC cell periphery	GO:0071944	< 0.01	ACVR2B,AMOT,ATP1A3,BAIAP2L2,CD44,CD96,CLCN5,CLDND2,COL8A2,CPNE8,CRACR2A,DGKH,EPN1,F2R,FBN1,GAPDH,IGKV2-24,INPP5B,KIAA1549,KLRG1,LAMA5,LINGO3,LPO,MGST1,MRPL42,PAQR6,PFKM,PI4KA,PLXNA1,PPP1R9B,RACK1,SCAMP5,SLC29A2,STXBP5,TBCD,TIAM1,TPM1,TRIP10,TRO,VSTM1,XKR6,ABCB6,ADAMTS10,ADGRG1,ADGRL1,ANXA1,AOC1,ARHGEF1,CACNA2D4,CALCRL,CALR,CARMIL1,CD109,CEACAM21,CLIC2,CPNE3,DLG3,DMPK,DNAAF4,DOCK1,DOCK4,FHL1,GRAMD1B,HLA-E,HTR1F,IFITM2,ITPR3,KMT2E,LAG3,LPAR4,LRATD2,LSP1,MICALL2,MYLK,NDFIP1,PALS2,PECAM1,PLCG1,PLIN2,PLXNC1,PREX1,PRKD

			3,PTK2,PTP4A3,RAB11FIP4,RAB39B,RASGRP1,SH2D3C,SIDT1,SLC38A1,SLC41A1,SLC44A2,SORL1,SRPX2,SYNE1,SYTL4,TJP2,TMEM163,TMEM163A,TRPM2,TSPAN13,TSPAN14,VIM,VTN,ZNF804A
--	--	--	---

**Table S12. Gene ontology analysis for older group**

source - term name	term id	adjusted p-value	genes
GO:MF peptide antigen binding	GO:0042605	0.04	HLA-B,HLA-DPA1,HLA-DPB1,HLA-DRA,TAP1,HLA-E
GO:BP cell adhesion	GO:0007155	<0.01	CASK,CD34,CD74,CELSR1,COL6A2,CTTN,CYFIP2,DUSP1,EFNB1,EGFL7,ESAM,FBLN5,GATA3,HLA-DPA1,HLA-DPB1,HLA-DRA,IL6ST,KANK1,LAPTM5,LMLN,LRP6,PAWR,PEAR1,PRICKLE1,RDX,ROBO4,SCARF1,SCN1B,SLC39A8,ADGRG1,ADGRL1,ANXA1,AZU1,CALR,CARMIL1,DLG3,DOCK1,GAS6,HLA-E,LAG3,MAD2L2,MICALL2,NDFIP1,PECAM1,PLXNC1,PPM1F,PTK2,RASGRP1,RREB1,RUNX3,SORBS3,SRPX2,TJP2,VTN
GO:BP antigen processing and presentation of peptide antigen	GO:0048002	<0.01	CD74,HLA-B,HLA-DPA1,HLA-DPB1,HLA-DRA,LGMN,TAP1,CALR,HLA-E
GO:BP regulation of protein metabolic process	GO:0051246	0.02	ABCG1,BIN1,BMAL1,C8ORF88,CD74,CUL4A,CYFIP2,DUSP1,FBXO22,FIS1,FOXO1,GNL3,GPRC5B,IFI16,IGF2BP3,IL6ST,INHBA,LAPTM5,LGMN,LRP6,NECAB1,NEDD4L,PAK1,PAWR,PIWIL4,PRICKLE1,PRR5,PRR5L,PSMB8,PSMB9,RAMP1,RDX,RNF217,RTN1,SERPINB6,SLC6A9,SNX33,TIMP3,YEATS2,AZU1,CALR,CCNK,CD109,DLG3,DNAAF4,GAS6,MAD2L2,NCOA7,NDFIP1,NECAB3,OXR1,PECAM1,PPM1F,PTK2,RASAL2,RASGRP1,SH2D3A,SH2D3C,SORL1,VIM,VTN
GO:BP organonitrogen compound metabolic process	GO:1901564	0.03	AAK1,ABCG1,ADCY3,AGBL2,AMPD3,ANPEP,APOL3,BIN1,BMAL1,C8ORF88,CASK,CAT,CD74,CDC42BPA,CUL4A,CYFIP2,DGAT1,DUSP1,ELOVL7,ERG,FBXO22,FIS1,FOXO1,GATA3,GGTA1,GNL3,GPRC5B,HAL,HIBCH,HSD17B4,IFI16,IGF2BP3,IL6ST,INHBA,LAPTM5,LGMN,LMLN,LRP6,MAN1A1,MYL9,MMP19,MMP28,MRPL33,MTAP,NECAB1,NEDD4L,NEK11,NOXRED1,PAK1,PAWR,PIWIL4,PREP,PREPL,PRICKLE1,PRKCH,PRR5,PRR5L,PSMB8,PSMB9,QPRT,RAMP1,RBKS,RDX,RNF130,RNF217,RORA,RTN1,SCRN1,SERPINB6,SLC2A9,SLC39A8,SLC6A9,SNX33,ST3GAL3,TBCK,TIMP3,TRH,USP54,YEATS2,ABCB6,ADAMTS10,AKT3,ALDH1A1,ANXA1,AOC1,AZU1,C1R,C1RL,CALR,CARMIL1,CCNK,CD109,CFD,DLG3,DMPK,DNAAF4,DTX3,EXT2,FBXO44,GALNT12,GAS6,GSTA4,KLHL3,KRTCAP2,MAD2L2,MBOAT2,MYLK,NCOA7,NDFIP1,NECAB3,OSBPL5,OXR1,PECAM1,PFKFB2,PPM1F,PPP3CC,PRKD3,PTK2,PTP4A3,RASAL2,RASGRP1,RUNX3,SARM1,SH2D3A,SH2D3C,SLC38A1,SLC44A2,SOCS2,SORL1,ST6GALNAC4,ST8SIA4,TJP2,TSPAN14,USP30,VIM,VTN
GO:BP cardiac fibroblast cell development	GO:0060936	0.04	INHBA,SPRY1,TGFBR3
GO:BP regulation of cellular component organization	GO:0051128	0.04	AAK1,ARHGEF10,ATP10A,BIN1,CELSR1,CTTN,CYFIP2,DUSP1,EFNB1,ESAM,F2RL1,FIS1,GATA3,GBP5,GNAI1,GNL3,IFT140,INHBA,KANK1,MAP6D1,NEDD4L,PAK1,PPHLN1,PRKCH,PRR5,RDX,SCARF1,SCN1B,SHANK3,SNX33,SYNE2,TAL1,TGFBR3,TMC8,ADGRL1,ANXA1,AZU1,CALR,CARMIL1,DMPK,EVL,FHL1,GAS6,ITPKA,MAD2L2,MEF2C,MLLT3,PLXNC1,PPM1F,PPP3CC,PREX1,PTK2,RREB1,SARM1,SOCS2,SORBS3,SORL1,SREBF1,SRPX2,SYNE1,SYTL4,TRIM46,TRPM2,USP30,VIM,VTN,ZNF804A
GO:CC MHC protein complex	GO:0042611	<0.01	CD74,HLA-B,HLA-DPA1,HLA-DPB1,HLA-DRA,HLA-E
GO:CC plasma membrane bounded cell projection	GO:0120025	<0.01	AAK1,ABLIM1,ADCY3,AGBL2,BAALC,BIN1,CASK,CATSPER2,CDC42BPA,CNGB3,CRHBP,CTTN,CYFIP2,F2RL1,FANK1,HOMER3,IFT140,IL6ST,IQGK,KANK1,KCNA3,MYRIP,NAV1,PAK1,PEAR1,PROM1,RAB8A,RABGAP1L,RDX,ROBO4,RTN1,SCN1B,SHANK3,SPAG16,SV2A,SYNE2,ACAD9,ADGRG1,ADGRL1,ALDH1A1,ANXA1,CARMIL1,DLG3,DMWD,DNAAF4,DOCK4,EVL,HTR1F,ITPKA,MICALL2,MYLK,NDFIP1,PLCG1,PREX1,PTK2,RAB39B,SARM1,SH2D3C,SLC38A1,SYNE1,TRIM46,TSGA10,VIM,ZNF804A
GO:CC lysosomal membrane	GO:0005765	0.02	ANPEP,CD74,GNAI1,HLA-DPA1,HLA-DPB1,HLA-DRA,LAPTM5,MITF,MREG,SLC37A3,SLC39A8,STOM,ABCB6,AZU1,CPNE3,IFITM2,SLC44A2,TMEM163,TMEM63A,TRPM2