SUPPLEMENTARY APPENDIX

Integrative prognostic models predict long-term survival after immunochemotherapy in chronic lymphocytic leukemia patients

Johannes Bloehdorn,[1] Julia Krzykalla,[2] Karlheinz Holzmann,[3] Andreas Gerhardinger,[3] Billy Michael Chelliah Jebaraj,[1] Jasmin Bahlo,[4] Kathryn Humphrey,[5] Eugen Tausch,[1] Sandra Robrecht,[4] Daniel Mertens,[1,6] Christof Schneider,[1] Kirsten Fischer,[4] Michael Hallek,[4] Hartmut Döhner,[1] Axel Benner[2#] and Stephan Stilgenbauer[1#]

[1]Department of Internal Medicine III, University of Ulm, Ulm, Germany; [2]Division of Biostatistics, German Cancer Research Center, Heidelberg, Germany; [3]Genomics Core Facility, University of Ulm, Ulm, Germany; [4]Department I for Internal Medicine and Center for Integrated Oncology, University of Cologne, Cologne, Germany, [5]Clinical Development Oncology, Roche Products Ltd, Welwyn Garden City, UK and [6]German Cancer Research Center (DKFZ), Heidelberg, Germany

[#]AB and SS contributed equally as co-senior authors.

 doi:10.3324/haematol.2020.251561

Received: March 1, 2020.

Accepted: March 11, 2021.

Pre-published: March 18, 2021.

Correspondence: *STEPHAN STILGENBAUER* - stephan.stilgenbauer@uniklinik-ulm.de

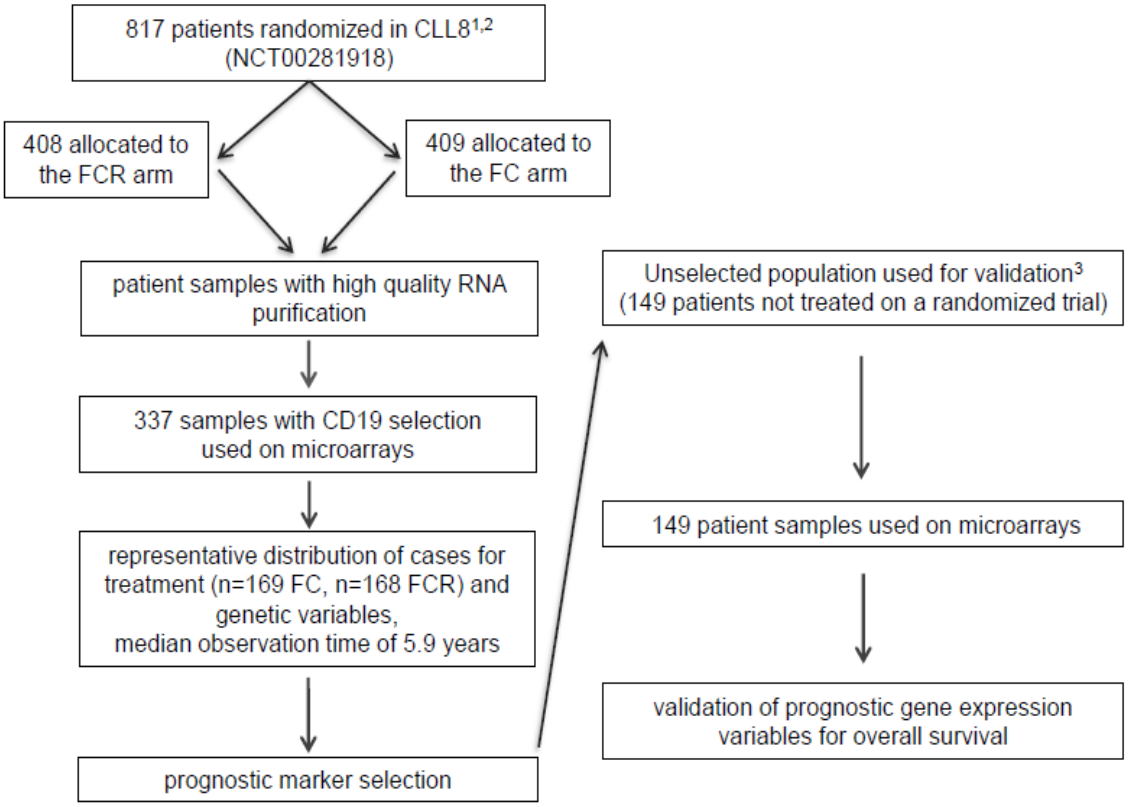
DATA SUPPLEMENTS

I. Supplementary Info, Tables and Figures

CONSORT diagram for the trial population.
Table S1) Confirmed prognostic variables included in the non-penalized Cox model for OS (S1A) and PFS (S1B) which serve as "reference model" for the comparison with penalized models including prognostic GEP variables. Data is based on 337 cases of the CLL8 cohort with existing gene expression data.
Table S2) Shown are the results of the application of a SCAD penalized Cox model including only gene expression data. S2A) Gene expression signature containing prognostic GEP variables which were selected in the penalized Cox model for OS and S2B) GEP variables selected in the penalized Cox model for PFS. Data is based on 337 cases of the CLL8 cohort with existing gene expression data.
Table S3) Shown are the results for the "equally penalized model" on all variables from the reference model and gene expression data for OS (S3A) and PFS (S3B). Due to the fact that we now include additional variables in the regression model a different result is expected if one or more variables from the reference model would remain after penalization. This is the case here. Interestingly, the differences between Tables S2 and Tables S3 are small, suggesting that the selected gene expression variables add considerable prognostic information to the one using the reference model alone. Besides genes shown in the Table S3B a transcript previously annotated as *LOC100510059* was selected. Respective cDNA shows high sequence homology to the HLA class II histocompatibility antigen, DQ(3) alpha chain precursor. Spearman`s rank correlation for *LOC100510059* and other transcripts represented on the array did not identify significantly correlated transcripts, therefore cross-hybridization of the *LOC100510059* transcript to other probes is highly unlikely.
Table S4) A) Association of *LDOC1* log2 expression levels <6 vs. >6 with categorical and B) continuous prognostic variables. C) Association of *L3MBTL4* log2 expression levels <7 vs. >7 with categorical and D) continuous prognostic variables.

Figure S1) Kaplan-Meier estimates for the lowest, the median, and the highest observed values of the prognostic variable combinations illustrating OS (left) and PFS (right) with regard to the "reference model" (confirmed prognostic variables only) and the "equally penalized model" (confirmed prognostic variables and GEP equally penalized). Here, analyses were focused on patient subgroups with A) age <60 years and B) treated with FCR.
Figure S2) Comparison of prediction error curves for OS based on the prognostic gene expression signatures established on two independent cohorts. The blue curve represents the prediction error curve of the prognostic gene expression signature established on 337 cases of the CLL8 trial (Table S2) which is here validated on the independent cohort. The red curve represents the prognostic gene signature established on the dataset of the validation cohort which was reported in the original publication. The Kaplan-Meier based prediction was used as reference (black curve).
Figure S3) Distribution for log2 expression of *RGS1*, *LDOC1* and *L3MBTL4* for the population of 337 cases from the CLL8 cohort with existing gene expression data.

CONSORT diagram for the trial population



1) Hallek et al., Lancet. 2010 376(9747):1164-74
2) Fischer et al., Blood. 2016 127(2):208-15.
3) Herold, T. et al., Leukemia 2011 25,1639

Abbreviations: FC = chemotherapy with fludarabine and cyclophosphamide, FCR = combined chemoimmunotherapy with FC plus rituximab

Table S1A) Reference model for OS

	HR	lower 95 CI	upper 95 CI	p.value
age [10 years]	1.26	0.98	1.61	0.07
sex	1.65	0.98	2.78	0.06
study medication	0.59	0.41	0.86	0.01
ECOG	1.42	0.97	2.07	0.07
log(white blood cell count)	0.81	0.64	1.03	0.09
log(thymidin kinase)	1.11	0.88	1.41	0.38
β2-microglobulin	1.13	0.99	1.29	0.06
IGHV mutation status	0.47	0.29	0.77	0.002
del(11q)	1.32	0.86	2.01	0.2
del(13q)	1.15	0.76	1.72	0.51
del(17p)	4.3	2.14	8.63	<0.001
trisomy 12	1.12	0.58	2.16	0.73
TP53 mutation status	2.4	1.31	4.23	0.004
NOTCH1 mutation status	0.66	0.31	1.42	0.29
SF3B1 mutation status	1.41	0.9	2.18	0.13

Table S1B) Reference model for PFS

	HR	lower 95 CI	upper 95 CI	p.value
age [10 years]	0.88	0.75	1.04	0.13
sex	1.72	1.21	2.44	0.002
study medication	0.45	0.34	0.6	<0.001
ECOG	1.12	0.86	1.5	0.42
log(white blood cell count)	1.01	0.85	1.2	0.91
log(thymidin kinase)	1.01	0.86	1.2	0.94
β2-microglobulin	1.1	0.99	1.22	0.07
IGHV mutation status	0.77	0.56	1.07	0.12
del(11q)	1.78	1.31	2.41	<0.001
del(13q)	1.05	0.78	1.42	0.75
del(17p)	5.36	2.8	10.3	<0.001
trisomy 12	0.97	0.6	1.6	0.91
TP53 mutation status	1.82	1.1	3.02	0.02
NOTCH1 mutation status	1.03	0.62	1.71	0.92
SF3B1 mutation status	1.6	1.15	2.23	0.005

Table S2A)

Gene	Coefficient	Effect for increased GEP	Localisation
CLEC2B	0.0550	adverse prognosis	12p13.31
RGS1	0.0882	adverse prognosis	1q31.2
LDOC1	0.0144	adverse prognosis	Xq27.1
L3MBTL4	0.1393	adverse prognosis	18p11.31
ZNF804A	-0.0678	favorable prognosis	2q32.1
PRKCA	0.1208	adverse prognosis	17q24.2
SGCE	0.0358	adverse prognosis	7q21.3
PYHIN1	-0.0028	favorable prognosis	1q23.1
DCLK2	0.0173	adverse prognosis	4q31.23
MLLT3	0.1361	adverse prognosis	9p21.3
IQGAP2	-0.0003	favorable prognosis	5q13.3
RBMS2	0.009	adverse prognosis	12q13.3
VSIG1	-0.0774	favorable prognosis	Xq22.3
BCAT1	0.0368	adverse prognosis	12p12.1
C4orf34	0.0938	adverse prognosis	4p14
MAP4K4	-0.0189	favorable prognosis	2q11.2
IL18	0.0855	adverse prognosis	11q23.1
CYBRD1	0.0121	adverse prognosis	2q31.1
MS4A6A	0.0296	adverse prognosis	11q12.2
CD72	0.0411	adverse prognosis	9p13.3

Table S2B)

Gene	Coefficient	Effect for increased GEP	Localisation
RGS1	0.0016	adverse prognosis	1q31.2
EIF1AY	0.0719	adverse prognosis	Yq11.223
LDOC1	0.0835	adverse prognosis	Xq27.1
L3MBTL4	0.1175	adverse prognosis	18p11.31
PHEX	0.007	adverse prognosis	Xp22.11
GPR34	-0.0104	favorable prognosis	Xp11.4
CD80	-0.015	favorable prognosis	3q13.33
FLT3	0.001	adverse prognosis	13q12.2
DCAF12	0.0631	adverse prognosis	9p13.3
PLD5	0.031	adverse prognosis	1q43
NSMCE1	0.0254	adverse prognosis	16p12.1
BCAT1	0.0419	adverse prognosis	12p12.1
PSD3	0.0057	adverse prognosis	8p22
MAP4K4	-0.001	favorable prognosis	2q11.2
GM2A	0.0314	adverse prognosis	5q33.1
CYBRD1	0.0044	adverse prognosis	2q31.1
CD72	0.0012	adverse prognosis	9p13.3
CCDC144B	0.0241	adverse prognosis	17p11.2
SORL1	-0.0102	favorable prognosis	11q24.1
TLE1	0.0006	adverse prognosis	9q21.32

Table S3A)

Gene	Coefficient	Effect for presence or increased GEP	Localisation
FCR treatment	-0.009	favorable prognosis	
B2M	0.0189	adverse prognosis	
del(17p)	1.9642	adverse prognosis	
CLEC2B	0.0154	adverse prognosis	12p13.31
RGS1	0.0221	adverse prognosis	1q31.2
LDOC1	0.1417	adverse prognosis	Xq27.1
L3MBTL4	0.0045	adverse prognosis	18p11.31
PRKCA	0.0353	adverse prognosis	17q24.2
FHL1	0.0143	adverse prognosis	Xq26.3
SGCE	0.0401	adverse prognosis	7q21.3
DCLK2	0.001	adverse prognosis	4q31.23
VSIG1	-0.043	favorable prognosis	Xq22.3
CD72	0.068	adverse prognosis	9p13.3

Table S3B)

Gene	Coefficient	Effect for presence or increased GEP	Localisation
FCR treatment	-0.5324	favorable prognosis	
del(11q)	0.0843	adverse prognosis	
del(17p)	2.0739	adverse prognosis	
SF3B1 mutation	0.0905	adverse prognosis	
RGS1	0.0464	adverse prognosis	1q31.2
EIF1AY	0.1299	adverse prognosis	Yq11.223
LDOC1	0.1972	adverse prognosis	Xq27.1
L3MBTL4	0.0064	adverse prognosis	18p11.31
DCAF12	0.0341	adverse prognosis	9p13.3
PLD5	0.0674	adverse prognosis	1q43
GTSF1L	0.0304	adverse prognosis	20q13.12
NIPAL2	0.0375	adverse prognosis	8q22.2
CYBRD1	0.02	adverse prognosis	2q31.1
ANXA1	-0.028	favorable prognosis	9q21.13
LOC100510059	-0.0077	favorable prognosis	

Table S4A)

Variable	Levels	n_{high}	$\%_{high}$	n_{low}	$\%_{low}$	n_{all}	$\%_{all}$
Sex	F	37	20.7	44	28.2	81	24.2
	M	142	79.3	112	71.8	254	75.8
$p = 0.13$	all	179	100.0	156	100.0	335	100.0
Treatment	FC	89	49.2	80	51.3	169	50.1
	FCR	92	50.8	76	48.7	168	49.9
$p = 0.74$	all	181	100.0	156	100.0	337	100.0
ECOG	0	93	52.8	71	48.6	164	50.9
	1,2	83	47.2	75	51.4	158	49.1
$p = 0.50$	all	176	100.0	146	100.0	322	100.0
IGHV (Immunoglobulin Heavy-Chain Variable) status	um	167	93.3	48	32.4	215	65.8
	m	12	6.7	100	67.6	112	34.2
$p < 0.0001$	all	179	100.0	148	100.0	327	100.0
del(11q)	ABSENCE	112	62.2	127	81.9	239	71.3
	PRESENCE	68	37.8	28	18.1	96	28.7
$p < 0.0001$	all	180	100.0	155	100.0	335	100.0
del(13q)	ABSENCE	87	48.3	41	26.4	128	38.2
	PRESENCE	93	51.7	114	73.5	207	61.8
$p < 0.0001$	all	180	100.0	155	100.0	335	100.0
del(17p)	ABSENCE	158	87.8	149	96.1	307	91.6
	PRESENCE	22	12.2	6	3.9	28	8.4
$p = 0.0057$	all	180	100.0	155	100.0	335	100.0
Trisomy 12	ABSENCE	159	88.3	138	89.0	297	88.7
	PRESENCE	21	11.7	17	11.0	38	11.3
$p = 0.86$	all	180	100.0	155	100.0	335	100.0
TP53 mutation status	ABSENCE	148	82.2	140	92.7	288	87.0
	PRESENCE	32	17.8	11	7.3	43	13.0
$p = 0.0051$	all	180	100.0	151	100.0	331	100.0
NOTCH1 mutation status	0	158	89.3	143	96.6	301	92.6
	1	19	10.7	5	3.4	24	7.4
$p = 0.02$	all	177	100.0	148	100.0	325	100.0
SF3B1 mutation status	0	133	75.1	122	83.0	255	78.7
	1	44	24.9	25	17.0	69	21.3
$p = 0.10$	all	177	100.0	147	100.0	324	100.0

Table S4B)

Variable	Levels	n	Min	q_1	$\tilde{x}$	$\bar{x}$	q_3	Max	s	IQR	#NA
Age	high	181	36.0	52.0	60.0	58.9	66.0	81.0	9.2	14.0	0
	low	156	35.0	54.8	62.0	60.3	65.0	78.0	8.0	10.2	0
$p = 0.23$	all	337	35.0	53.0	61.0	59.5	65.0	81.0	8.7	12.0	0
White blood cell count	high	181	8.6	55.1	96.2	113.1	145.0	498.4	81.9	89.9	0
	low	156	8.9	45.2	95.2	115.1	147.4	1500.0	133.7	102.2	0
$p = 0.52$	all	337	8.6	50.3	96.2	114.0	146.6	1500.0	108.8	96.3	0
Thymidine kinase	high	173	3.1	14.1	23.9	35.3	42.6	276.0	37.8	28.5	8
	low	150	2.7	8.8	15.4	40.9	33.8	970.0	111.1	25.0	6
$p = 0.00018$	all	323	2.7	10.9	20.0	37.9	37.4	970.0	80.5	26.4	14
Beta-2 microglobuline	high	173	1.1	2.2	2.8	3.2	3.8	8.0	1.3	1.6	8
	low	150	0.9	2.1	2.7	3.0	3.7	9.2	1.3	1.7	6
$p = 0.20$	all	323	0.9	2.2	2.8	3.1	3.8	9.2	1.3	1.6	14

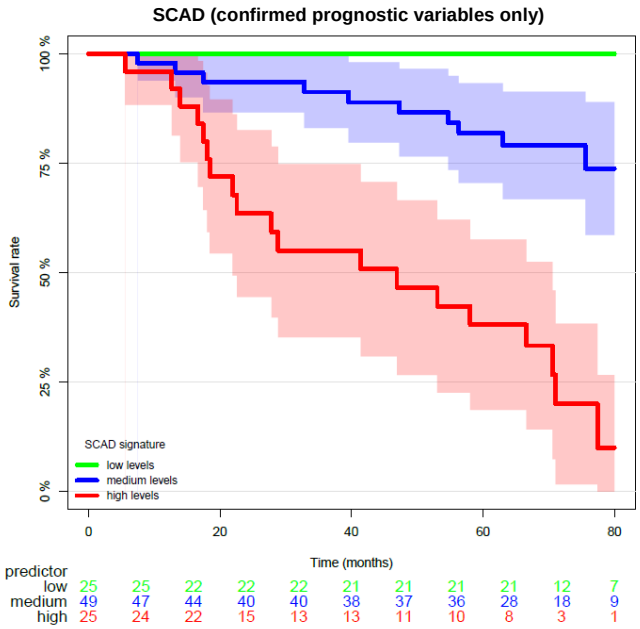
Table S4C)

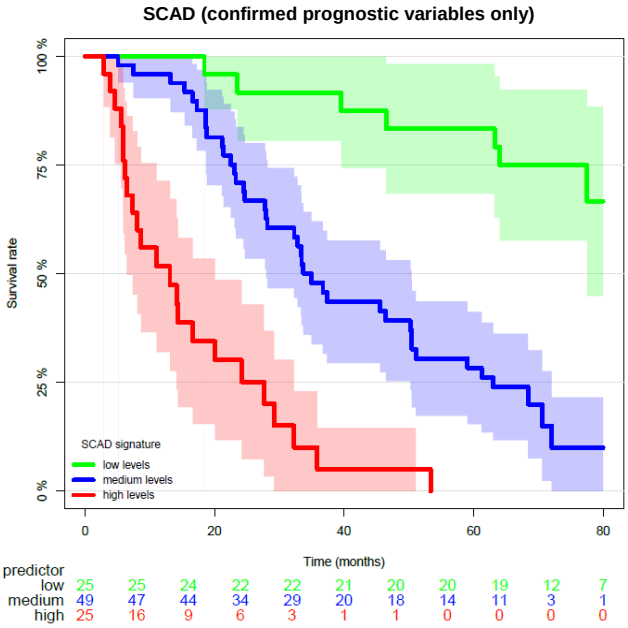
Variable	Levels	n_{high}	$\%_{high}$	n_{low}	$\%_{low}$	n_{all}	$\%_{all}$
Sex	F	37	19.9	44	29.5	81	24.2
	M	149	80.1	105	70.5	254	75.8
$p = 0.05$	all	186	100.0	149	100.0	335	100.0
Treatment	FC	94	50.0	75	50.3	169	50.1
	FCR	94	50.0	74	49.7	168	49.9
$p = 1.00$	all	188	100.0	149	100.0	337	100.0
ECOG	0	90	49.2	74	53.2	164	50.9
	1,2	93	50.8	65	46.8	158	49.1
$p = 0.50$	all	183	100.0	139	100.0	322	100.0
IGHV (Immunoglobulin Heavy-Chain Variable) status	um	167	89.8	48	34.0	215	65.8
	m	19	10.2	93	66.0	112	34.2
$p < 0.0001$	all	186	100.0	141	100.0	327	100.0
del(11q)	ABSENCE	120	64.2	119	80.4	239	71.3
	PRESENCE	67	35.8	29	19.6	96	28.7
$p = 0.0015$	all	187	100.0	148	100.0	335	100.0
del(13q)	ABSENCE	82	43.9	46	31.1	128	38.2
	PRESENCE	105	56.1	102	68.9	207	61.8
$p = 0.02$	all	187	100.0	148	100.0	335	100.0
del(17p)	ABSENCE	163	87.2	144	97.3	307	91.6
	PRESENCE	24	12.8	4	2.7	28	8.4
$p = 0.00064$	all	187	100.0	148	100.0	335	100.0
Trisomy 12	ABSENCE	163	87.2	134	90.5	297	88.7
	PRESENCE	24	12.8	14	9.5	38	11.3
$p = 0.39$	all	187	100.0	148	100.0	335	100.0
TP53 mutation status	ABSENCE	154	83.2	134	91.8	288	87.0
	PRESENCE	31	16.8	12	8.2	43	13.0
$p = 0.02$	all	185	100.0	146	100.0	331	100.0
NOTCH1 mutation status	0	167	91.8	134	93.7	301	92.6
	1	15	8.2	9	6.3	24	7.4
$p = 0.53$	all	182	100.0	143	100.0	325	100.0
SF3B1 mutation status	0	140	76.9	115	81.0	255	78.7
	1	42	23.1	27	19.0	69	21.3
$p = 0.41$	all	182	100.0	142	100.0	324	100.0

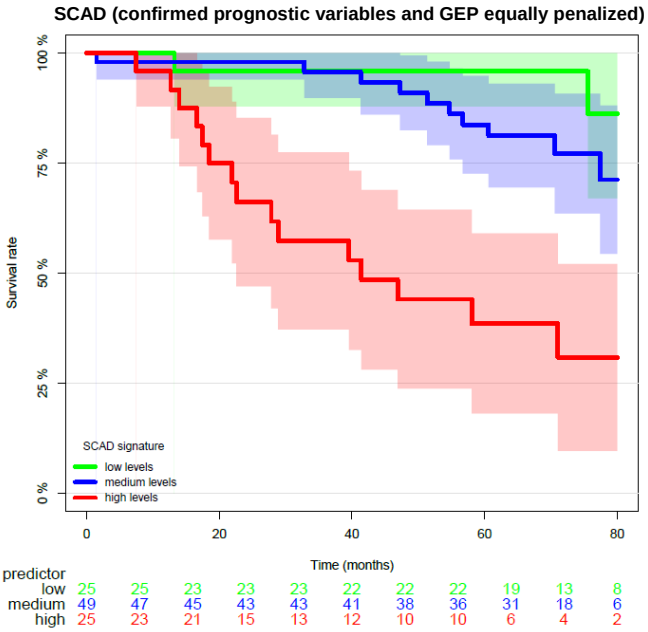
Table S4D)

Variable	Levels	n	Min	q_1	$\tilde{x}$	$\bar{x}$	q_3	Max	s	IQR	#NA
Age	high	188	35.0	53.0	61.0	59.5	66.2	81.0	9.2	13.2	0
	low	149	39.0	53.0	62.0	59.6	65.0	78.0	8.0	12.0	0
$p = 0.80$	all	337	35.0	53.0	61.0	59.5	65.0	81.0	8.7	12.0	0
White blood cell count	high	188	8.9	52.3	98.0	114.0	150.2	498.4	82.8	97.9	0
	low	149	8.6	46.9	90.1	114.0	141.3	1500.0	135.0	94.4	0
$p = 0.45$	all	337	8.6	50.3	96.2	114.0	146.6	1500.0	108.8	96.3	0
Thymidine kinase	high	182	2.8	12.7	21.7	34.0	39.6	289.0	38.1	26.8	6
	low	141	2.7	8.9	16.9	42.9	34.2	970.0	114.0	25.3	8
$p = 0.02$	all	323	2.7	10.9	20.0	37.9	37.4	970.0	80.5	26.4	14
Beta-2 microglobuline	high	182	1.1	2.2	2.8	3.2	3.9	9.2	1.4	1.6	6
	low	141	0.9	2.1	2.7	3.0	3.8	6.6	1.2	1.7	8
$p = 0.17$	all	323	0.9	2.2	2.8	3.1	3.8	9.2	1.3	1.6	14

Figure S1A) Age <60 years OS and PFS







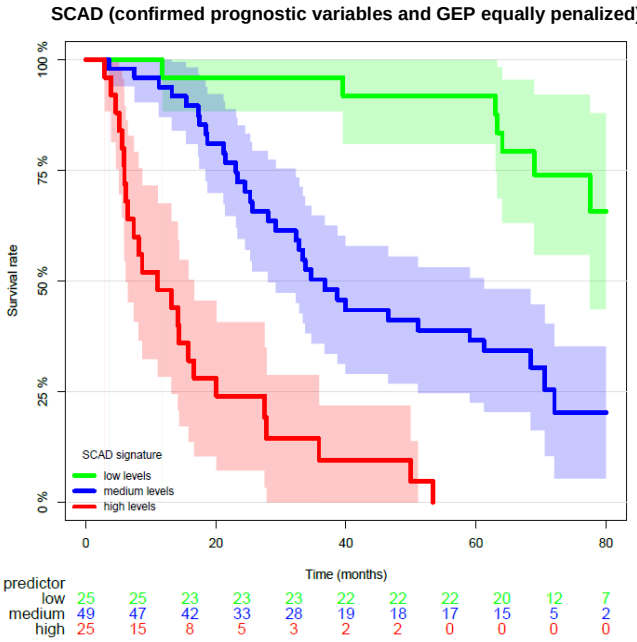
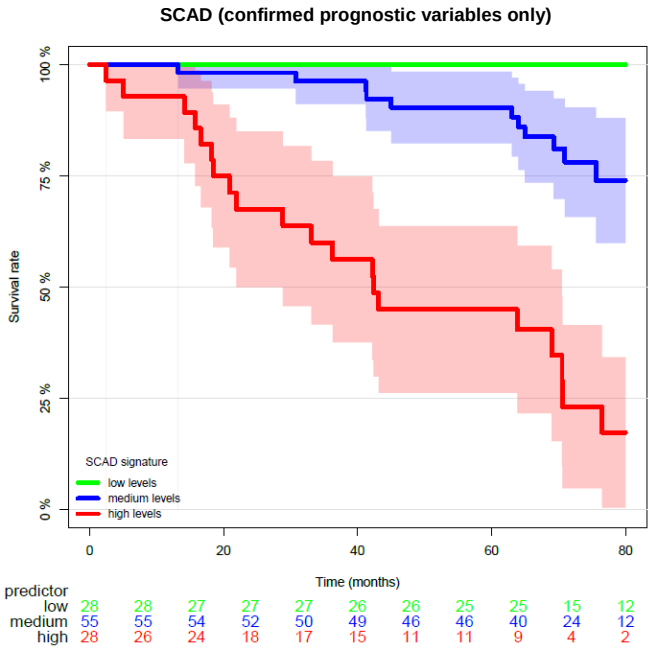
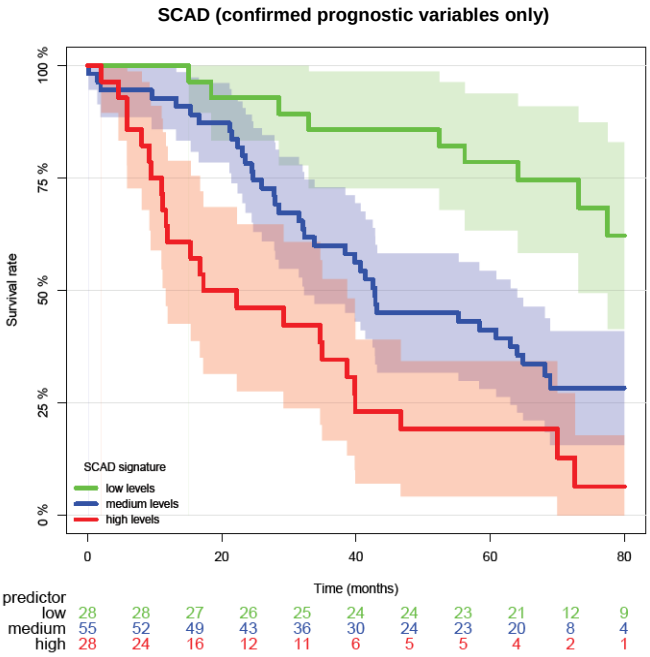
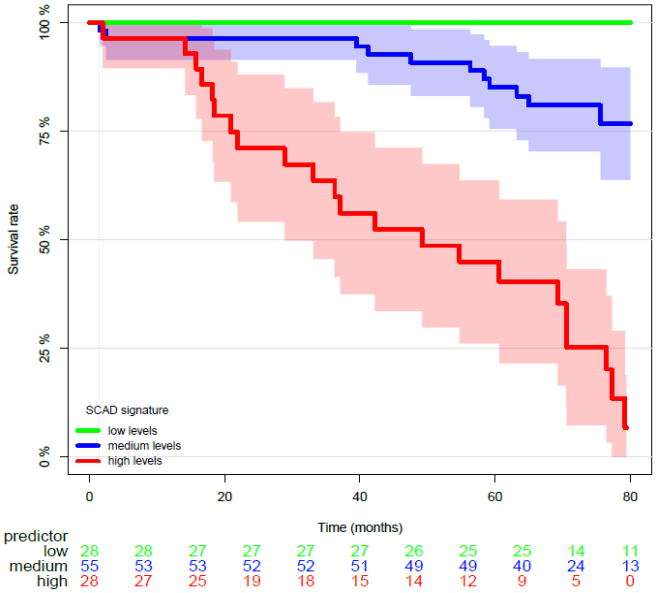


Figure S1B) FCR OS and PFS







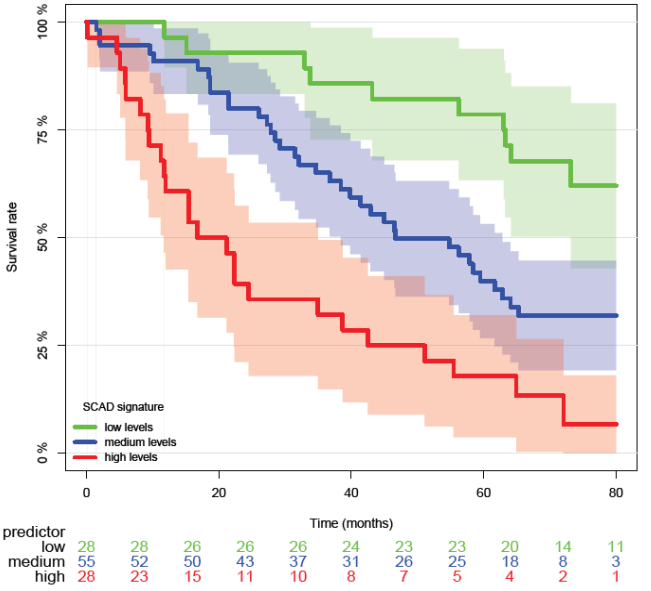


Figure S2)

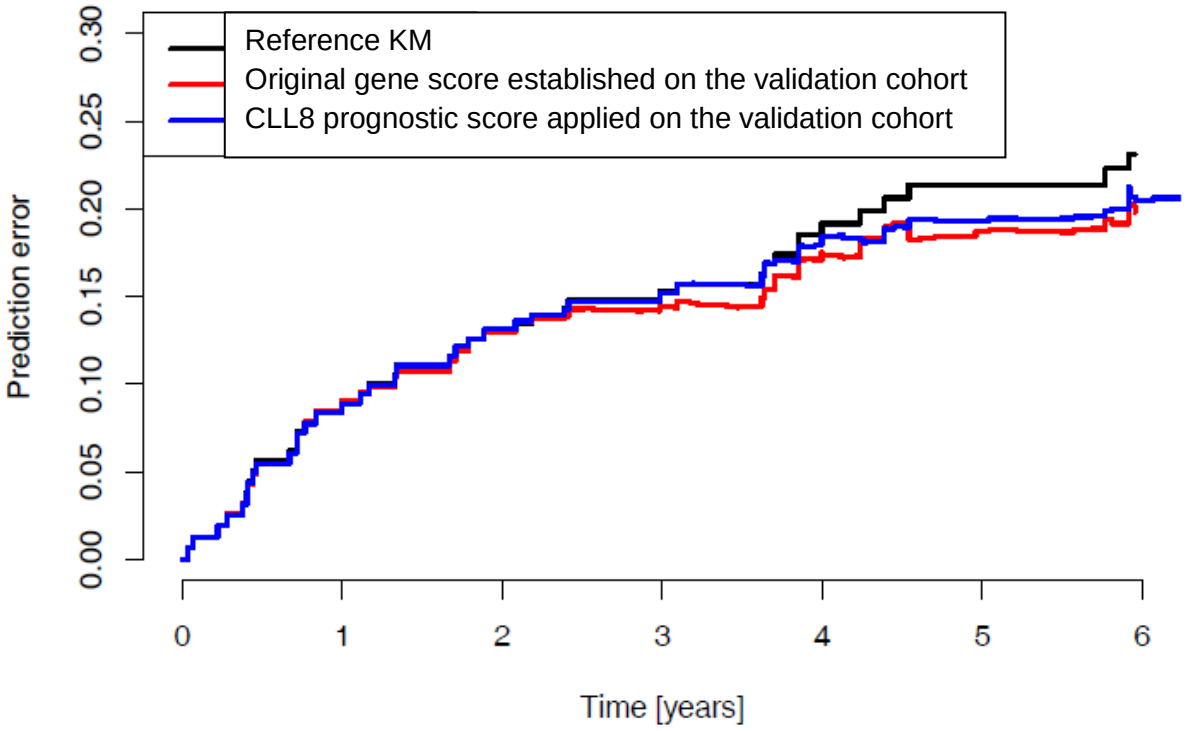
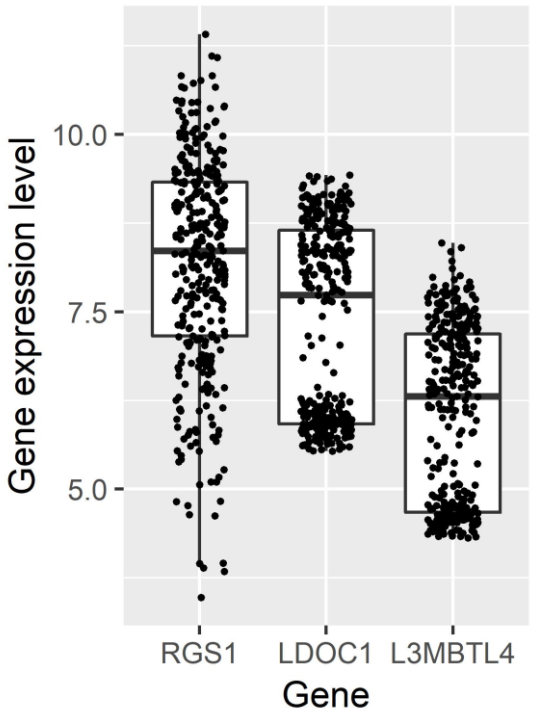


Figure S3)



Supplementary experimental information:

RNA isolation, quality assessment and GEP on Exon ST 1.0 Arrays: The experiment was conducted according to the manufacturer's protocol. In brief, 250 ng RNA per sample were amplified, transcribed to cDNA, fragmented and subsequently labeled with Biotin. Array hybridization was performed at 45°C for 16-18h in the Affymetrix GeneChip® Hybridization Oven 640, arrays were subsequently washed in the Fluidics Station 450 and scanned on the GeneChip scanner 3000 7G.
Expression data has been stored in GEO with an assigned analysis ID, raw data files include sort-status (S1 for CD19 sorted) and digit code of the registry (GSE126595). Data (n=149) of the validation cohort has been deposited previously (GSE22762) as described in the original publication[22].