Supplementary Material: Development and validation of a 6-gene signature for the prognosis of loco-regional control in patients with HPV-negative locally advanced HNSCC treated by postoperative radio(chemo)therapy

Shivaprasad Patil, Annett Linge, Marianne Grosser, Fabian Lohaus, Volker Gudziol, Max Kemper, Alexander Nowak, Dominik Haim, Inge Tinhofer, Volker Budach, Maja Guberina, Martin Stuschke, Panagiotis Balermpas, Claus Rödel, Henning Schäfer, Anca-Ligia Grosu, Amir Abdollahi, Jürgen Debus, Ute Ganswindt, Claus Belka, Steffi Pigorsch, Stephanie E. Combs, Simon Boeke, Daniel Zips, Gustavo B. Baretton, Michael Baumann, Mechthild Krause, Steffen Löck for the DKTK-ROG

Supplementary Table 1. List of 34 genes used to identify the gene signature after a series of filtering steps. The first 31 genes were selected after differential gene expression analysis and Cox regression. The last 3 differentially expressed genes were added as they were part of the 838 genes from the cancer gene census COSMIC database. COSMIC: Catalogue of Somatic Mutations in Cancer.

Number	Gene symbol
1	BGN
2	CAV1
3	CD248
4	CD44
5	COL12A1
6	COL1A1
7	COL1A2
8	COL6A3
9	FN1
10	GPX8
11	IGKV2-28
12	IGLV3-25
13	INHBA
14	ITGA6
15	KRT14
16	KTN1
17	LAMC2
18	MMP1
19	MMP10
20	MMP13
21	NDRG1
22	POSTN
23	RNU7-72P
24	SERPINE1
25	SFRP2
26	SPARC
27	SULF1
28	TGFBI
29	THBS1
30	TNC
31	VCAN
32	EGFR
33	IGKV1D-39
34	S100A7

Supplementary Table 2. Selection of genes using six feature selection methods (genes are ranked based on their occurrences).

Genes that were selected by 3 or more methods were included in the 6-gene signature. These genes are marked in bold.

	Feature selection methods 1000 bootstraps					
Rank	MRMR	MIM	Pearson	Spearman	Elastic net Cox	Univariate regression
1	***GPX8***	***GPX8***	***IGLV3-25***	***INHBA***	***GPX8***	***GPX8***
2	***CAV1***	*COL6A3*	*FN1*	***IGLV3-25***	***CAV1***	***CAV1***
3	***IGLV3-25***	***INHBA***	***INHBA***	*LAMC2*	*CD248*	*COL6A3*
4	*KTN1*	***SERPINE1***	***SERPINE1***	*SULF1*	*RNU7-72P*	***INHBA***
5	*RNU7-72P*	***TGFBI***	*SULF1*	***TGFBI***	***TGFBI***	***SERPINE1***

Supplementary Table 3. Univariable Cox regression of prognostic factors for loco-regional control, overall survival and freedom from distant metastases. HR: hazard ratio, CI: confidence interval, R: Residual tumour, ECE: Extracapsular extension, Grading: based on tumour abnormality, UICC: Union for International Cancer Control.

Parameter	**Loco-regional control**		**Overall survival**		**Freedom from distant metastases**	
	HR (95 % CI)	**p-value**	**HR (95 % CI)**	**p-value**	**HR (95 % CI)**	**p-value**
Tumour localisation (oral cavity vs others)	2.37 (1.11-5.06)	**0.027**	2.08 (1.21-3.59)	**0.008**	1.87 (0.93-3.79)	0.081
Gender (female vs male)	1.46 (0.64-3.33)	0.37	1.06 (0.56-1.98)	0.87	0.47 (0.95-1.36)	0.16
Age (years)	0.95 (0.91-0.99)	**0.018**	0.99 (0.96-1.02)	0.41	0.99 (0.95-1.03)	0.67
R status (1 vs 0)	1.21 (0.57-2.57)	0.62	1.05 (0.61-1.80)	0.86	0.98 (0.48-2.01)	0.96
ECE status (1 vs 0)	1.50 (0.69-3.25)	0.30	1.90 (1.08-3.35)	**0.025**	2.92 (1.30-6.52)	**0.009**
UICC stage (4 vs 1,2,3)	1.24 (0.47-3.29)	0.67	0.97 (0.51-1.84)	0.92	1.42 (0.55-3.71)	0.47
T stage (3,4 vs 1,2)	2.53 (1.17-5.48)	**0.019**	1.82 (1.06-3.12)	**0.030**	1.55 (0.77-3.15)	0.22
N stage (2,3 vs 0,1)	1.19 (0.75-1.87)	0.47	1.27 (0.69-2.38)	0.46	1.92 (1.06-3.48)	**0.032**
Grading (3 vs 1,2)	0.65 (0.27-1.53)	0.32	0.49 (0.26-0.93)	**0.029**	0.79 (0.36-1.72)	0.56
Dose (Gy)	1.10 (0.93-1.30)	0.29	1.09 (0.97-1.24)	0.16	0.92 (0.80-1.05)	0.20
Treatment time (days)	1.00 (0.93-1.10)	0.82	1.01 (0.95-1.07)	0.85	0.98 (0.91-1.06)	0.66
Smoking status (current or former vs never)	0.87 (0.20-3.82)	0.85	2.03 (0.48-8.60)	0.34	*	*
Alcohol status (current or former vs never)	0.72 (0.24-2.15)	0.55	0.89 (0.37-2.14)	0.79	1.87 (0.45-8.04)	0.40

*Cox model failed to converge

Supplementary Table 4. Comparison of gene expressions between the training and validation cohort for genes of the extended gene signature based on Affymetrix data.

Genes	Training cohort Mean (standard deviation)	Validation cohort Mean (standard deviation)
CAV1	6.13 (0.60)	6.12 (0.57)
GPX8	5.87 (0.54)	5.88 (0.52)
IGLV3-25	7.10 (1.44)	7.09 (1.22)
Metagene INHBA-SERPINE1	6.69 (0.95)	6.70 (0.93)
TGFBI	6.37 (0.64)	6.37 (0.59)
CD44	10.59 (1.13)	10.57 (1.16)
ADM	4.99 (0.22)	4.99 (0.20)
ALDOA	5.76 (0.15)	5.76 (0.14)
ANKRD37	4.63 (0.16)	4.63 (0.15)
BNIP3	6.11 (0.76)	6.12 (0.76)
BNIP3L	7.90 (0.39)	7.90 (0.40)
EGLN3	4.71 (0.23)	4.71 (0.21)
FAM162A	6.58 (0.54)	6.58 (0.52)
KCTD11	5.19 (0.31)	5.19 (0.28)
LOX	5.18 (0.32)	5.18 (0.30)
NDRG1	8.41 (0.88)	8.41 (0.86)
P4HA1	5.58 (0.23)	5.58 (0.22)
P4HA2	5.26 (0.13)	5.26 (0.11)
PDK1	5.72 (0.28)	5.72 (0.25)
PFKFB3	5.96 (0.18)	5.96 (0.17)
SLC2A1	6.33 (0.69)	6.34 (0.65)

Supplementary Table 5. Multivariable Cox regression of overall survival for the 6-gene signature, the extended gene signature, clinical parameters and their combinations. HR: hazard ratio, CI: confidence interval.

Parameter	coefficient (ß)	Overall survival HR (95 % CI)	p-value	ci Training (95 % CI)	ci Validation (95 % CI)
6-gene signature				0.74 (0.68-0.80)	0.58 (0.50-0.66)
CAV1	0.18	1.20 (0.81-1.76)	0.37		
GPX8	0.20	1.23 (0.88-1.71)	0.23		
IGLV3-25	-0.52	0.59 (0.42-0.84)	**0.004**		
Metagene *INHBA+SERPINE1*	0.32	1.38 (0.86-2.22)	0.19		
TGFBI	0.037	1.04 (0.64-1.68)	0.88		
Extended gene signature				0.69 (0.61-0.77)	0.61 (0.55-0.67)
6-gene classifier (≥0.53 vs <0.53)	1.35	3.84 (2.18-6.77)	**<0.001**		
CD44 classifier (≥-0.04 vs <-0.04)	0.37	1.46 (0.79-2.67)	0.23		
Hypoxia classifier (high vs low hypoxia)	0.31	1.37 (0.77-2.43)	0.29		
Clinical parameters				0.65 (0.57-0.73)	0.64 (0.58-0.70)
Tumour localization (oral cavity vs others)	0.67	1.96 (1.13-3.38)	**0.016**		
T stage (3,4 vs 1,2)	0.52	1.68 (0.98-2.90)	0.061		
6-gene signature and clinical parameters				0.72 (0.64-0.80)	0.64 (0.58-0.70)
6-gene classifier (≥0.53 vs <0.53)	1.36	3.88 (2.19-6.93)	**<0.001**		
Tumour localization (oral cavity vs others)	0.24	1.28 (0.72-2.27)	0.41		
T stage (3,4 vs 1,2)	0.50	1.64 (0.95-2.83)	0.074		
Extended gene signature and clinical parameters				0.72 (0.64-0.80)	0.65 (0.59-0.71)
6-gene classifier (≥0.53 vs <0.53)	1.23	3.43 (1.86-6.31)	**<0.001**		
CD44 classifier (≥-0.04 vs <-0.04)	0.42	1.52 (0.82-2.80)	0.18		
Hypoxia classifier (high vs low hypoxia)	0.27	1.31 (0.71-2.42)	0.39		
Tumour localization (oral cavity vs others)	0.29	1.34 (0.73-2.45)	0.34		
T stage (3,4 vs 1,2)	0.48	1.61 (0.92-2.84)	0.097		

Supplementary Table 6. Multivariable Cox regression of freedom from distant metastases for the 6-gene signature, the extended gene signature, clinical parameters and their combinations. HR: hazard ratio, CI: confidence interval.

Parameter	**coefficient (ß)**	**Distant metastases HR (95 % CI)**	**p-value**	**ci Training (95 % CI)**	**ci Validation (95 % CI)**
6-gene signature				0.64 (0.54-0.74)	0.60 (0.48-0.72)
CAV1	0.13	1.20 (0.81-1.76)	0.62		
GPX8	0.26	1.23 (0.88-1.71)	0.25		
IGLV3-25	-0.14	0.59 (0.42-0.84)	0.52		
Metagene *INHBA+SERPINE1*	0.047	1.38 (0.86-2.22)	0.88		
TGFBI	0.15	1.04 (0.64-1.68)	0.62		
Extended gene signature				0.69 (0.59-0.79)	0.61 (0.51-0.71)
6-gene classifier (≥0.63 vs <0.63)	1.35	3.86 (1.74-8.58)	**<0.001**		
CD44 classifier (≥-0.05 vs <-0.05)	0.93	2.52 (1.05-6.05)	**0.038**		
Hypoxia classifier (high vs low hypoxia)	-0.02	0.98 (0.46-2.10)	0.96		
Clinical parameters				0.61 (0.51-0.71)	0.60 (0.50-0.70)
Tumour localization (oral cavity vs others)	0.60	1.82 (0.90-3.70)	0.098		
T stage (3,4 vs 1,2)	0.40	1.49 (0.73-3.02)	0.28		
6-gene signature and clinical parameters				0.66 (0.54-0.78)	0.60 (0.50-0.70)
6-gene classifier (≥0.63 vs <0.63)	1.44	4.24 (1.89-9.47)	**<0.001**		
Tumour localization (oral cavity vs others)	0.43	1.54 (0.75-3.17)	0.24		
T stage (3,4 vs 1,2)	0.09	1.09 (0.52-2.30)	0.82		
Extended gene signature and clinical parameters				0.71 (0.61-0.71)	0.63 (0.53-0.73)
6-gene classifier (≥0.63 vs <0.63)	1.25	3.49 (1.55-7.86)	**0.003**		
CD44 classifier (≥-0.05 vs <-0.05)	0.95	2.59 (1.08-6.19)	**0.032**		
Hypoxia classifier (high vs low hypoxia)	-0.003	1.00 (0.44-2.24)	0.99		
Tumour localization (oral cavity vs others)	0.46	1.59 (0.77-3.26)	0.21		
T stage (3,4 vs 1,2)	0.19	1.21 (0.56-2.62)	0.63		

Supplementary Table 7. Multivariable Cox regression of overall survival for 6-gene signature, the extended gene signature with clinical parameters significant for overall survival and their combinations. HR: hazard ratio, CI: confidence interval, ECE: extracapsular extension.

Parameter	**coefficient (ß)**	**Overall survival HR (95 % CI)**	**p-value**	**ci Training (95 % CI)**	**ci Validation (95 % CI)**
6-gene signature				0.74 (0.68-0.80)	0.58 (0.50-0.66)
CAV1	0.18	1.20 (0.81-1.76)	0.37		
GPX8	0.20	1.23 (0.88-1.71)	0.23		
IGLV3-25	-0.52	0.59 (0.42-0.84)	**0.004**		
Metagene *INHBA+SERPINE1*	0.32	1.38 (0.86-2.22)	0.19		
TGFBI	0.037	1.04 (0.64-1.68)	0.88		
Extended gene signature				0.69 (0.61-0.77)	0.61 (0.55-0.67)
6-gene classifier (≥0.53 vs <0.53)	1.35	3.84 (2.18-6.77)	**<0.001**		
CD44 classifier (≥-0.04 vs <-0.04)	0.37	1.46 (0.79-2.67)	0.23		
Hypoxia classifier (high vs low hypoxia)	0.31	1.37 (0.77-2.43)	0.29		
Clinical parameters				0.67 (0.59-0.75)	0.61 (0.55-0.67)
Tumour localization (oral cavity vs others)	0.53	1.70 (0.97-2.97)	0.064		
T stage (3,4 vs 1,2)	0.48	1.62 (0.93-2.83)	0.089		
ECE status (1 vs 0)	0.65	1.92 (1.08-3.41)	**0.027**		
Grading (3 vs 1,2)	-0.63	0.53 (0.28-1.03)	0.062		
6-gene signature and clinical parameters				0.76 (0.70-0.82)	0.62 (0.54-0.70)
6-gene classifier (≥0.53 vs <0.53)	1.30	3.68 (2.06-6.57)	**<0.001**		
Tumour localization (oral cavity vs others)	0.15	1.16 (0.64-2.07)	0.62		
T stage (3,4 vs 1,2)	0.53	1.69 (0.98-2.94)	0.061		
ECE status (1 vs 0)	0.52	1.69 (0.94-3.02)	0.078		
Grading (3 vs 1,2)	-0.62	0.54 (0.28-1.03)	0.061		
Extended gene signature and clinical parameters				0.75 (0.69-0.81)	0.64 (0.58-0.70)
6-gene classifier (≥0.53 vs <0.53)	1.13	3.11 (1.68-5.74)	**<0.001**		
CD44 classifier (≥-0.04 vs <-0.04)	0.25	1.28 (0.66-2.49)	0.46		
Hypoxia classifier (high vs low hypoxia)	0.42	1.52 (0.81-2.86)	0.19		
Tumour localization (oral cavity vs others)	0.25	1.28 (0.69-2.38)	0.43		
T stage (3,4 vs 1,2)	0.47	1.60 (0.91-2.81)	0.10		
ECE status (1 vs 0)	0.55	1.73 (0.96-3.13)	0.067		
Grading (3 vs 1,2)	-0.60	0.55 (0.28-1.09)	0.084		

Supplementary Table 8. Multivariable Cox regression of freedom from distant metastases for 6-gene signature, the extended gene signature with clinical parameters significant for freedom from distant metastases and their combinations. HR: hazard ratio, CI: confidence interval, ECE: extracapsular extension.

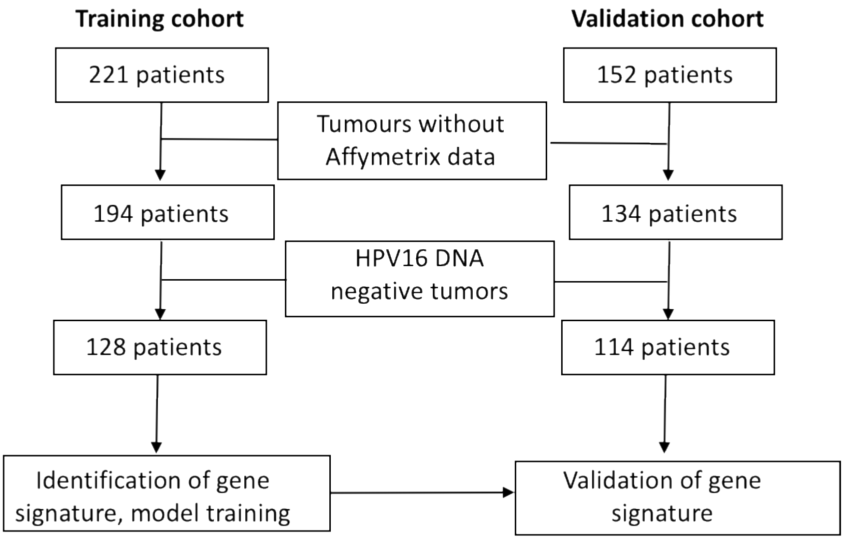
Parameter	**coefficient (ß)**	**Distant metastases HR (95 % CI)**	**p-value**	**ci Training (95 % CI)**	**ci Validation (95 % CI)**
6-gene signature				0.64 (0.54-0.74)	0.60 (0.48-0.72)
CAV1	0.13	1.20 (0.81-1.76)	0.62		
GPX8	0.26	1.23 (0.88-1.71)	0.25		
IGLV3-25	-0.14	0.59 (0.42-0.84)	0.52		
Metagene *INHBA+SERPINE1*	0.047	1.38 (0.86-2.22)	0.88		
TGFBI	0.15	1.04 (0.64-1.68)	0.62		
Extended gene signature				0.69 (0.59-0.79)	0.61 (0.51-0.71)
6-gene classifier (≥0.63 vs <0.63)	1.35	3.86 (1.74-8.58)	**<0.001**		
CD44 classifier (≥-0.05 vs <-0.05)	0.93	2.52 (1.05-6.05)	**0.038**		
Hypoxia classifier (high vs low hypoxia)	-0.02	0.98 (0.46-2.10)	0.96		
Clinical parameters				0.66 (0.58-0.74)	0.57 (0.45-0.69)
ECE status	0.91	2.48 (1.10-5.60)	**0.029**		
N stage (2,3 vs 0,1)	1.08	2.94 (0.88-9.83)	0.079		
6-gene signature and clinical parameters				0.72 (0.62-0.82)	0.58 (0.46-0.70)
6-gene classifier (≥0.63 vs <0.63)	1.71	5.55 (2.55-12.10)	**<0.001**		
ECE status (1 vs 0)	0.93	2.52 (1.09-5.85)	**0.031**		
N stage (2,3 vs 0,1)	1.10	3.01 (0.87-10.40)	0.081		
Extended gene signature and clinical parameters				0.75 (0.65-0.85)	0.61 (0.49-0.73)
6-gene classifier (≥0.63 vs <0.63)	1.38	3.97 (1.67-9.48)	**0.002**		
CD44 classifier (≥-0.05 vs <-0.05)	1.05	2.86 (1.15-7.12)	**0.024**		
Hypoxia classifier (high vs low hypoxia)	-0.12	0.88 (0.40-1.96)	0.76		
ECE status (1 vs 0)	1.06	2.90 (1.26-6.67)	**0.012**		
N stage (2,3 vs 0,1)	1.06	2.88 (0.85-9.73)	0.089		

Supplementary Table 9. Multivariable Cox regression of loco-regional tumour control for the 6-gene signature, the extended gene signature, clinical parameters and their combinations based on nanoString nCounter data. HR: hazard ratio, CI: confidence interval.

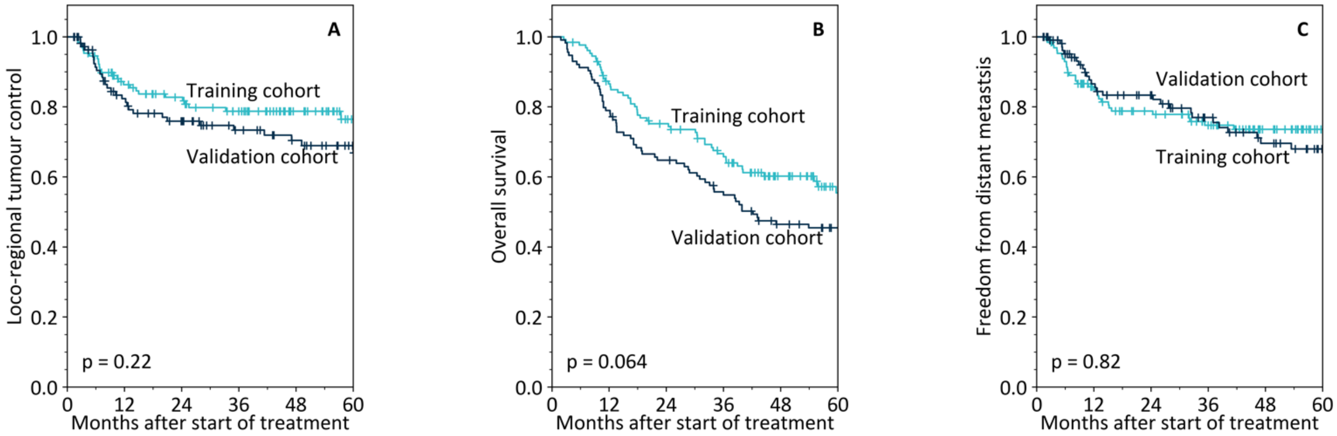
Parameter	**coefficient (ß)**	**Loco-regional control HR (95 % CI)**	**p-value**	**ci Training (95 % CI)**	**ci Validation (95 % CI)**
6-gene signature				0.79 (0.71-0.87)	0.62 (0.52-0.72)
CAV1	0.40	1.49 (0.88-2.51)	0.14		
GPX8	0.39	1.48 (0.91-2.39)	0.11		
IGLV3-25	-0.032	0.97 (0.63-1.49)	0.88		
Metagene *INHBA+SERPINE1*	0.32	1.38 (0.64-2.95)	0.41		
TGFBI	0.34	1.40 (0.76-2.59)	0.28		
Extended gene signature				0.81 (0.75-0.87)	0.67 (0.59-0.75)
6-gene classifier (≥1.44 vs <1.44)	1.89	6.62 (2.97-14.73)	**<0.001**		
CD44 classifier (≥0.01 vs <0.01)	0.53	1.70 (0.75-3.84)	0.20		
Hypoxia classifier (high vs low hypoxia)	0.93	2.53 (0.93-6.83)	0.068		
Clinical parameters				0.68 (0.58-0.78)	0.62 (0.52-0.72)
Tumour localization (oral cavity vs others)	0.74	2.10 (0.98-4.51)	0.053		
T stage (3,4 vs 1,2)	0.86	2.36 (1.09-5.14)	**0.032**		
6-gene signature and clinical parameters				0.74 (0.64-0.84)	0.66 (0.56-0.76)
6-gene classifier (≥1.44 vs <1.44)	1.71	5.52 (2.32-13.16)	**<0.001**		
Tumour localization (oral cavity vs others)	0.37	1.44 (0.63-3.28)	0.38		
T stage (3,4 vs 1,2)	0.53	1.70 (0.74-3.88)	0.21		
Extended gene signature and clinical parameters				0.82 (0.74-0.90)	0.69 (0.61-0.77)
6-gene classifier (≥1.44 vs <1.44)	1.65	5.23 (2.07-13.2)	**<0.001**		
CD44 classifier (≥0.01 vs <0.01)	0.47	1.59 (0.70-3.63)	0.27		
Hypoxia classifier (high vs low hypoxia)	0.92	2.52 (0.92-6.93)	0.073		
Tumour localization (oral cavity vs others)	0.41	1.51 (0.66-3.44)	0.33		
T stage (3,4 vs 1,2)	0.24	1.27 (0.52-3.11)	0.60		

Supplementary Table 10. Comparison of gene expressions between the training and validation cohort for genes of the extended gene signature based on nanoString nCounter data.

Genes	**Training cohort** **Mean (standard deviation)**	**Validation cohort** **Mean (standard deviation)**
CAV1	-2.01 (1.36)	-0.83 (1.22)
GPX8	-5.96 (0.80)	-5.86 (0.93)
IGLV3-25	-2.23 (1.95)	-2.69 (1.76)
Metagene *INHBA-SERPINE1*	-1.61(1.51)	-1.28 (1.5)
TGFBI	-0.15 (1.13)	-0.001 (1.21)
CD44	-6.18 (0.63)	-5.89 (0.85)
ADM	-2.19 (0.99)	-2.07 (1.08)
ALDOA	1.03 (0.51)	1.23 (0.51)
ANKRD37	-4.57 (0.77)	-3.76 (0.97)
BNIP3	-3.66 (0.89)	-3.73 (0.94)
BNIP3L	-1.95 (0.46)	-1.98 (0.52)
EGLN3	-3.61 (1.11)	-3.53 (1.19)
FAM162A	-2.62 (0.68)	-2.62 (0.68)
KCTD11	-2.06 (0.85)	-1.57 (0.82)
LOX	-3.56 (1.11)	-3.65 (1.16)
NDRG1	0.96 (1.15)	1.03 (1.28)
P4HA1	-3.35 (0.58)	-3.73 (0.69)
P4HA2	-2.80 (0.78)	-2.69 (0.77)
PDK1	3.78 (0.62)	-3.55 (0.68)
PFKFB3	-1.68 (0.61)	-1.26 (0.61)
SLC2A1	-0.46 (1.14)	-0.87 (1.37)



Supplementary Figure 1. Study design.



Patients at risk

Training cohort	128	99	86	71	50	30
Validation cohort	114	79	67	54	46	34

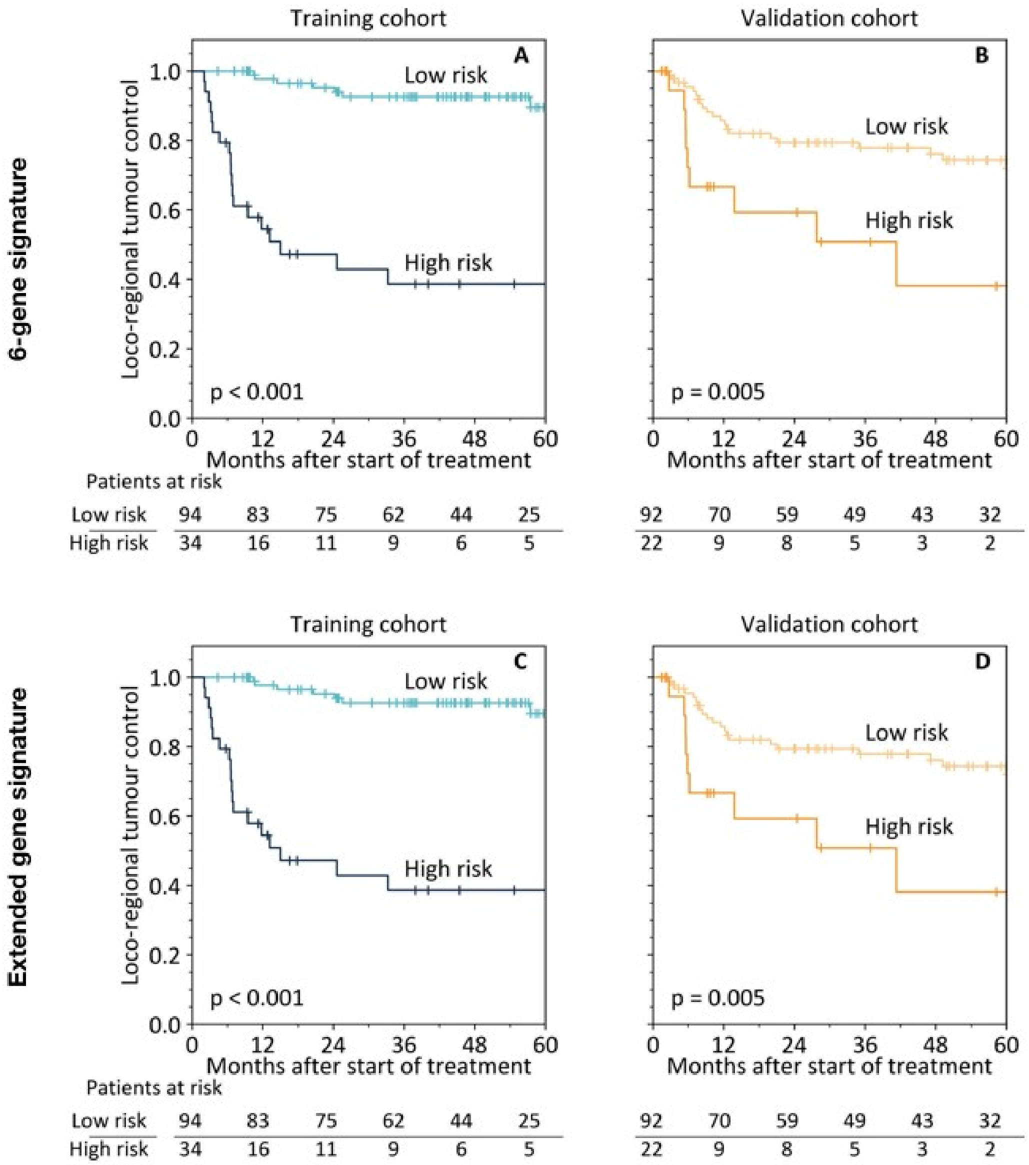
Patients at risk

Training cohort	128	107	91	75	52	31
Validation cohort	114	90	73	60	47	39

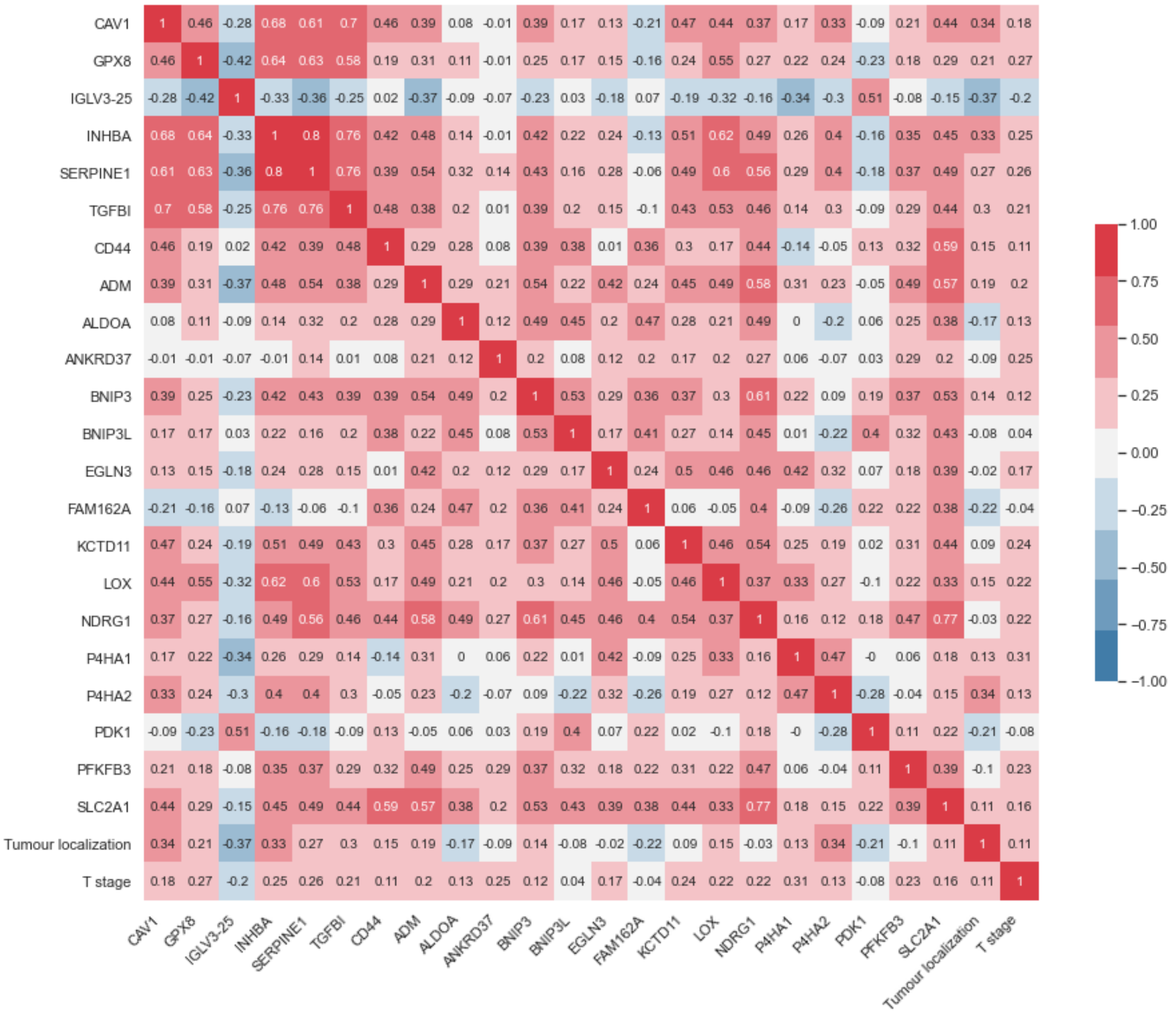
Patients at risk

Training cohort	128	98	83	69	48	29
Validation cohort	114	80	69	55	45	33

Supplementary Figure 2. Loco-regional tumour control **(A)**, overall survival **(B)** and freedom from distant metastasis **(C)** on the training and the validation cohort.

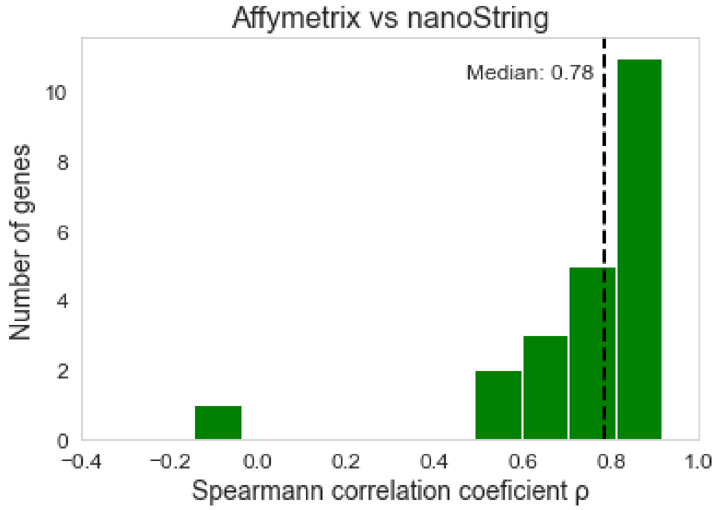


Supplementary Figure 3. Patient stratification by the 6-gene signature **(A, B)** and for the extended gene signature **(C, D)** for loco-regional tumour control (LRC) for the training cohort (left) and the validation cohort (right). Both signatures lead to the same stratification.

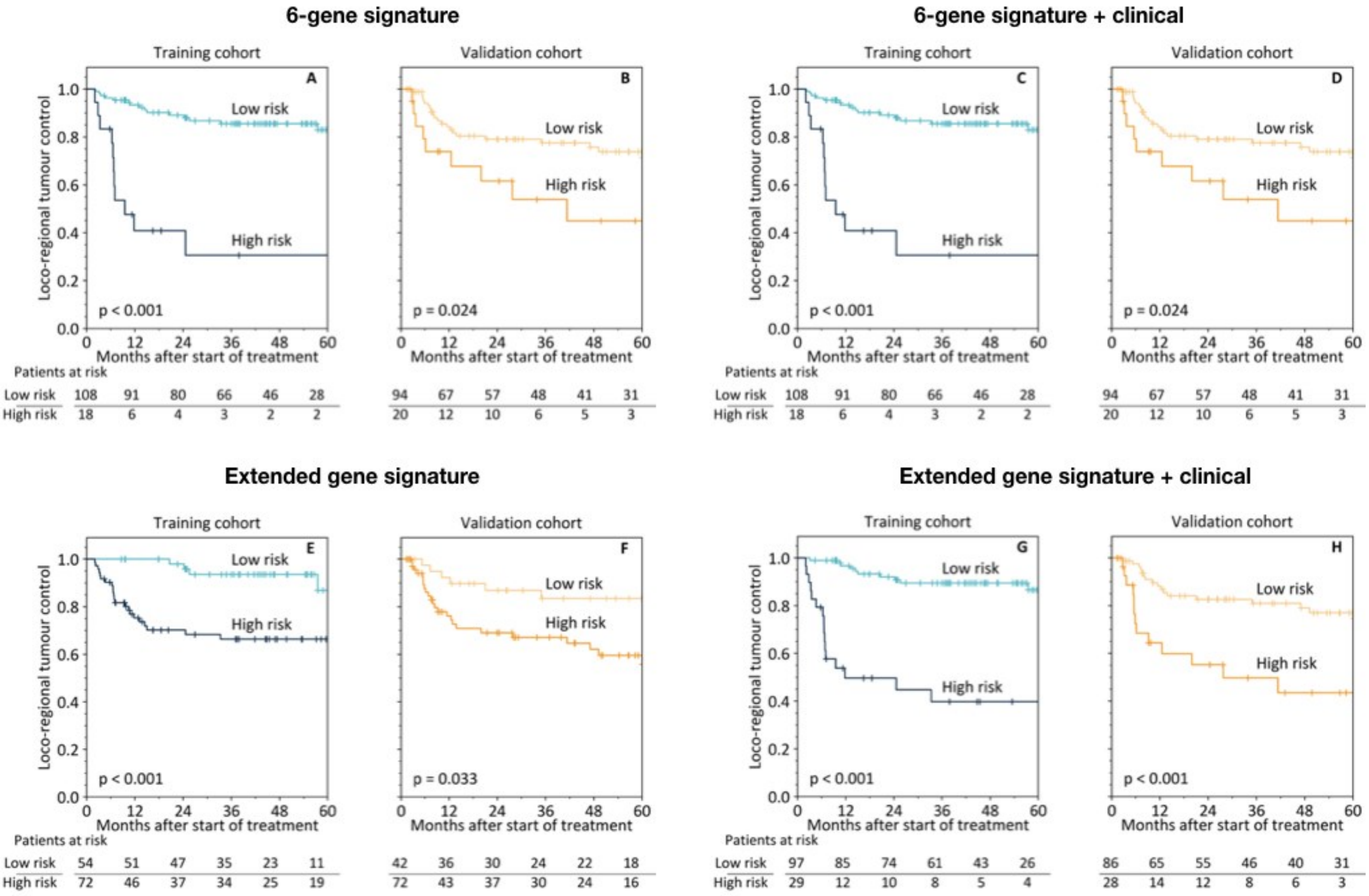


	CAV1	GPX8	IGLV3-25	INHBA	SERPINE1	TGFBI	CD44	ADM	ALDOA	ANKRD37	BNIP3	BNIP3L	EGLN3	FAM162A	KCTD11	LOX	NDRG1	P4HA1	P4HA2	PDK1	PFKFB3	SLC2A1	Tumour localization	T stage
CAV1	1	0.46	-0.28	0.68	0.61	0.7	0.46	0.39	0.08	-0.01	0.39	0.17	0.13	-0.21	0.47	0.44	0.37	0.17	0.33	-0.09	0.21	0.44	0.34	0.18
GPX8	0.46	1	-0.42	0.64	0.63	0.58	0.19	0.31	0.11	-0.01	0.25	0.17	0.15	-0.16	0.24	0.55	0.27	0.22	0.24	-0.23	0.18	0.29	0.21	0.27
IGLV3-25	-0.28	-0.42	1	-0.33	-0.36	-0.25	0.02	-0.37	-0.09	-0.07	-0.23	0.03	-0.18	0.07	-0.19	-0.32	-0.16	-0.34	-0.3	0.51	-0.08	-0.15	-0.37	-0.2
INHBA	0.68	0.64	-0.33	1	0.8	0.76	0.42	0.48	0.14	-0.01	0.42	0.22	0.24	-0.13	0.51	0.62	0.49	0.26	0.4	-0.16	0.35	0.45	0.33	0.25
SERPINE1	0.61	0.63	-0.36	0.8	1	0.76	0.39	0.54	0.32	0.14	0.43	0.16	0.28	-0.06	0.49	0.6	0.56	0.29	0.4	-0.18	0.37	0.49	0.27	0.26
TGFBI	0.7	0.58	-0.25	0.76	0.76	1	0.48	0.38	0.2	0.01	0.39	0.2	0.15	-0.1	0.43	0.53	0.46	0.14	0.3	-0.09	0.29	0.44	0.3	0.21
CD44	0.46	0.19	0.02	0.42	0.39	0.48	1	0.29	0.28	0.08	0.39	0.38	0.01	0.36	0.3	0.17	0.44	-0.14	-0.05	0.13	0.32	0.59	0.15	0.11
ADM	0.39	0.31	-0.37	0.48	0.54	0.38	0.29	1	0.29	0.21	0.54	0.22	0.42	0.24	0.45	0.49	0.58	0.31	0.23	-0.05	0.49	0.57	0.19	0.2
ALDOA	0.08	0.11	-0.09	0.14	0.32	0.2	0.28	0.29	1	0.12	0.49	0.45	0.2	0.47	0.28	0.21	0.49	0	-0.2	0.06	0.25	0.38	-0.17	0.13
ANKRD37	-0.01	-0.01	-0.07	-0.01	0.14	0.01	0.08	0.21	0.12	1	0.2	0.08	0.12	0.2	0.17	0.2	0.27	0.06	-0.07	0.03	0.29	0.2	-0.09	0.25
BNIP3	0.39	0.25	-0.23	0.42	0.43	0.39	0.39	0.54	0.49	0.2	1	0.53	0.29	0.36	0.37	0.3	0.61	0.22	0.09	0.19	0.37	0.53	0.14	0.12
BNIP3L	0.17	0.17	0.03	0.22	0.16	0.2	0.38	0.22	0.45	0.08	0.53	1	0.17	0.41	0.27	0.14	0.45	0.01	-0.22	0.4	0.32	0.43	-0.08	0.04
EGLN3	0.13	0.15	-0.18	0.24	0.28	0.15	0.01	0.42	0.2	0.12	0.29	0.17	1	0.24	0.5	0.46	0.46	0.42	0.32	0.07	0.18	0.39	-0.02	0.17
FAM162A	-0.21	-0.16	0.07	-0.13	-0.06	-0.1	0.36	0.24	0.47	0.2	0.36	0.41	0.24	1	0.06	-0.05	0.4	-0.09	-0.26	0.22	0.22	0.38	-0.22	-0.04
KCTD11	0.47	0.24	-0.19	0.51	0.49	0.43	0.3	0.45	0.28	0.17	0.37	0.27	0.5	0.06	1	0.46	0.54	0.25	0.19	0.02	0.31	0.44	0.09	0.24
LOX	0.44	0.55	-0.32	0.62	0.6	0.53	0.17	0.49	0.21	0.2	0.3	0.14	0.46	-0.05	0.46	1	0.37	0.33	0.27	-0.1	0.22	0.33	0.15	0.22
NDRG1	0.37	0.27	-0.16	0.49	0.56	0.46	0.44	0.58	0.49	0.27	0.61	0.45	0.46	0.4	0.54	0.37	1	0.16	0.12	0.18	0.47	0.77	-0.03	0.22
P4HA1	0.17	0.22	-0.34	0.26	0.29	0.14	-0.14	0.31	0	0.06	0.22	0.01	0.42	-0.09	0.25	0.33	0.16	1	0.47	-0	0.06	0.18	0.13	0.31
P4HA2	0.33	0.24	-0.3	0.4	0.4	0.3	-0.05	0.23	-0.2	-0.07	0.09	-0.22	0.32	-0.26	0.19	0.27	0.12	0.47	1	-0.28	-0.04	0.15	0.34	0.13
PDK1	-0.09	-0.23	0.51	-0.16	-0.18	-0.09	0.13	-0.05	0.06	0.03	0.19	0.4	0.07	0.22	0.02	-0.1	0.18	-0	-0.28	1	0.11	0.22	-0.21	-0.08
PFKFB3	0.21	0.18	-0.08	0.35	0.37	0.29	0.32	0.49	0.25	0.29	0.37	0.32	0.18	0.22	0.31	0.22	0.47	0.06	-0.04	0.11	1	0.39	-0.1	0.23
SLC2A1	0.44	0.29	-0.15	0.45	0.49	0.44	0.59	0.57	0.38	0.2	0.53	0.43	0.39	0.38	0.44	0.33	0.77	0.18	0.15	0.22	0.39	1	0.11	0.16
Tumour localization	0.34	0.21	-0.37	0.33	0.27	0.3	0.15	0.19	-0.17	-0.09	0.14	-0.08	-0.02	-0.22	0.09	0.15	-0.03	0.13	0.34	-0.21	-0.1	0.11	1	0.11
T stage	0.18	0.27	-0.2	0.25	0.26	0.21	0.11	0.2	0.13	0.25	0.12	0.04	0.17	-0.04	0.24	0.22	0.22	0.31	0.13	-0.08	0.23	0.16	0.11	1

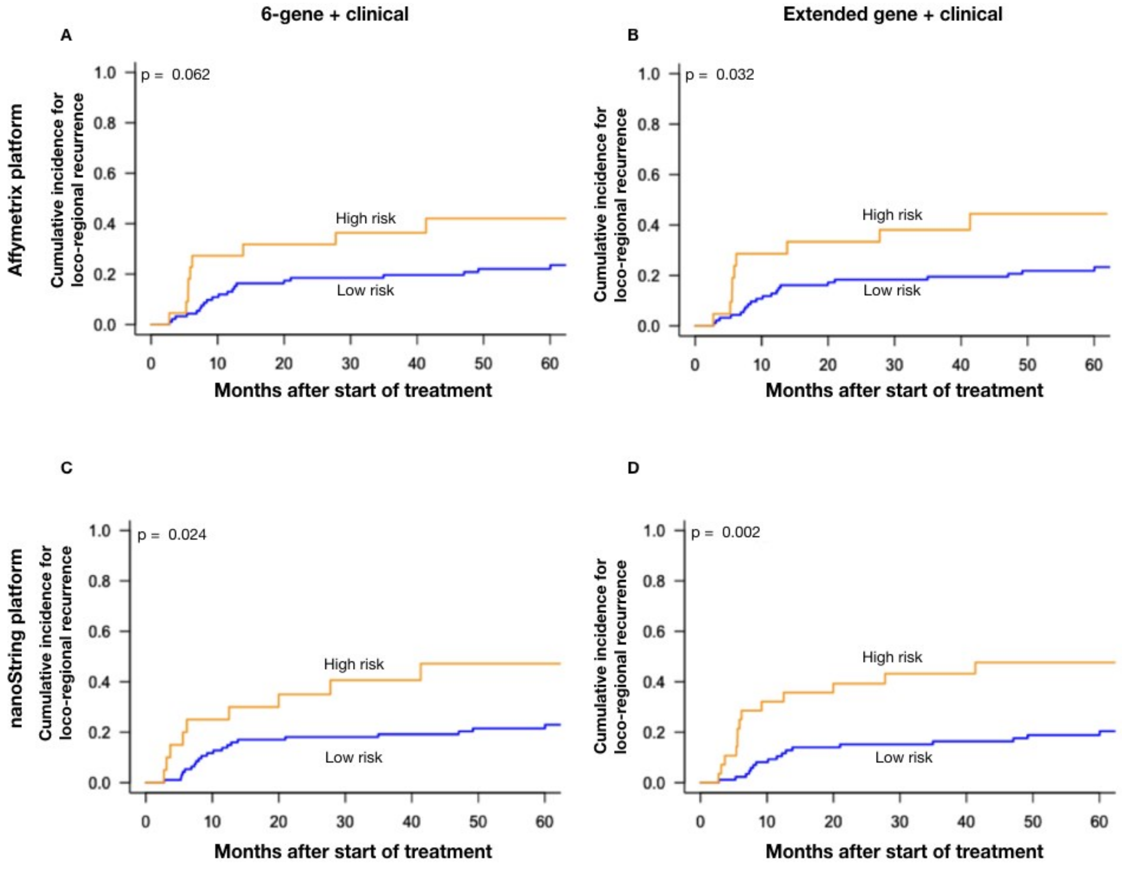
Supplementary Figure 4. Correlation matrix (Spearman correlation coefficient) between the genes from the 6-gene signature, *CD44*, the 15-gene hypoxia-associated signature and the clinical features (tumour localization and T stage) in the training cohort.



Supplementary Figure 5. Correlation of gene expressions between nanoString nCounter and Affymetrix GeneChip measurements of the 22 considered genes (6-gene signature, *CD44*, 15-gene hypoxia-associated signature). The dashed line indicates the median correlation. The outlier with low correlation of -0.15 is the *CD44* gene.



Supplementary Figure 6. Patient stratification by the 6-gene signature **(A, B)**, the six-gene signature combined with the clinical parameters tumour localization and T stage **(C, D)**, the extended gene signature (6-gene signature, *CD44* and the 15-gene hypoxia-associated signature) **(E, F)**, and the extended gene signature combined with the same clinical parameters **(G, H)** for loco-regional tumour control (LRC) based on nanoString nCounter data.



Supplementary Figure 7. Cumulative incidence for loco-regional recurrence accounting for death of all causes as competing risk. Results of the validation cohort are presented for the 6-gene signature combined with the clinical parameters tumour localization and T stage **(A, C)** and for the extended gene signature (6-gene signature, *CD44* and the 15-gene hypoxia-associated signature) combined with the same clinical parameters **(B, D)** based on Affymetrix data **(A, B)** and nanoString nCounter data **(C, D)**. The p-values originate from Gray's test.