

SUPPLEMENTARY INFO

Supplementary Table 1 GSM accession numbers from GEO datasets GSE6532 and GSE7390 used in the TRANSBIG cohort.

Supplementary Table 2 Variable expressed genes in our Mainz cohort. List of 2579 genes classified according to visual appearance of cluster within the heat map (**Fig. 1**).

Supplementary Table 3 GeneOntology terms overrepresented in individual gene clusters.

Supplementary Table 4 Median gene values in Mainz and calculated metagene values

Supplementary Table 5 Cox regression analyses of B-cell metagene expression and clinical parameter in relation to time to metastasis in the stratified Mainz cohort.

Supplementary Table 6 Cox regression analyses of metagene expression in the Mainz cohort excluding six medullary breast cancers

Supplementary Fig. 1 Gene expression and clinical features visualized in the Mainz cohort study using PCA.

Supplementary Fig. 2 Metagene expression and time to metastasis visualized in the Rotterdam cohort study by PCA.

Supplementary Fig. 3 Metagene expression and time to metastasis visualized in the TRANSBIG cohort by PCA.

Supplementary Fig. 4 Distribution of a) ESR1 and b) ERBB2 signal in the combined cohorts as well as the combined low and high proliferation subcohorts.

Supplementary Fig. 5 Survival analyses according to B-cell metagene expression

Supplementary Fig. 6 Distribution of metagenes within the patient cohorts

GSM65880	GSM65837	GSM65793	GSM178078	GSM178018	GSM177971	GSM177928
GSM65879	GSM65836	GSM65792	GSM178077	GSM178017	GSM177970	GSM177927
GSM65878	GSM65835	GSM65791	GSM178076	GSM178012	GSM177969	GSM177926
GSM65877	GSM65834	GSM65790	GSM178075	GSM178011	GSM177968	GSM177925
GSM65876	GSM65833	GSM65789	GSM178074	GSM178010	GSM177967	GSM177924
GSM65875	GSM65832	GSM65788	GSM178073	GSM178009	GSM177966	GSM177923
GSM65874	GSM65831	GSM65787	GSM178072	GSM178008	GSM177965	GSM177922
GSM65873	GSM65830	GSM65786	GSM178071	GSM178007	GSM177964	GSM177921
GSM65872	GSM65829	GSM65785	GSM178070	GSM178006	GSM177963	GSM177920
GSM65871	GSM65828	GSM65784	GSM178069	GSM178005	GSM177962	GSM177919
GSM65870	GSM65827	GSM65783	GSM178068	GSM178004	GSM177961	GSM177918
GSM65869	GSM65826	GSM65782	GSM178067	GSM178003	GSM177960	GSM177917
GSM65868	GSM65825	GSM65781	GSM178066	GSM178002	GSM177959	GSM177916
GSM65867	GSM65824	GSM65780	GSM178065	GSM178001	GSM177958	GSM177915
GSM65866	GSM65823	GSM65779	GSM178064	GSM178000	GSM177957	GSM177914
GSM65865	GSM65822	GSM65776	GSM178063	GSM177999	GSM177956	GSM177913
GSM65864	GSM65821	GSM65775	GSM178062	GSM177998	GSM177955	GSM177912
GSM65863	GSM65820	GSM65774	GSM178061	GSM177997	GSM177954	GSM177911
GSM65862	GSM65819	GSM65773	GSM178060	GSM177996	GSM177953	GSM177910
GSM65861	GSM65818	GSM65772	GSM178059	GSM177995	GSM177952	GSM177909
GSM65860	GSM65817	GSM65771	GSM178058	GSM177994	GSM177951	GSM177908
GSM65859	GSM65816	GSM65770	GSM178057	GSM177993	GSM177950	GSM177907
GSM65858	GSM65815	GSM65769	GSM178056	GSM177992	GSM177949	GSM177906
GSM65857	GSM65814	GSM65768	GSM178055	GSM177991	GSM177948	GSM177905
GSM65856	GSM65813	GSM65767	GSM178054	GSM177990	GSM177947	GSM177904
GSM65855	GSM65812	GSM65766	GSM178053	GSM177989	GSM177946	GSM177903
GSM65854	GSM65811	GSM65765	GSM178052	GSM177988	GSM177945	GSM177902
GSM65853	GSM65810	GSM65764	GSM178051	GSM177987	GSM177944	GSM177901
GSM65852	GSM65808	GSM65763	GSM178050	GSM177986	GSM177943	GSM177900
GSM65851	GSM65807	GSM65762	GSM178049	GSM177985	GSM177942	GSM177899
GSM65850	GSM65806	GSM65761	GSM178048	GSM177984	GSM177941	GSM177898
GSM65849	GSM65805	GSM65760	GSM178047	GSM177983	GSM177940	GSM177897
GSM65848	GSM65804	GSM65758	GSM178046	GSM177982	GSM177939	GSM177896
GSM65847	GSM65803	GSM65757	GSM178038	GSM177981	GSM177938	GSM177895
GSM65846	GSM65802	GSM65756	GSM178037	GSM177980	GSM177937	GSM177894
GSM65845	GSM65801	GSM65755	GSM178036	GSM177979	GSM177936	GSM177893
GSM65844	GSM65800	GSM65754	GSM178025	GSM177978	GSM177935	GSM177892
GSM65843	GSM65799	GSM65753	GSM178024	GSM177977	GSM177934	GSM177891
GSM65842	GSM65798	GSM65752	GSM178023	GSM177976	GSM177933	GSM177890
GSM65841	GSM65797	GSM178082	GSM178022	GSM177975	GSM177932	GSM177889
GSM65840	GSM65796	GSM178081	GSM178021	GSM177974	GSM177931	GSM177888
GSM65839	GSM65795	GSM178080	GSM178020	GSM177973	GSM177930	GSM177887
GSM65838	GSM65794	GSM178079	GSM178019	GSM177972	GSM177929	GSM177886
						GSM177885

Supplementary Table 1 GSM accession numbers from samples used in the TRANSBIG cohort

Cluster name	Published marker genes	Top enriched GO terms (p-value)
Basal like	KRT5, KRT17	<ul style="list-style-type: none"> intermediate filament ($< 10^{-10}$) epidermis development ($< 10^{-5}$) structural constituent of cytoskeleton ($< 10^{-4}$) morphogenesis ($< 10^{-3}$) central nervous system development ($< 10^{-3}$) zinc ion binding ($< 10^{-3}$)
T-cell	TRA@, TRB@, TRD@, CD8A	<ul style="list-style-type: none"> immune response ($< 10^{-14}$) cellular defense response ($< 10^{-13}$) inflammatory response ($< 10^{-6}$) chemotaxis ($< 10^{-6}$) MHC protein binding ($< 10^{-5}$) intracellular signaling cascade ($< 10^{-5}$) T-cell activation ($< 10^{-4}$)
B-cell	IGHM, IGHG, IGKC	<ul style="list-style-type: none"> antigen binding ($< 10^{-30}$) immune response ($< 10^{-30}$) membrane fraction ($< 10^{-13}$)
Interferon	IFI27, IFI35, IFIT1	<ul style="list-style-type: none"> immune response ($< 10^{-9}$) response to virus ($< 10^{-8}$) nucleotidyltransferase activity ($< 10^{-5}$) RNA binding ($< 10^{-5}$) Cytoplasm ($< 10^{-4}$)
Proliferation	CCNB2, UBE2C, STK6, MKI67	<ul style="list-style-type: none"> Mitosis ($< 10^{-30}$) Cytokinesis ($< 10^{-21}$) Nucleus ($< 10^{-20}$) cell cycle ($< 10^{-14}$)
Luminal	ESR1, MAPT, PGR, SCUBE2	<ul style="list-style-type: none"> DNA binding ($< 10^{-3}$)
ERBB2	ERBB2, GRB7	<ul style="list-style-type: none"> Chromosome 17*
Stromal	COL1A1, COL5A1, COL6A1	<ul style="list-style-type: none"> extracellular matrix structural constituent ($< 10^{-28}$) extracellular matrix (sensu Metazoa) ($< 10^{-27}$) phosphate transport ($< 10^{-17}$) collagen ($< 10^{-16}$) cell adhesion ($< 10^{-15}$)
Normal like	FABP4, PPARG	<ul style="list-style-type: none"> muscle development ($< 10^{-5}$) insulin-like growth factor receptor binding ($< 10^{-4}$) trans-1,2-dihydrobenzene-1,2-diol dehydrogenase activity ($< 10^{-4}$) neurogenesis ($< 10^{-4}$)
Jun Fos	JUN, FOS	<ul style="list-style-type: none"> hemoglobin complex ($< 10^{-13}$) oxygen transporter activity ($< 10^{-12}$) oxygen transport ($< 10^{-12}$) heme binding ($< 10^{-10}$) oxygen binding ($< 10^{-7}$) transcription factor activity ($< 10^{-5}$)
Transcription		<ul style="list-style-type: none"> nuclear mRNA splicing, via spliceosome ($< 10^{-7}$) RNA binding ($< 10^{-6}$) nucleic acid binding ($< 10^{-5}$) RNA processing ($< 10^{-5}$)

Supplementary Table 3 GeneOntology terms overrepresented in individual gene clusters.

Mainz cohort

univariate	All Data (n = 200, 47 events)			High Proliferation (n = 95, 31 events)			Low Proliferation (n = 105, 16 events)		
	HR	P value	95% CI	HR	P value	95% CI	HR	P value	95% CI
Parameter									
B-Cell Meta	0.84	0.095	0.68-1.03	0.70	0.018	0.53-0.94	1.16	0.395	0.83-1.62
Age	0.99	0.501	0.97-1.02	1.00	0.995	0.97-1.03	0.99	0.713	0.95-1.03
Grading	1.86	0.015	1.13-3.07	1.20	0.592	0.61-2.37	2.12	0.141	0.78-5.77
ER Status	1.04	0.883	0.58-1.87	1.58	0.222	0.76-3.27	1.04	0.936	0.38-2.82
PR Status	0.80	0.499	0.41-1.54	0.49	0.327	0.12-2.05	2.07	0.160	0.75-5.69
T-Size	1.10	0.891	0.27-4.55	0.93	0.924	0.22-3.92	>10	0.995	0.00-Inf

multivariate	All Data			High Proliferation			Low Proliferation		
	HR	P value	95% CI	HR	P value	95% CI	HR	P value	95% CI
Parameter									
B-Cell Meta	0.80	0.043	0.64-0.99	0.68	0.018	0.50-0.94	1.15	0.399	0.83-1.58
Age	1.00	0.822	0.97-1.03	1.01	0.778	0.97-1.04	0.99	0.721	0.95-1.04
Grading	2.21	0.007	1.24-3.94	1.56	0.245	0.74-3.31	2.66	0.106	0.81-8.69
ER Status	1.11	0.757	0.57-2.16	1.50	0.335	0.66-3.44	1.26	0.668	0.43-3.69
PR Status	0.81	0.561	0.41-1.63	0.31	0.123	0.07-1.38	1.86	0.235	0.67-5.15
T-Size	1.64	0.509	0.38-7.13	1.03	0.971	0.24-4.37	>10	0.995	0.00-Inf

Supplementary Table 5 Cox regression analyses of B-cell metagene expression and clinical parameters in relation to time to metastasis in the Mainz cohort study (Age: continuous, Grade: ordered, ER: ≤ 2 vs. > 2 , PR: ≤ 2 vs. > 2 , T-Size: < 2 cm vs. ≥ 2 cm).

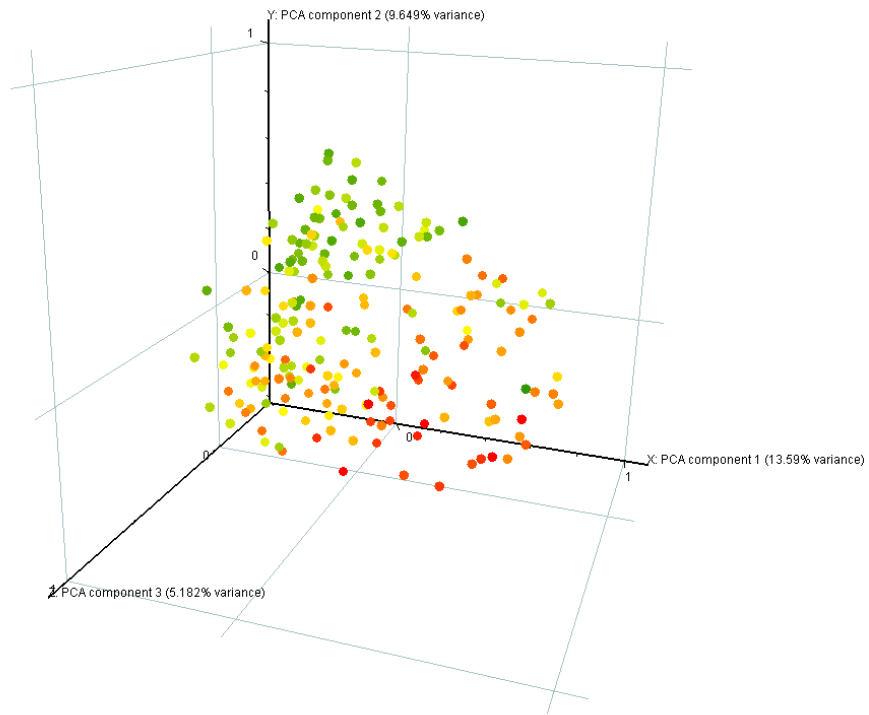
Mainz cohort non-medullary carcinomas only

univariate	All Data (n = 194, 46 events)			High Proliferation (n = 89, 30 events)			Low Proliferation (n = 105, 16 events)		
	HR	P value	95% CI	HR	P value	95% CI	HR	P value	95% CI
Metagene									
T-Cell	0.85	0.456	0.56-1.29	0.67	0.108	0.41-1.09	0.99	0.986	0.38-2.6
B-Cell	0.86	0.17	0.70-1.06	0.73	0.038	0.55-0.98	1.16	0.395	0.83-1.62
Proliferation	2.36	< 0.001	1.68-3.33	2.2	0.001	1.37-3.55	3.6	0.313	0.3-43.46
ER	0.42	0.016	0.21-0.85	0.68	0.41	0.27-1.70	0.62	0.571	0.12-3.20

multivariate	All Data			High Proliferation			Low Proliferation		
	HR	P value	95% CI	HR	P value	95% CI	HR	P value	95% CI
Metagene									
T-Cell	0.85	0.533	0.51-1.41	0.94	0.824	0.54-1.62	0.47	0.269	0.12-1.80
B-Cell	0.82	0.137	0.63-1.07	0.69	0.054	0.48-1.01	1.41	0.136	0.90-2.20
Proliferation	2.19	< 0.001	1.40-3.43	1.99	0.018	1.12-3.51	7.06	0.151	0.49-101
ER	0.62	0.311	0.24-1.56	0.58	0.338	0.19-1.77	0.54	0.519	0.08-3.58

Supplementary Table 6 Cox regression analyses of metagene expression in relation to time to metastasis in the Mainz cohort study of non-medullary carcinomas.

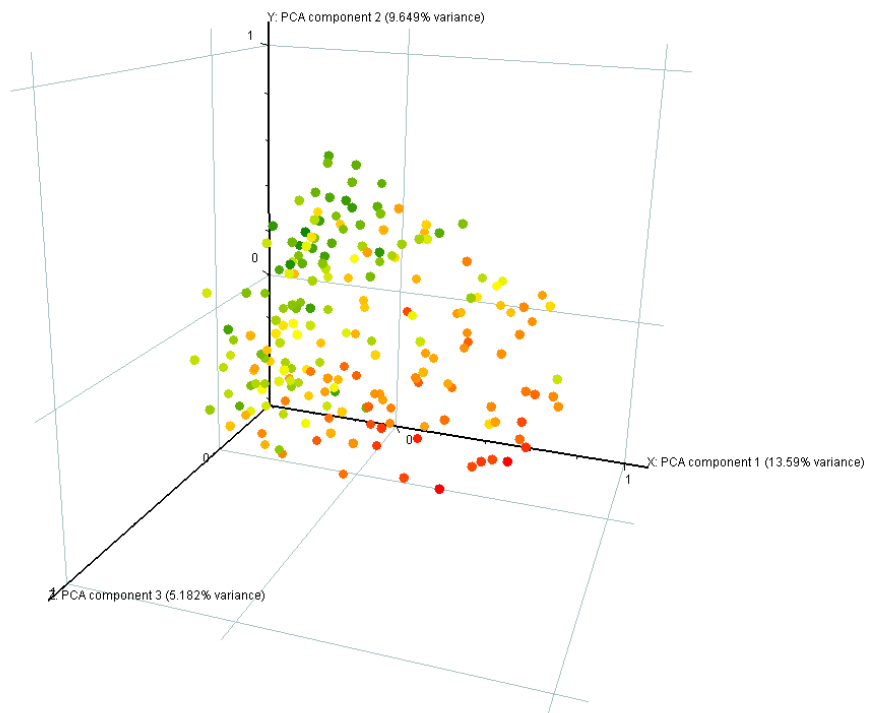
a)



X-axis: PCA component 1 (13.59% variance)
Y-axis: PCA component 2 (9.649% variance)
Z-axis: PCA component 3 (5.182% variance)

Conditions: Mainz Prognosis, Default Interpretation
Colored by: Gene 205034_at CCNE2

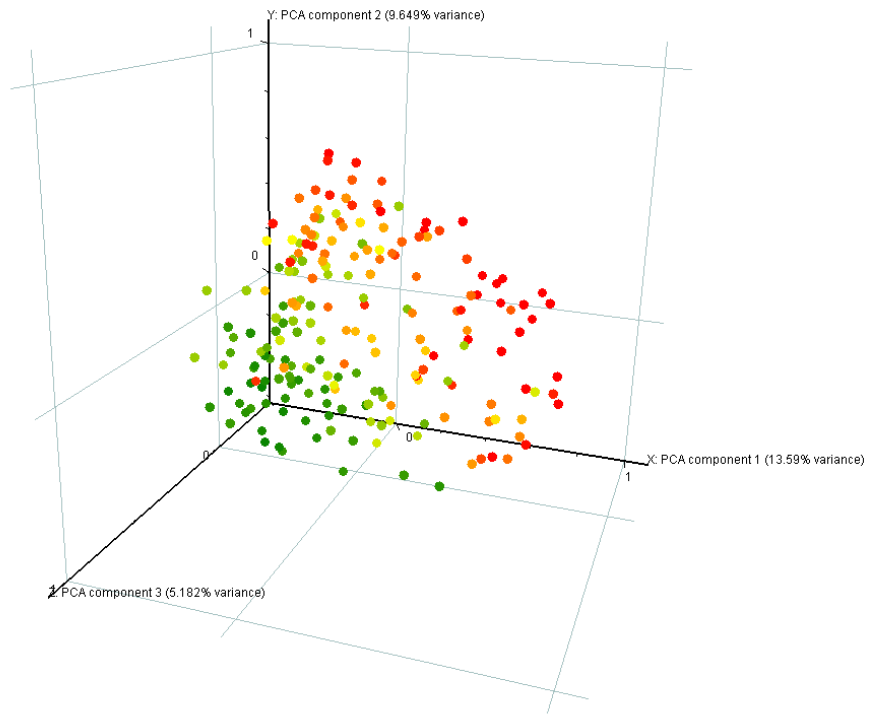
b)



X-axis: PCA component 1 (13.59% variance)
Y-axis: PCA component 2 (9.649% variance)
Z-axis: PCA component 3 (5.182% variance)

Conditions: Mainz Prognosis, Default Interpretation
Colored by: Gene 212022_s_at MK367

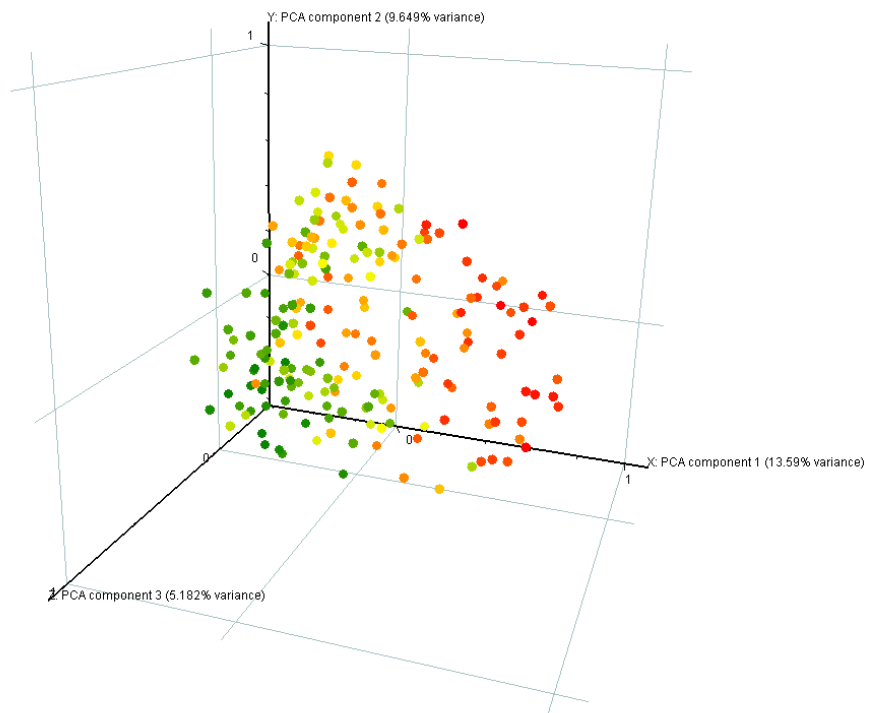
c)



X-axis: PCA component 1 (13.59% variance)
Y-axis: PCA component 2 (9.649% variance)
Z-axis: PCA component 3 (5.182% variance)

Conditions: Mainz Prognosis, Default Interpretation
Colored by: Gene 209374_s_at IGHM

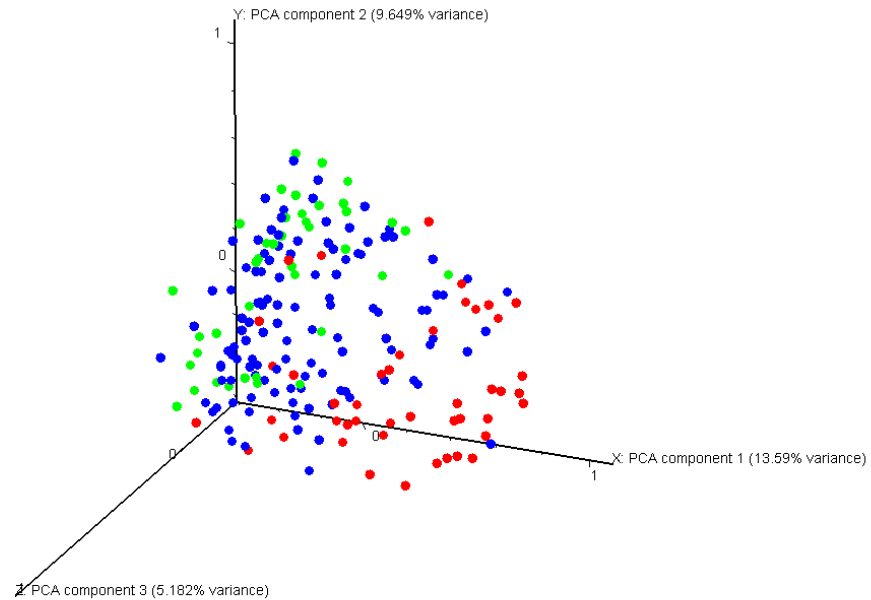
d)



X-axis: PCA component 1 (13.59% variance)
Y-axis: PCA component 2 (9.649% variance)
Z-axis: PCA component 3 (5.182% variance)

Conditions: Mainz Prognosis, Default Interpretation
Colored by: Gene 211430_s_at IGHG3

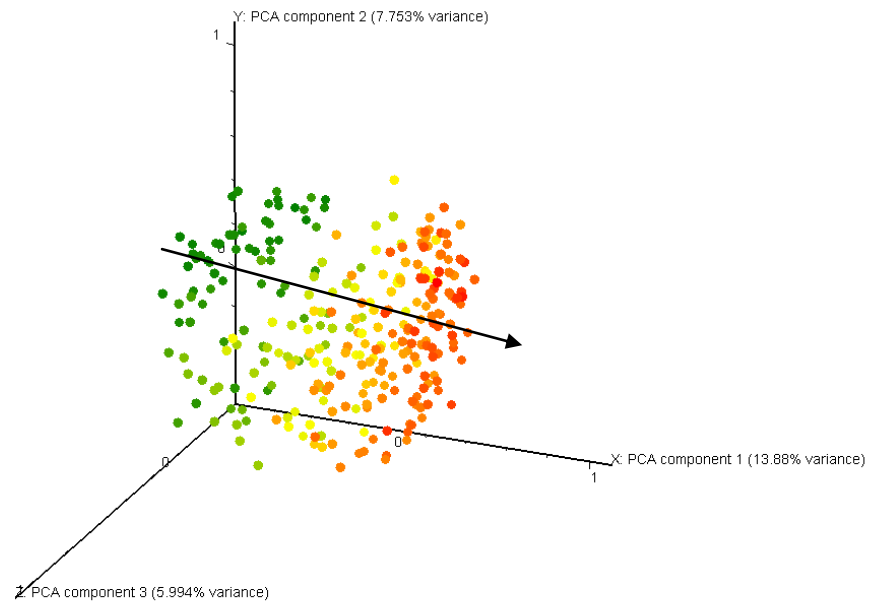
e)



Colored by: Parameter Grading

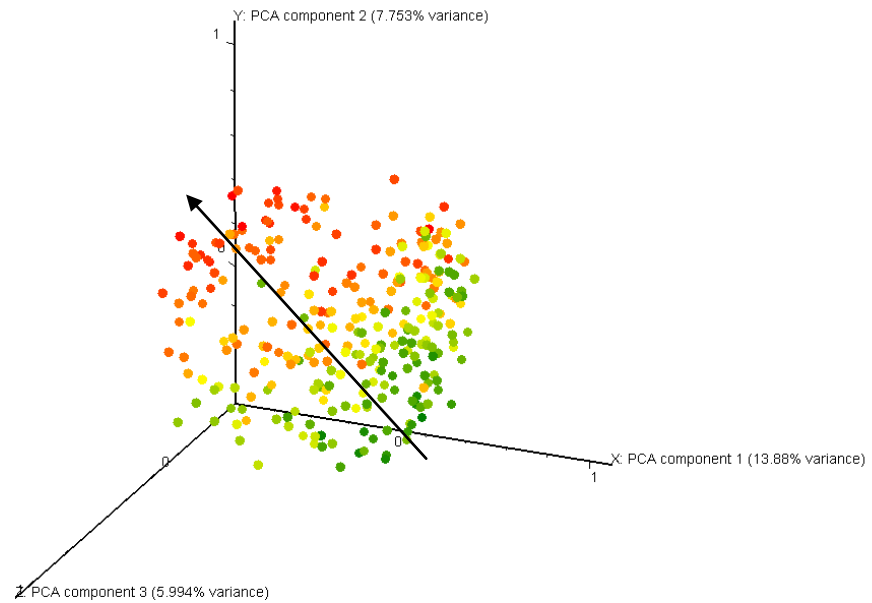
Supplementary Figure 1 Gene expression and grade of tumor visualized in the Mainz cohort using PCA. Relative expression of **a)** MKI67, **b)** CCNE2, **c)** IGHM and **d)** IGHG is shown in 200 samples distributed in the three dimensional space of PC1 - 3. High expression is indicated by red, intermediate by yellow and low by green color. In **e)** tumors of grade 1 are colored in green, grade 2 in blue and grade 3 in red.

a)



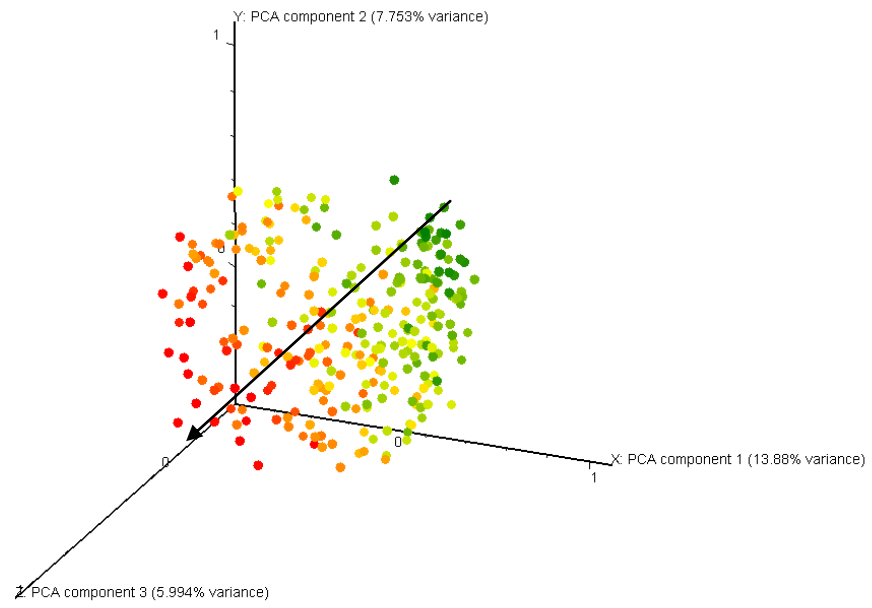
Colored by: Gene Median Cluster 6a

b)



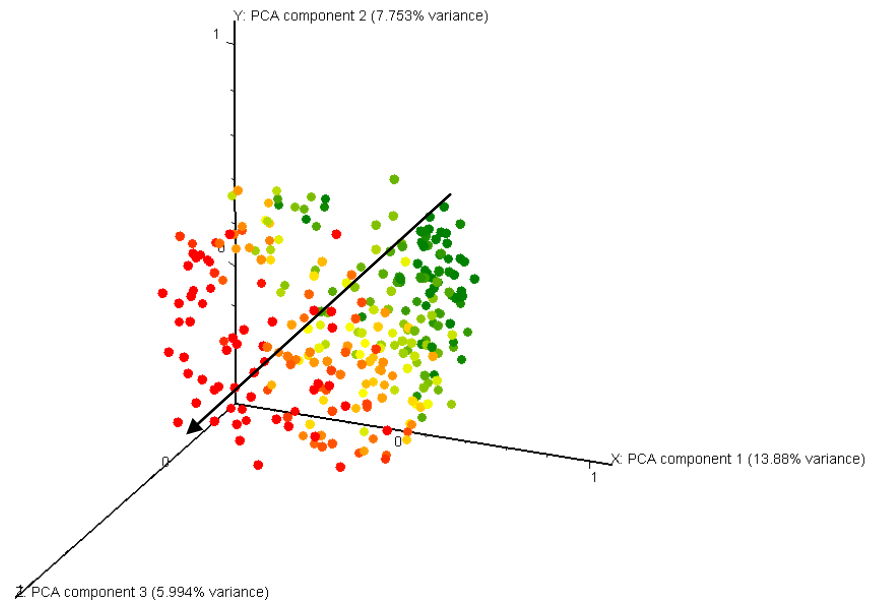
Colored by: Gene Median Cluster 5a

c)



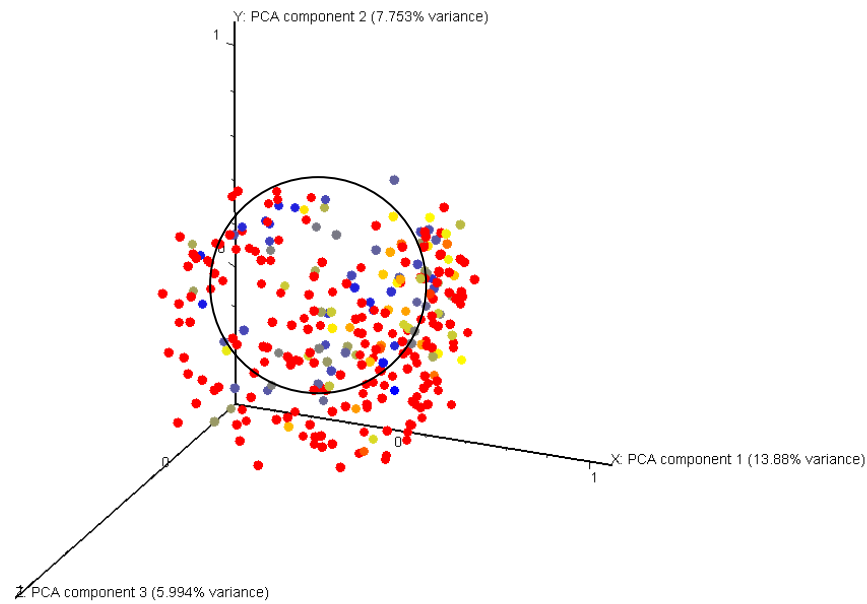
Colored by: Gene Median Cluster 2

d)



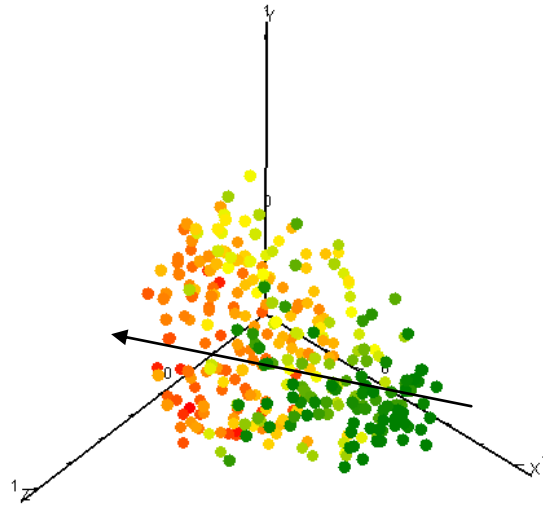
Colored by: Gene Median Cluster 3

e)



Supplementary Figure 2 Metagene expression and time to metastasis visualized in the Rotterdam cohort by PCA. Relative expression of **a)** ER, **b)** proliferation **c)** T-cell **d)** B-cell metagenes as well as **e)** time to distant metastasis in 286 samples (Rotterdam cohort) distributed in the three dimensional space of PC1 - 3 in the same genes used in our finding cohort. High expression is indicated by red, intermediate by yellow and low expression by green. The arrow indicates the direction of the “metagene axis” from low values to high values. All samples of patients who developed a distant metastasis are colored on a continuous scale from early (blue) to late (dark orange). Samples from patients that remained disease free until end of follow up were colored red, regardless of time. A region of accumulated early metastases is marked by a circle.

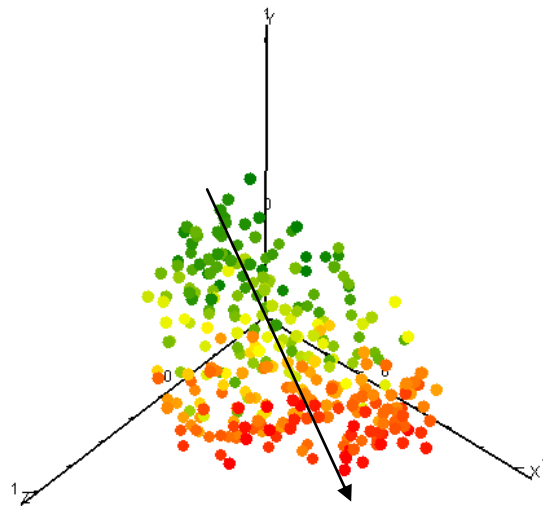
a)



X-axis: PCA component 1 (14.55% variance)
Y-axis: PCA component 2 (7% variance)
Z-axis: PCA component 3 (5.939% variance)

Conditions: TRANSBIG, Default Interpretation
Colored by: Gene ER Metagene

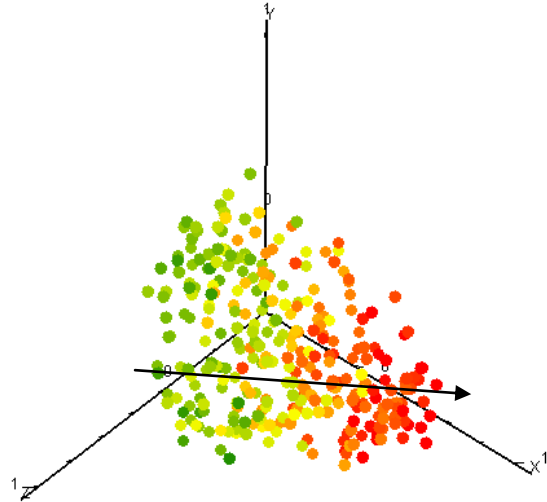
b)



X-axis: PCA component 1 (14.55% variance)
Y-axis: PCA component 2 (7% variance)
Z-axis: PCA component 3 (5.939% variance)

Conditions: TRANSBIG, Default Interpretation
Colored by: Gene Proliferation Metagene

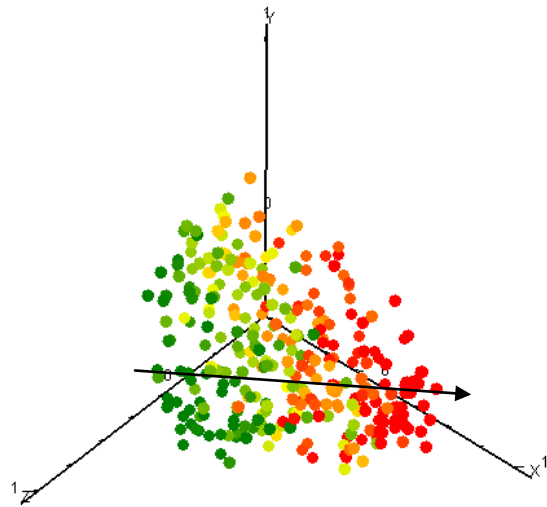
c)



X-axis: PCA component 1 (14.55% variance)
Y-axis: PCA component 2 (7% variance)
Z-axis: PCA component 3 (5.939% variance)

Conditions: TRANSBIG, Default Interpretation
Colored by: Gene T-cell Metagene

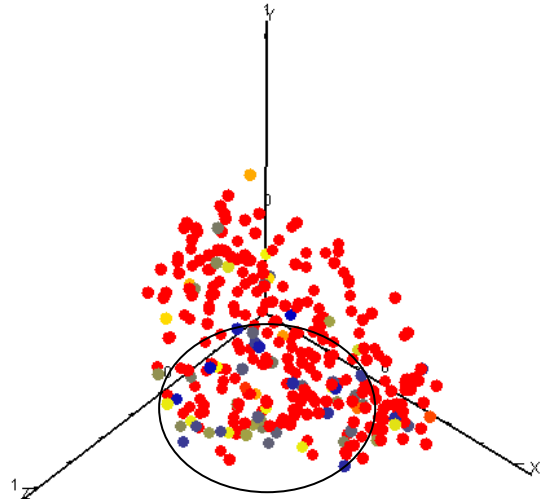
d)



X-axis: PCA component 1 (14.55% variance)
Y-axis: PCA component 2 (7% variance)
Z-axis: PCA component 3 (5.939% variance)

Conditions: TRANSBIG, Default Interpretation
Colored by: Gene B-cell Metagene

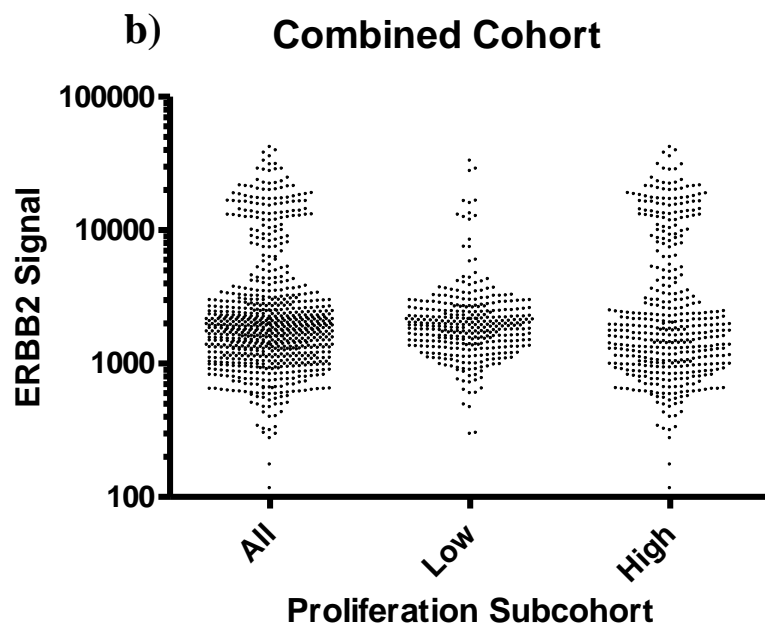
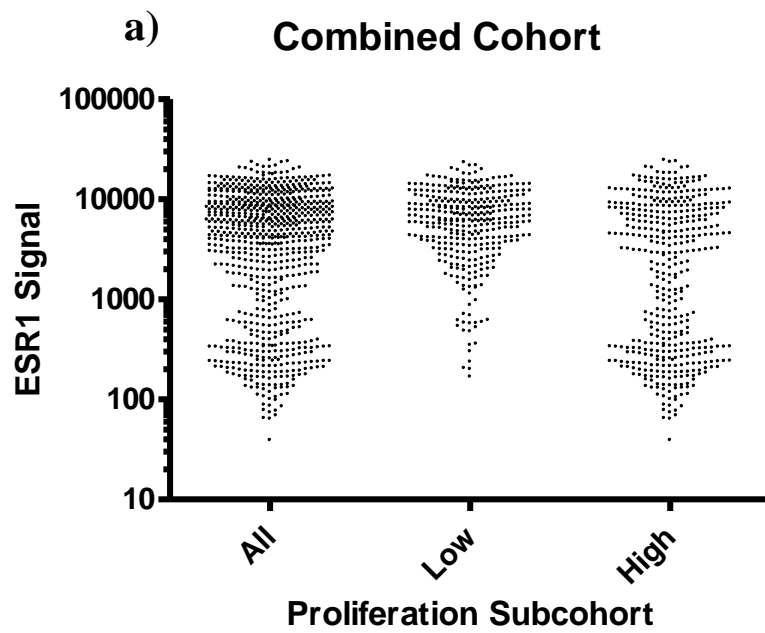
e)



X-axis: PCA component 1 (14.55% variance)
Y-axis: PCA component 2 (7% variance)
Z-axis: PCA component 3 (5.939% variance)

Conditions: TRANSBIG, Default Interpretation
Colored by: Parameter TTM_10Y

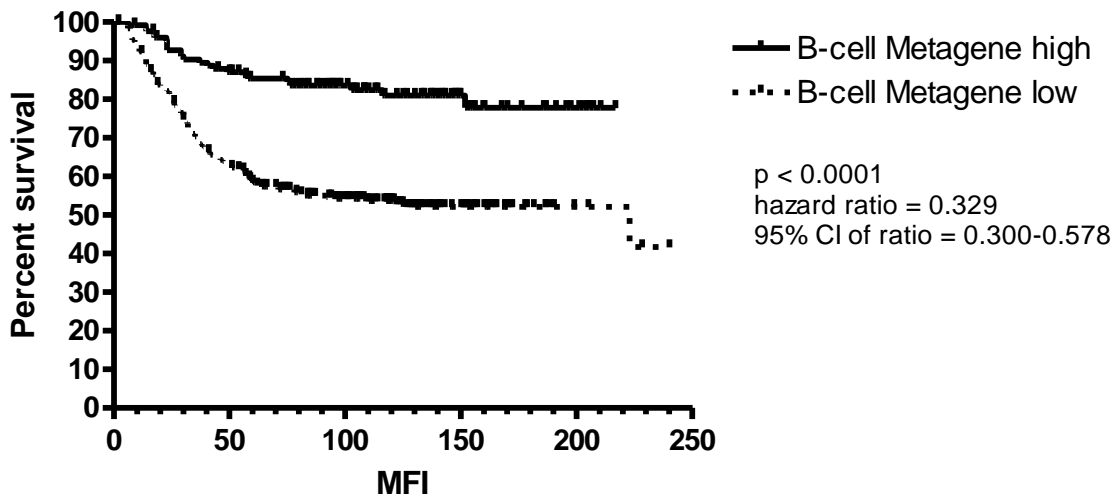
Supplementary Figure 3 Metagene expression and time to metastasis visualized in the TRANSBIG cohort by PCA. Relative expression of **a)** ER, **b)** proliferation **c)** T-cell **d)** B-cell metagenes as well as **e)** time to distant metastasis in 302 samples distributed in the three dimensional space of PC1 - 3 in the same genes used in our finding cohort. High expression is indicated by red, intermediate by yellow and low expression by green. The arrow indicates the direction of the “metagene axis” from low values to high values. All samples of patients who developed a distant metastasis are colored on a continuous scale from early (blue) to late (dark orange). Samples from patients that remained disease free until end of follow up or at least 10 years were colored red. A region of accumulated early metastases is marked by a circle.



Supplementary Figure 4 Distribution of a) ESR1 and b) ERBB2 signal in the combined cohorts as well as the combined low and high proliferation subcohorts. At a cut off of 1000 for ESR1 and 6000 for ERBB2, 89.5 % of all ESR1 negative and 87.4% of all ERBB2 positive tumors belong to the high proliferation subtype.

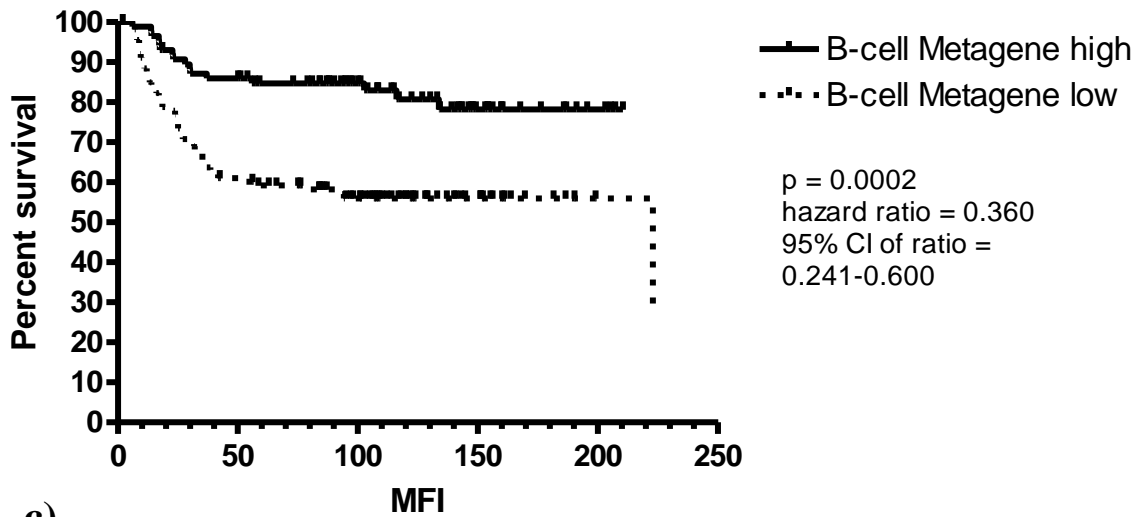
a)

Combined Cohorts Fast Proliferation



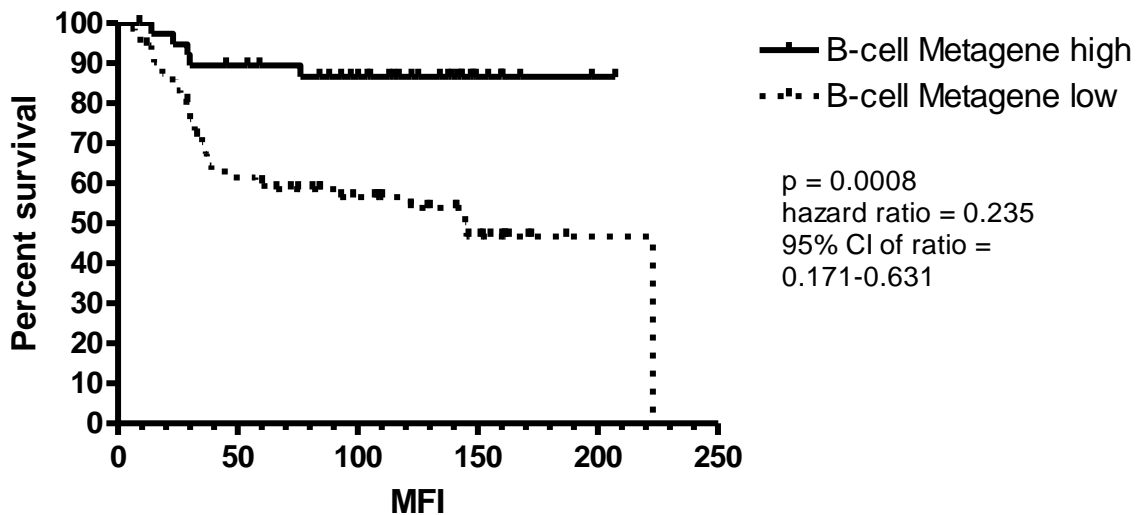
b)

Combined Cohorts ESR1- tumors



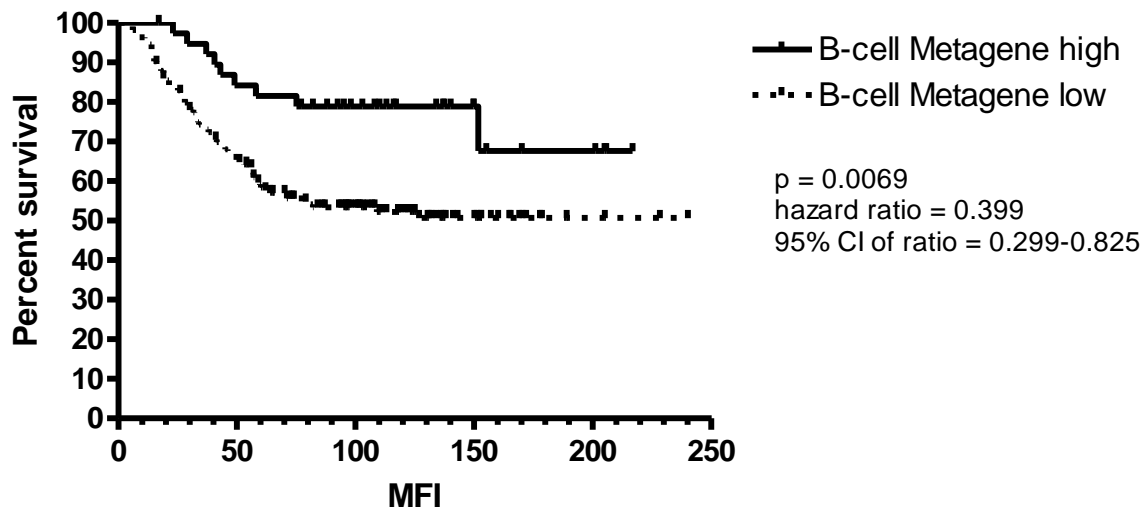
c)

Combined Cohorts ERBB2+ tumors



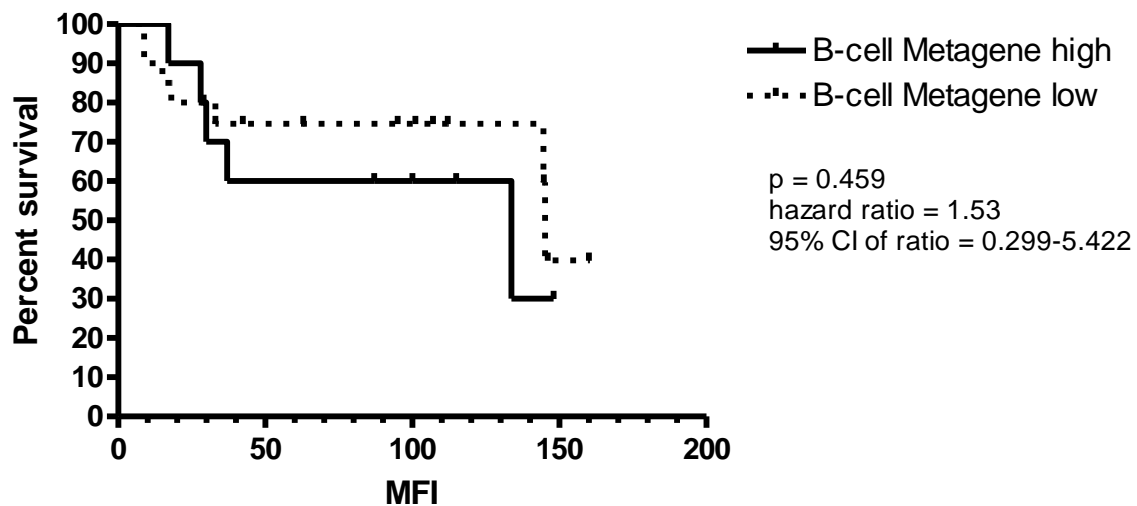
d)

Combined Cohorts Fast Proliferation (ESR1+ /ERBB2-)

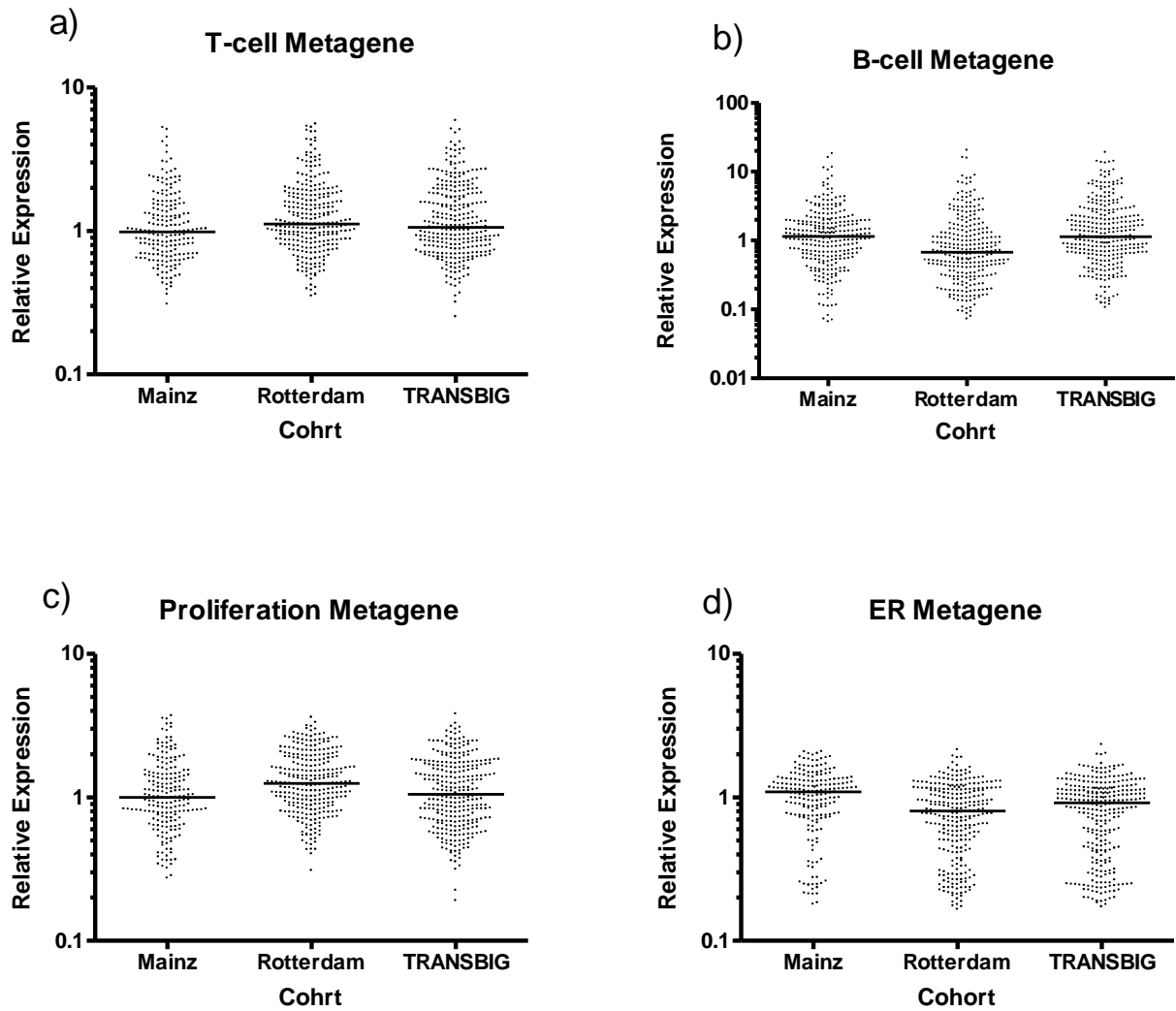


e)

Combined Low Proliferation (ESR1- or ERBB2+)



Supplementary Figure 5 Survival analyses according to B-cell metagene expression. Metastasis-free survival in patient subgroups of the combined cohort study cohorts stratified according to high or low expression of the B-cell metagene. A) tumors with high expression of the proliferation metagene, b) ESR1 negative tumors, c) ERBB2 positive tumors, d) ESR1 positive/ERBB2 negative fast proliferating tumors and e) ESR1 negative or ERBB2 positive low proliferating tumors. Tumor classification: ESR1 negative if signal < 1000 and ERBB2 positive if signal > 6000.



Supplementary Fig. 6 Comparison of the distribution of the a) T-cell metagene, b) B-cell metagene, c) proliferation metagene, and d) ER metagene within the three patient cohorts. B-cell related transcripts are expressed at a lower and proliferation genes at a higher level in the Rotterdam cohort. Mainz samples have an overall higher expression of ER related transcripts.