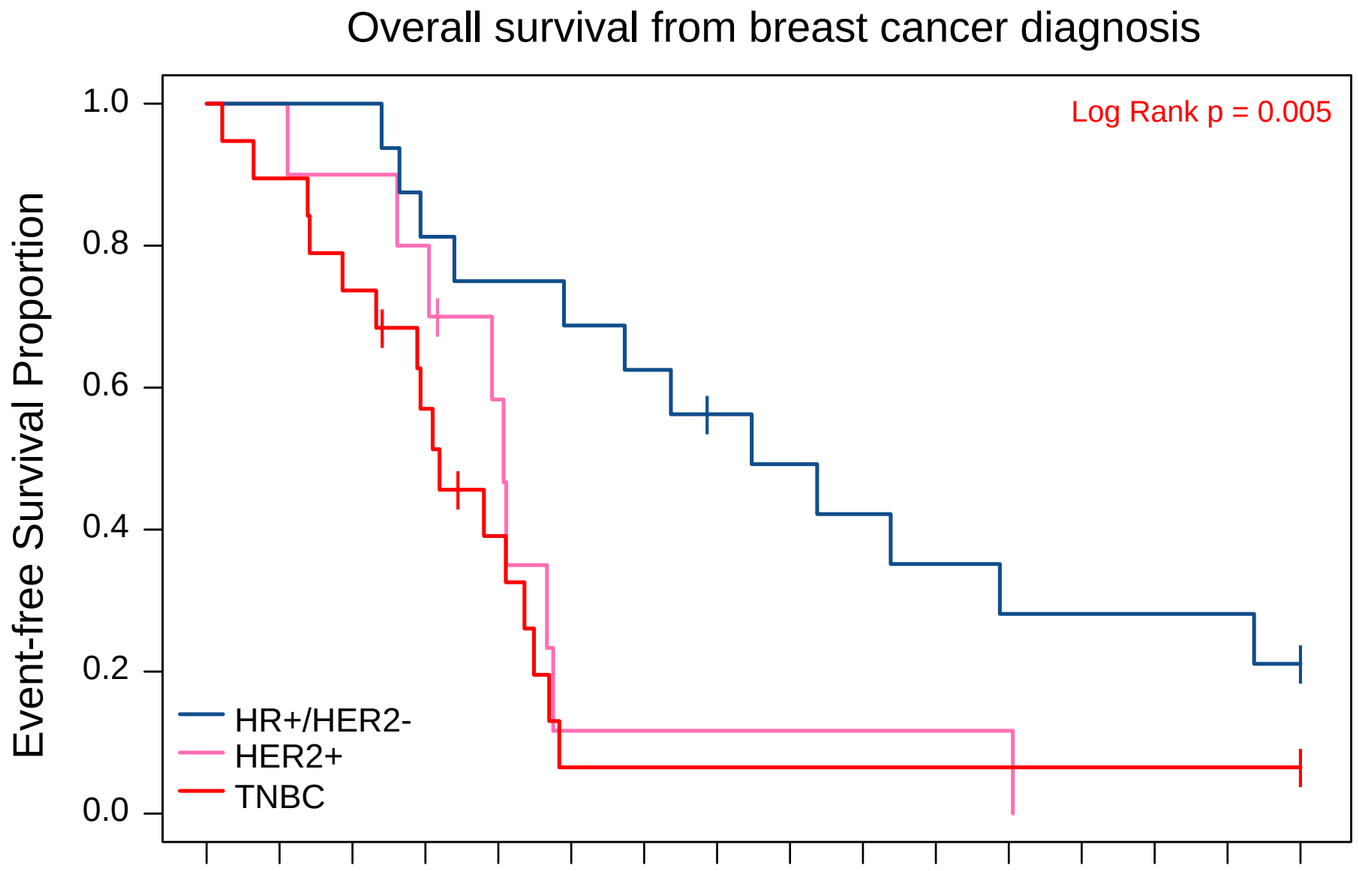
Extended Data Fig.1

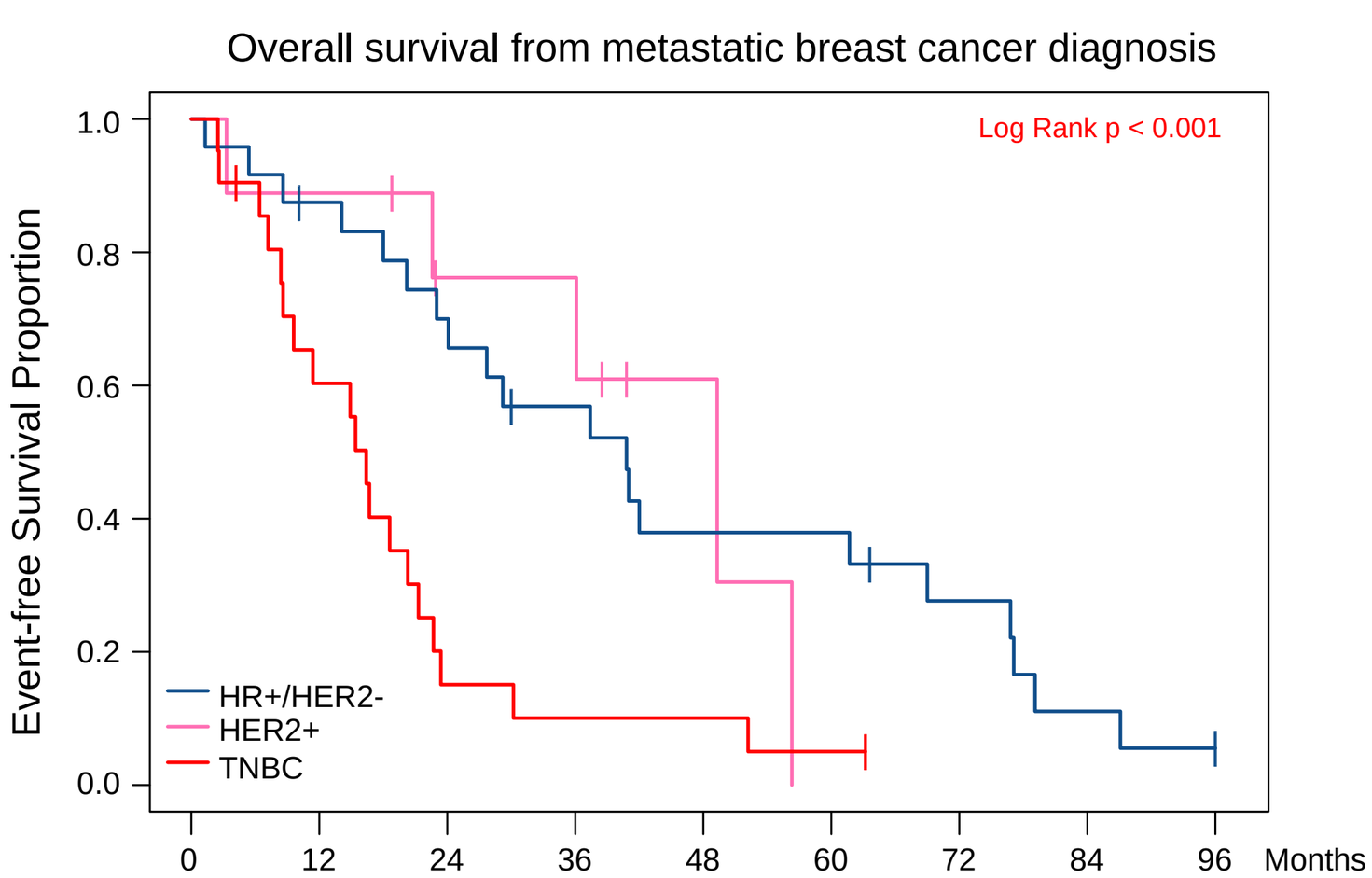
a



No. at risk																
HR+/HER2- EBC	17	17	16	13	12	11	10	8	7	6	5	4	4	4	4	3
HER2+ EBC	10	10	9	8	5	1	1	1	1	1	1					
TN EBC	19	17	14	9	6	1	1	1	1	1	1	1	1	1	1	1

	No. of events/ No. of patinets	Median Overall Survival (95% CI), months	Hazard ratio for death (95% CI)
HR+/HER2- EBC	14/17	88.5 (40.2 - 170.0)	1
HER2pos EBC	9/10	48.2 (13.1 - 56.2)	2.19 (0.81 - 5.92)
TN EBC	17/19	36.7 (22.1 - 51.6)	2.44 (1.08 - 5.57)

b



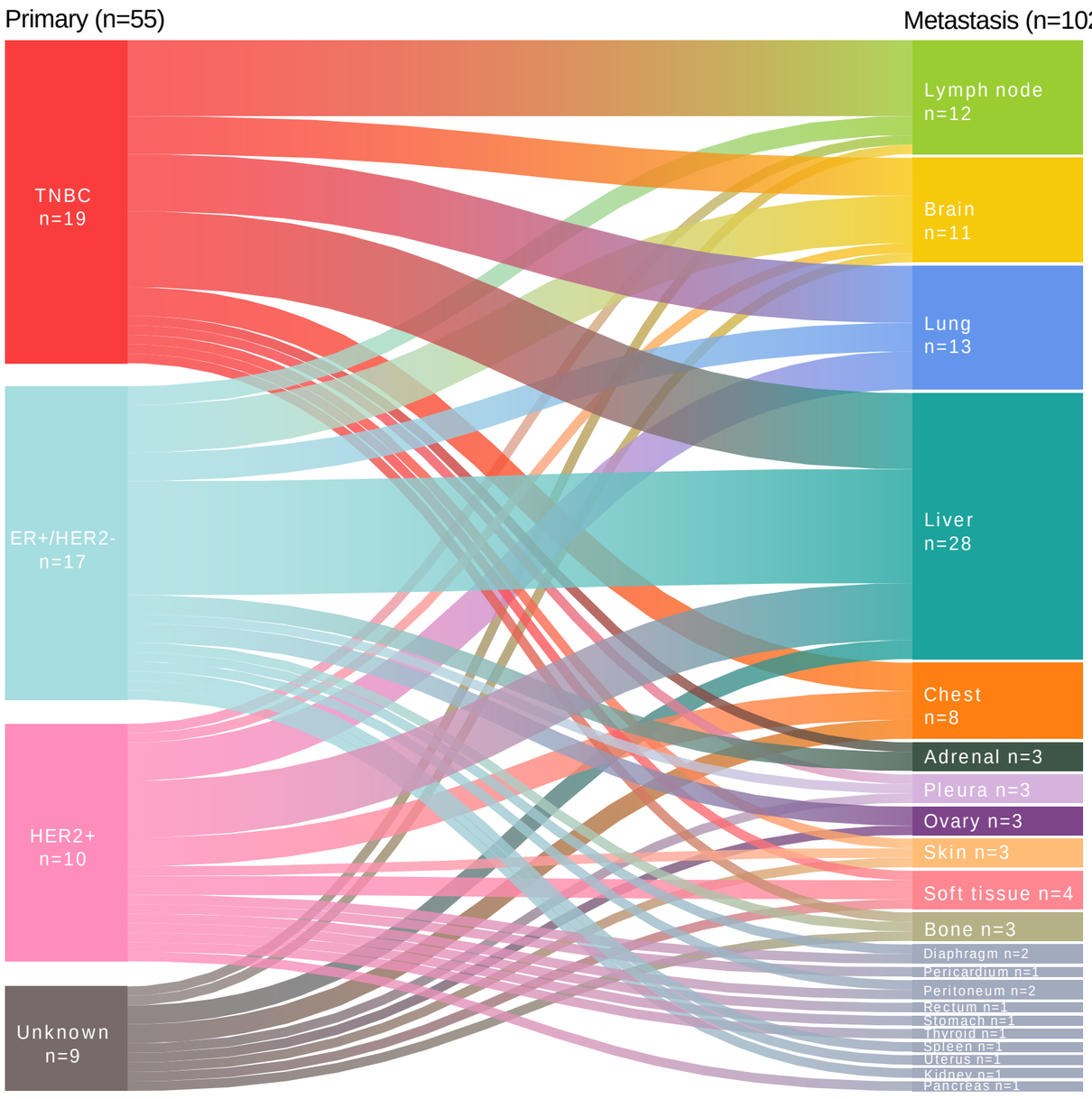
No. at Risk									
HR+/HER2-MBC	24	20	16	12	8	8	5	2	1
HER2+ MBC	9	8	5	5	2				
TN MBC	21	12	3	2	2	1			

	No. of events/ No. of patinets	Median Overall Survival (95% CI), months	Hazard ratio for death (95% CI)
HR+/HER2- MBC	21/24	40.8 (23.3 - 69.0)	1
HER2+ MBC	5/9	49.3 (3.3 - 56.3)	0.91 (0.41 - 2.05)
TN MBC	19/21	16.4 (8.6 - 21.3)	2.82 (1.34- 5.92)

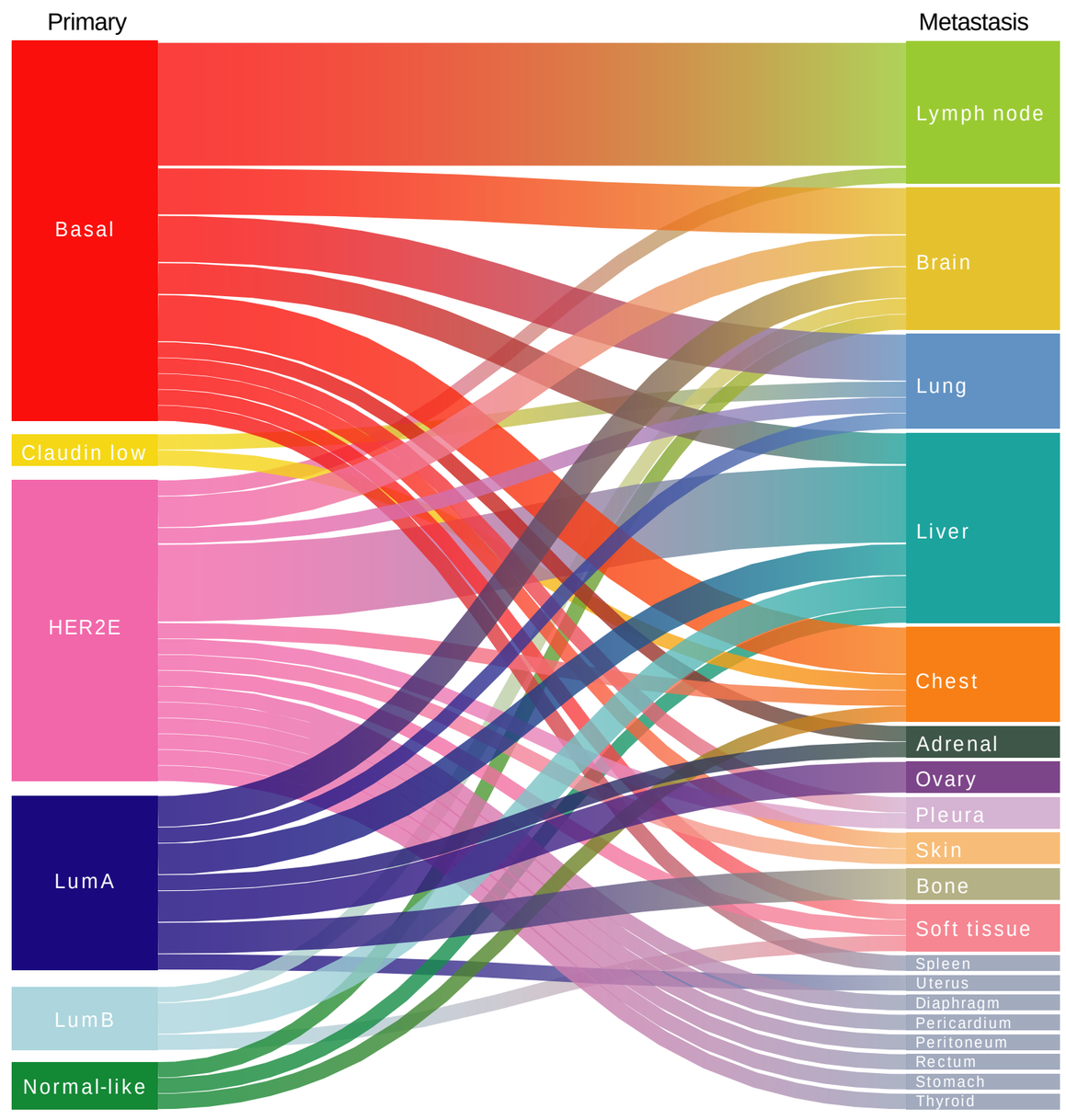
Extended Data Fig.1. **Survival outcomes according to clinical subtypes of AURORA cohort.** **a.** Kaplan-Meier, log rank test and Cox proportional hazards regression model methods were used to study the overall survival from breast cancer diagnosis ("First Primary Receptor at diagnosis" column of Supplementary table 2) in HER2 positive (HER2+), Hormone receptor positive and HER2 negative (HR+/HER2-) and TNBC (triple negative breast cancers). **b**. Kaplan-Meier, the log rank test and Cox proportional hazards regression model to study the overall survival from metastatic breast cancer diagnosis ("Metastasis original receptors" column of Supplementary table 2) in HER2+, HR+/HER2- and TNBC. In absence of HR/HER2 status in the metastatic relapse we used the data from the most recent biopsy. EBC, early breast cancer; MBC, metastatic breast cancer; confidence interval (CI). Statistically significant values are highlighted in red.

Extended Data Fig.2

a



b



*More than one site of metastais from a given primary have been collected in some patients

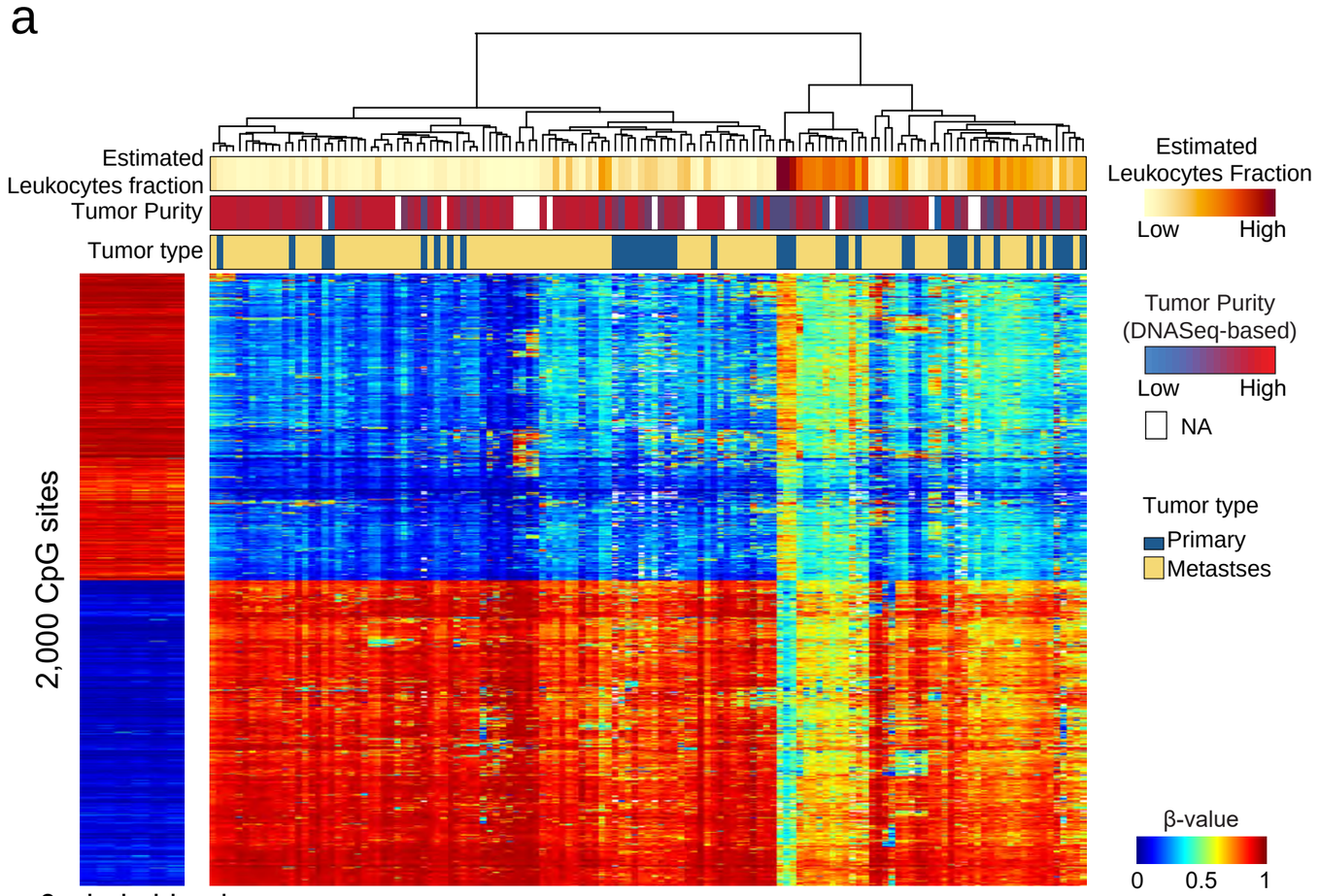
*5 patients failed on DNA/RNA sequencing in primary tumor (IHC information added)

*2 patients failed on DNA/RNA sequencing in metastatic tumor (site of metastasis not added)

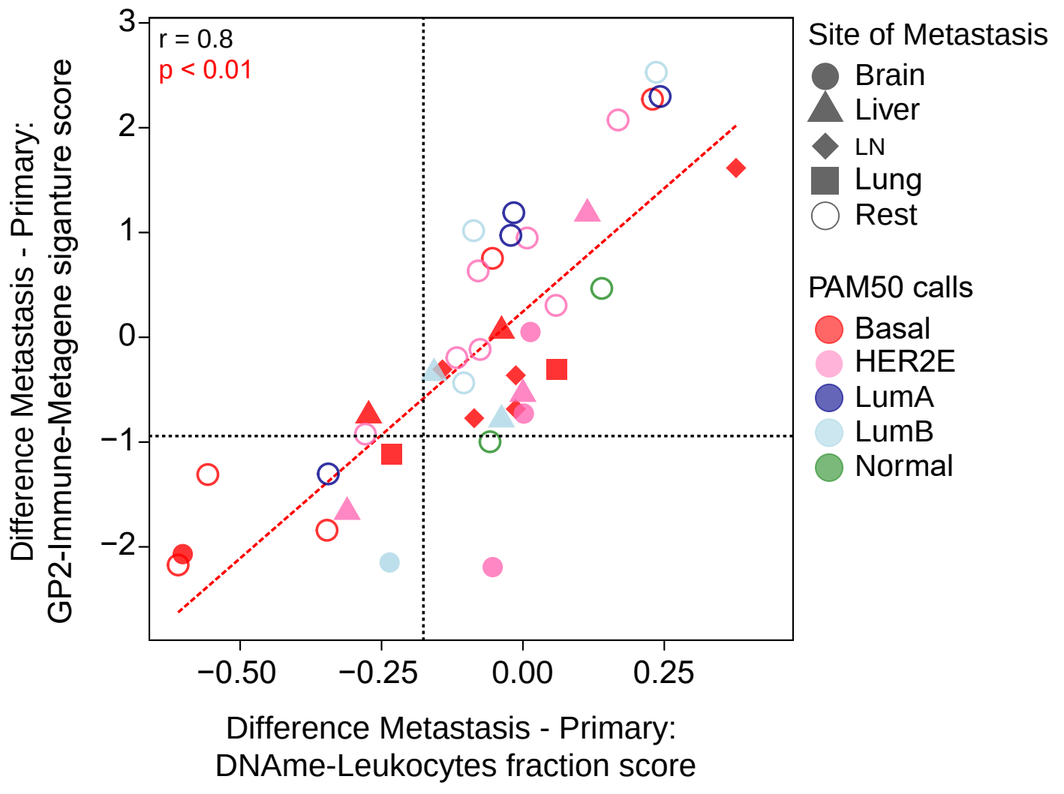
Extended Data Fig.2. **Clinical subtype and molecular subtype distribution according to site of metastasis. a.** Distribution of the 55 diagnosed primary tumors by clinical receptor status (TNBC, ER+/HER2-, HER2+, and unknown, left side) linked to their anatomic sites of metastasis (right). Clinical receptor status at the time of first primary diagnosis ("First Primary Receptors" column of Supplementary table 2). **b**. Distribution of 39 diagnosed primary tumors by gene expression based intrinsic molecular subtype when available (left) linked to their anatomic sites of metastasis (right).

Extended Data Fig.3

a



b

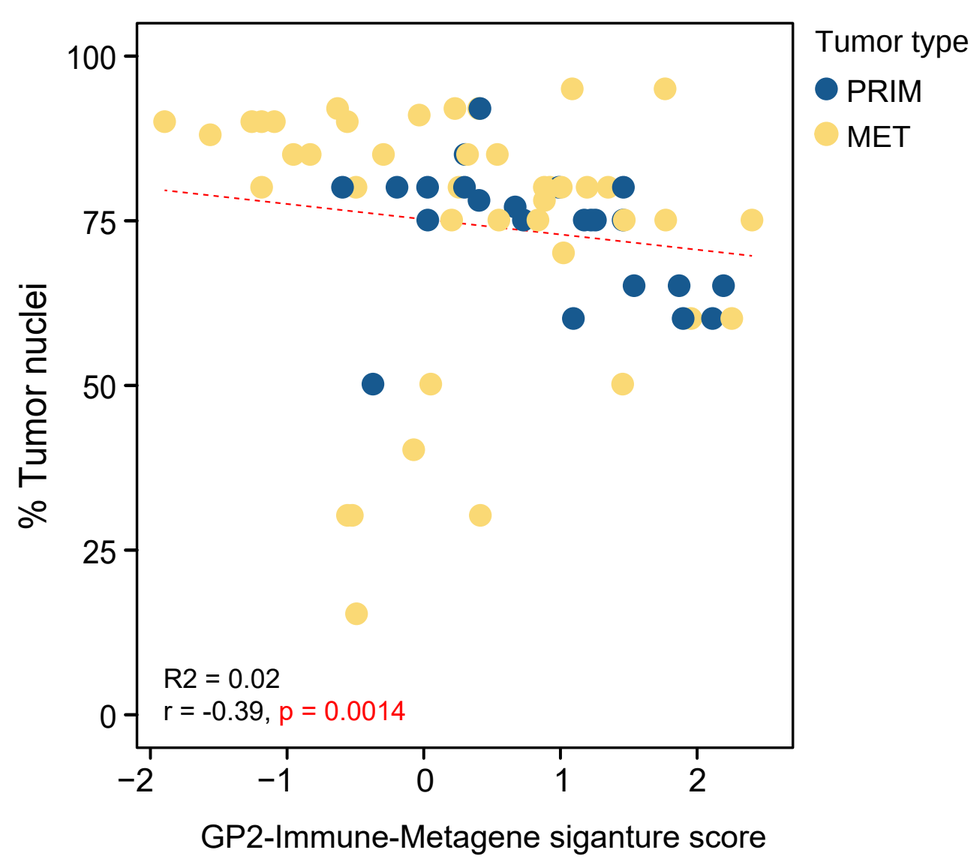


c

GP2-Immune-Metagene siganture score
vs Pathology report and Genomics assesments

Pathology report	Rho	p value
% Tumor nuclei	-0.39	0.00
% Normal cells	0.18	0.16
% Stromal cells	0.10	0.41
% Lymphocyte infiltration	0.34	0.01
Tumor cellularity	-0.32	0.02
Genomics	**Rho**	**p value**
Estimate-Stromal Score	0.69	0.00
Estimate-Immune Score	0.91	0.00
Estimate Score	0.84	0.00
DNAme-Leukocytes Fraction	0.86	0.00

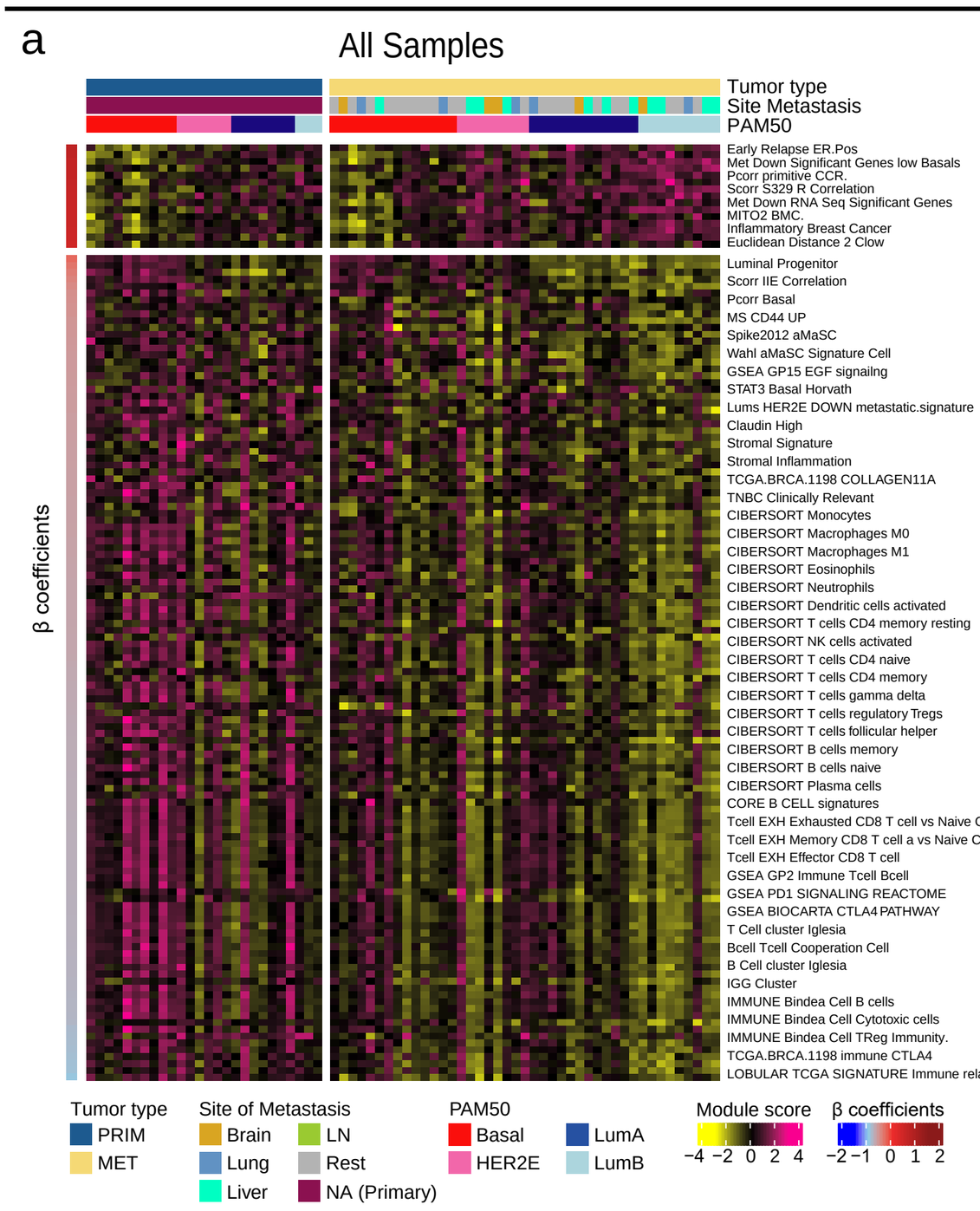
d

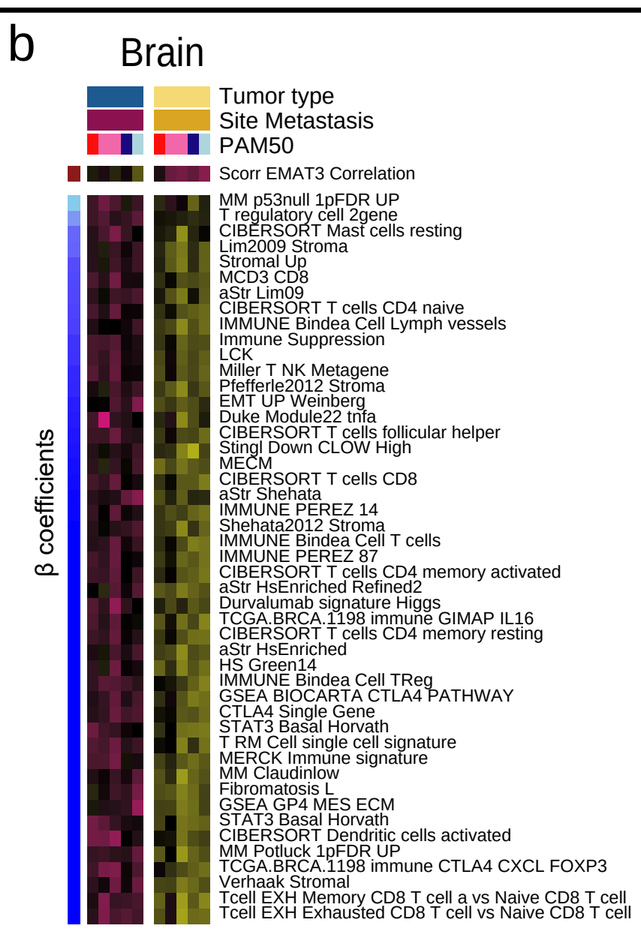


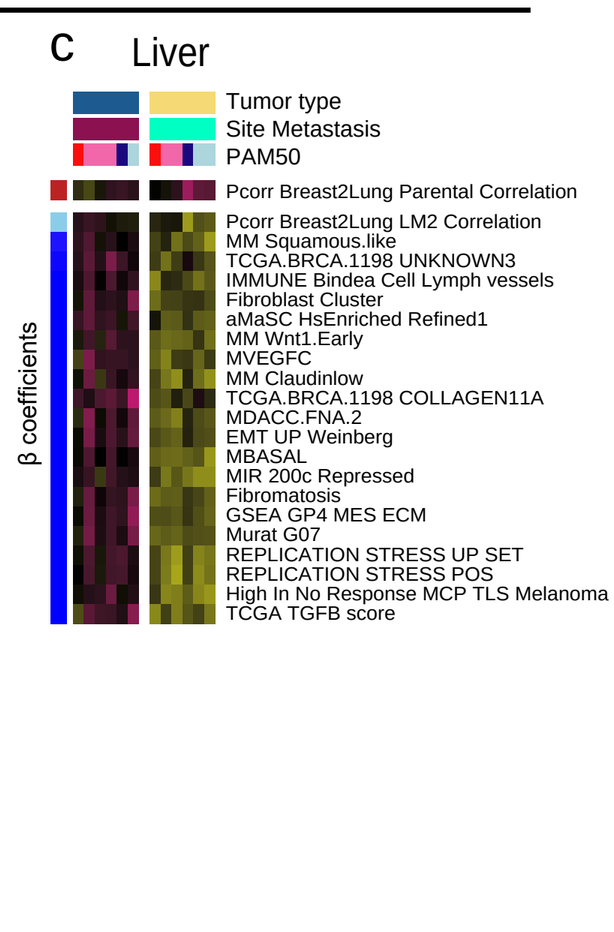
Extended Data Fig.3. **Correlation between tumor cellularity metrics and immune signatures. a.** Supervised hierarchical clustering of the top 1,000 leukocyte-specifically methylated probes and the bottom 1,000 tumor tissue-specifically methylated probes, after ranking all probes based on the mean leukocytes - mean tumor tissues**.** For the 133 tumors, association is shown with tumor type, tumor purity, and estimated leukocyte fraction scores[19,20]. **b**. Pearson correlation between the difference (Metastasis – Primary gene expression values) of the Leukocyte fraction scores and GP2-Immune-Metagene signature scores (calculated from the Level 4 RNAseq data). Higher scores mean higher expression in metastasis compared to primary tumors. Correlation was measured using the Pearson correlation coefficient. **c**. Spearman correlations between GP2-Immune-Metagene signature scores and several pathology determined scores (% Tumor nuclei, % of normal cells, % of Stromal cells, % Lymphocyte infiltration and tumor cellularity) or genomic scores (Estimate-Stromal scores, Estimate-Immune Score and Estimate Score using ESTIMATE method[42]. Rho (spearman correlation coefficient, ρ). **d**. Pearson correlation of GP2-Immune-Metagene signature score and % of Tumor nuclei form pathology report. Statistically significant values are highlighted in red. GP2-Immune-Metagene signature scores were calculated from the Level 4 RNAseq data (see methods).

Extended Data Fig.4

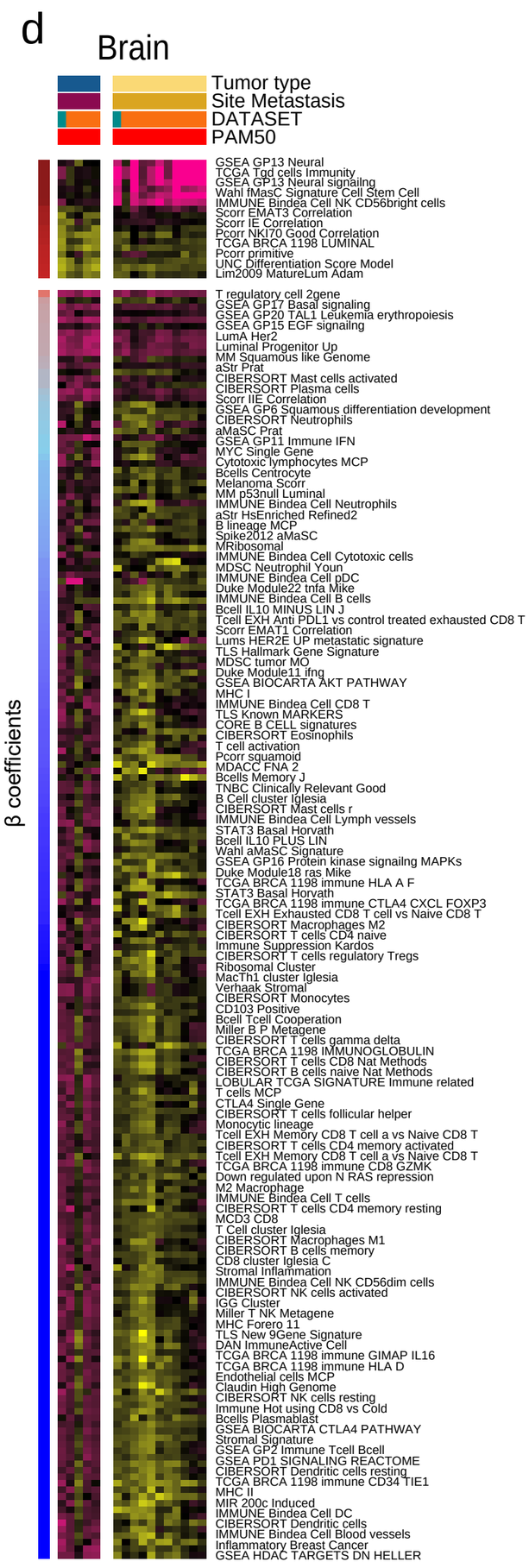
AURORA

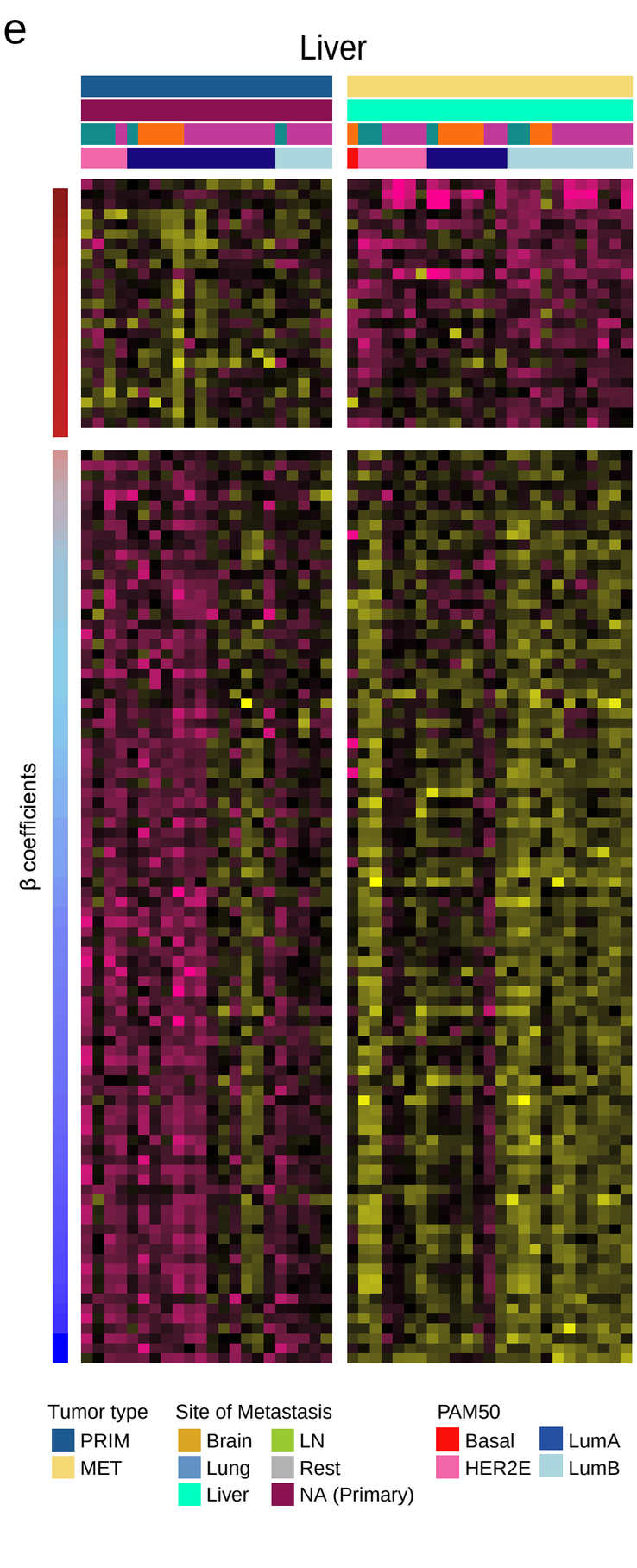






AURORA-RAP-GEICAM





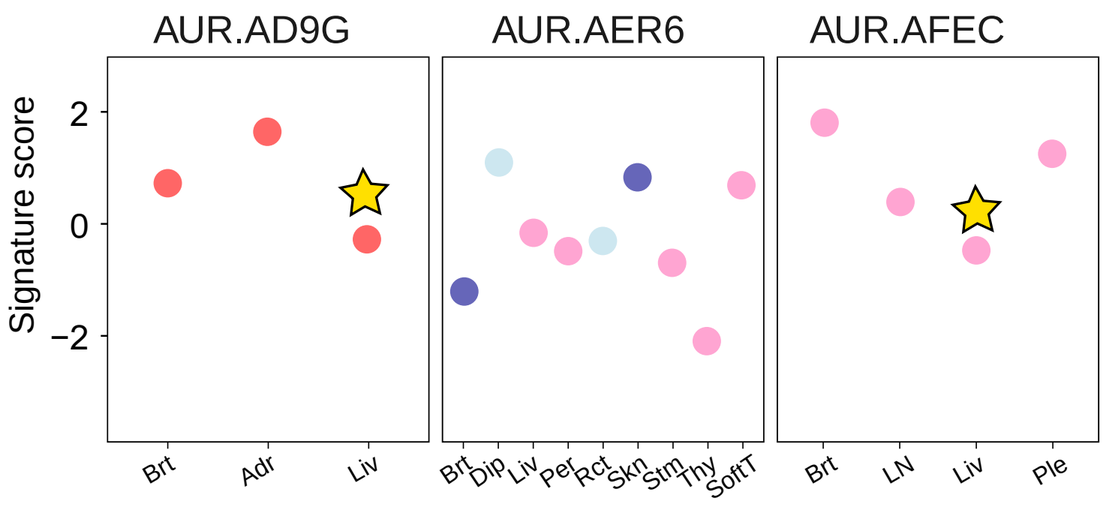
Extended Data Fig.4. Supervised analysis of gene expression signatures according to site of metastasis in AURORA or combined AURORA-RAP-GEICAM cohorts. a. Heatmap depicting the differentially expressed (DE) signatures between primary (n=26) and metastasis (n=69) in the AURORA cohort using all samples. **b**. Heatmap depicting the DE signatures between paired primary (n=5) and brain metastasis (n=5) in the AURORA cohort. **c**. Heatmap depicting the DE signatures between paired primary (n=6) and liver metastasis (n=6) in the AURORA cohort. **d**. Heatmap depicting the DE signatures between Basal-like paired primary (n=5) and brain metastasis (n=11) in the AURORA-RAP-GEICAM cohort. **d**. Heatmap depicting the DE signatures between Luminals (LumA, LumB, and HER2E) paired primary (n=22) and liver metastasis (n=25) in the AURORA-RAP-GEICAM cohort. Significance of the differences between primary and metastasis was calculated using linear mixed models ($q<0.01$). Significant signatures are row ordered from high to low according to β-coefficients (or regression coefficients) and divided according to upregulated (positive) or downregulated (negative) in metastasis. Patients are column ordered according to PAM50 molecular subtype and divided according to primary and metastasis. Signatures scores were calculated in the Level 4 RNAseq data (see methods). Normal-like tumors and post-treatment primaries were removed from the analysis in the AURORA cohort.

Extended Data Fig.5

GP2-Immune-Metagene siganture

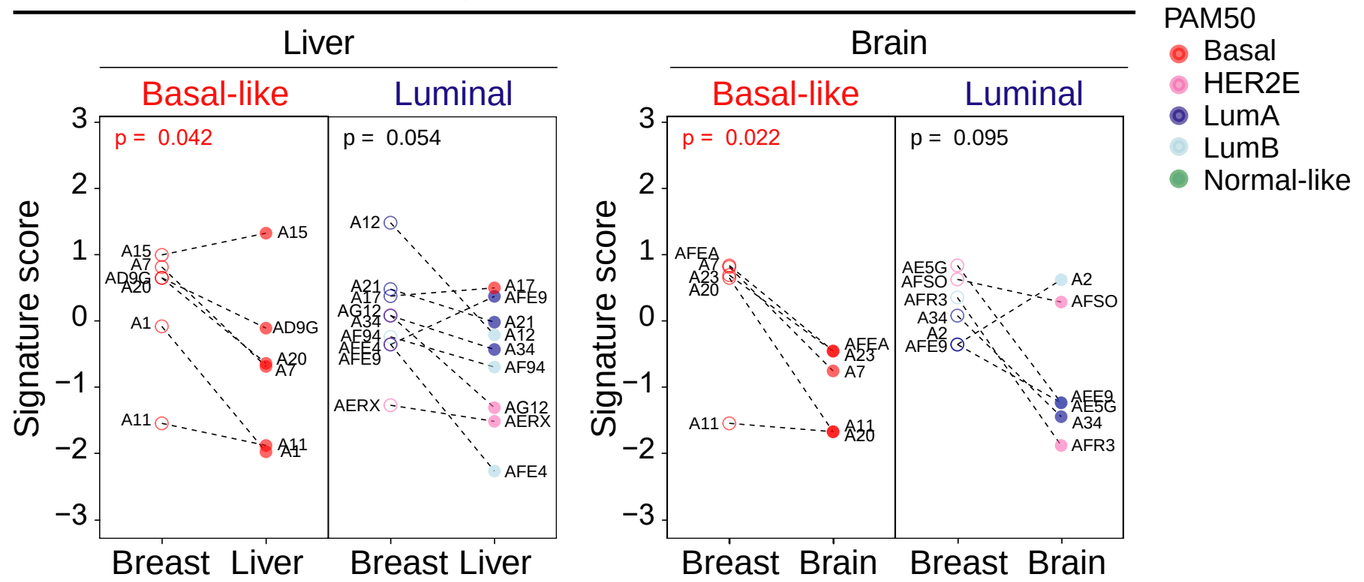
a

Multimetastatic samples AURORA-RAP dataset



b

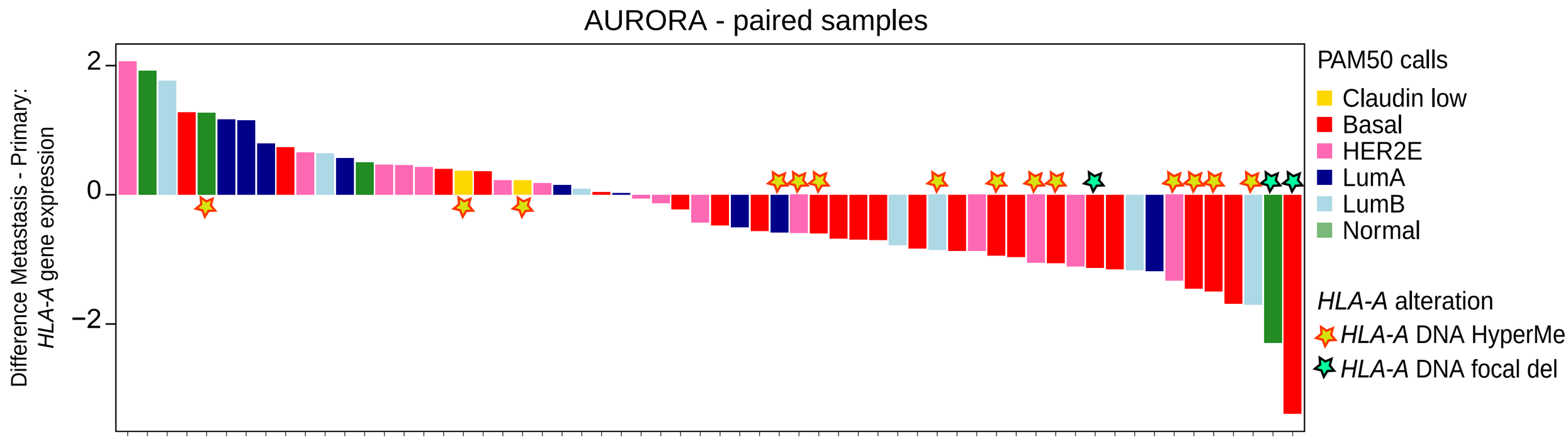
Paired samples AURORA-RAP dataset



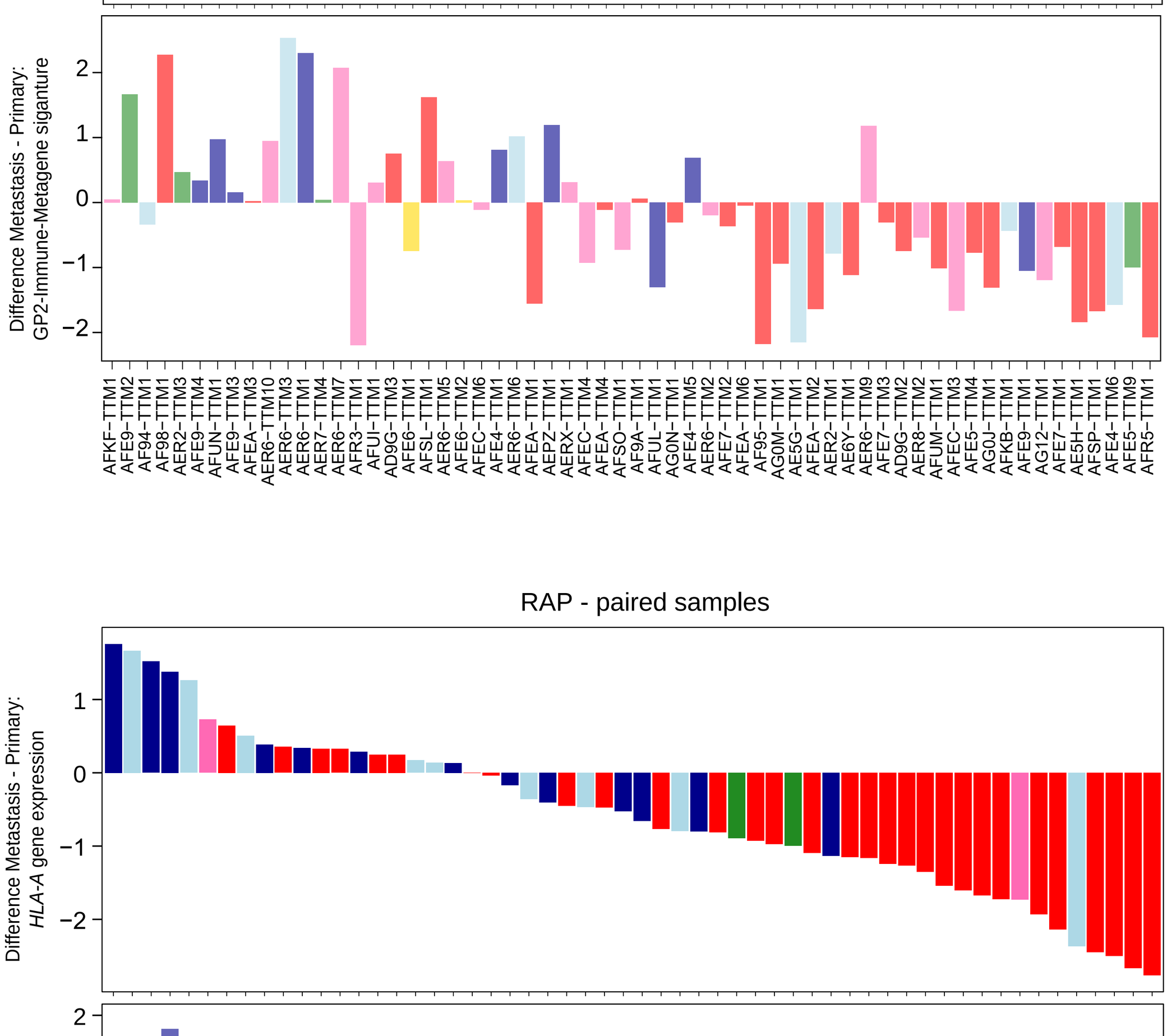
Extended Data Fig.5. Patients with multiple metastases examined for immune features in AURORA-RAP combined cohort. a. gene expression signature scores of GP2-Immune-Metagene are shown according to individual specimens coming from patients with at least 2 metastases analyzed by RNAseq data (n=16). The star indicates liver specimens with the lowest expression of signature. **b.** Signature expression changes between paired primary and liver, brain, lung or rest of metastasis of GP2-Immune-Metagene signature divided by Basal-like or Luminal groups (Luminal A, B and HER2E). Comparisons between 2 paired groups were performed by paired, 2-tailed t-test Statistically significant values are highlighted in red. Brt, breast; Adr, adrenal; Liv, liver; Dip, Diaphragm; Per, peritoneum; Rct, rectum, Skn, skin; Stm, stomach; Thy, Thyroid; SoftT, soft tissue; LN, lymph node; Ple, pleura; Lun, lung, Brn, brain; Bon, bone; Kid, kidney; Che, chest; Spl, spleen; Mes, mesentery; Pan, pancreas; LumA, luminal A; LumB, luminal B.

Extended Data Fig.6

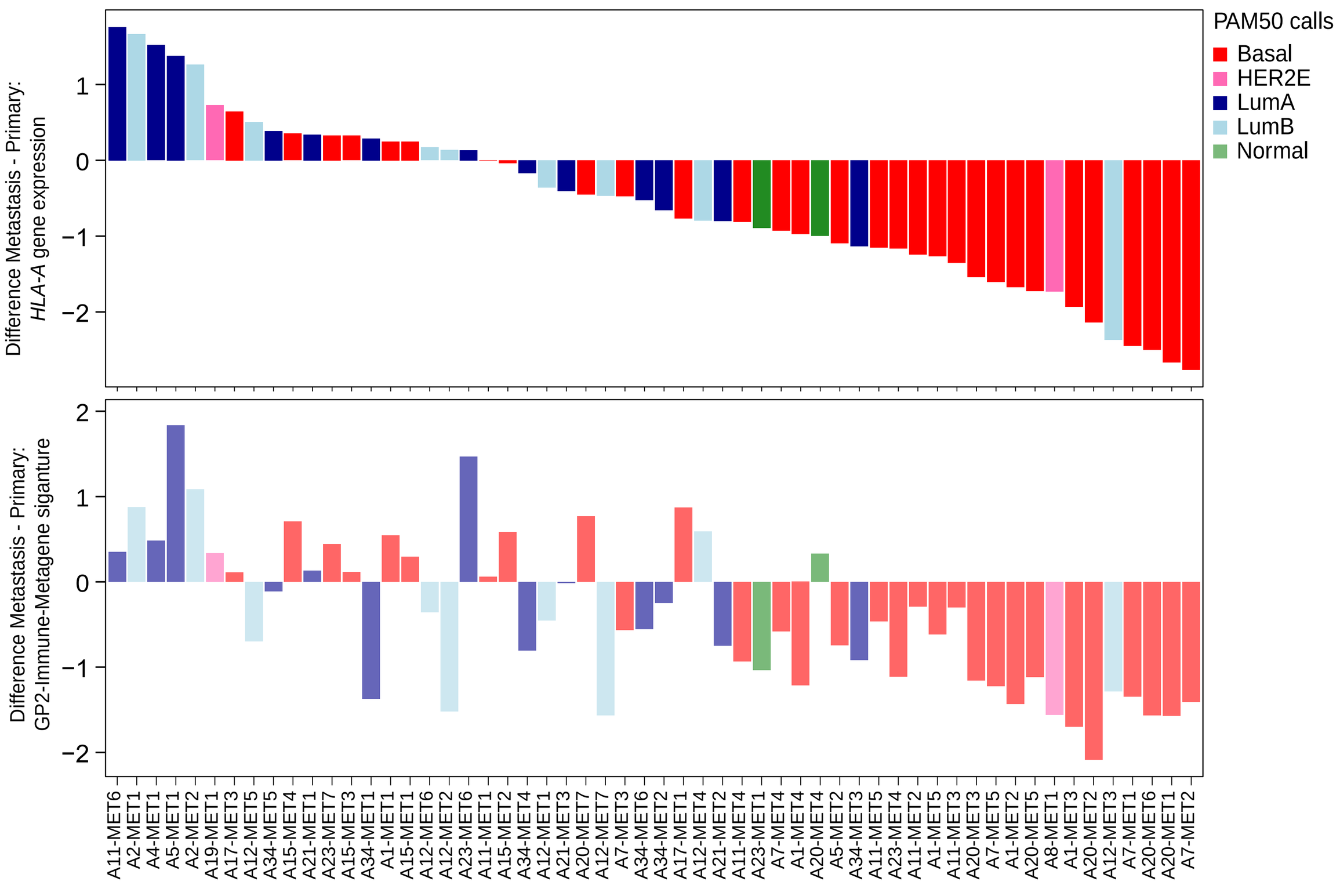
a



b



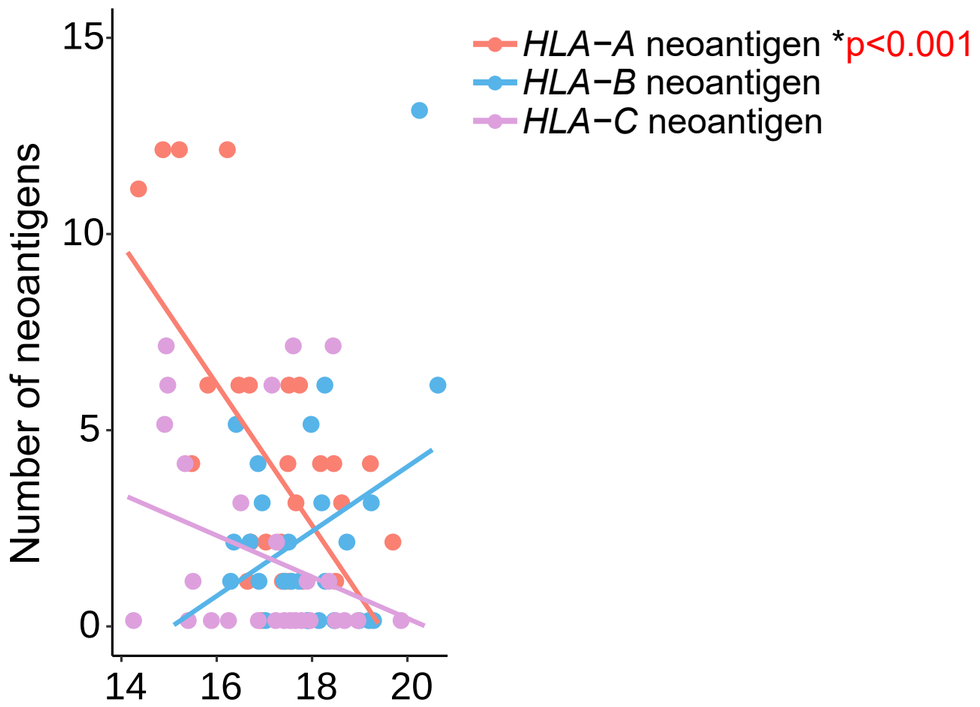
c



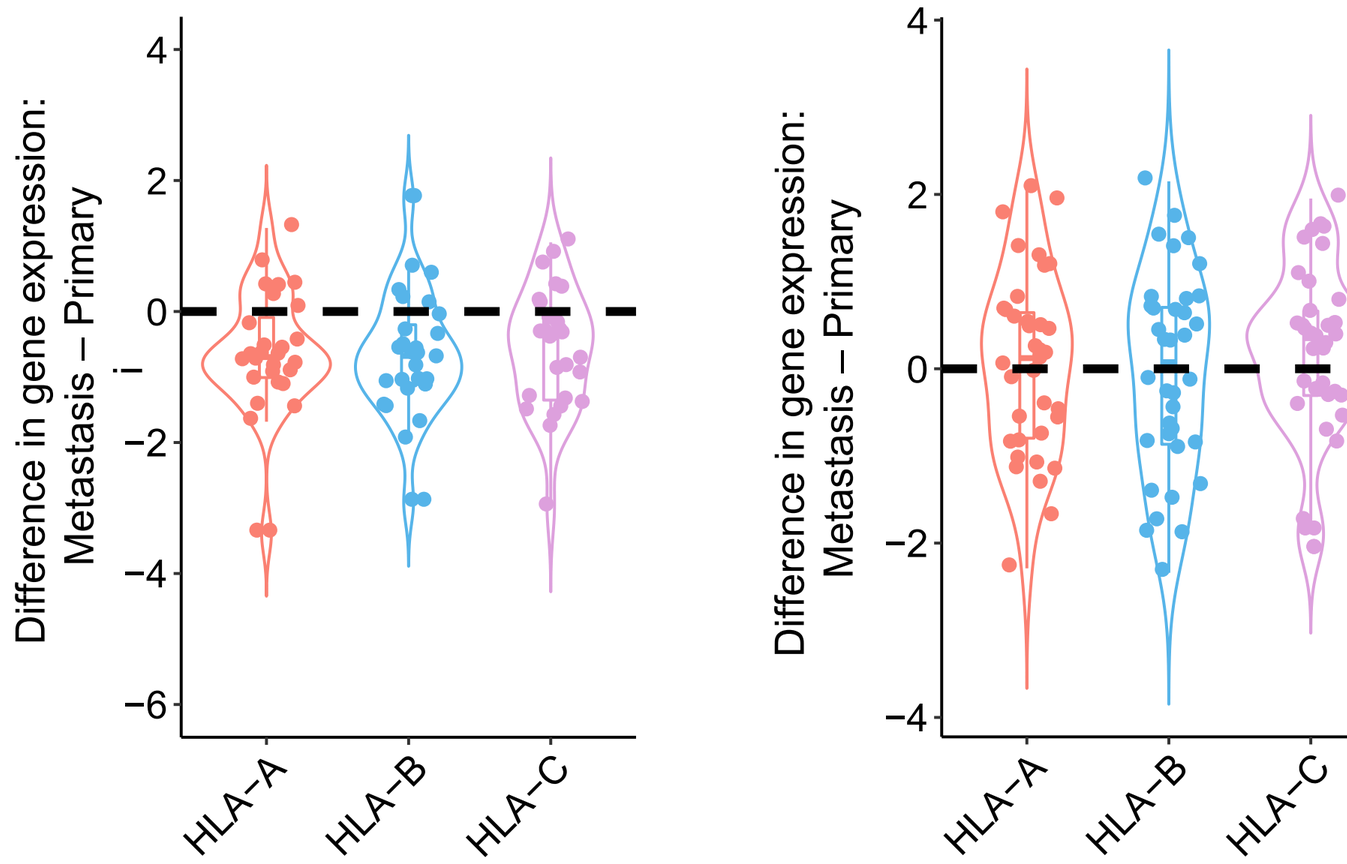
d

Extended Data Fig.6. Difference in *HLA-A* and immune-signature expression between primary and metastatic tumors. a. Waterfall plot of AURORA cases showing the difference between primary (n=36) and metastasis (n=60) (Difference: Metastasis – Primary gene expression value) ordered from the highest (left) to the lowest (right) signature score for *HLA-A* mRNA expression (bottom panel). The top panel show the difference of primary versus metastases for GP2-Immune-Metagene values (top panel). Yellow stars highlight *HLA-A* Hypermethylated cases and green stars highlight the samples with DNA *HLA-A* focal deletions. **b.** Waterfall plot of RAP cases showing the difference between primary (n=15) and metastasis (n=56) (Difference: Metastasis – Primary gene expression value) ordered from the highest (left) to the lowest (right) signature score for *HLA-A* mRNA expression (bottom panel). The top panel show the difference of primary versus metastases for GP2-Immune-Metagene values (top panel). Pairs with a Normal-like primary tumor were removed from the analysis. LumA, luminal A; LumB, luminal B.

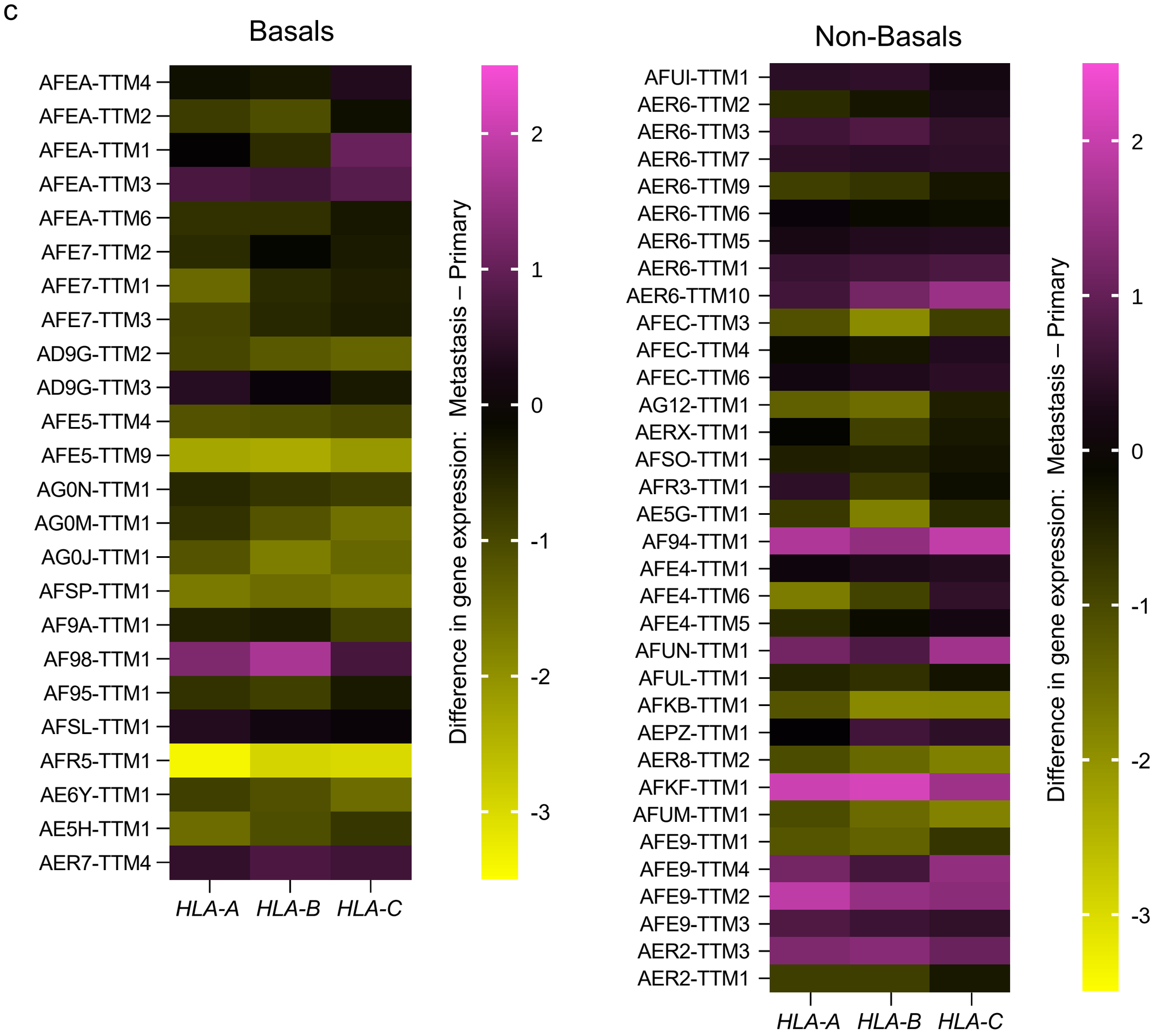
a


Basals (Prim and Mets)
HLA−A neoantigen *p<0.001
HLA−B neoantigen
HLA−C neoantigen
Number of neoantigens
15
10
5
0
14
16
18
20
HLA−specific expression

b

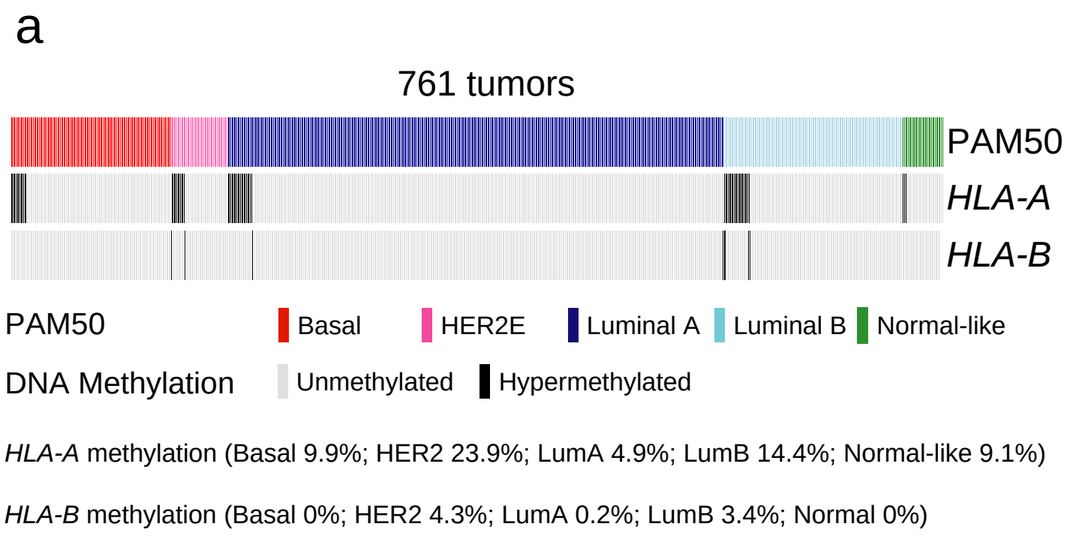

Basals
Non-Basals
Difference in gene expression: Metastasis – Primary
4
2
0
−2
−4
−6
HLA−A
HLA−B
HLA−C

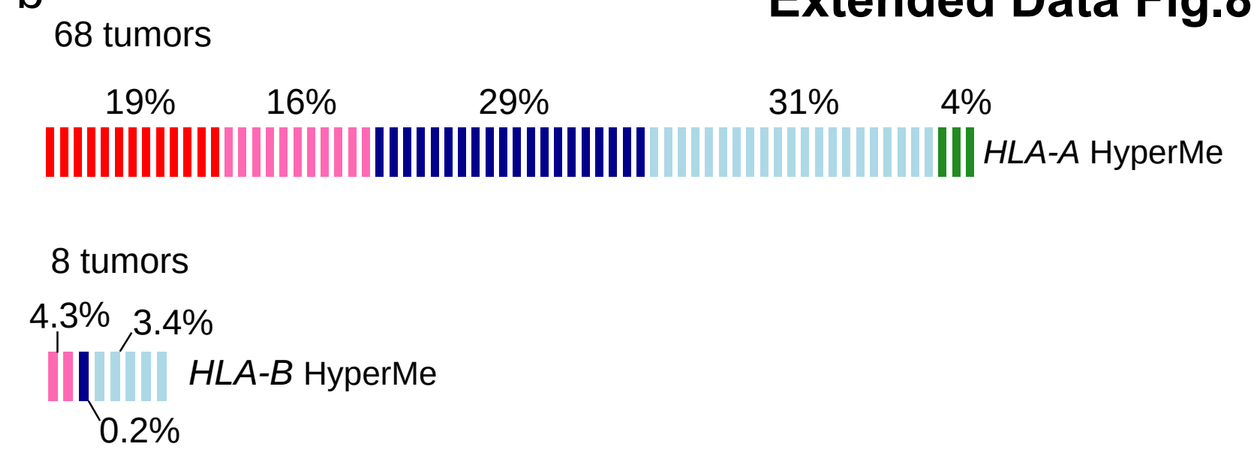
c

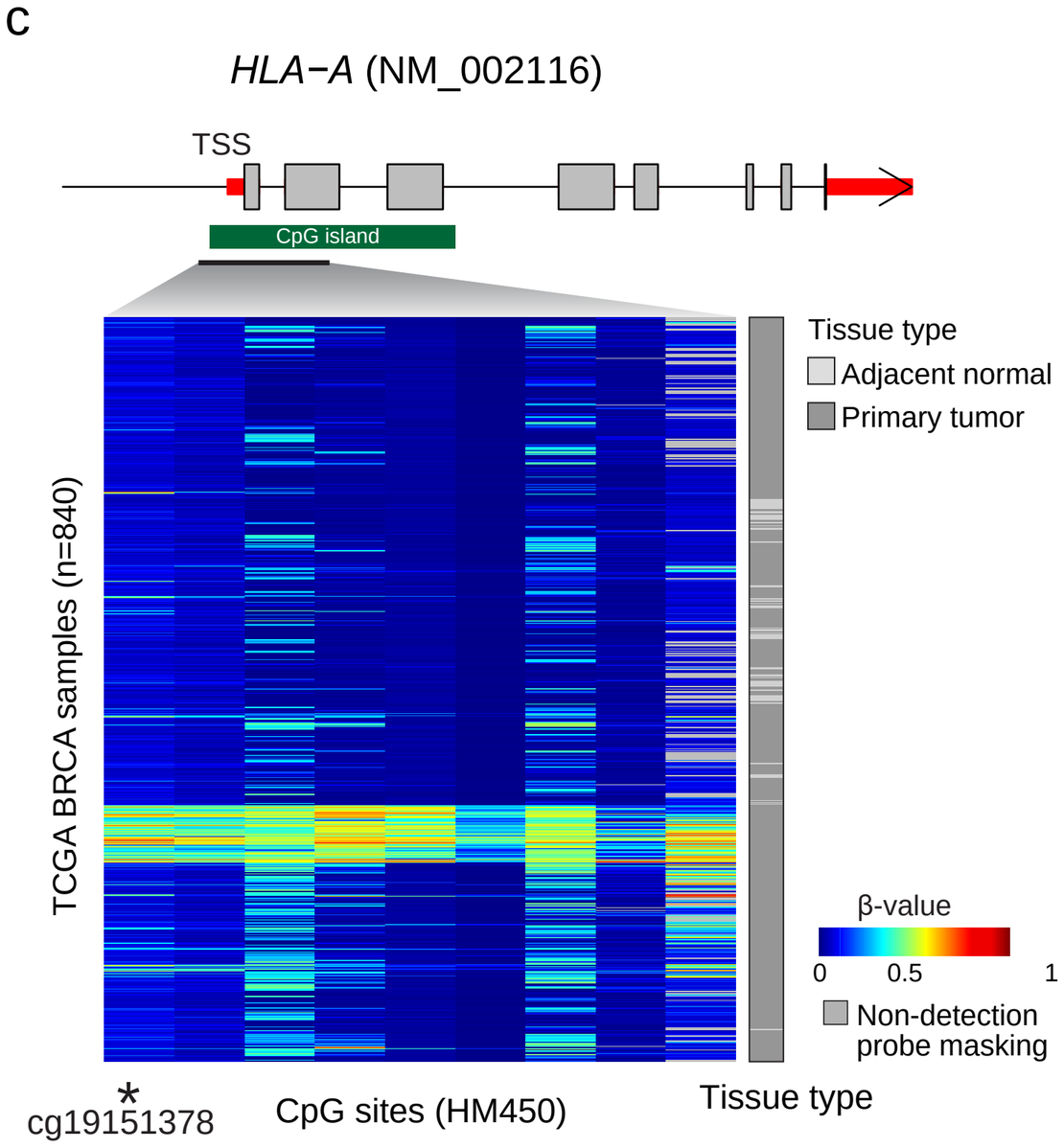

Basals
AFEA-TTM4
AFEA-TTM2
AFEA-TTM1
AFEA-TTM3
AFEA-TTM6
AFE7-TTM2
AFE7-TTM1
AFE7-TTM3
AD9G-TTM2
AD9G-TTM3
AFE5-TTM4
AFE5-TTM9
AG0N-TTM1
AG0M-TTM1
AG0J-TTM1
AFSP-TTM1
AF9A-TTM1
AF98-TTM1
AF95-TTM1
AFSL-TTM1
AFR5-TTM1
AE6Y-TTM1
AE5H-TTM1
AER7-TTM4
HLA-A
HLA-B
HLA-C
Difference in gene expression: Metastasis – Primary
2
1
0
-1
-2
-3
Non-Basals
AFUI-TTM1
AER6-TTM2
AER6-TTM3
AER6-TTM7
AER6-TTM9
AER6-TTM6
AER6-TTM5
AER6-TTM1
AER6-TTM10
AFEC-TTM3
AFEC-TTM4
AFEC-TTM6
AG12-TTM1
AERX-TTM1
AFSO-TTM1
AFR3-TTM1
AE5G-TTM1
AF94-TTM1
AFE4-TTM1
AFE4-TTM6
AFE4-TTM5
AFUN-TTM1
AFUL-TTM1
AFKB-TTM1
AEPZ-TTM1
AER8-TTM2
AFKF-TTM1
AFUM-TTM1
AFE9-TTM1
AFE9-TTM4
AFE9-TTM2
AFE9-TTM3
AER2-TTM3
AER2-TTM1
HLA-A
HLA-B
HLA-C
Difference in gene expression: Metastasis – Primary
2
1
0
-1
-2
-3

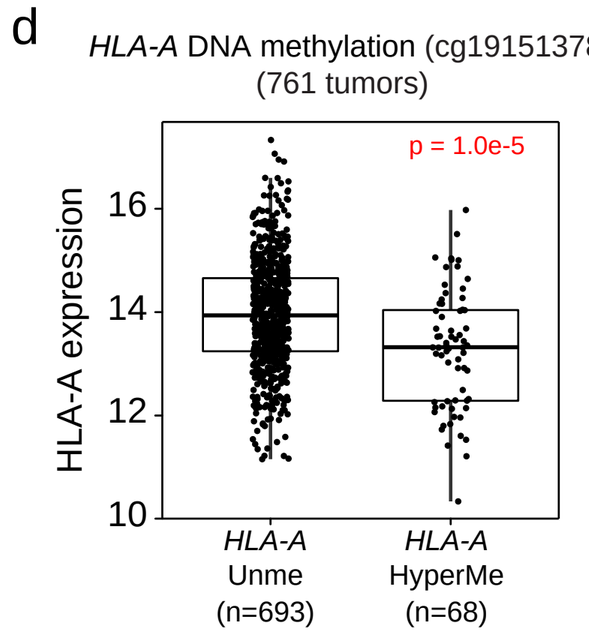
Extended Data Fig.7. *HLA-A* gene expression correlation with number of neoantigens. a. Linear relationship between number of neoantigens and *HLA-A*, -*B* and *C* gene expression Level 4 RNAseq data (see methods) of Basal-like only primary and metastatic tumors. Correlation was measured using the Pearson correlation coefficient. Statistically significant values are highlighted in red. **b**. Violin plots showing changes in gene expression for *HLA-A*, -*B*, and -*C* between primary and metastatic samples (Difference: Metastasis – Primary gene expression values) in Basals and Luminals/HER2E metastatic tumors. **c**. Patient-specific changes in gene expression for *HLA-A*, -*B*, and -*C* between primary and metastatic samples (Difference: Metastasis – Primary gene expression values) in Basal-likes (n=24 metastasis) and Luminals/HER2E (n=34 metastasis) metastatic tumors of AURORA cohort. Normal-like paired and unpaired tumors were removed from this analysis (Paired Normal and unpaired group from the “Pairs-PAM50-Prim” column of the Supplementary table 2).

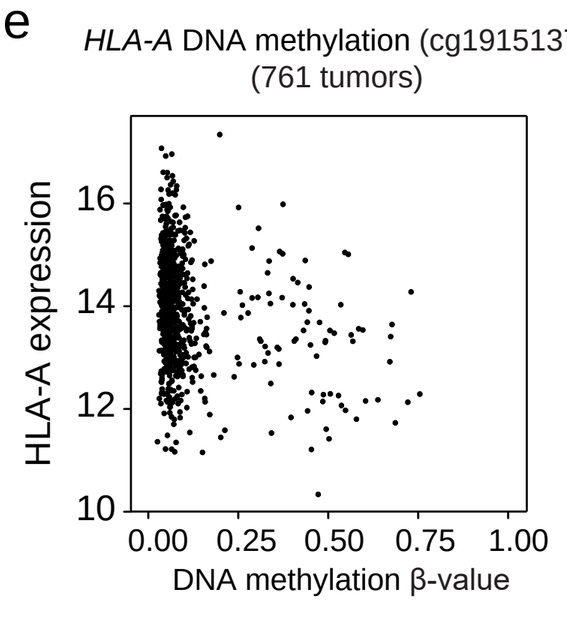
Extended Data Fig.8

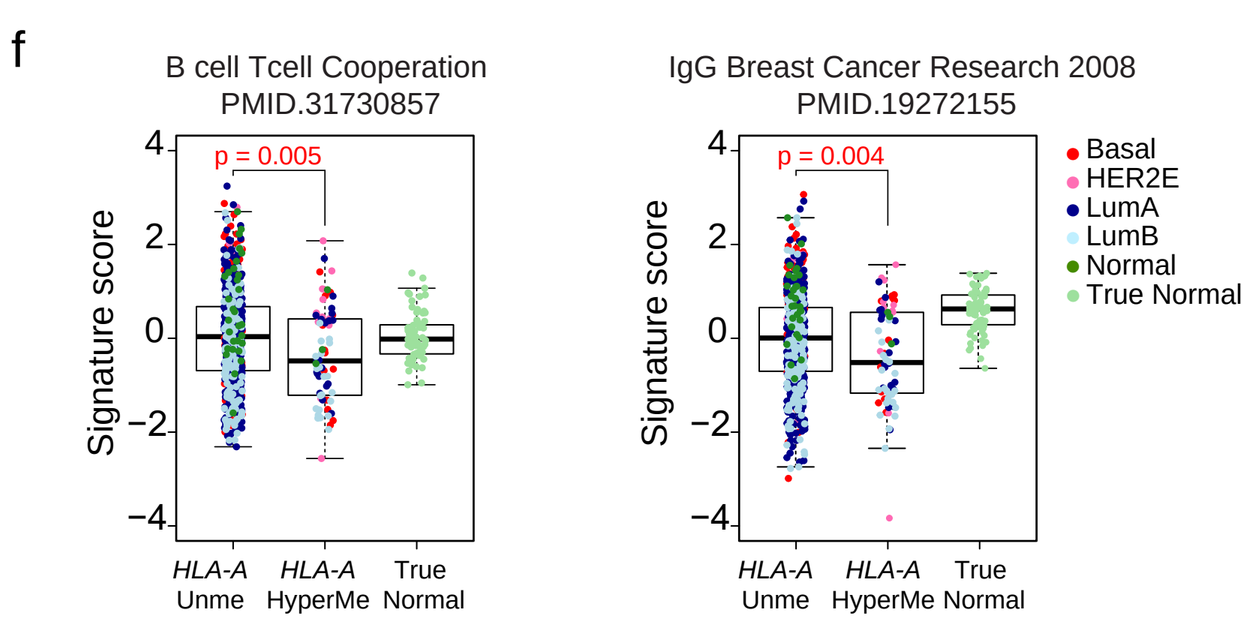


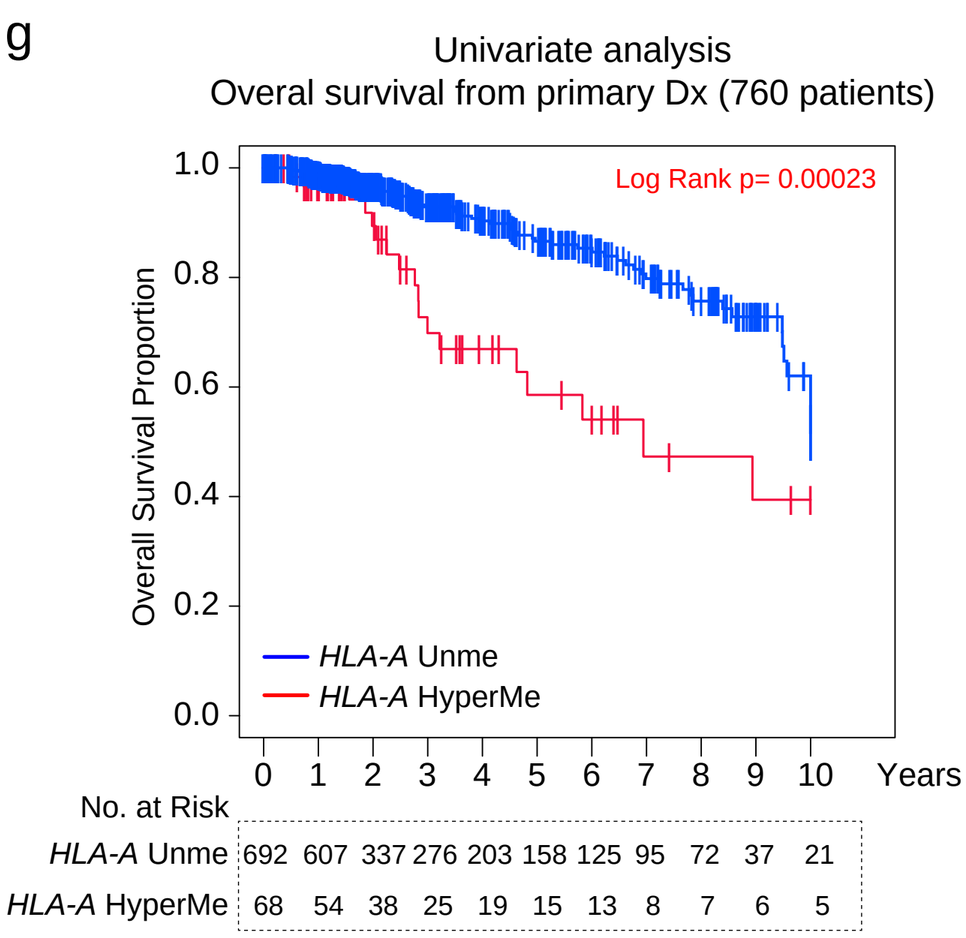












h

Multivarite analsyis
Cox Proportional Hazards Model (744 patients)

	No. of patients	Hazard ratio for death (95% IC)	P.value
HLA-A HyperMe			
HLA-A Unme	678	1	-
HLA-A HyperMe	66	2.06 (1.19, 3.55)	0.009
PAM50call			
LumA	399	1	-
LumB	142	1.68 (0.97 , 2.92)	0.665
HER2E	44	2.20 (0.99 , 4.87)	0.051
Basal	127	1.80 (0.99 , 3.26)	0.051
Normal	32	1.69 (0.59 , 4.81)	0.321
AJCC Stage			
Stage I	123	1	-
Stage II	423	1.66 (0.81 , 3.43)	0.169
Stage III	198	3.03 (1.44 , 6.37)	0.003

Extended Data Fig.8. *HLA-A* methylated primary tumors and prognostic value of *HLA-A* in TCGA data. a. Oncoprint diagram depicting *HLA-A* and *HLA-B* methylated cases using 761 primary tumors of TCGA BRCA dataset according to PAM50 molecular subtype. **b**. Proportion of each molecular subtypes found in *HLA-A* (68) and *HLA-B* (8) methylated tumors. **c**. Hypermethylated CpG sites in *HLA-A* (9 CpG sites) using 761 TCGA primary breast tumors and 74 tumor-adjacent breast tissues. **d**. Boxplots of *HLA-A* mRNA gene expression levels according to DNA methylation status. Comparisons between 2 paired groups were performed by t-test. **e**. Scatter plot showing the correlation between *HLA-A* mRNA expression values and DNA methylation levels (β-values). **f**. Boxplots of gene expression signature B cell/T cell cooperation and IgG scores according to DNA methylation status in tumors and tumor-adjacent breast tissues in TCGA-BRCA. Box-and-whisker plots display the median value on each bar, showing the lower and upper quartile range of the data and data outliers. The whiskers represent the interquartile range. Comparison between 2 groups was performed by ANOVA with post hoc Tukey's test. Statistically significant values are highlighted in red. Each mark represents the value of a single sample. **g**. Kaplan-Meier plots using the log rank test of overall survival from primary tumors according to *HLA-A* methylation status. **h**. Multivariable Cox proportional hazards analyses of TCGA BRCA patients for overall survival prediction using the covariates of *HLA-A* methylation status, PAM50 subtypes and tumor stage (10 Stage IV patients were removed from the analysis). Hazard ratio (HR) = 1: no effect. HR < 1: reduction in hazard. HR > 1: increase in hazard. Statistically significant values are highlighted in red. Unme, unmethlylated; HyperMe, hypermethylated.

a
JAM3
Paired data
p = 0.0064
Promoter DNA methylation (cg24625128)
1.0
0.8
0.6
0.4
0.2
0.0
P
M

b
JAM3
Paired data
p = 1.45e-05
p = 0.07734
1
2
Promoter DNA methylation (cg24625128)
P-Unme
M
P-Me
M
Paired data
Paired data

c
JAM3
All
Luminals/HER2E
Basals
p = 0.1
p = 0.038
p = 0.88
JAM3 mRNA expression
4
2
0
−2
−4
P
M

d
YBX3
Paired data
p = 0.0035
Promoter DNA methylation (cg05406791)
1.0
0.8
0.6
0.4
0.2
0.0
P
M

e
YBX3
Paired data
p = 1.45e-04
p = 0.6
1
2
Promoter DNA methylation (cg05406791)
P-Unme
M
P-Me
M
Paired data
Paired data

f
YBX3
All
Luminals/HER2E
Basals
p = 0.01
p = 0.001
p = 0.96
YBX3 mRNA expression
4
2
0
−2
−4
−6
P
M

g
SYNDIG1
Paired data
p = 0.15
Promoter DNA methylation (cg19654195)
1.0
0.8
0.6
0.4
0.2
0.0
P
M

h
SYNDIG1
Paired data
p = 0.036
p =0.071
1
2
Promoter DNA methylation (cg19654195)
P-Unme
M
P-Me
M
Paired data
Paired data

i
SYNDIG1
All
Luminals/HER2E
Basals
p = 0.028
p = 0.001
p = 0.56
SYNDIG1 mRNA expression
6
4
2
0
−2
−4
−6
P
M

Extended Data Fig. 9. DNA methylation alterations associated with metastatic tumors. **a-i**. Analysis of metastasis-associated promoter DNA hypermethylation of three genes (*JAM3, YBX3 and SYNDIG1*) encoding components of tight junctions or regulation of adhesion molecules. For each gene, a comparison of promoter CpG DNA methylation between primary and metastatic tumors is shown on the left (a, d, g), a second comparison of promoter CpG DNA methylation between ①unmethylated primaries (β-value of <0.3) and their paired metastasis and ②methylated primaries (β-value of > 0.3) with their paired metastasis (b, c, e) is shown in the middle, and a third comparison of gene expression between primary and metastatic tumors based on all samples (All) , Luminal A-B and HER2E only (Luminals/HER2E), and basal-like subtype only (Basals) is shown on the right (c, f, i). P, primary; M, metastasis; P-Unme, Unmethylated primary; P-Me, Methylated primary.