

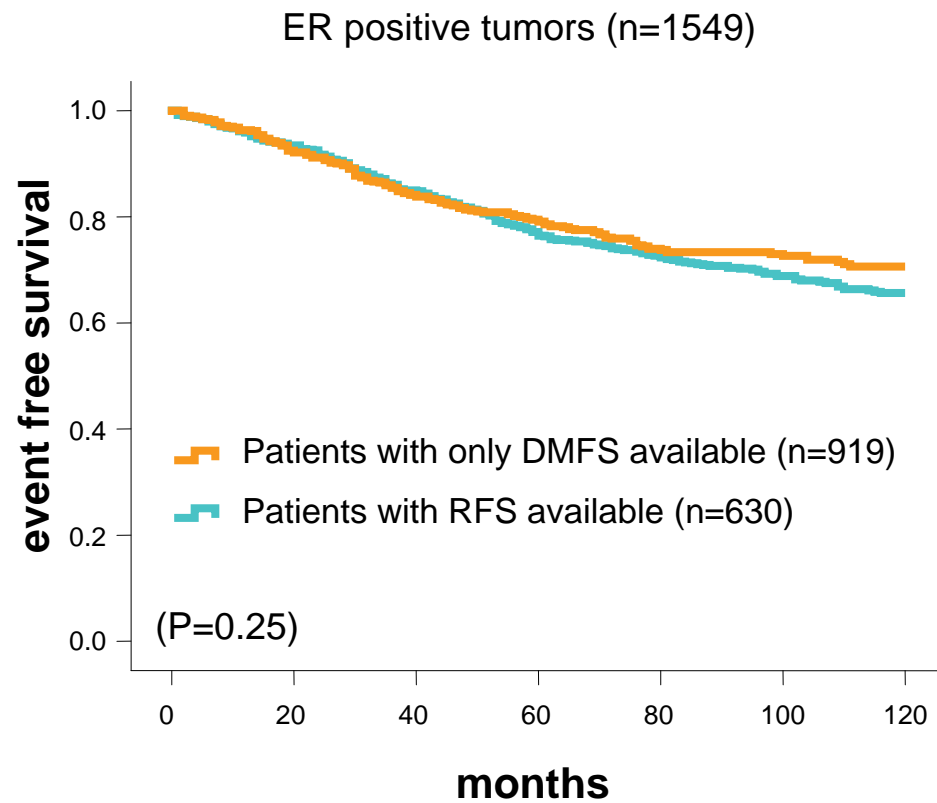
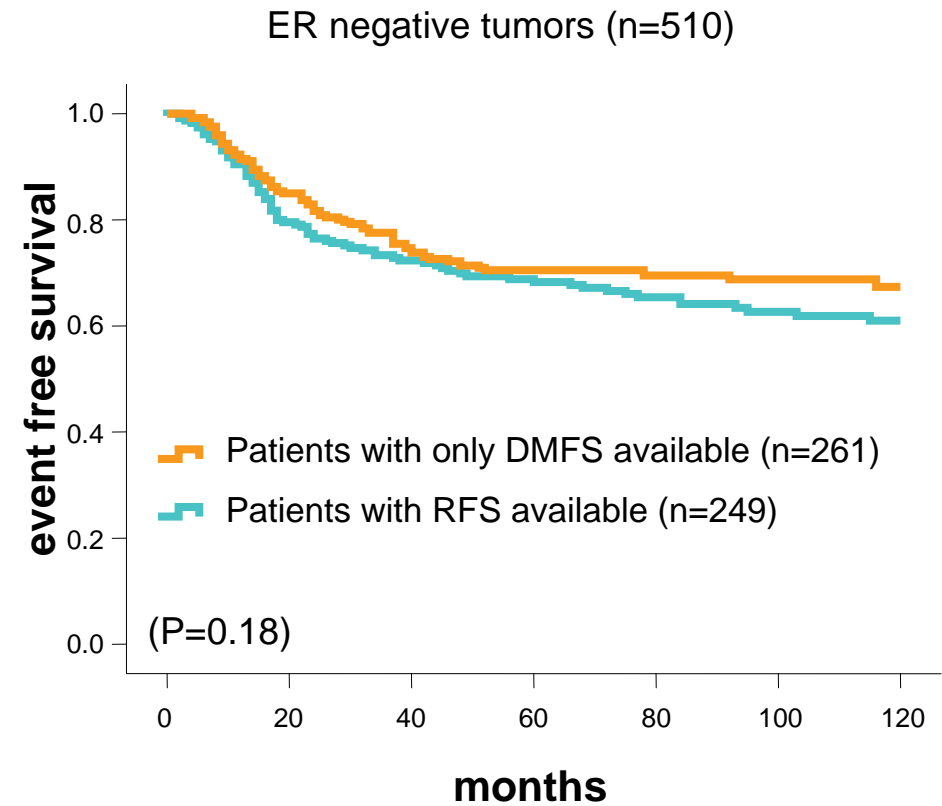
**Supplementary Table S1: Methods applied to determine the biochemical status of ER, PgR, and HER2 in the different datasets.**

Dataset	Biochemical Estrogen receptor status			Biochem. Progesterone receptor status			HER2 status			Reference
	n=	Method	cutoff	n=	Method	Cutoff	n=	Method	Cutoff	
<b>Rotterdam</b>	344	LBA, EIA, (IHC n=9)	10 fmol/mg, (IHC: 10% pos. cells)	256	LBA, EIA	4 fmol/mg,	n.a.	n.a.	n.a.	14, 15, 16
<b>TransBIG</b>	198	IHC	not given	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	18
<b>Oxford-Untreated</b>	63	not given	not given	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	19
<b>London</b>	87	not given	not given	85	not given	not given	n.a.	n.a.	n.a.	20
<b>London-2</b>	77	not given	not given	77	not given	not given	n.a.	n.a.	n.a.	21
<b>Oxford-Tamoxifen</b>	109	not given	not given	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	20
<b>Veridex-Tam</b>	136	IHC, LBA	10% pos. cells, 10 fmol/mg,	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	22
<b>Frankfurt-3</b>	50	IHC	10% pos. tumor cells	46	IHC	10% pos. tumor cells	19	IHC / FISH	IHC3+ OR FISH >2.0	12
<b>Uppsala</b>	249	EIA	>0.05 fmol/μg DNA	251	EIA	>0.05 fmol/μg DNA	n.a.	n.a.	n.a.	25
<b>San Francisco</b>	118	not given	not given	117	not given	not given	79	not given	not given	26
<b>New York</b>	99	not given	not given	98	not given	not given	88	n.a.	n.a.	27
<b>Frankfurt</b>	114	IHC	10% pos. tumor cells	112	IHC	10% pos. tumor cells	65	IHC / FISH	IHC3+ OR FISH >2.0	10
<b>Frankfurt-2</b>	65	IHC	10% pos. tumor cells	65	IHC	10% pos. tumor cells	57	IHC / FISH	IHC3+ OR FISH >2.0	13
<b>MDA133</b>	133	IHC	10% pos. tumor cells	133	IHC	10% pos. tumor cells	132	IHC / FISH	IHC3+ OR FISH >2.0	8
<b>EORTC</b>	48	IHC	not given	47	IHC	not given	n.a.	n.a.	n.a.	29
<b>Edinburgh</b>	116	IHC	not given	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	30
<b>expO</b>	153	not given	not given	151	not given	not given	141	not given	not given	31
<b>Boston</b>	39	not given	not given	39	not given	not given	37	not given	not given	34
<b>TOTAL</b>	<b>2198</b>			<b>1474</b>			<b>618</b>			

**Supplementary Table S2: Univariate and multivariate Cox regression analysis of disease free survival according to microarray based ER, PgR, and HER2 status among 2058 breast cancer patients.**

Parameter	n=	Univariate			Multivariate			
		P Value*	Hazard Ratio	95 % CI	P Value*	Hazard Ratio	95 % CI	
ER status	negative vs positive	510 vs. 1548	<b>0.002</b>	1.34	(1.12-1.60)	0.78	1.03	(0.84-1.27)
PgR status	negative vs positive	854 vs. 1204	<b>&lt;0.001</b>	1.55	(1.32-1.82)	<b>&lt;0.001</b>	1.48	(1.23-1.78)
HER2 status	positive vs negative	238 vs. 1820	<b>0.002</b>	1.43	(1.13-1.80)	0.083	1.24	(0.97-1.58)

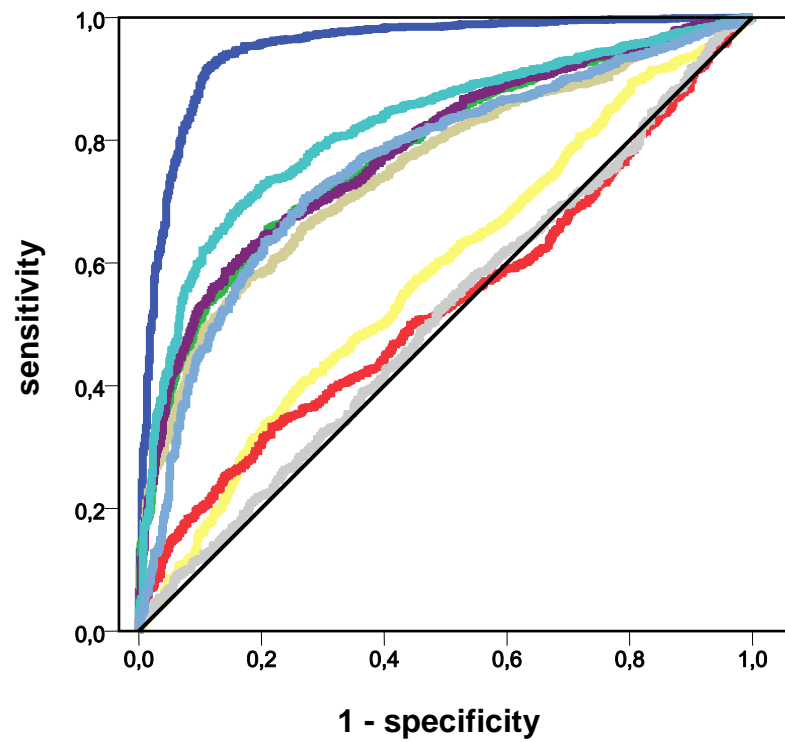
\*significant P values are given in bold










**A****B**

**Supplementary Figure S1: Event free survival of patients according to the available endpoint.**

A) Event free survival of ER positive breast cancers (n=1549) with either RFS (n=919) or DMFS (n=630) event status available.

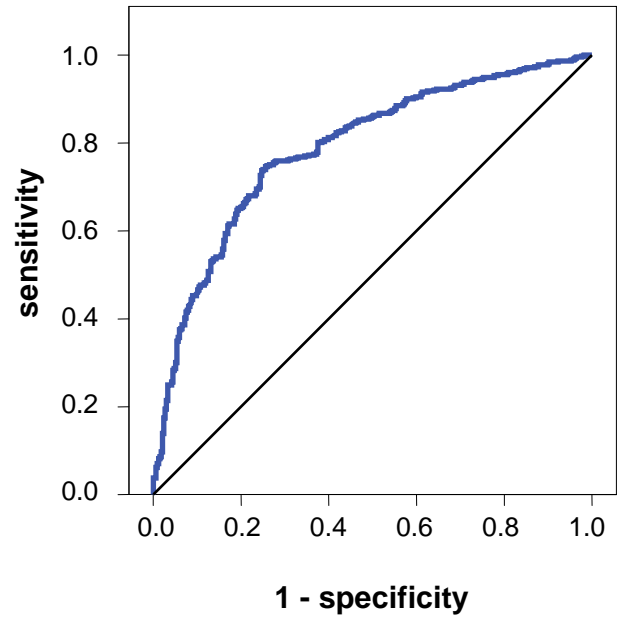
B) Event free survival ER negative breast cancers (n=510) with either RFS (n=261) or DMFS (n=249) event status available.



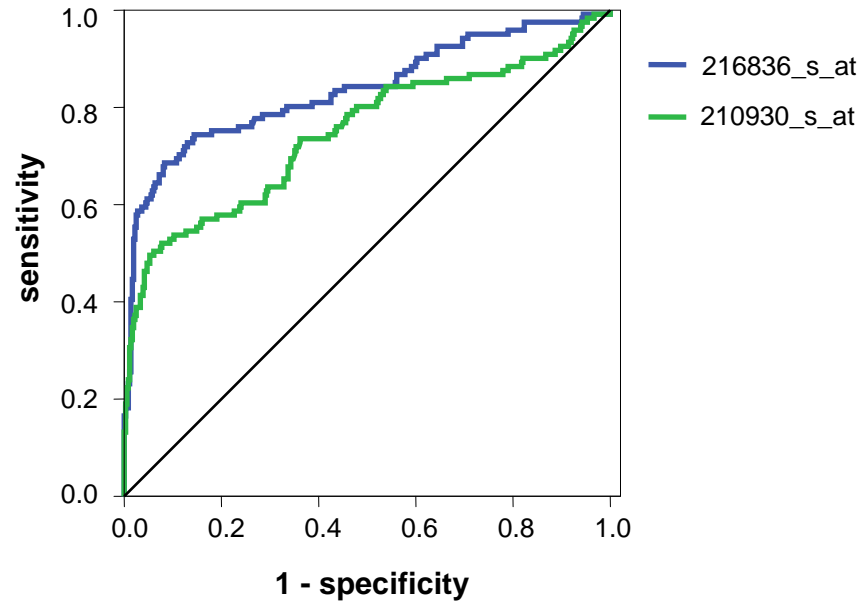
	Probe Set	AUC	95% CI	
	205225_at	0.949	0.938	0.960
	215552_s_at	0.820	0.802	0.839
	211235_s_at	0.783	0.763	0.802
	211233_x_at	0.779	0.759	0.799
	217190_x_at	0.759	0.738	0.781
	211234_x_at	0.752	0.731	0.772
	211627_x_at	0.580	0.554	0.606
	215551_at	0.533	0.508	0.558
	217163_at	0.513	0.486	0.539

**Supplementary Figure S2: ROC curves demonstrating the relationship of sensitivity and specificity of different Affymetrix ProbeSets for the estrogen receptor with biochemical data of the estrogen receptor status.**

# A

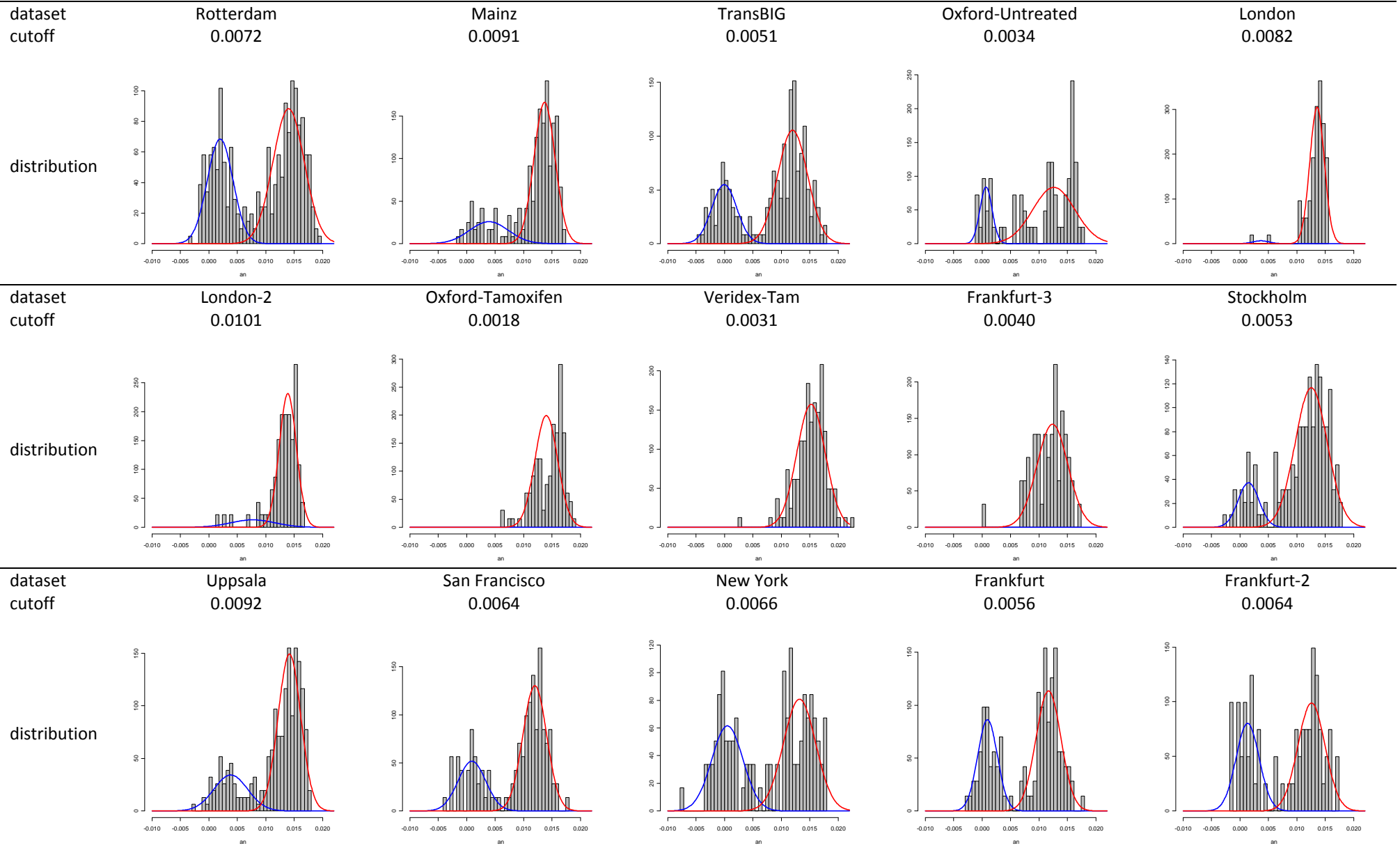


# B



**Supplementary Figure S3: ROC curves demonstrating the relationship between sensitivity and specificity of Affymetrix ProbeSets for the PgR (A) and HER2 genes (B) with the results from immunohistochemistry of the progesterone receptor and HER2 (3+ IHC or FISH positive).**

## Supplementary Figure S4: Distribution of ER expression values in the individual datasets.

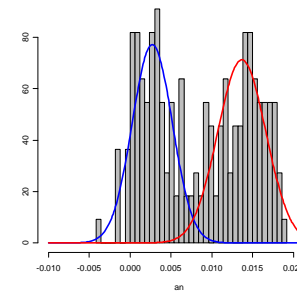
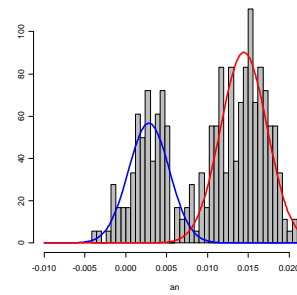
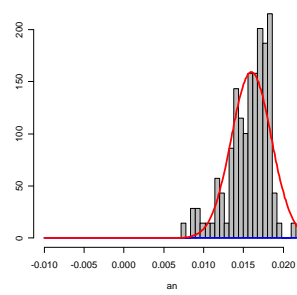
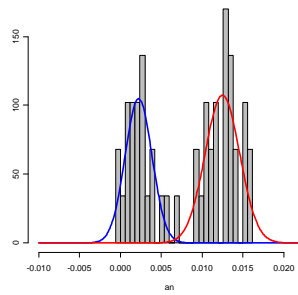
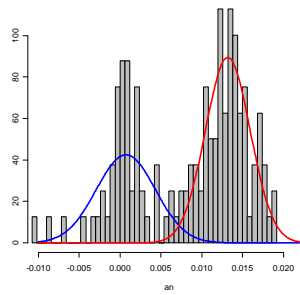


---

dataset	MDA133	EORTC	Edinburgh	expO	Singapore
cutoff	0.0072	0.0067	0.0061	0.0080	0.0078

---

distribution

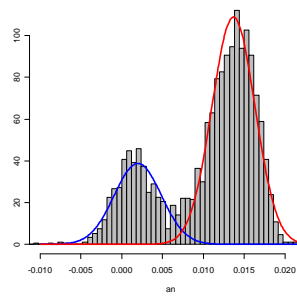
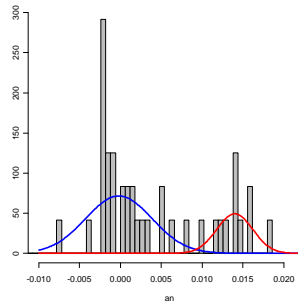
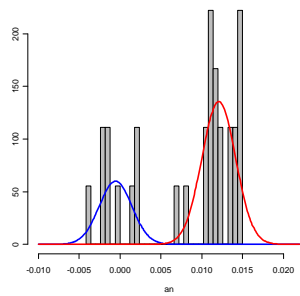


---

dataset	Genentech	Boston	combined
cutoff	0.0054	0.0093	0.0075

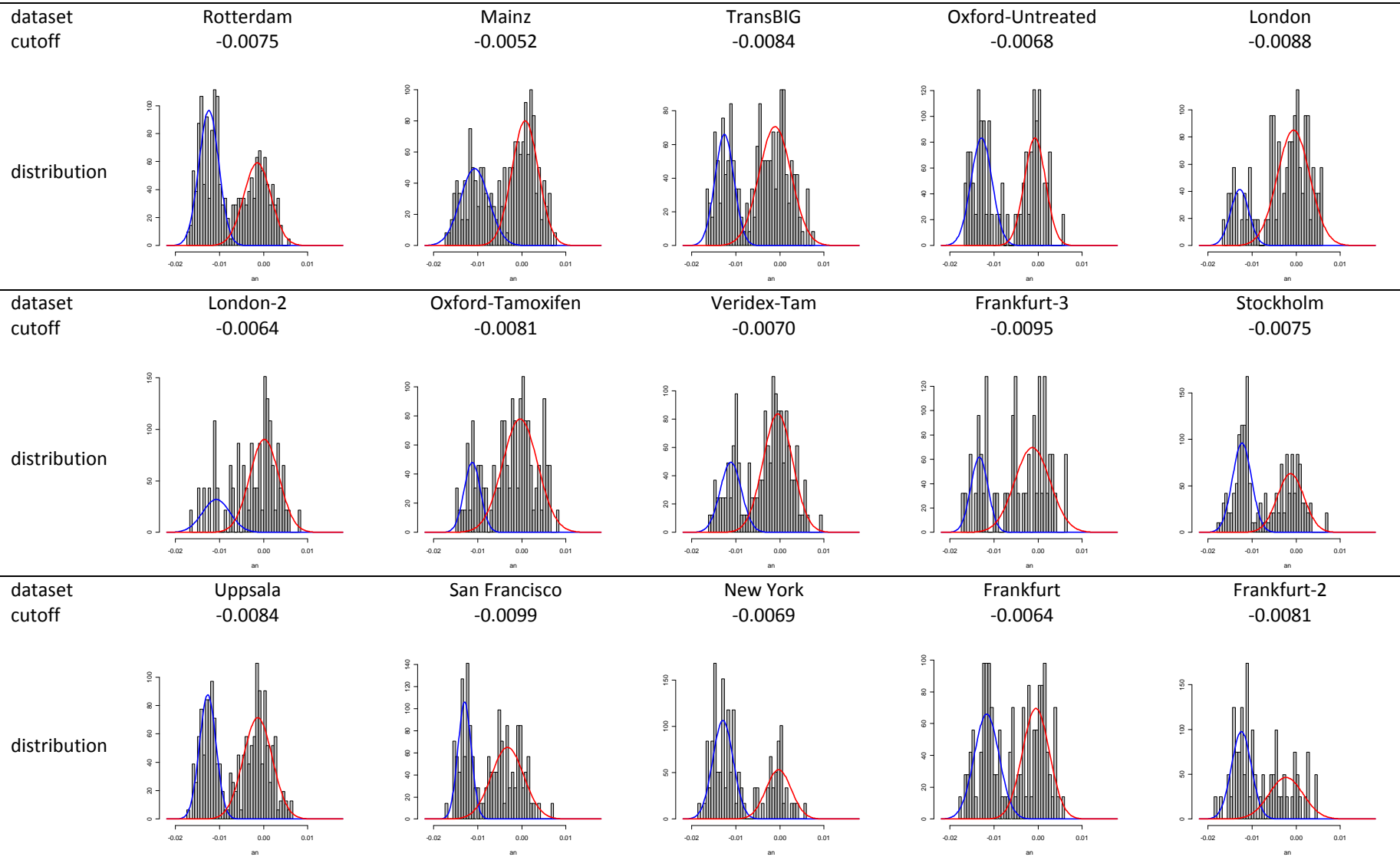
---

distribution



---

## Supplementary Figure S5: Distribution of PgR expression values in the individual datasets.

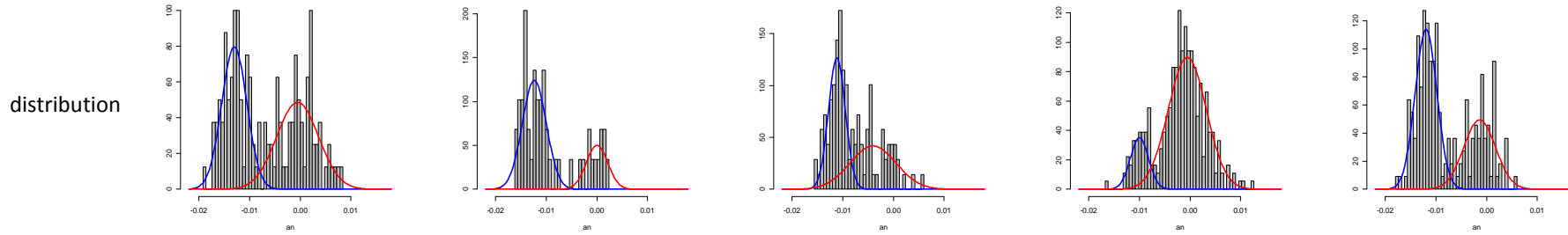




---

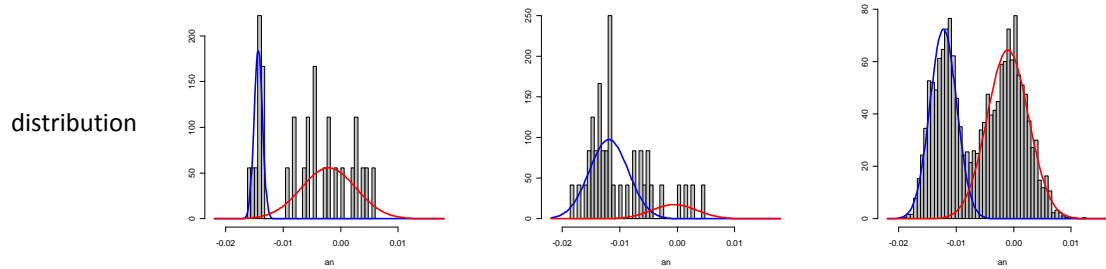
dataset	MDA133	EORTC	Edinburgh	expO	Singapore
cutoff	-0.0079	-0.0054	-0.0082	-0.0076	-0.0071

---

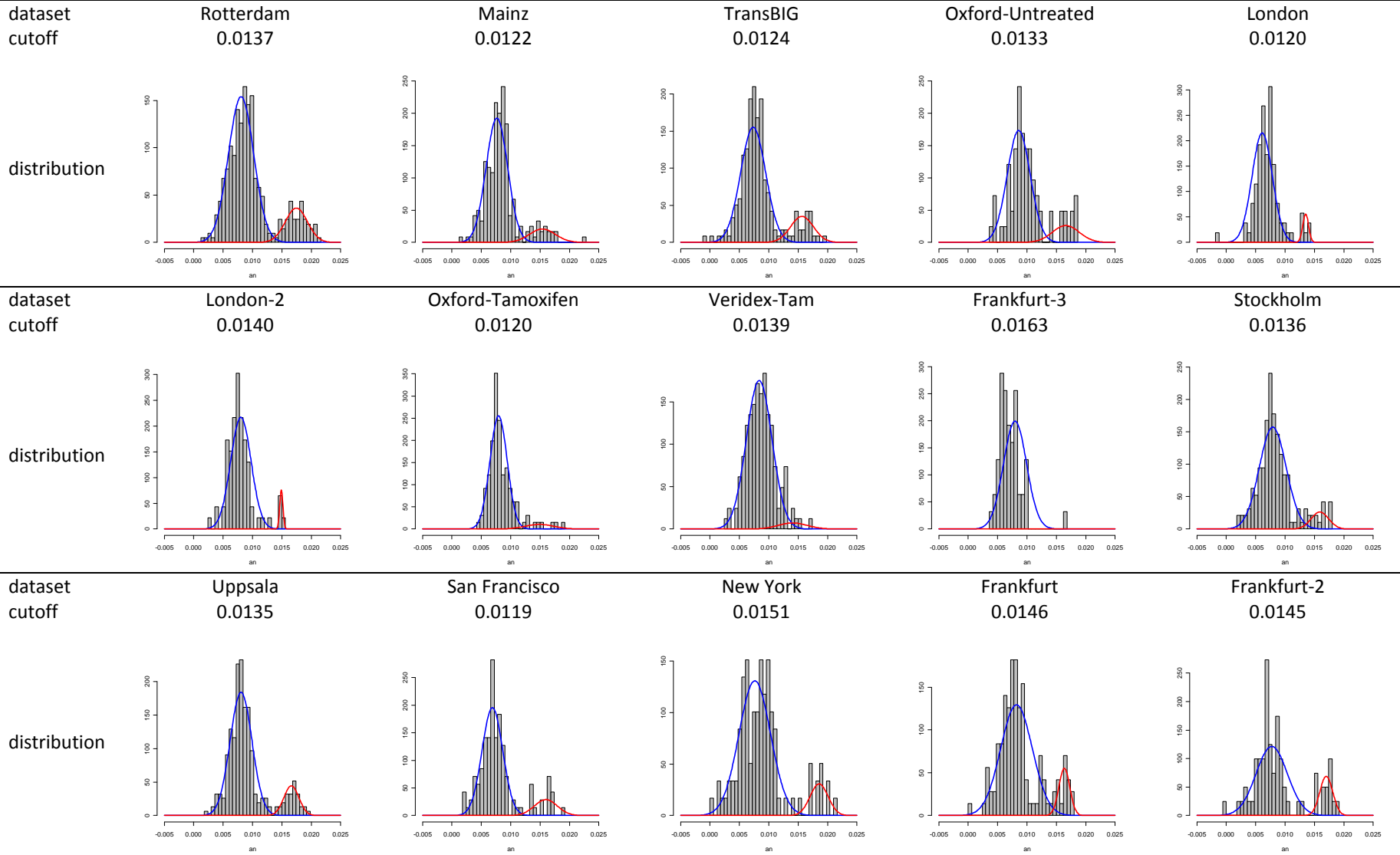


dataset	Genentech	Boston	combined
cutoff	-0.0123	-0.0047	-0.0078

---

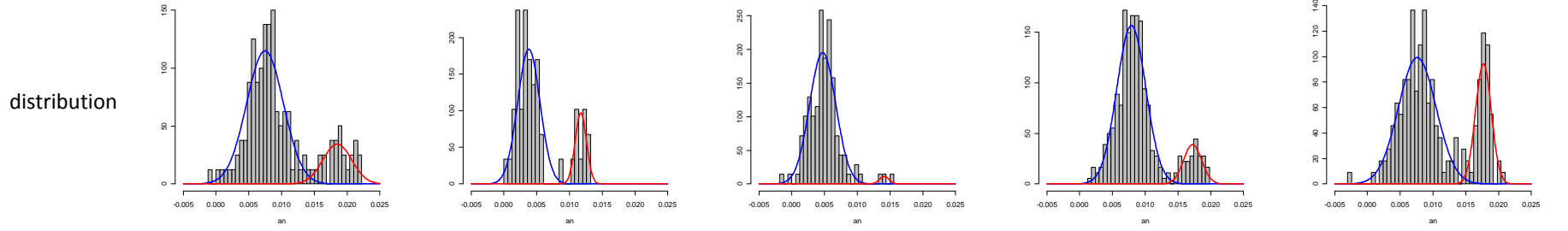


## Supplementary Figure S6: Distribution of HER2 expression values in the individual datasets.



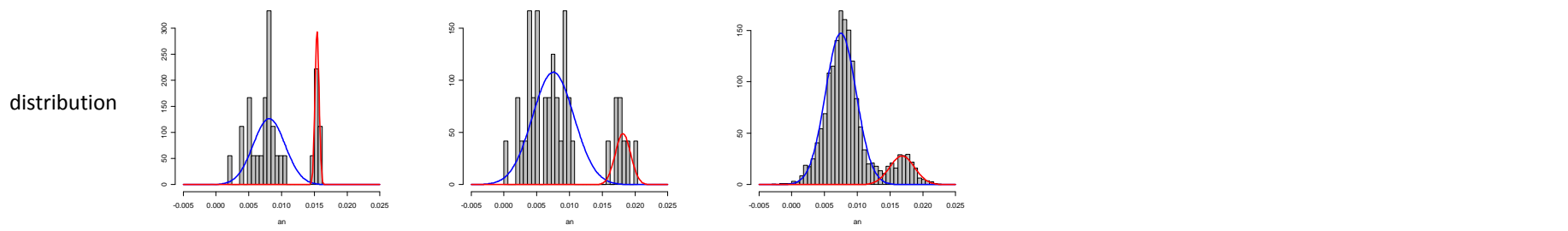
---

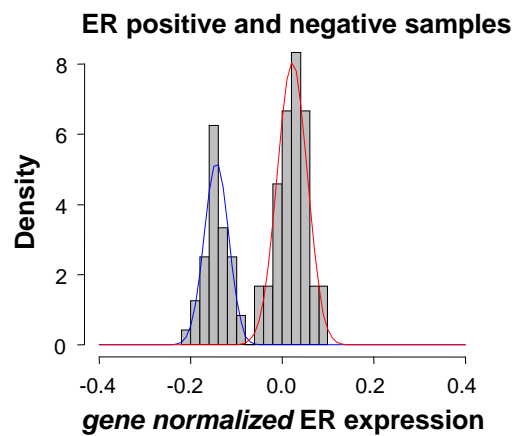
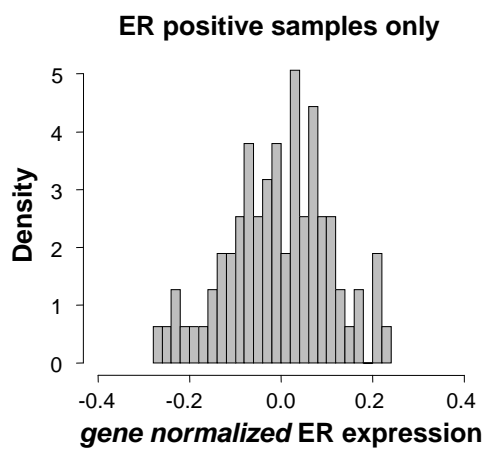
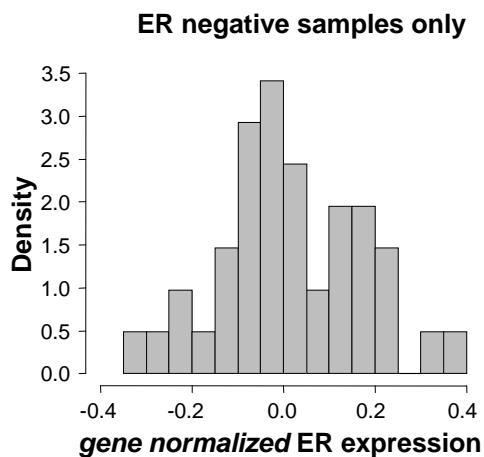
dataset	MDA133	EORTC	Edinburgh	expO	Singapore
cutoff	0.0144	0.0092	0.0121	0.0141	0.0147



---

dataset	Genentech	Boston	combined
cutoff	0.0145	0.0155	0.0135



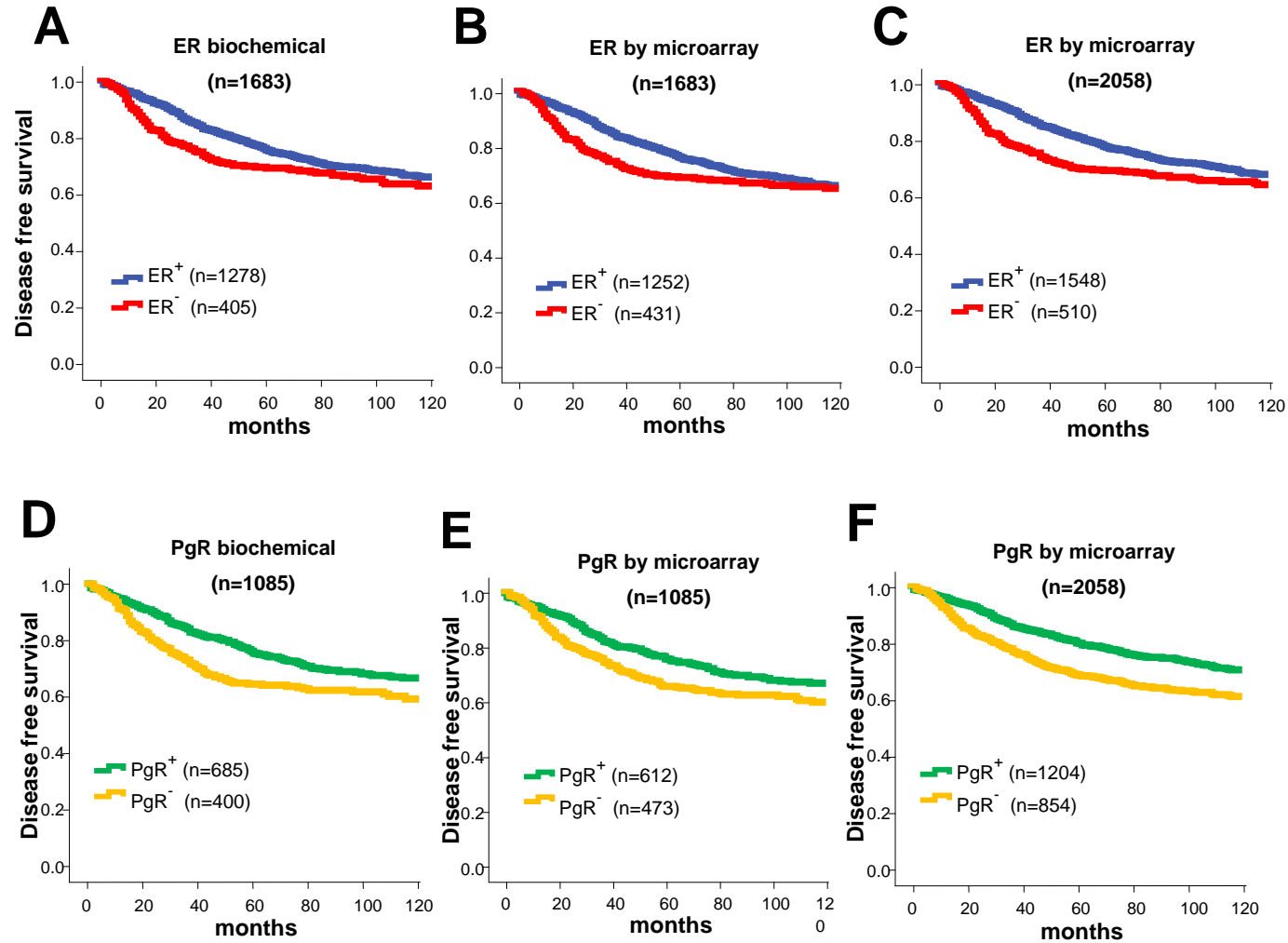
**A****B****C**

**Suppl. Figure S7: Influence of *gene normalization* on the distribution of ER expression values.**

A) Distribution of *gene normalized ER expression* values (ProbeSet 205225\_at) in the Frankfurt dataset (n=120 samples). Two normal distributions were fitted to the data allowing separation of ER positive and ER negative samples.

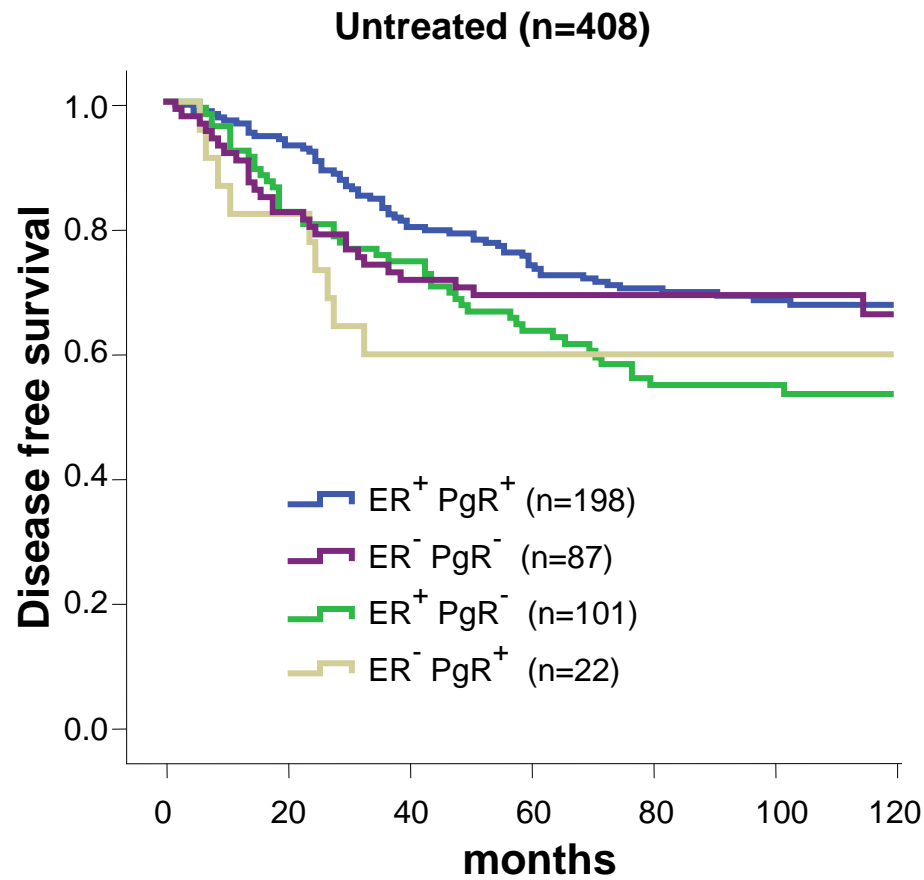
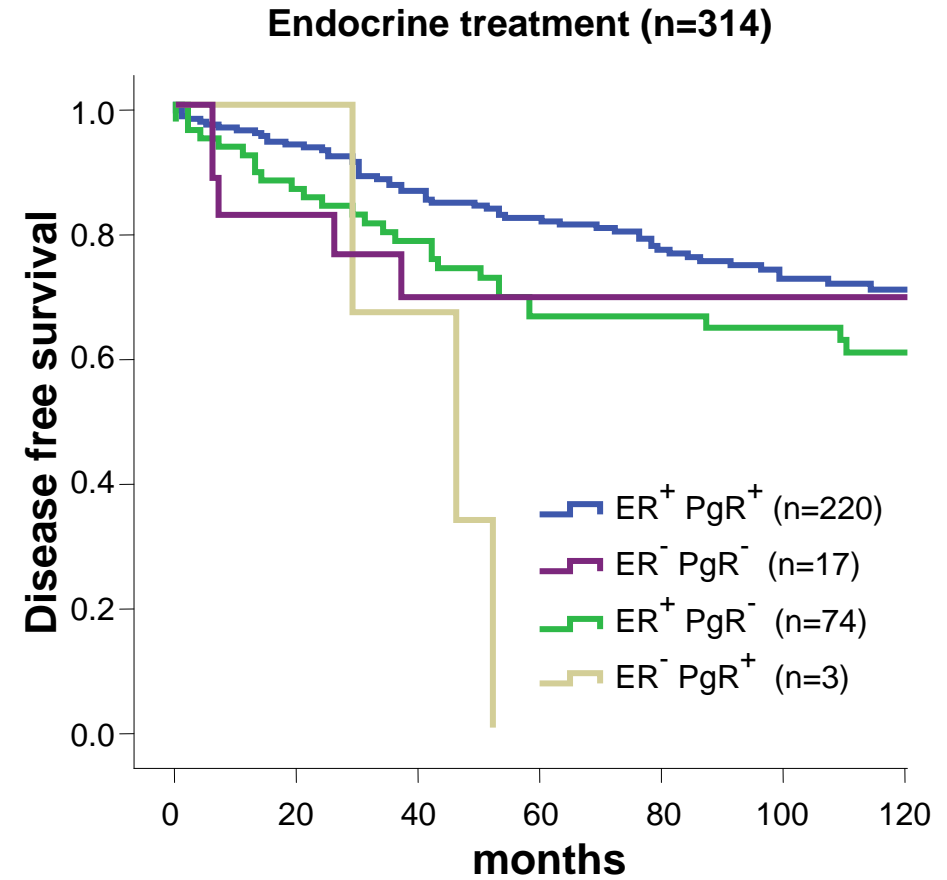
B) Distribution of ER expression values after *gene normalization* of only ER positive samples (n=79) from the Frankfurt dataset. The same scale as in (A) has been used for the X axis.

C) Distribution of ER expression values after *gene normalization* of only ER negative samples (n=41) from the Frankfurt dataset. The same scale as in (A) has been used for the X axis.



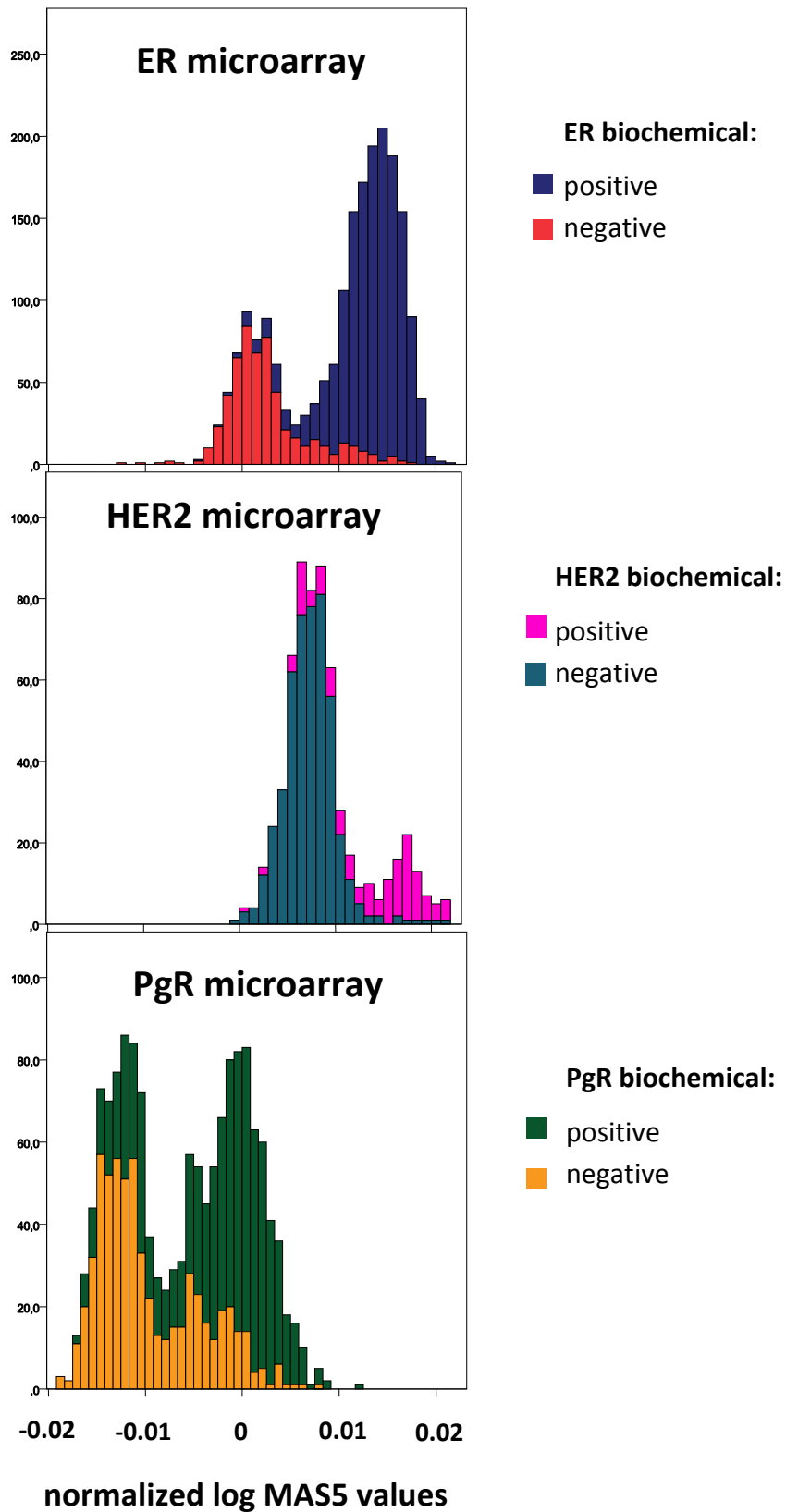
**Supplementary Figure S8: Disease free survival of patients according to biochemical and microarray derived status for ER and PgR.**

Kaplan Meier analysis of disease free survival of patients stratified according biochemical (A,D) or microarray (B,C,E,F) status of ER (A,B,C) and PgR (D,E,F). Either only samples with biochemical derived status (A,B,D,E) or all 2058 samples with follow up data (C,F) were included.

**A****B**

**Supplementary Figure S9: Disease free survival of untreated (A) and endocrine treated (B) patients stratified according to ER and PgR status based on microarray.**

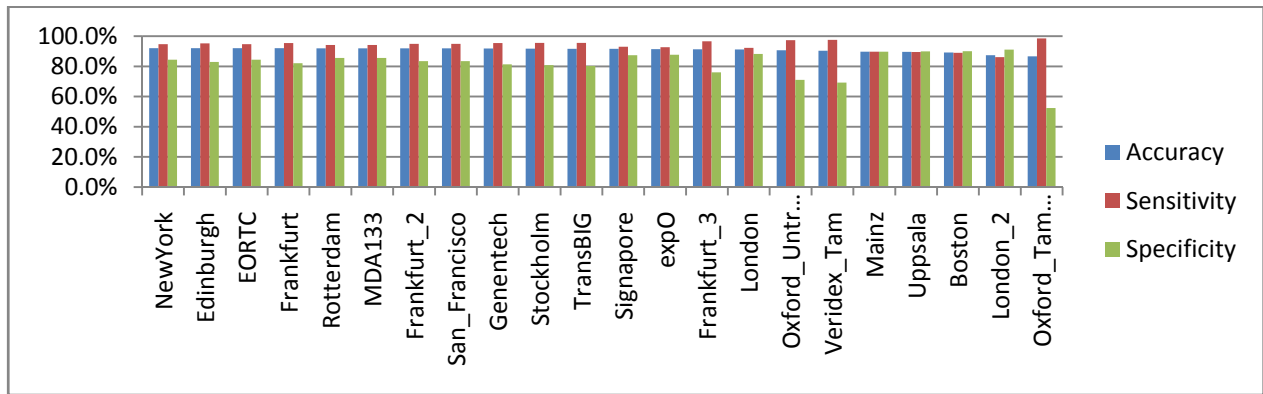
All patients were with available follow up information were selected which were either untreated or treated only with adjuvant endocrine therapy and which have both ER and PgR status available based on microarray as well as biochemical assay (n=722).



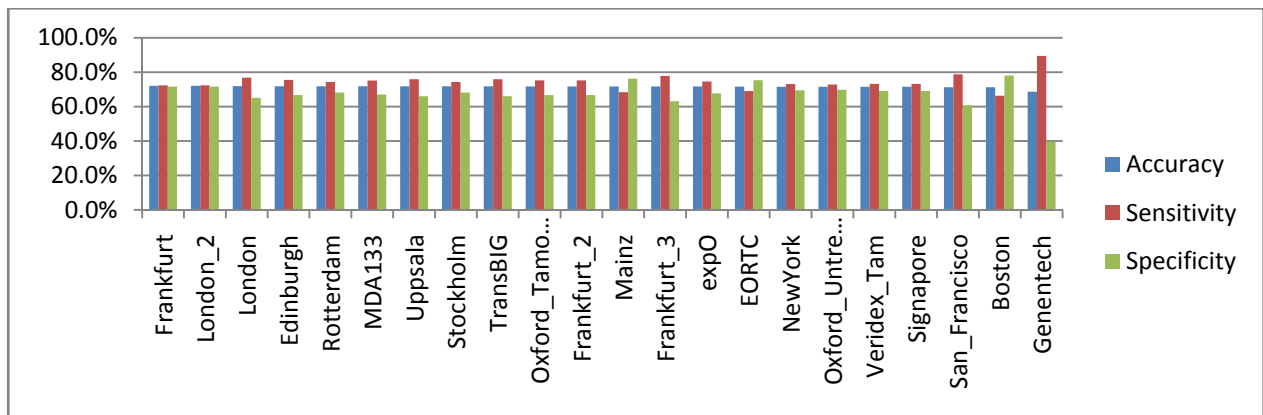
**Supplementary Figure S10: Comparison of Affymetrix expression values for ER, HER2 and PgR.**

Distribution of expression values are given on an identical scale on the x-axis (ProbeSets 205225\_at, 216836\_s\_at, and 208305\_at for ER, HER2 and PgR, respectively). Color codes represent the biochemical derived status of the samples for ER, HER2, and PgR.

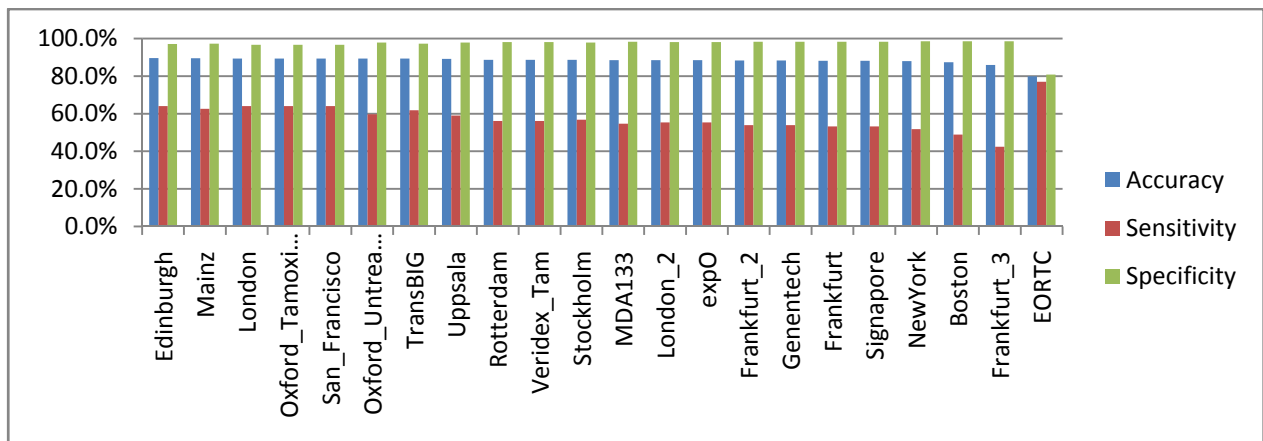
### ER status:



### PgR status:



### HER2 status:



**Supplementary Figure S11: Comparison of biochemical ER, HER2 and PgR status with microarray derived status using individual dataset specific cutoff values in a training-validation approach.**

Specific cutoff values derived from fitting in individual datasets (as training sets) were validated using all remaining samples with information on biochemical status (validation sets). The observed values for overall accuracy sensitivity and specificity are given for each cutoff derived from a single dataset.