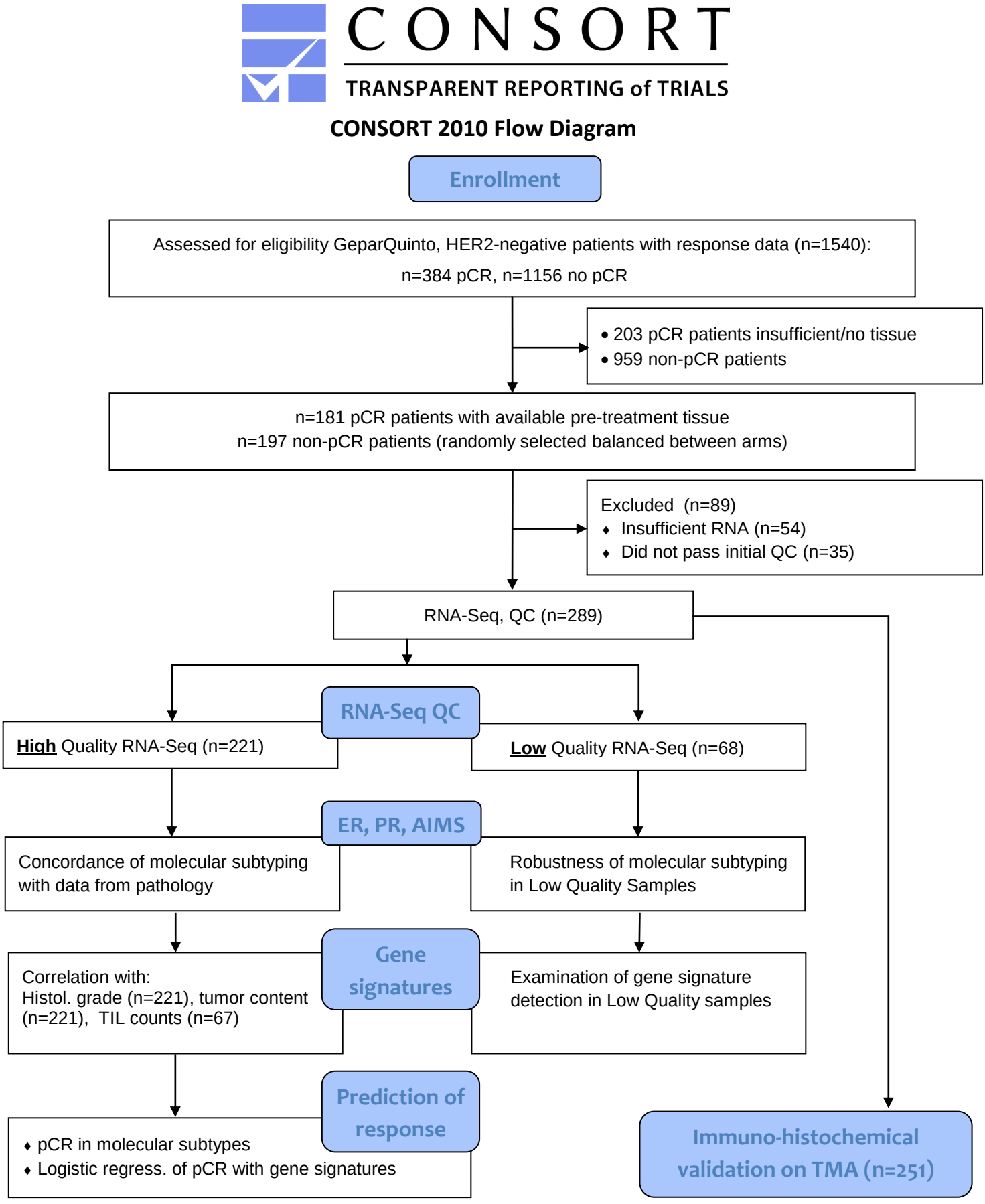
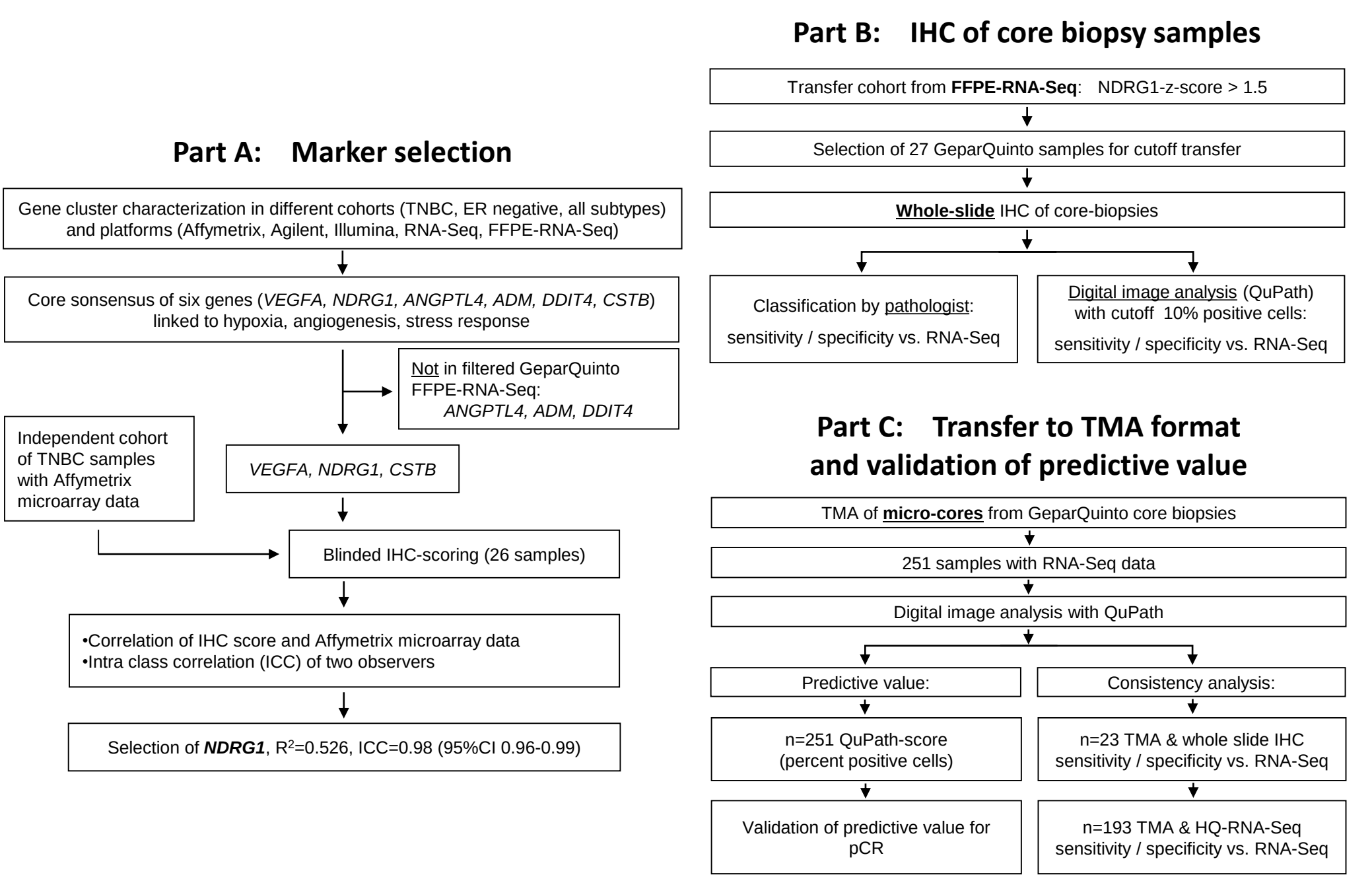
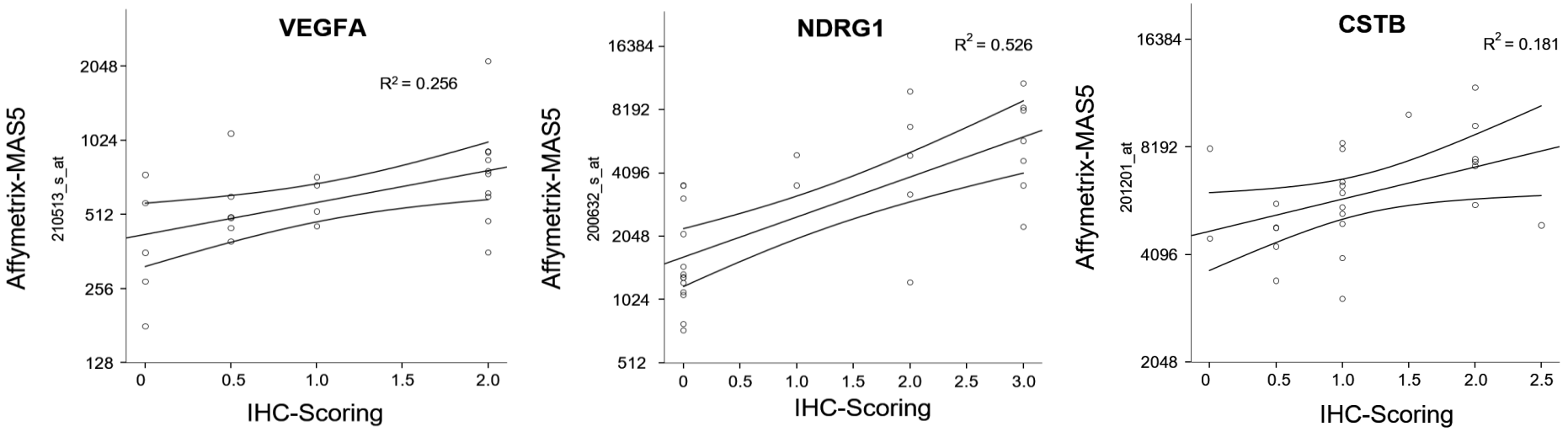


Supplementary Figure S1: Flow of samples through the study

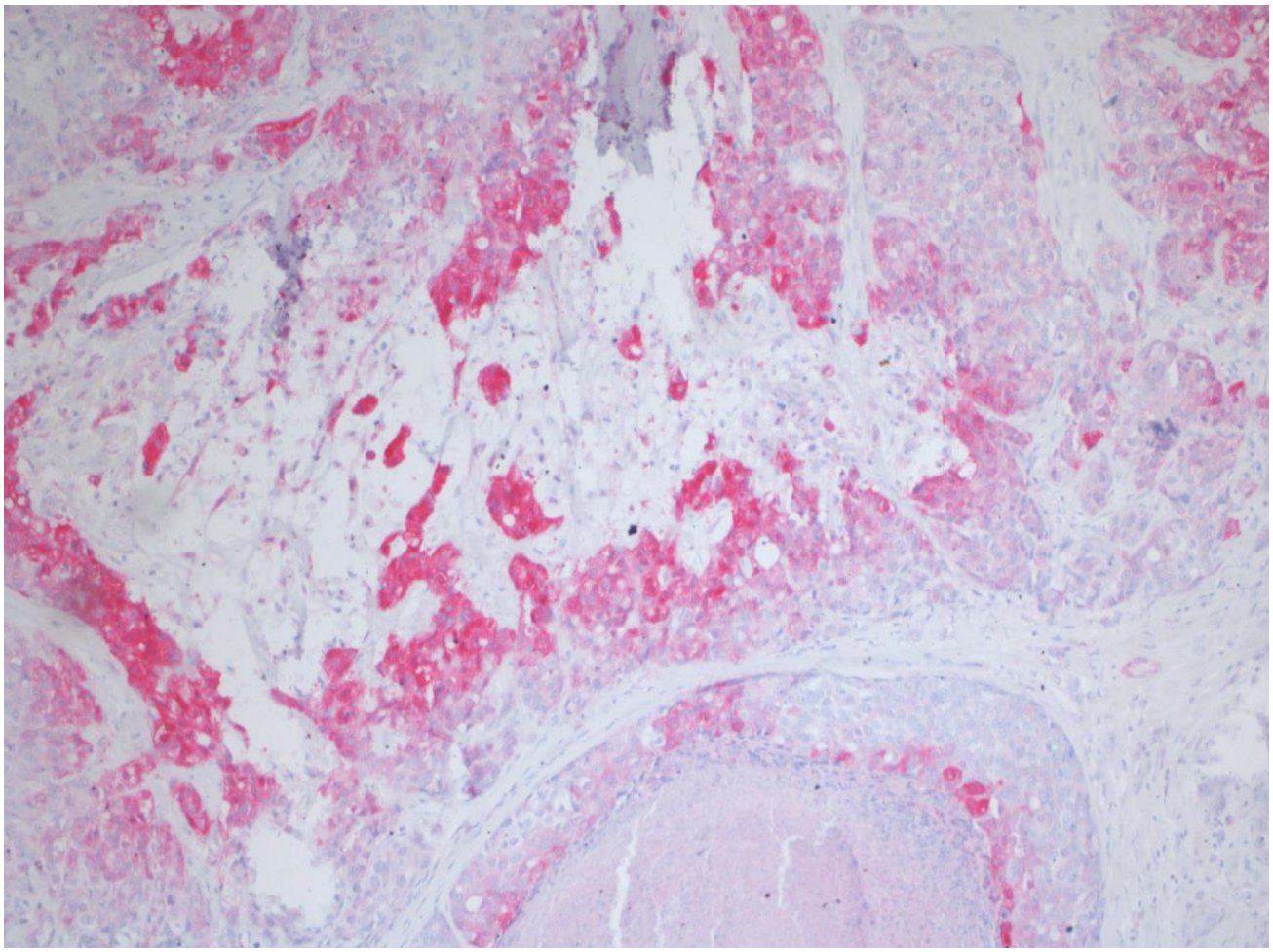


Supplementary Figure S2: Immunohistochemical validation of hypoxia signature



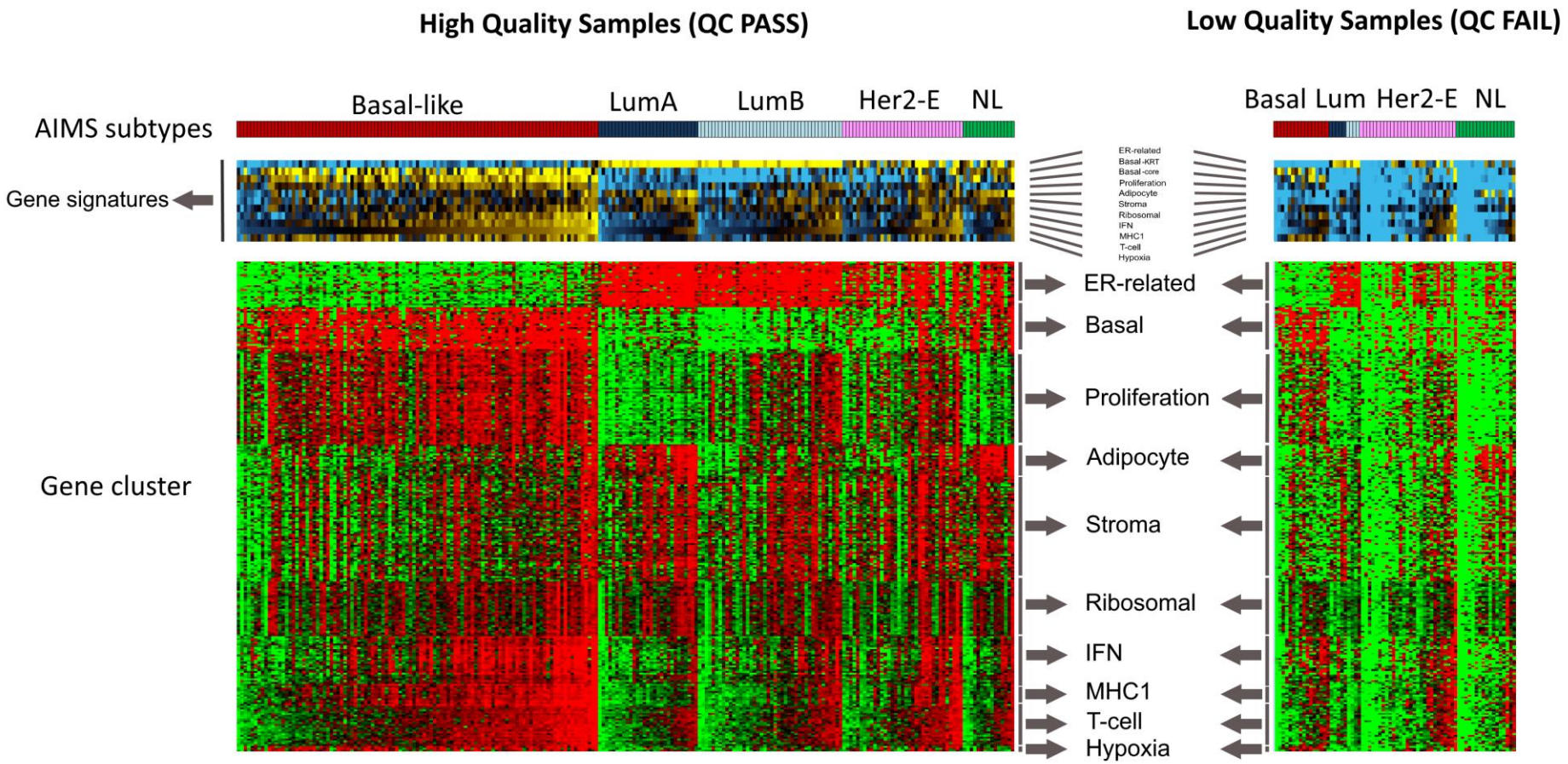
Supplementary Figure S3: Correlation of Affymetrix microarray expression data and IHC scoring of three genes in independent TNBC dataset

RNA expression from Affymetrix on y-axis is compared with immunohistochemical scoring for three individual genes (VEGFA, NDRG1, CSTB) from the hypoxia signature. R^2 values are from Spearman rank correlation.



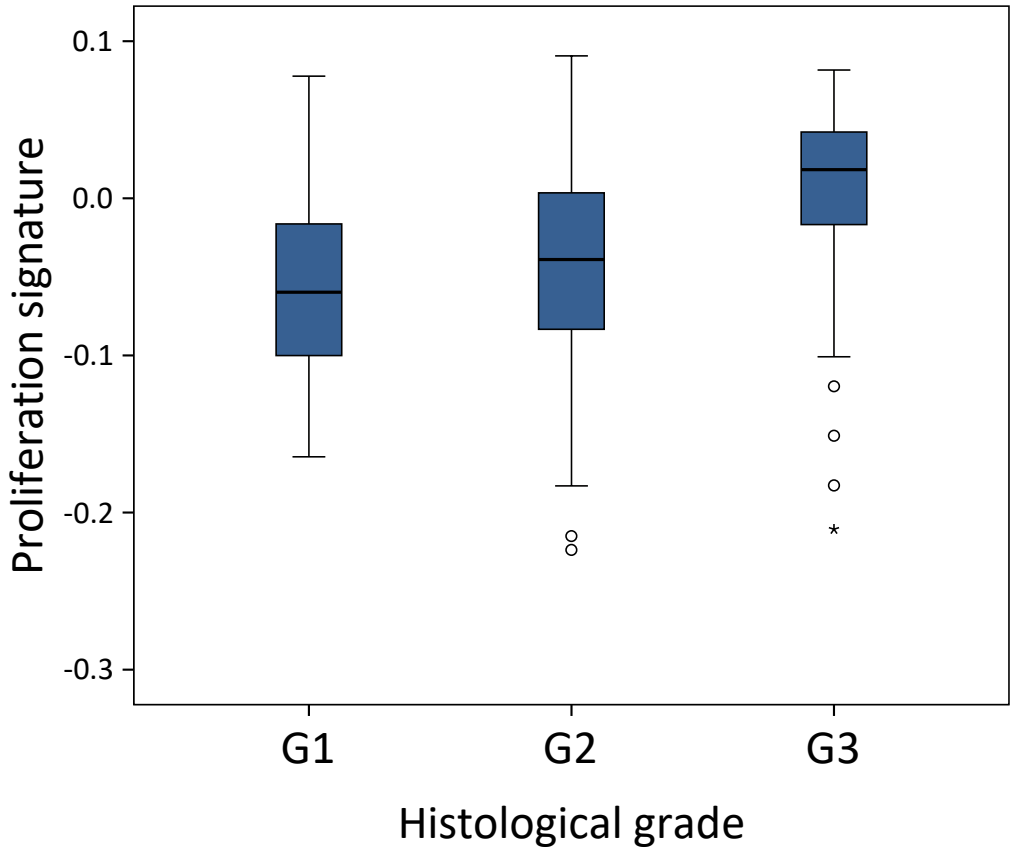
Supplementary Figure S4: Para-necrotic expression of NDRG1 in TNBC

Strong expression of NDRG1 is detected in regions of tumor necrosis (upper part) as well as DCIS necrosis (lower part).



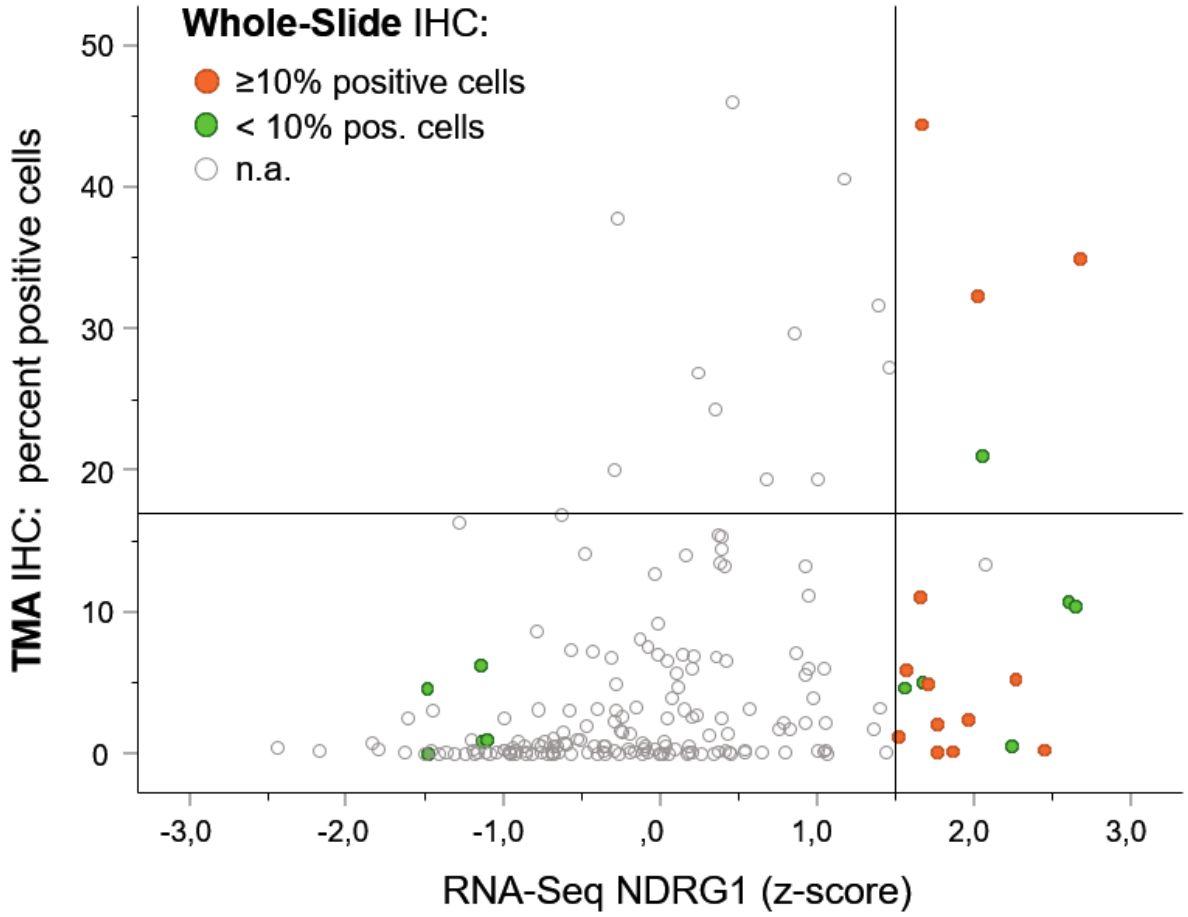
Supplementary Figure S5: Comparison of signature detection in High Quality and Low Quality Samples

Upper row presents results from AIMS subtyping for 221 *High Quality Samples* (left) and 68 *Low Quality samples* (right). Log_2 count data of the 268 genes shown in the lower (red/green) heatmaps were normalized by median centering across all 289 samples together for the analysis shown here. Gene signatures in the above (yellow/blue) heatmaps were calculated as mean values of the respective gene clusters in the lower panels. The *Low Quality samples* on the right show a larger proportion of samples classified as Her2-enriched and Normal-Like (NL) subtypes (Supplementary Table S8). The heatmaps of the *Low Quality samples* reveal that no effective detection was obtained for most individual genes (green in the lower right panel) and signatures (blue in the upper right panel). (Sample sorting from left to right follows AIMS subtype and subsequently T-cell signature expression).



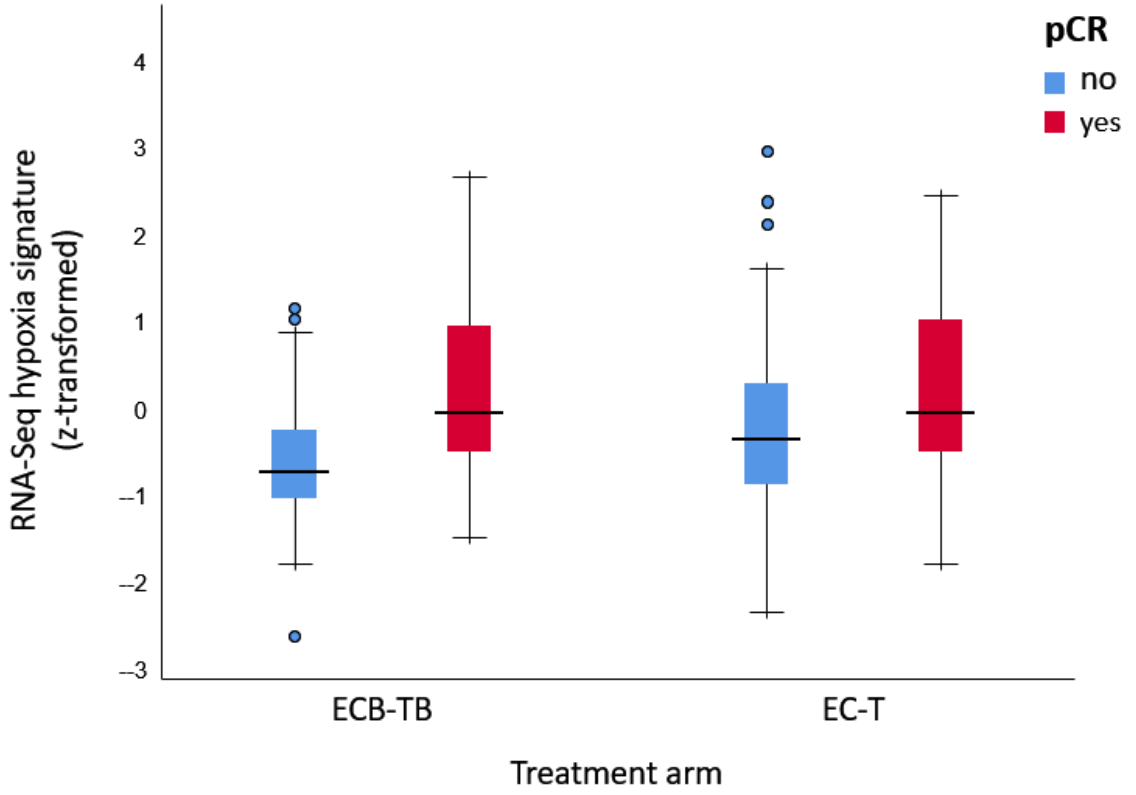
Supplementary Figure S6: Correlation of proliferation signature and histological grade

Box-plot demonstrating expression of proliferation signature in 221 high quality samples stratified by histological grading (median -0.73, -0.39, and 0.53 in G1, G2, and G3, respectively; P<0.001).



Supplementary Figure S7: Poor detection of NDRG1 expression signal in TMA analysis

Scatterplot comparing NDRG1-RNA-Seq expresssion (x-axis) and IHC detection on TMA (y-axis) as percent positive cells from from digital image analysis by QuPath. Horizontal and vertical lines represent 1.5 z-score cutoff values for RNA-Seq and TMA-IHC, respectively. In general, samples with high percentage of positive cell in the TMA analysis are associated with higher RNA-Seq values (Spearman rank correlation = 0.429). But for many samples with high RNA expression the signal was lost in TMA analysis. 23 of the samples, which were also analyzed by whole-slide IHC of core biopsies, are coloured in orange (≥10% positive cells) and green (<10% positive cells) according to the whole-slide result. The majority of the positive samples from whole slide IHC (orange) did not reach the 10% value in the TMA analysis.



Supplementary Figure S8: Hypoxia signature and response by treatment arm

Box-plots of expression of the hypoxia signature from RNA-Seq in samples showing a pCR (red) or not (blue). The box-plots are given separately for the patients from the two treatment arms of the trial either with bevacizumab (ECB-TB) or without (EC-T). Univariate logistic regression of pCR by the hypoxia signature was significant in the bevacizumab group (ECB-TB: OR 3.52, 95% CI 1.91-6.49, P<0.001) and only showed a trend in the group without bevacizumab (EC-T: OR 1.40, 95% CI 0.97-2.02, P=0.076).

Supplementary Table S1: Specific analysis roles and blinding status of contributing teams

Order of analysis	Role	Team members	Location/Affiliation	Blinding status
1	Pathological analyses, tissue banking, sample provision	CD, KE, BS, CSo, IS, TF	Local trial units, central pathology	blinded to molecular and clinical study data
2	RNA preparation and sequencing	BMY, BLJ	Avera Cancer Institute	fully blinded
3	Primary RNA-Seq raw data analysis and QC	TM, JB	Avera Cancer Institute	fully blinded
4	Sample coding, dataset assembly and distribution	KW, VN	GBG-Statistics Dept.	unblinded
5	Blinded gene expression analysis, development of statistical analysis plan and SPSS script	TK, UH	Goethe University	fully blinded
6	Correlation with patient data according to predeveloped SPSS script	KW, VN	GBG-Statistics Dept.	unblinded
7	Blinded interpretation of summary results	TK, UH	Goethe University	no patient level data
8	Monitoring and review of results	SL, MU, PAF, FM, VM, BG, CSch, CH, ES, JH, MvM	GBG-Boards	no patient level data

Supplementary Table S2: Pre-defined analytical aims of the study:

Pre-defined analytical aims of the study:	
1.	Concordance of RNA-Seq-derived genomic ER-/PR-status, proliferation, immune signature expression, and molecular subtype with pathology-derived IHC-based ER-/PR-status, histological grading, and tumor-infiltrating-lymphocyte (TIL)-scoring, respectively.
2.	Robustness of the above concordances with regard to sample quality (QC class).
3.	Univariate predictive value for pCR of RNA-Seq-derived molecular subtypes, and signatures for proliferation, stroma, T-cell signature, and hypoxia signature.
4.	Multivariate logistic regression of pCR including the following predictor variables: a) Hormone receptor status, treatment arm (+/- Bev), hypoxia signature, and the interaction between hypoxia signature and treatment arm. b) All predictor variables from (a), with additional clinical variables (age, cT, cN, histological grade) as predictors.

Supplementary Table S3: Comparison of clinical parameters of the complete trial cohort and the RNA-Seq cohorts

Parameter	Category	Total cohort 1540[#] (100.0%)	RNA-Seq data 289 (100.0%)	*P*-Value[§]	High Quality RNA-Seq data 221 (100.0%)	*P*-Value[§]
Age	median	48	46	<0.001	46	0.009
clin. tumor status	T1	261 (17.0%)	45 (15.6%)	0.028	37 (16.8%)	0.035
	T2	883 (57.5%)	188 (65.3%)		145 (65.9%)	
	T3	224 (14.6%)	33 (11.5%)		23 (10.5%)	
	T4a-c	79 (5.1%)	8 (2.8%)		8 (3.6%)	
	T4d	89 (5.8%)	14 (4.9%)		7 (3.2%)	
	missing	4	1		1	
clin. lymph node status	negative	772 (51.0%)	148 (51.6%)	n.s.	115 (52.5%)	n.s.
	LN1-3	668 (44.1%)	124 (43.2%)		93 (42.5%)	
	LN4-9	57 (3.8%)	11 (3.8%)		7 (3.2%)	
	LN>=10	17 (1.1%)	4 (1.4%)		4 (1.8%)	
	missing	26	2		2	
HER2 status	negative	1540 (100.0%)	289 (100.0%)		221 (100.0%)	
Hormone receptor status	negative	558 (36.2%)	133 (46.0%)	<0.001	102 (46.2%)	0.001
	positive	982 (63.8%)	156 (54.0%)		119 (53.8%)	
Histological grade	G1	53 (3.5%)	9 (3.1%)	<0.001	7 (3.2%)	<0.001
	G2	781 (51.0%)	119 (41.5%)		86 (39.3%)	
	G3	697 (45.5%)	159 (55.4%)		126 (57.5%)	
	missing	9	2		2	
Histological subtype	Ductal invasive	1241 (80.8%)	238 (82.4%)	n.s.	179 (81.0%)	n.s.
	Lobular invasive	162 (10.5%)	26 (9.0%)		23 (10.4%)	
	other	133 (8.7%)	25 (8.7%)		19 (8.6%)	
	missing	4	0		0	
Treatment arm	EC-T	743 (48.2%)	150 (51.9%)	n.s.	117 (52.9%)	n.s.
	ECB-TB	797 (51.8%)	139 (48.1%)		104 (47.1%)	
pCR	no	1156 (75.1%)	149 (51.6%)	<0.001	112 (50.7%)	<0.001
	yes	384 (24.9%)	140 (48.4%)		109 (49.3%)	

[§] *P*-values are the result of Fisher's exact tests for binary variables, of chi-square tests for variables with three or more levels, and of Wilcoxon test for continous variables, respectively.
[#] 1540 total patients with HER2 negative disease with response data from treatment arms out of the ITT population of 2572 patients from the GeparQuinto trial.

Supplementary Table S4: Gene lists of signatures

Signature	Gene	EntrezID
Adipocyte	AASS	10157
Adipocyte	ABCA10	
Adipocyte	ABCA6	23460
Adipocyte	ABCA8	10351
Adipocyte	ABCA9	
Adipocyte	ADH1B	125
Adipocyte	ANK2	287
Adipocyte	APOD	347
Adipocyte	COL14A1	7373
Adipocyte	FAT4	79633
Adipocyte	FREM1	
Adipocyte	IGF1	3479
Adipocyte	LAMA2	3908
Adipocyte	PLEKHH2	
Adipocyte	SPARCL1	8404
Adipocyte	SVEP1	79987
Adipocyte	TNXB	7148

Signature	Gene	EntrezID
Basal	ABCA13	
Basal	COL27A1	
Basal	CRYAB	1410
Basal	ELF5	2001
Basal	FOXC1	2296
Basal	GABRP	2568
Basal	KRT23	25984
Basal	PROM1	8842
Basal	RGMA	
Basal	SFRP1	6422
Basal	SLC34A2	10568
Basal	SLC6A14	11254
Basal	SOX10	6663
Basal	TCF7L1	83439
Basal	ZNF462	
Basal	DSC3	1825
Basal	DSG3	1830
Basal	FAT2	2196
Basal	KRT14	3861
Basal	KRT16	
Basal	KRT17	3872
Basal	KRT5	3852
Basal	KRT6A	
Basal	KRT6B	3854
Basal	TRIM29	23650

Signature	Gene	EntrezID
ER-related	ABCC8	6833
ER-related	AFF3	3899
ER-related	AGR2	10551
ER-related	AR	367
ER-related	CA12	771
ER-related	CAPN8	
ER-related	CCDC170	80129
ER-related	CYP2B7P1	
ER-related	ERBB4	2066
ER-related	ESR1	2099
ER-related	FOXA1	3169
ER-related	FSIP1	
ER-related	GATA3	2625
ER-related	GFRA1	2674
ER-related	GREB1	9687
ER-related	GRPR	2925
ER-related	KDM4B	23030
ER-related	MAPT	4137
ER-related	MLPH	79083
ER-related	NAT1	9
ER-related	PGR	5241
ER-related	SCUBE2	57758
ER-related	SLC44A4	80736
ER-related	TBC1D9	23158
ER-related	THSD4	79875

Signature	Gene	EntrezID
Hypoxia	CSTB	1476
Hypoxia	NDRG1	10397
Hypoxia	VEGFA	7422

Signature	Gene	EntrezID
IFN	CMPK2	
IFN	DDX58	23586
IFN	DDX60	55601
IFN	DDX60L	
IFN	EIF2AK2	5610
IFN	EPSTI1	
IFN	HERC5	51191
IFN	HERC6	55008
IFN	IFI44	10561
IFN	IFI44L	10964
IFN	IFIH1	64135
IFN	IFIT1	3434
IFN	IFIT2	3433
IFN	MX1	4599
IFN	MX2	4600
IFN	OAS1	4938
IFN	OAS2	4939
IFN	OAS3	4940
IFN	PARP12	64761
IFN	PLSCR1	5359
IFN	RSAD2	91543
IFN	SP100	6672
IFN	SP110	3431
IFN	UBA7	7318
IFN	XAF1	54739
IFN	ZNFX1	

Signature	Gene	EntrezID
MHC1	GBP1	2633
MHC1	GBP2	2634
MHC1	HLA-A	3105
MHC1	HLA-B	3106
MHC1	HLA-C	3107
MHC1	HLA-E	3133
MHC1	HLA-F	3134
MHC1	NLRC5	
MHC1	PSMB8	5696
MHC1	TAP1	6890
MHC1	TAP2	6891
MHC1	WARS	7453

Signature	Gene	EntrezID
Proliferation	ANLN	
Proliferation	ARHGAP11A	9824
Proliferation	ASPM	259266
Proliferation	AURKA	6790
Proliferation	BRIP1	83990
Proliferation	BUB1	699
Proliferation	BUB1B	701
Proliferation	CCNB1	891
Proliferation	CDC20	991
Proliferation	CDCA2	
Proliferation	CENPE	1062
Proliferation	CENPF	1063
Proliferation	CIT	11113
Proliferation	DEPDC1	55635
Proliferation	DIAPH3	81624
Proliferation	ECT2	1894
Proliferation	ESPL1	9700
Proliferation	EXO1	9156
Proliferation	FANCA	2175
Proliferation	FANCI	55215
Proliferation	FOXM1	2305
Proliferation	GMPS	8833
Proliferation	GTSE1	51512
Proliferation	HIST1H2BO	8348
Proliferation	HJURP	55355
Proliferation	IQGAP3	
Proliferation	KIF14	9928
Proliferation	KIF23	9493
Proliferation	KIF2C	11004
Proliferation	KIFC1	3833
Proliferation	KPNA2	3838
Proliferation	LMNB1	4001
Proliferation	MCM10	55388
Proliferation	MELK	9833
Proliferation	MKI67	4288
Proliferation	MYBL2	4605
Proliferation	NCAPD2	9918
Proliferation	ORC6	23594
Proliferation	POLQ	10721
Proliferation	PRC1	9055
Proliferation	PRR11	
Proliferation	RACGAP1	29127
Proliferation	RRM2	6241
Proliferation	SMC4	10051
Proliferation	SPAG5	10615
Proliferation	STIL	6491
Proliferation	TICRR	
Proliferation	TOP2A	7153
Proliferation	TPX2	22974
Proliferation	TTK	7272

Signature	Gene	EntrezID
Ribosomal	CSDE1	7812
Ribosomal	EEF1A1	1915
Ribosomal	EEF1G	1937
Ribosomal	HSPD1	3329
Ribosomal	LRPPRC	10128
Ribosomal	NACA	4666
Ribosomal	PABPC1	26986
Ribosomal	PRKDC	5591
Ribosomal	RPL10	6134
Ribosomal	RPL11	6135
Ribosomal	RPL12	6136
Ribosomal	RPL19	6143
Ribosomal	RPL23A	
Ribosomal	RPL27A	6157
Ribosomal	RPL3	6122
Ribosomal	RPL32	6161
Ribosomal	RPL4	6124
Ribosomal	RPL5	6125
Ribosomal	RPL9	6133
Ribosomal	RPLP0	6175
Ribosomal	RPS11	6205
Ribosomal	RPS18	6222
Ribosomal	RPS20	6224
Ribosomal	RPS24	6229
Ribosomal	RPS27A	6233
Ribosomal	RPS4X	6191
Ribosomal	RPS6	6194
Ribosomal	RPS7	6201
Ribosomal	RPSA	
Ribosomal	TPT1	7178

Signature	Gene	EntrezID
Stroma	ADAM12	8038
Stroma	ADAMTS12	81792
Stroma	BNC2	54796
Stroma	CACNA1C	775
Stroma	COL10A1	1300
Stroma	COL11A1	1301
Stroma	COL5A3	50509
Stroma	COL8A1	
Stroma	COMP	1311
Stroma	CRISPLD2	83716
Stroma	CTGF	1490
Stroma	DCHS1	8642
Stroma	DPYSL3	1809
Stroma	FAP	2191
Stroma	FBLN2	2199
Stroma	FLNC	2318
Stroma	FNDC1	
Stroma	GAS6	2621
Stroma	GAS7	8522
Stroma	GLIS2	
Stroma	IGFBP4	3487
Stroma	ITGA11	
Stroma	ITGA5	3678
Stroma	ITGB5	3693
Stroma	ITGBL1	9358
Stroma	KANK2	25959
Stroma	KIAA1462	57608
Stroma	KIF26B	55083
Stroma	LMOD1	25802
Stroma	LRRC15	131578
Stroma	MAP1A	4130
Stroma	MICAL2	9645
Stroma	MMP11	4320
Stroma	MMP14	4323
Stroma	MRC2	9902
Stroma	MXRA8	54587
Stroma	MYH1	4619
Stroma	MYL9	10398
Stroma	NID1	4811
Stroma	NID2	22795
Stroma	PCOLCE	5118
Stroma	PDLIM7	9260
Stroma	PHLDB1	23187
Stroma	PMEPA1	56937
Stroma	PMP22	5376
Stroma	PODN	
Stroma	RIN2	54453
Stroma	SFRP2	
Stroma	SFRP4	6424
Stroma	SPON1	10418
Stroma	SSCSD	
Stroma	TAGLN	6876
Stroma	TENM4	26011
Stroma	THY1	7070
Stroma	TIMP2	7077
Stroma	TIMP3	7078
Stroma	UNC5B	219699
Stroma	ZFHX4	79776

Signature	Gene	EntrezID
T-Cell	B2M	567
T-Cell	CD74	972
T-Cell	CIITA	4261
T-Cell	CYBB	1536
T-Cell	DOCK10	55619
T-Cell	DOCK2	1794
T-Cell	DOCK8	
T-Cell	HLA-DPA1	3113
T-Cell	HLA-DRA	3122
T-Cell	HLA-DRB1	3123
T-Cell	IL10RA	3587
T-Cell	NCKAP1L	3071
T-Cell	PARP14	
T-Cell	PARP9	
T-Cell	PTPRC	5788
T-Cell	SAMD9	54809
T-Cell	SAMD9L	
T-Cell	SAMHD1	25939
T-Cell	STAT1	6772
T-Cell	STAT2	6773
T-Cell	TRIM22	10346
T-Cell	XRN1	

Supplementary Table S5: Core genes of the hypoxia signature cluster from different datasets with correlated expression

Gene symbol	Gene name	Category	Details	FFPE-RNA-Seq data
VEGFA	Vascular endothelial growth factor	Angiogenesis	Growth factor active in angiogenesis, vasculogenesis and endothelial cell growth. Induces endothelial cell proliferation, promotes cell migration, inhibits apoptosis and induces permeabilization of blood vessels. Binds to the FLT1/VEGFR1 and KDR/VEGFR2 receptors, heparan sulfate and heparin.	**yes**
NDRG1	N-myc downstream regulated gene 1	Stress response	Involved in stress responses, hormone responses, cell growth, and differentiation. Necessary for p53-mediated caspase activation and apoptosis.	**yes**
ANGPTL4	Angiopoietin-like 4	Angiogenesis, hypoxia	Hypoxia-induced expression in endothelial cells. May act as a regulator of angiogenesis and modulate tumorigenesis. In response to hypoxia, the unprocessed form of the protein accumulates in the subendothelial extracellular matrix.	no
ADM	Adrenomedullin	Angiogenesis	Adrenomedullin functions include vasodilation, regulation of hormone secretion, and promotion of angiogenesis.	no
DDIT4	DNA damage induced transcript 4	Stress response	Regulates cell growth, proliferation and survival via inhibition of mTORC1. Important role in responses to cellular energy levels and cellular stress, including responses to hypoxia and DNA damage.	no
CSTB	Cystatin-B	Proteinase inhibitor	Intracellular thiol proteinase inhibitor thought to play a role in protecting against proteases leaking from lysosomes.	**yes**

Supplementary Table S6: Accuracy of IHC detection of NDRG1 as marker of the hypoxia signature

Cohort	Finding (*N*=23)	Finding (*N*=23)	Finding (*N*=23)	Full (*N*=193)
Source	Whole slide	Whole slide	TMA	TMA
Method	pathological scoring	digital image analysis	digital image analysis	digital image analysis
Cutoff	pos/neg	>10% positive cells	>10% positive cells	>10% positive cells
Positive by RNA-Seq*	19	19	19	20
Negative by RNA-Seq*	4	4	4	173
Accuracy	91.3 %	73.9 %	47.8 %	81.9 %
Sensitivity	89.5 %	68.4 %	**36.8 %**	**40.0 %**
Specificity	100.0 %	100.0 %	100.0 %	86.7 %
PPV	100.0 %	100.0 %	100.0 %	**25.8 %**
NPV	66.7 %	40.0 %	25.0 %	92.6 %

* based on cutoff z-score 1.5 from RNA-Seq

Supplementary Table S7: Comparison of hormone receptor status from RNA-Seq and IHC

High Quality samples	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Accuracy**
$ER_{RNA\text{-}Seq}$ vs. ER_{IHC} (*N*=221)	75.7 %	93.4 %	92.6 %	78.0 %	84.2 %
$PR_{RNA\text{-}Seq}$ vs. PR_{IHC} (*N*=221)	76.6 %	83.5 %	77.4 %	82.8 %	80.5 %
Low Quality samples	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**Accuracy**
$ER_{RNA\text{-}Seq}$ vs. ER_{IHC} (*N*=68)	72.7 %	88.6 %	85.7 %	77.5 %	80.9 %
$PR_{RNA\text{-}Seq}$ vs. PR_{IHC} (*N*=67)	61.3 %	86.1 %	79.2 %	72.1 %	74.6 %

Supplementary Table S8: Comparison of Molecular Subtyping between High Quality and Low Quality Samples

Group	Basal-like	Her2-enrich.	LumA	LumB	Normal-like
Total (*N*=289)	119 (41.2%)	60 (20.8%)	33 (11.4.%)	46 (15.9%)	31 (10.7%)
HQ no dupl. (*N*=221, %)	103 (46.6%)	33 (14.9%)	28 (12.7%)	42 (19.0%)	15 (6.8%)
LQ no dupl. (*N*=68, %)	16 (23.5%)	27 (39.7%)	5 (7.4%)	4 (5.9%)	16 (23.5%)

The distributions of molecular subtypes according to AIMS differ significantly ($P = 8.6 \times 10^{-9}$, Fisher's Exact Test) between samples with high and low quality.

Supplementary Table S9: Univariate logistic regression of pCR by molecular markers (N=221 High Quality samples)

Molecular marker	OR	95% CI	P-value
Basal-like*	8.88	2.34-33.6	**0.001**
HER2-enriched*	3.33	0.79-14.1	0.10
Lum-A*	0.87	0.18-4.28	0.86
Lum-B*	2.22	0.54-9.14	0.27
T-cell signature#	1.60	1.21-2.12	**0.001**
Proliferation signature#	2.88	2.00-4.16	**<0.001**
Hypoxia signature#	1.92	1.41-2.60	**<0.001**

* vs. Normal-like subtype, # z-score

Supplementary Table S10: Multivariate logistic regression of pCR with NDRG1 from TMA analysis

	OR	95% CI	*P*-value
Hormone receptor (neg. vs. pos.)	4.35	2.34-8.07	**<0.001**
NDRG1-TMA-IHC (z-score >1.5)	3.79	0.85-16.9	0.080
Bevacizumab therapy	0.92	0.53-1.60	0.766
Interaction NDRG1-TMA-IHC * bevacizumab	1.31	0.78-2.21	0.309
cN (≥10 vs 4-9 vs 1-3 vs 0 positive nodes)	0.80	0.52-1.24	0.323
cT (T4d vs T4a-c vs T3 vs T2 vs T1)	0.80	0.59-1.0	0.173
Grading (G3 vs G2 vs G1)	1.43	0.83-2.48	0.197