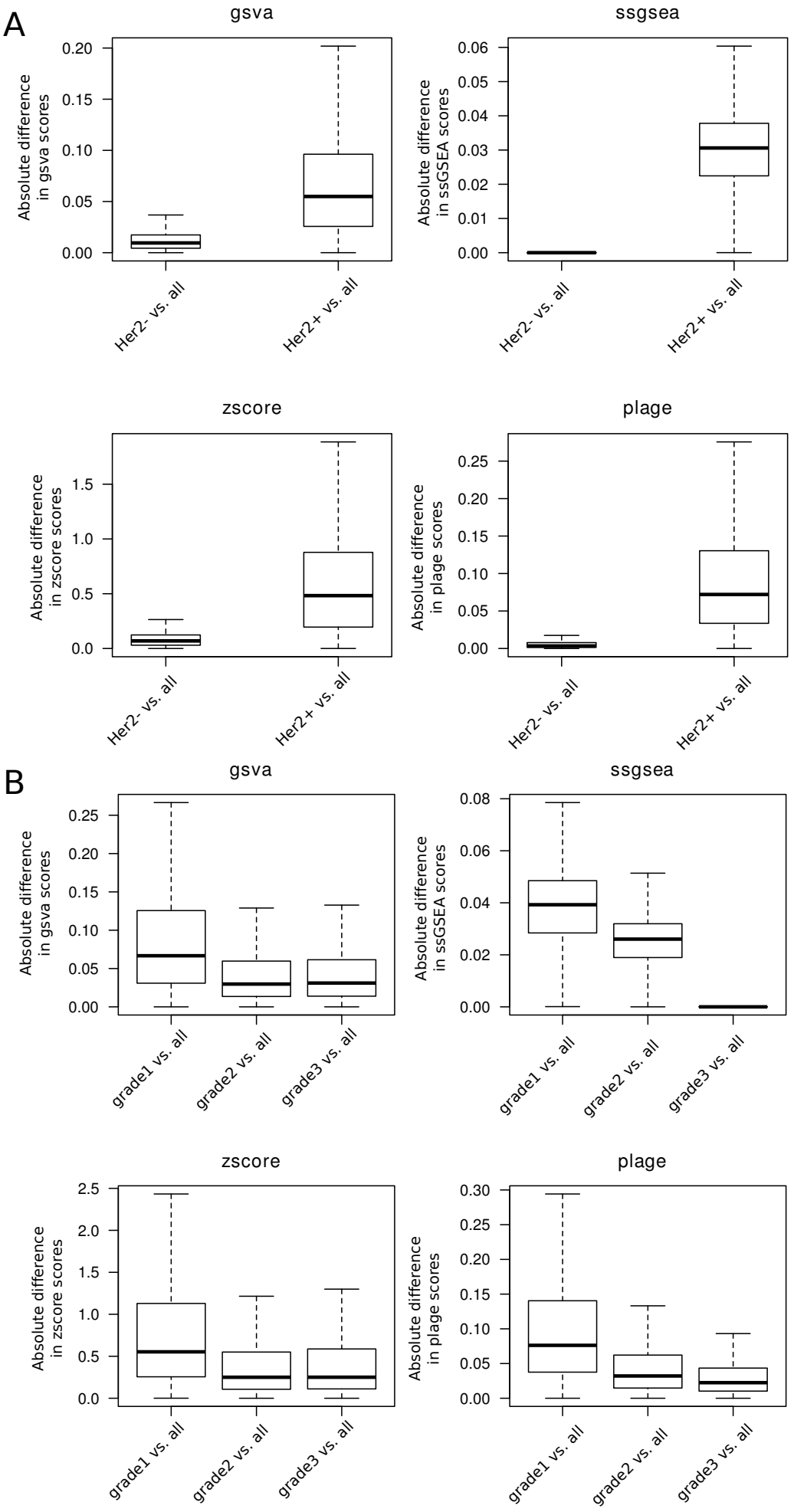
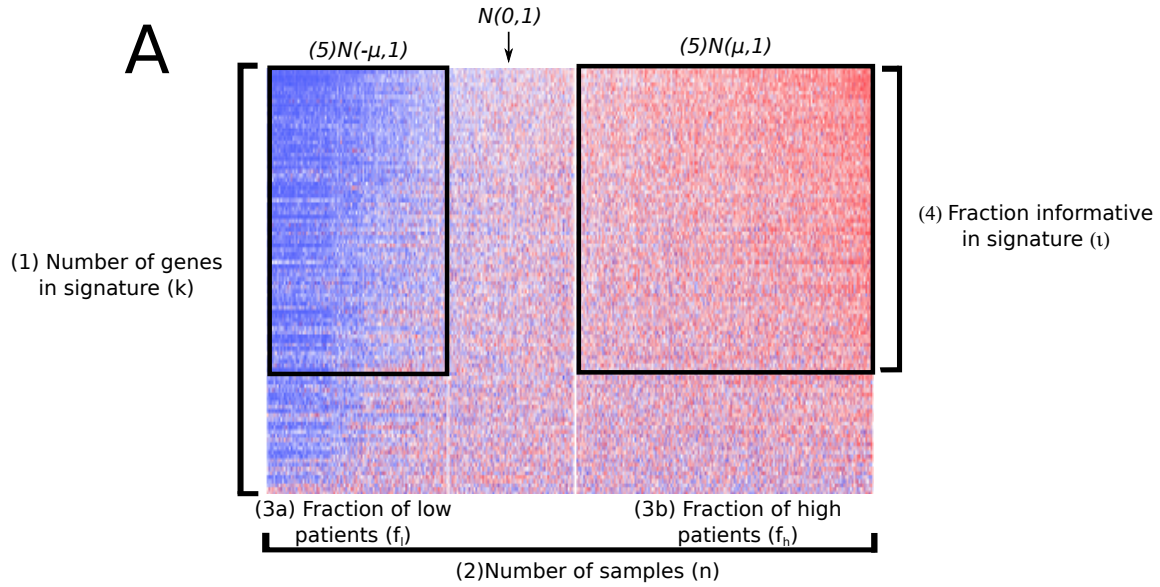
Supp. Fig. 1. Paquet et al.



Supplementary Figure S1. Instability of current pathway activation tools in function of grade and Her2 (A) Boxplots represent the absolute difference between the score obtained from a specific pathway activation tool (GSVA, ssGSEA, zscore and plage) using all patients from the METABRIC dataset, and the score obtained when METABRIC is restricted to either Her2+ samples (left) or Her2- samples (right). **(B)** Boxplots represent the absolute difference between the score obtained from a specific pathway activation tool (GSVA, ssGSEA, zscore and plage) using all patients from the METABRIC dataset, and the score obtained when METABRIC is restricted to either Grade 1 (left), Grade 1 (middle), or Grade 3 (left) samples.



B

(1) Number of genes in signature (k)[50]
(2) Number of samples (n) [200]
(3) Fraction (low,independent,high) [0.1,0.8,0.1]

(5) μ \ (4) Fraction informative (ι)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.05	35	33	35	34	34	34	37*	32	35	34
0.1	35	33	35	34	35	34	39*	35	39*	37
0.2	35	34	35	34	35	40*	39*	47*	43*	50*
0.3	35	34	38*	33	39*	40*	45*	52*	54*	67*
0.4	37	37	39*	37	42*	48*	53*	60*	73*	74*
0.5	37	37	39*	40*	47*	64*	73*	70*	89*	90*
0.6	37	40*	44*	53*	60*	67*	85*	89*	89*	95*
0.7	37	40*	43*	57*	70*	80*	95*	97*	97*	97*
0.8	37	40*	50*	59*	82*	90*	95*	97*	99*	99*
0.9	37	40*	52*	70*	90*	95*	97*	99*	99*	99*
1	37	42*	55*	73*	95*	99*	99*	99*	99*	99*
1.5	38*	50*	77*	95*	99*	99*	99*	99*	99*	100*
2	39*	58*	89*	99*	99*	99*	100*	100*	100*	100*

C

(2) Number of samples (n) [200]
(3) Fraction (low,independent,high) [0.1,0.8,0.1]
(5) μ [.5]

(4) Fraction informative (ι) \ (1) Number of genes in signature (k)	10	20	30	40	50	60	70	80	90	100	200	300	500	1000
0.1	34	33	33	34	37	32	35	35	33	40*	34	33	38	54*
0.2	39*	43*	41	39	37	32	39*	35	40	43*	50*	51*	65*	85*
0.3	40*	48*	45*	42	39*	34	46*	40	42	53*	62*	81*	89*	96*
0.4	37	46*	55*	44*	40*	44*	57*	50*	61*	68*	85*	94*	99*	99*
0.5	44*	46*	53*	54*	47*	52*	61*	68*	68*	77*	92*	99*	99*	99*
0.6	44*	51*	56*	59*	64*	68*	76*	77*	81*	93*	97*	99*	99*	99*
0.7	50*	53*	63*	67*	73*	77*	81*	88*	91*	93*	99*	99*	99*	99*
0.8	47*	61*	73*	80*	71*	89*	91*	93*	96*	95*	99*	99*	99*	99*
0.9	53*	68*	81*	89*	88*	92*	91*	97*	98*	98*	99*	99*	99*	99*
1	52*	70*	81*	92*	90*	94*	94*	98*	98*	98*	98*	99*	99*	99*

D

(1) Number of genes in signature (k)[50]
(2) Number of samples (n) [200]
(5) μ [.5]

low	Ind.	high	(3) Fraction (low,independent,high) \ (4) Fraction informative (ι) 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.05	0.9	0.05	35	35	35	48*	54*	64*	78*	71*	78*	91*
0.1	0.8	0.1	37	37	39*	40*	47*	64*	73*	70*	89*	90*
0.2	0.6	0.2	35	36	38	43*	51*	65*	72*	78*	87*	91*
0.3	0.4	0.3	36	36	39*	46*	53*	68*	76*	79*	88*	92*
0.4	0.2	0.4	35	37	39	46*	56*	65*	72*	79*	88*	92*
0.45	0.1	0.45	37	41*	43*	51*	56*	69*	77*	83*	89*	94*
0.05	0.45	0.5	37	38	37	41*	54	58*	69*	72*	67*	72*
0.1	0.45	0.45	36*	37*	36	44*	56*	59*	71*	73*	75*	79*
0.2	0.4	0.4	37	38	38	45*	55*	66*	72*	75*	83*	88*
0.3	0.35	0.35	36	37	40*	47*	52*	67*	77*	82*	86*	93*
0.4	0.3	0.3	35	38	39	45*	53*	65*	72*	79*	88*	92*
0.5	0.25	0.25	34	35	40	47*	52*	63*	71*	74*	79*	79*
0.6	0.2	0.2	35	33	40*	43*	47*	58*	63*	69*	69*	70*
0.7	0.15	0.15	34	33	39	44*	54*	58*	64*	62*	63*	62*
0.8	0.1	0.1	36	40*	45*	52*	60*	61*	62*	57*	54*	52*
0.9	0.05	0.05	28	38	44*	45*	51*	54*	55*	45*	42	42

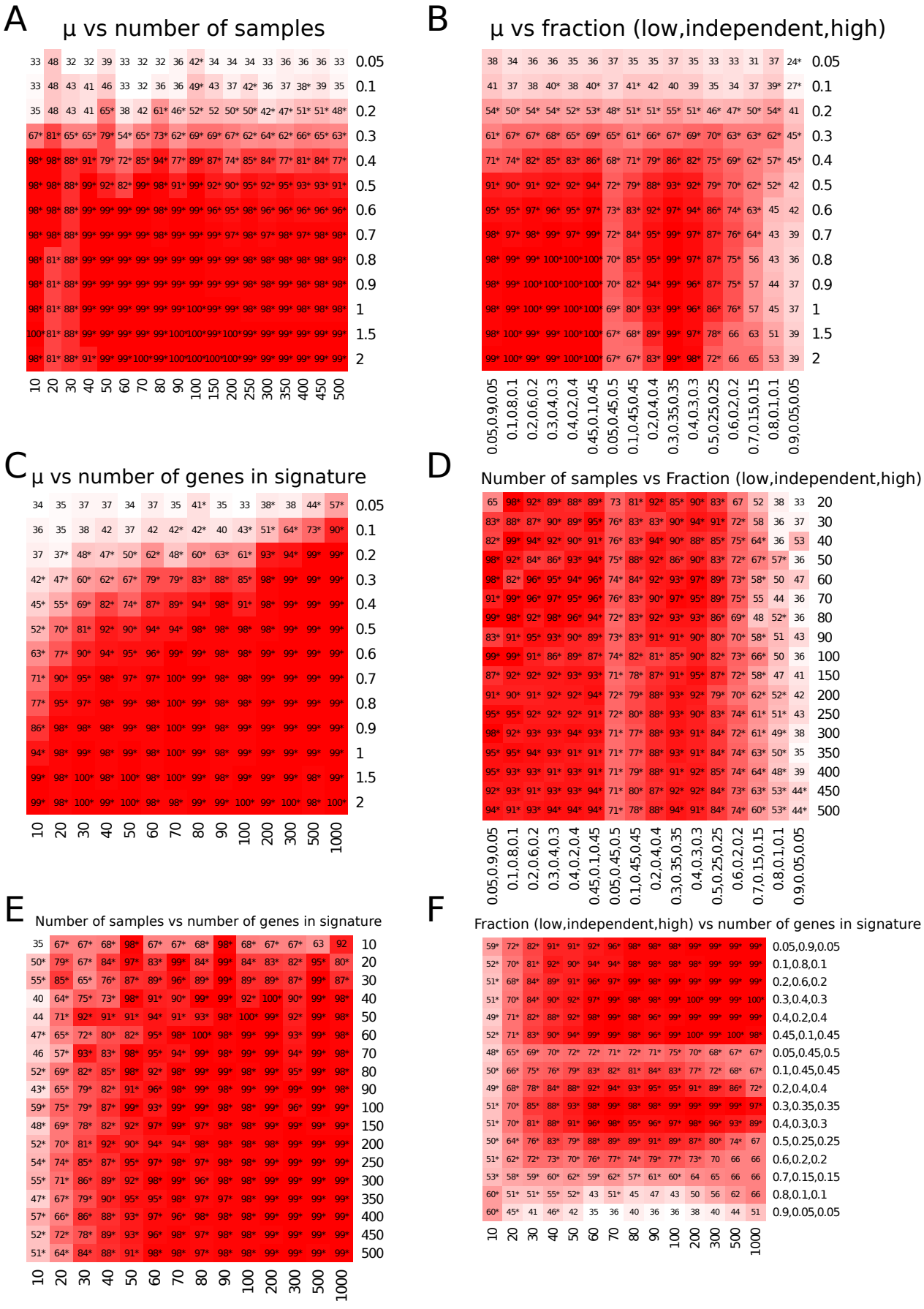
E

(1) Number of genes in signature (k)[50]
(3) Fraction (low,independent,high) [0.1,0.8,0.1]
(5) μ [.5]

(2) Number of samples (n) \ (4) Fraction informative (ι) 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
10	33	33	33	33	33	33	62*	62*	100*	100*
20	32	31	31	31	31	48	65*	65*	65*	98*
30	32	32	32	32	32	76*	88*	88*	76*	88*
40	32	32	32	32	40	57*	74*	91*	91*	99*
50	32	46	46	39	45	46	59*	72*	72*	92*
60	33	44	49*	55*	60*	54*	71*	71*	71*	82*
70	36	42	46*	42	51*	65*	56*	79*	80*	99*
80	37	41	36	45	48*	66*	73*	86*	95*	98*
90	36	36	47*	47*	58*	62*	58*	81*	88*	91*
100	36	42*	52*	49*	59*	69*	79*	76*	85*	99*
150	34	39	43*	48*	49*	65*	74*	85*	90*	92*
200	37	37	39*	40*	47*	64*	73*	70*	89*	90*
250	36	36	39	43*	51*	63*	73*	85*	87*	95*
300	34	33	40*	47*	57*	68*	76*	85*	89*	92*
350	36	39*	40	45*	56*	69*	73*	76*	85*	95*
400	35	36	40*	44*	47*	65*	73*	82*	90*	93*
450	36	36	40*	48*	56*	68*	75*	80*	88*	93*
500	35	36	41*	51*	62*	67*	72*	80*	87*	91*

Supplementary Figure S2. The ROI_q method is able to identify samples with either low or high activation (**A**) Description of the parameters utilized for the generation of the synthetic datasets. (**B-E**) Agreement (in percentages) between the ground truth and the estimated assignment obtained using the ROI_{95} for different pairwise comparisons of parameters: (**B**) [4] Percentage of informative (ι) vs. μ [5], (**C**) [1] Size of signature (k) vs. [4] Percentage of informative (ι), (**D**) [4] Percentage of informative (ι) vs. [3] Size of ROI and (**E**) [4] Percentage of informative (ι) vs. [2] Number of samples (n). Stars denote parameter configurations that had statistically significant agreement between the ground truth and the estimation using the ROI_{95} (Cohen's kappa statistics $P < 0.05$).

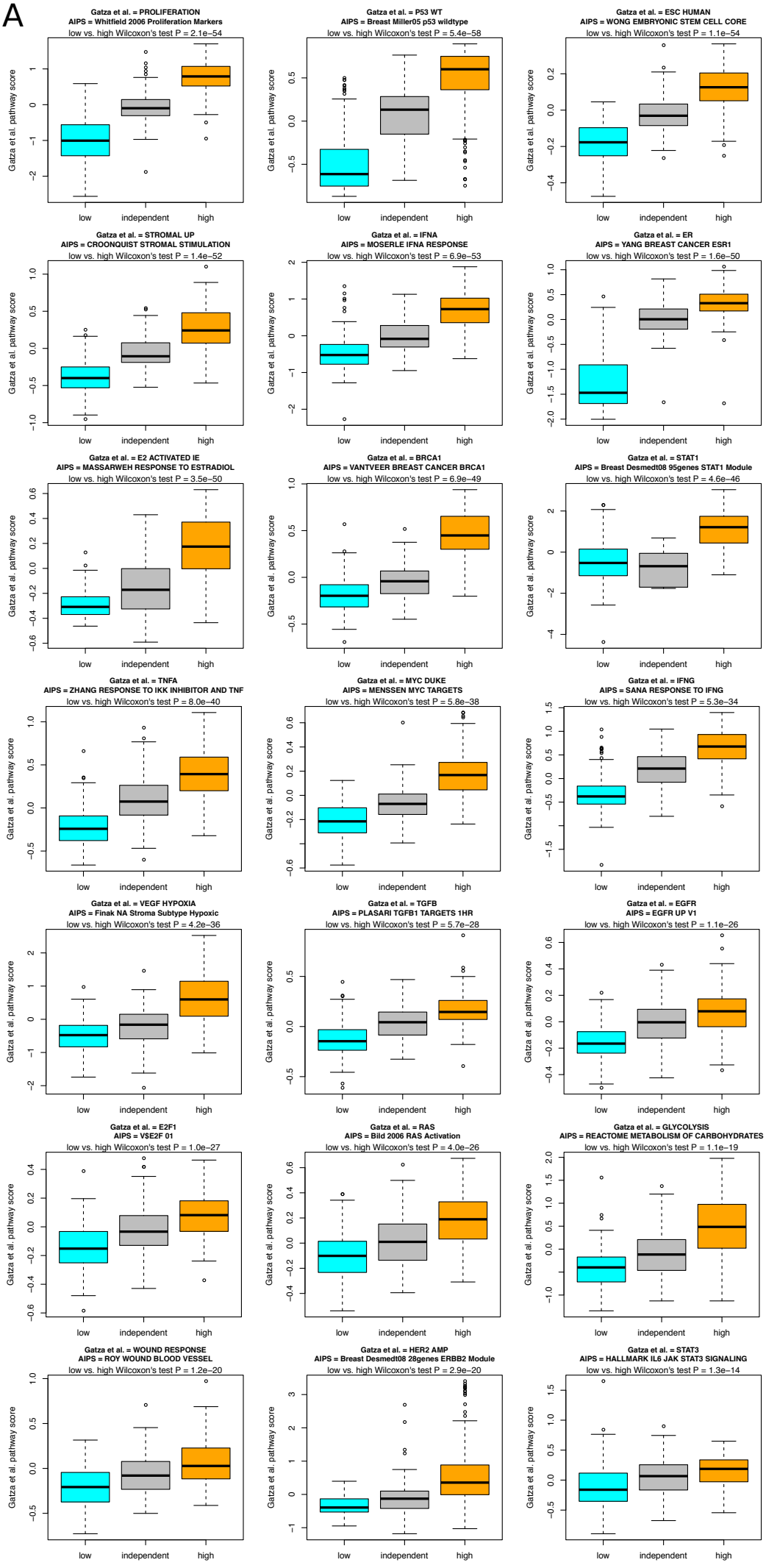
Supp. Fig. 3 - Paquet et al.



Supplementary Figure S3. The ROI_q method is able to identify samples with either low or high activation (**A-F**) Presentation of the agreement percentages between the imputed synthetic assignments and the one obtained using the ROI for different pairwise comparisons of two parameters: (**A**) μ vs. number of samples(n). (**B**) μ vs. size of the ROI. (**C**) μ vs. size of the signature (k). (**D**) Number of samples (n) vs. the size of the ROI. (**E**) Number of samples (n) vs. size of the signature(k). (**F**) Size of the ROI vs. size of the signature(k). Star following the agreement estimates highlight comparisons for which the agreement between the ROI and the real assignments are significant using a Cohen's kappa statistics $P < 0.05$.

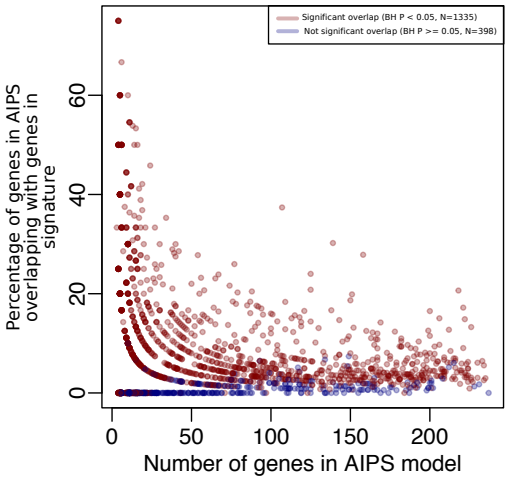
Supp. Fig. 4. Paquet et al.

A



Supplementary Figure S4. Comparing pathway scores from Gatza et al. to AIPS assignments. (**A**) Boxplots displaying for every pathway the distribution of scores obtained for the AIPS low, independent and high assignments. The name of the Gatza et al. pathway is written on top of the boxplots as well as the corresponding AIPS signature and the result of a Wilcoxon's test comparing the scores distribution for the low and high AIPS assignments.

Supp. Fig. 5. Paquet et al.



Supplementary Figure S5. Comparing genes selected in AIPS models with genes in the original gene signature. Plot showing the percentage of genes retained in AIPS models that overlap with the original signature in function of the number of genes in the AIPS models. The red dots represent the AIPS models for which we detect a significant overlap (Benjamini–Hochberg (BH) adjusted P < 0.05, 1335/1733 (77%)) and the blue dots are the AIPS models for which we detected no significant overlap (Benjamini–Hochberg (BH) adjusted P >= 0.05, 398/1733 (23%))