*NPJBCANCER-00433 revised version - Brief Communication*

***‘The therapeutic response of ER-positive breast cancers differs according to the molecular Basal- or Luminal-type’***

Supplemental methods

Here are described the classifiers used in the study of Bertucci et al., NPJ Breast Cancer 2019 and related to custom code included in the R script file ‘Breast\_cancer\_classifications.r’ and the R object ‘Breast\_cancer\_classifications.Rda’. R functions were built in respect of methodology and tools of each original source.

**Custom code, R ‘*functions*’**

`*BluePrint*`1, A nearest-centroid classification model built utilizing the 80-gene profile defined by authors for ‘Basal’, ‘HER2’ and ‘Luminal’ molecular subtypes.

`*PAM50*`2, A nearest-centroid classification model built utilizing the 50-gene profile defined by authors for ‘Basal-like’, ‘ERBB2-enriched’, ‘LuminalA’, ‘LuminalB’ and ‘Normal-like’ molecular subtypes.

`*Mammaprint*`3, A Pearson correlation score utilizing the prognostic 70-gene profile defined by authors to classify samples in ‘low/good-risk’ and ‘high/poor-risk’ groups.

`*EndoPredict*`4, A prognostic score built utilizing 8 genes of interest and 3 references genes associated with clinical risk factors based on a ER+/HER2- breast cancer population to classify samples in ‘low-risk’ and ‘high-risk’ groups.

`*breastHierarch*`5, A metagene score model built utilizing subpopulations gene expression signatures of the breast epithelial cell hierarchy defined by authors.

`*CD44CD24*`6, A Pearson correlation score utilizing the gene profile of 493 probes defined by authors in mammospheres (MS) forming cells to classify samples in ‘CD44+/CD24- MS-like ’ and ‘noCD44+/CD24- MS-like’ groups.

`*DLDA30*`7, A multigene predictor model of pathological complete response (pCR) built utilizing the 30-gene profile defined authors and their associated data with a Diagonal Linear Discriminant Analysis (DLDA, R package *sfsmisc* v1.1) to classify samples in ‘pCR-like’ and ‘RD-like’ groups.

`*E2F4*`8, A metagene Z-score built utilizing 24 gene-profile of E2F4 target activation defined by authors.

`*ICR*`9, An unsupervised classification model built utilizing the 20-gene profile of immune phenotypes defined by authors performed with consensus clustering (R package *ConsensusClusterPlus* v1.46) to classify samples in 4 ICR groups.

`*Immunome*`10, A metagene score model built utilizing gene expression signatures of 28 different immune cell types of infiltrating tumors defined by authors.

`*RBloss*`11, A metagene Z-score model built utilizing the 159-gene profile of RB-pathway loss defined by authors to classify samples in ‘low’, intermediate’ and ‘high’ RB-loss groups.

**Additional molecular variables used without using custom code**

`Gatza`12, A prediction model based on Bayesian Factor Regression Modeling to define activation probability of multiple molecular pathway using the web-tool SIGNATURE13.

PDL1 , mRNA expression (log2) of PD-L1 (CD274) exported from each used data sets as previously described14. Briefly, prior to extraction of PD-L1 in each data set, data were standardized using the luminal A population as reference to make PD-L1 comparable across all sets.

PIK3CA & TP53 mut, mutational status of PIK3CA and TP53 were retrieved from TCGA BRCA data15 using National Cancer Institue (NCI) Genomic Data Commons (GDC) web-tool where silent mutations were excluded.

HRD score, Homozygous Recombination Deficience (HRD) score was retrieved from TCGA BRCA data15 using NCI GDC web-tool.

**Reference**

1. Krijgsman, O. *et al.* A diagnostic gene profile for molecular subtyping of breast cancer associated with treatment response. *Breast Cancer Res. Treat.* **133**, 37–47 (2012).

2. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).

3. van ’t Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).

4. Filipits, M. *et al.* A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res* **17**, 6012–6020 (2011).

5. Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913 (2009).

6. Creighton, C. J. *et al.* Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13820–13825 (2009).

7. Hess, K. R. *et al.* Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J. Clin. Oncol.* **24**, 4236–4244 (2006).

8. Guerrero-Zotano, A. L. *et al.* ER+ Breast Cancers Resistant to Prolonged Neoadjuvant Letrozole Exhibit an E2F4 Transcriptional Program Sensitive to CDK4/6 Inhibitors. *Clin. Cancer Res.* **24**, 2517–2529 (2018).

9. Bertucci, F. *et al.* The immunologic constant of rejection classification refines the prognostic value of conventional prognostic signatures in breast cancer. *Br J Cancer* **119**, 1383–1391 (2018).

10. Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795 (2013).

11. Ertel, A. *et al.* RB-pathway disruption in breast cancer: differential association with disease subtypes, disease-specific prognosis and therapeutic response. *Cell Cycle* **9**, 4153–4163 (2010).

12. Gatza, M. L. *et al.* A pathway-based classification of human breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6994–6999 (2010).

13. Chang, J. T. *et al.* SIGNATURE: A workbench for gene expression signature analysis. *BMC Bioinformatics* **12**, 443 (2011).

14. Sabatier, R. *et al.* Prognostic and predictive value of PDL1 expression in breast cancer. *Oncotarget* **6**, 5449–5464 (2015).

15. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).