DNA methylation-based immune response signature improves patient diagnosis in multiple cancers

Jana Jeschke et al.

SUPPLEMENTAL MATERIAL

1. SUPPLEMENTAL METHODS

Patient cohorts

Three BC patient cohorts were used in this study. **Cohort 1** and **cohort 2** consist of 118 and 119 retrospectively selected fresh frozen tumor samples from patients treated with adjuvant therapies, as per institutional recommendations, and diagnosed at the Jules Bordet Institute from 1995 to 2003 and 2004 to 2009, respectively. Both studies received approval by local ethics committees (Institute Jules Bordet, nr 1918). Samples for both cohorts were selected according to the following criteria: (i) sufficient presence of invasive cells as defined by a pathologist (> 90% of tumor area); (ii) > 2µg yield of DNA available; (iii) balanced distribution of the four main "BC expression subtypes" determined by IHC; and (iv) balanced distribution of patients with and without relapses within each subtype. Of note, cohort 2 was slightly enriched in basal-like tumors as compared to cohort 1. For survival analyses, we considered only samples for which the lymphocyte infiltration assessed by pathology (PaTIL) and BC subtypes defined based on IHC results of the hormone receptor (ER and/or PR) and HER2 were available resulting in cohort sizes of 105 and 100 for cohort 1 and cohort 2, respectively. 'The preoperative Trial of Principle' (**TOP**) cohort consists of 149 patients with estrogen receptor (ER) negative disease that were treated at the Jules Bordet Institute from 2003 to 2008 with neoadjuvant epirubicin monotherapy (100 mg/m2). Patients with operable BC were treated every 3 weeks for four cycles and patients with inflammatory or locally advanced BC every 2 weeks for six cycles. Pretreatment biopsies were obtained from the primary lesion. Pathologic complete response (pCR) was the primary endpoint of this trial. pCR was defined as the absence of residual invasive breast carcinoma in the breast and in the axillary nodes after completion of chemotherapy. Persistence of in situ carcinoma without an invasive component was also considered pCR. Fifty-eight fresh frozen samples with sufficient amount of tumor cells and high DNA yield were selected for Infinium methylation analyses. Patients provided written informed consent prior to study entry. The study received approval by local ethics committees. The trial, including methodology, baseline patient characteristics and results, has been extensively reported by Desmedt et al. (1). The clinical and pathological characteristics for the three patient cohorts are summarized in eTable 7. The primary outcome was overall survival, calculated from the date of surgery to the date of death or date of last follow-up. Disease-free survival was calculated as the time from the date of surgery to the time of recurrence or date of last follow-up. Median follow-up times for overall or disease-free survival were 89, 55 and 37 months for PNC1, PNC2 and TOP, respectively.

TCGA cancer data

DNA methylation data and PaTIL for twenty cancer types were downloaded from the TCGA data repository (March 2015) and clinical data and survival information from the 'firehose' website ('Clinical Pick Tier1' files January 2017). The clinical and pathological characteristics for the TCGA cohorts are summarized in eTable 11. Histopathological measurements of TILs for TCGA tumors were performed as summarized in eTable 3. Melanoma subtype information were obtained from a recent study by the Cancer Genome Atlas Network (2).

ENCODE and other public data

ENCODE data for different cell types typically found in a breast tumor biopsy (breast epithelial cells, lymphocytes, fibroblasts, muscle cells, endothelial cells, and adipocytes) were retrieved from GEO

(GSE40699 and GSE40700) and different types of sorted blood cells were also obtained from GEO (GSE35069, GSE39981, GSE49667 and GSE59796). The GSE40699, GSE35069, GSE49667 and GSE59796 series contains Infinium HumanMethylation450K raw data (.idat or raw intensities) that were quality checked and preprocessed similarly our in-house array data (see *Infinium HumanMethylation450K pre-processing*). The GSE40700 and GSE39981 series contains Infinium HumanMethylation27K data expressed as methylation scores ranging from 0 to 100 or 1000 (GSE40700) or raw intensities (GSE39981). These data were converted as Beta-values (0 to 1 range) and used as such. Beta-values of the five cytosines from the MeTIL signature were used to compute the MeTIL score using the formula described in *Establishment of the MeTIL score*.

Tumor preparation and flow cell sorting

Fresh breast tumor tissues were collected immediately following surgery and dissociated (without enzymatic digestion) using the GentleMacs Dissociator (Miltenyi Biotec, Begisch Gladbach, Germany) prior to Ab labeling, as previously described (3). Fluorescent-conjugated antibodies against CD45 (BD Biosciences, San Jose, CA, USA) and Epcam (Miltenyi Biotec, Netherlands) were used for surface staining of cells, according to manufacturer's protocol. Fluorescently labeled lymphocytes (CD45+Epcam-) and tumor cells (CD45-Epcam+) were sorted on a Moflo ASTRIOS EQ 12/4 sorter. The cell purity were controlled on a GALLIOS 10/3 cytometer and analyzed using Kaluza® 1.3 Flow Analysis Software (Beckman Coulter, Brea, CA, USA).

DNA methylation profiling

Infinium HumanMethylation450K

Genomic DNA was extracted with the Qiagen-DNeasy Blood &Tissue Kit (Qiagen, Hilden, Germany) or the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) as previously described (4). DNA methylation was analyzed on Infinium HumanMethylation450K bead-arrays. Genomic DNA (300 to 800 ng) was converted with sodium bisulfite using the Zymo EZ DNA Methylation Kit (Zymo Research, Orange, USA) following the alternative incubation conditions of the manufacturer recommended for Illumina Infinium HumanMethylation assays. Methylation assays were performed with 4 ml converted DNA at 50 ng/ml according to the manufacturer's protocol. GenomeStudio™ Methylation Module software (TOP cohort) or methylumi R package (other cohorts) were used to extract raw probe intensity values. The quality of array data was evaluated using the visualization tool from GenomeStudio™ (or in-house R reimplementation of this tool) by assessing the intensity level of the control probes. All samples that showed the expected profiles for the different control probes were utilized for further analyses. Infinium HumanMethylation 450K raw data were submitted to Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi? GSE72308). Breast epithelial cell and lymphocyte methylation and gene expression profiles were assessed using HumanMethylation27K and Affymetrix HG133 Plus 2.0 expression arrays in a previous study (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20713).

Bisulfite pyrosequencing

PyroMark Assay Design 2.0 and PyroMark Q24 software (Qiagen) were used for the design of PCR and sequencing primers and for analysis and visualization of pyrosequencing results, respectively. PCR primers

for each marker of the MeTIL signature were designed around the target cytosine from the Infinium probe. Thereby, the product size was kept as minimal as possible to anticipate amplification difficulties of fragmented bisulfite-treated FFPE DNA. Genomic DNA (275 ng) was bisulfite-converted with the EZ DNA Methylation^{TM} kit (Zymo Research) and 3 to 6 µl of converted DNA (corresponding to approximately 45 - 95 ng DNA) served as input for PCR. PCR assays were performed with HotStarTaq DNA polymerase (Qiagen) under the following cycle conditions: 95°C 15 min; 50 to 60 cycles of 95°C 1 min; 50 - 53°C 1 min; 72°C 1 min; 72°C 10 min. Amplification was confirmed on an agarose gel and pyrosequencing of successfully amplified PCR products was performed with the PyroMark^{TM} Q24 system (Qiagen). Primer sequences are provided in eTable 14.

Bioinformatics

Infinium HumanMethylation450K pre-processing: Raw data (uncorrected probe intensity values) from the Infinium Methylation arrays were processed with the following steps: Probes of low quality (detection p-value threshold of 0.05), cross-reactive probes (i.e. targeting several genomic locations) as well as probes containing SNPs based on the extended annotation of Price et al. (5, 6) were removed. Additionally, probes targeting X and Y-chromosomes were removed from the analysis. Beta-values were computed using the following formula: Beta-value = M/[U+M] where M and U are the raw "methylated" and "unmethylated" signals, respectively. Beta-values were corrected for type I and type II bias using the peak-based correction (6, 7).

Selection of markers for MeTIL signature: Given that among TILs, T-cell subsets are most abundant (8) and associated with clinical outcomes (9-12), we identified in a first step T-lymphocyte-associated cytosines by utilizing previously published genome-wide DNA methylation profiles (Infinium HumanMethylation) from eight normal or cancerous breast epithelial cell lines (MCF10A, MCF-7, T47D, SKBR3, BT20, MDA-MB-231, MDA-MB-361, ZR-75-1) and three T-lymphocyte samples (WEIS3E5, R12C9 and ex-vivo T-cells) (4). We computed the median Beta-value for each probe in the group of T-lymphocytes and in the group of breast epithelial cells (eFigure 1) and calculated the delta Beta-value for each probe between the median Beta-values of the two groups ($\text{Median}_{\text{T-lymphocytes}} - \text{Median}_{\text{Breast epithelial cells}}$). Since our previous work in BC revealed that TILs are reflected primarily through hypomethylation of immune-related genes (4), we selected probes that were hypomethylated in T-lymphocytes versus breast epithelial cells with a minimum difference in delta Beta of 0.8 (Delta Beta < -0.8). To identify the most discriminative probes, we further selected for probes that showed a low variability within both groups ($SD_{\text{T-lymphocytes}} \leq 0.1$ & $SD_{\text{Breast epithelial cells}} \leq 0.1$). This selection approach yielded 29 T-lymphocyte-associated markers (eTable 1).

MeTIL signature optimization (machine learning): We applied a machine learning approach to select for probes from the list of T-lymphocyte-associated cytosines (eTable 1) that reflect most accurately the quantity of PaTIL in patient samples (eFigure 2). We utilized Infinium HumanMethylation profiles of 105 breast primary tumors (cohort 1) for which H&E staining-based PaTIL (in %) were available. We defined three PaTIL categories based on the pathological intra-tumoral (iTU-Ly) PaTIL readings: 'PaTIL-Absent' (PaTIL < 1%), 'PaTIL-Low' (PaTIL ≥ 1% & PaTIL ≤ 20%) and 'PaTIL-High' (PaTIL > 20% & PaTIL ≤ 100%). The BC cohort was divided into three parts (two 'training' sets and one 'test' set) in order to apply a three-fold cross validation. To establish a small signature with minimal redundancy between probes we applied the mRMR feature selection method to the training set (13). To account for the unequal size

between the three TIL categories (PaTIL-High is smaller than PaTIL-Low or PaTIL-Absent), we integrated the 'EasyEnsemble' approach (14) into the model. This approach trained 20 different Random Forest models, each model on a different data set obtained through the inclusion of all samples from the PaTIL-High category while randomly selecting the same number of samples from the PaTIL-Low and the PaTIL-Absent category. The performance of the developed 'EasyEnsemble' models was assessed through prediction making in the test set and the computation of the 'Balanced-Error Rate' (BER). For a more robust estimation of the BER, a three-fold cross validation was applied 200 times and for each run randomized data were used as negative control. This entire process was repeated for a signature size ranging from a single cytosine to the entire set of T-lymphocyte-associated cytosines. The signature size, for which no more improvement of the BER was observed (five features signature size), was selected as the final size (eFigure 8). This process generated 3x200 output signatures. The distance between the signatures was defined as $1-\left(\sum_{i=1}^{k} cor_{spear}(F1_i,F2_i)\right)/k$ (cor_{spear} refers to the spearman correlation, $F1_i$ to the i^{th} feature from signature 1 and $F2_i$ to the i^{th} feature from signature 2 after sorting the features in order to maximize the sum of the spearman correlation). For each signature, the sum of its pairwise distance to all other output signatures was computed, and the signature with the smallest sum was assumed to be the most representative and chosen as final signature (named MeTIL signature) (eTable 4).

Establishment of the MeTIL score: We applied a "normalized PCA" (NPCA) approach to transform the individual methylation values of the probes of the machine learning-derived MeTIL signature into a score (MeTIL score), which reflects TIL percentage in a way that does not require complex algorithm and can easily be applied by any lab on any methylation dataset. Therefore, we first standardized ($\mu=0$, $\sigma=1$) the Beta-values of each CpG from the signature on the discovery dataset (cohort 1). We then applied a principal component analysis (PCA) and used the first component as final MeTIL score for the samples of the discovery set. The MeTIL score on any other datasets was obtained using the NPCA parameters derived from the discovery set with the following formula: $Sc\left(\mathbf{beta}|\vec{s},\vec{c}\right)\vec{v}$ (**beta** is the matrix of Beta-values for the MeTIL signature on the new dataset, $Sc\left(\mathbf{beta}|\vec{s},\vec{c}\right)$ are the data transformed using the scale *s* and the center *c* from the discovery cohort and *v* is the eigenvector from the discovery cohort. Note that this formula also allows any reader to generate the MeTIL score from its own methylation data, for example using the following command in R language: (scale(beta, center=c1, scale=s1)%*%v1)[,1] (beta is the matrix of Beta-values for the probes from the signature (with samples in rows and probes in column) and c1, s1 and v1 are parameters from our discovery cohort that can be found in eTable 15.

Simulations: Whole tumor samples were simulated using Infinium methylation profiles from cell lines. Breast epithelial cell line profiles where used to mimic tumor cells while T-lymphocyte profiles represented TILs. A random selection of methylation probes across all cell lines was used to simulate the presence of additional cells types typically represented in a tumor (fibroblasts, endothelial cells, macrophages etc.). These three components and additional noise were combined using the following formula: $M_c = f_1M_1 + f_2M_2 + f_3M_3 + e$ where M_c is the combined M-value, f_1, f_2 and f_3 are the fractions of TILs, tumor cells and other cell-type(s) in the simulated tissue, respectively, and M_1, M_2 and M_3 are their respective M-values and *e* a Gaussian noise of mean equal to 0 and standard deviation of *ws (s* is computed as the standard deviation of the delta between M-values of each possible pair of cell lines from the same type across all Infinium probes and *w* is a weight used to increase/decrease the noise level). M-values were used in this analysis since the heteroscedasticity and the bounded nature of Beta-values are incompatible

with a Gaussian noise while M-value are homoscedastic and unbounded (15). The conversion from Beta-values to M-values was achieved using a previously described formula (6). The combined M-value M_c was converted back into a Beta-value using the appropriate formula (6) and the MeTIL score was computed using the formula described in "Establishment of the MeTIL score". To verify the correlation between the MeTIL score and TILs, simulations were performed for increasing values of *f1* with fixed value of e and *f3,* and *f2* defined as 1-(*f1*+*f3*). 50 steps of *f1* were assessed (*i.e.* each step corresponding to a 2% increase in TILs if *f3*=0) and 200 simulations were calculated at each step allowing the spearman correlation to be computed from 10 000 dots. Figure 1G was obtained by performing this process with different values of *f3* (1% increased at each step) and *w* (0.05 increase by step).

Statistics

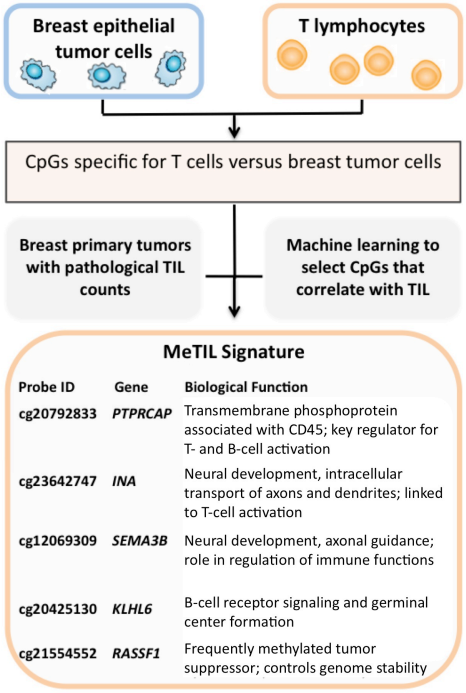
Statistical analyses were conducted with RStudio Version 0.94.110. Hierarchical clustering was applied using the Euclidian distance and complete linkage. Correlations were computed using the Spearman method and their significance were assessed using a correlation test. Differences between more than two groups were assessed with a one-way ANOVA (for continuous variable) or Chi^2 test (for discrete ones). Survival analyses were conducted with the 'survival' package. Cox proportional hazard regression analyses and Kaplan-Meier survival curves with log-rank tests, recording patients at the time of death or recurrence or last follow-up visit, were used to compare overall survival or disease free survival rates. Multivariate Cox regression models were established using a forward and backward variable selection based on the Akaike's information criterion (AIC) from the 'MASS' package. The model with the lowest AIC, including the variable of interest (MeTIL score or PaTIL), was chosen as the final multivariate model. Co-variables for multivariate Cox models in breast and other cancers were binarized as displayed in eTable 7 and 11. Missing data points for co-variables (if missed in < 10% of samples) were imputed by the simple mean matching method (16, 17). Odds ratios (ORs) were used to compare pCR rates. The area under the curve (AUC) was used to assess prediction performance. ROC and AUC were computed using the 'pROC' package. Generalized linear models (glm) and ORs were computed with the 'aod' package. All *P* values were two sided and *P* values < 0.05 were considered to be statistically significant.

2. SUPPLEMENTAL REFERENCES

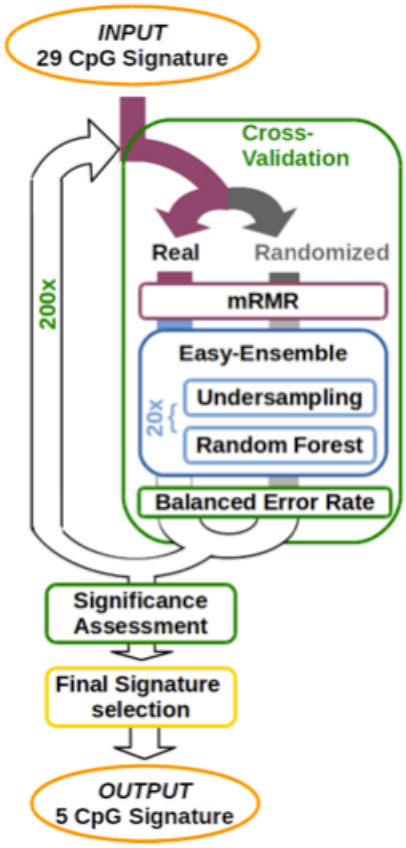
1. Desmedt, C., Di Leo, A., de Azambuja, E., Larsimont, D., Haibe-Kains, B., Selleslags, J., Delaloge, S., Duhem, C., Kains, J.P., Carly, B., et al. 2011. Multifactorial approach to predicting resistance to anthracyclines. *J Clin Oncol* 29:1578-1586.
2. Cancer Genome Atlas Network. Electronic address, i.m.o., and Cancer Genome Atlas, N. 2015. Genomic Classification of Cutaneous Melanoma. *Cell* 161:1681-1696.
3. Garaud, S., Gu-Trantien, C., Lodewyckx, J.N., Boisson, A., De Silva, P., Buisseret, L., Migliori, E., Libin, M., Naveaux, C., Duvillier, H., et al. 2014. A simple and rapid protocol to non-enzymatically dissociate fresh human tissues for the analysis of infiltrating lymphocytes. *J Vis Exp*.
4. Dedeurwaerder, S., Desmedt, C., Calonne, E., Singhal, S.K., Haibe-Kains, B., Defrance, M., Michiels, S., Volkmar, M., Deplus, R., Luciani, J., et al. 2011. DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol Med* 3:726-741.
5. Price, M.E., Cotton, A.M., Lam, L.L., Farre, P., Emberly, E., Brown, C.J., Robinson, W.P., and Kobor, M.S. 2013. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* 6:4.
6. Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., and Fuks, F. 2011. Evaluation of the Infinium Methylation 450K technology. *Epigenomics* 3:771-784.
7. Dedeurwaerder, S., Defrance, M., Bizet, M., Calonne, E., Bontempi, G., and Fuks, F. 2014. A comprehensive overview of Infinium HumanMethylation450 data processing. *Brief Bioinform* 15:929-941.
8. Ruffell, B., Au, A., Rugo, H.S., Esserman, L.J., Hwang, E.S., and Coussens, L.M. 2012. Leukocyte composition of human breast cancer. *Proc Natl Acad Sci U S A* 109:2796-2801.
9. Mahmoud, S.M., Paish, E.C., Powe, D.G., Macmillan, R.D., Grainge, M.J., Lee, A.H., Ellis, I.O., and Green, A.R. 2011. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol* 29:1949-1955.
10. Seo, A.N., Lee, H.J., Kim, E.J., Kim, H.J., Jang, M.H., Lee, H.E., Kim, Y.J., Kim, J.H., and Park, S.Y. 2013. Tumour-infiltrating CD8+ lymphocytes as an independent predictive factor for pathological complete response to primary systemic therapy in breast cancer. *Br J Cancer* 109:2705-2713.
11. Bates, G.J., Fox, S.B., Han, C., Leek, R.D., Garcia, J.F., Harris, A.L., and Banham, A.H. 2006. Quantification of regulatory T cells enables the identification of high-risk breast cancer patients and those at risk of late relapse. *J Clin Oncol* 24:5373-5380.
12. Gu-Trantien, C., Loi, S., Garaud, S., Equeter, C., Libin, M., de Wind, A., Ravoet, M., Le Buanec, H., Sibille, C., Manfouo-Foutsop, G., et al. 2013. CD4(+) follicular helper T cell infiltration predicts breast cancer survival. *J Clin Invest* 123:2873-2892.
13. Peng, H., Long, F., and Ding, C. 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 27:1226-1238.
14. Liu, X.Y., Wu, J., and Zhou, Z.H. 2009. Exploratory undersampling for class-imbalance learning. *IEEE Trans Syst Man Cybern B Cybern* 39:539-550.

15. Du, P., Zhang, X., Huang, C.C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. 2010. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11:587.
16. Little, R. 1988. Missing data adjustments in large surveys. *J Bus Econ Stat* 6:287-296.
17. Carpenter, J. 2007. Missing data in randomised controlled trials — a practical guide [Internet]. Available from: http://www.hta.nhs.uk/nihrmethodology/re- ports/1589.pdf.

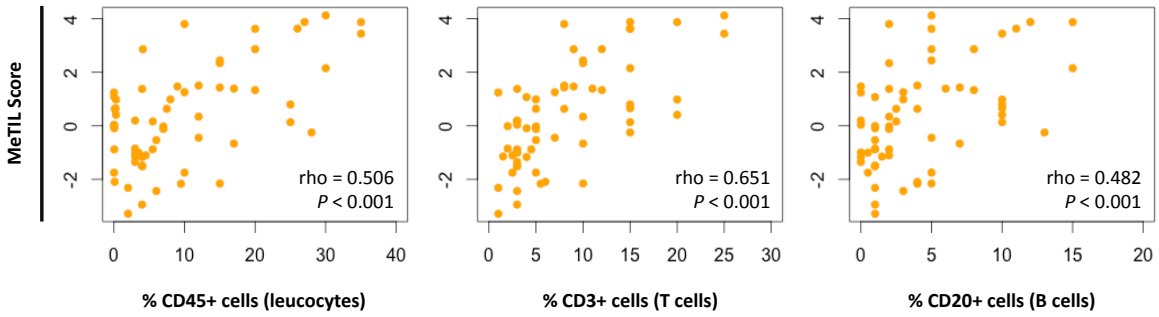
3. SUPPLEMENTAL FIGURES



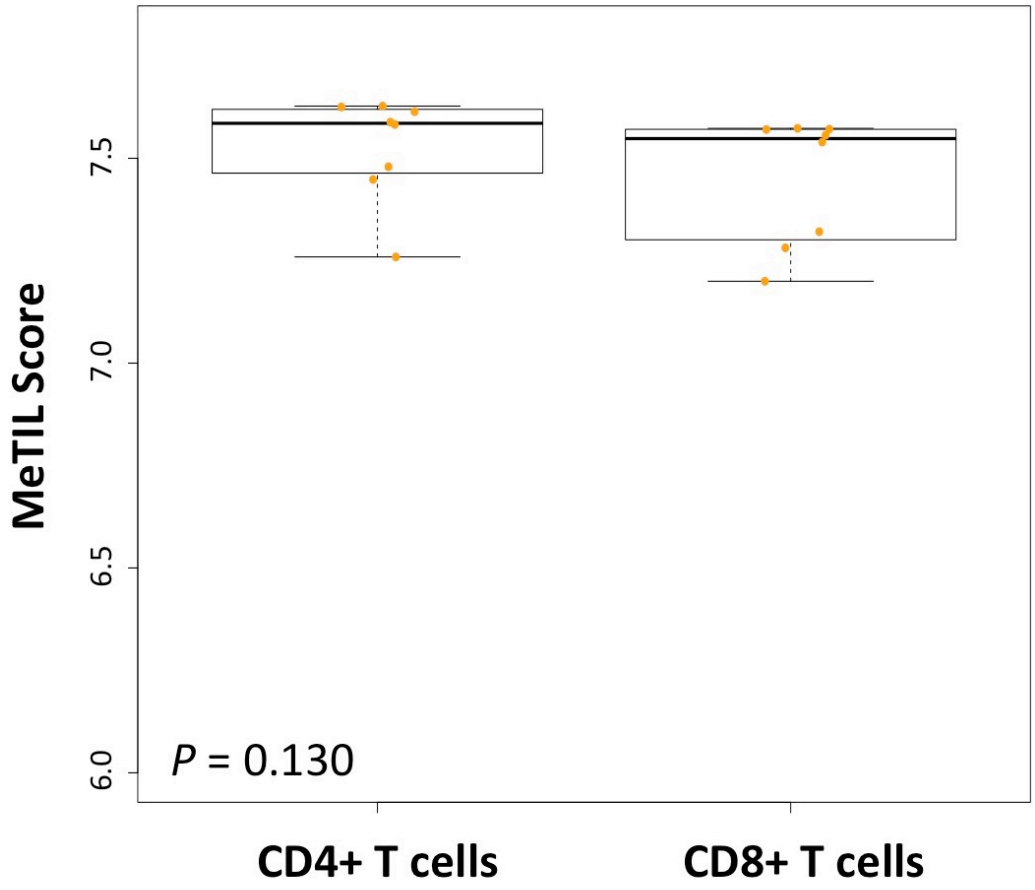
eFigure 1. Development of the MeTIL signature. MeTIL markers were identified in the following steps: 1) Filtering for T-cell-associated probes (CpG) from genome-wide DNA methylation profiles (Infinium Methylation) that are highly differentially methylated (Delta Beta < -0.8) between T-lymphocytes and breast epithelial cells. 2) Applying a random forest machine learning approach on DNA methylation profiles (Infinium HumanMethylation) from 105 breast primary tumors (cohort 1) for which H&E staining-based PaTIL were available to select for markers from our list of T-cell-associated CpGs that predicts most accurately the quantity of PaTIL in patient samples. The final signature included five markers and was named 'MeTIL' (**Me**thylation of **TIL**) signature.



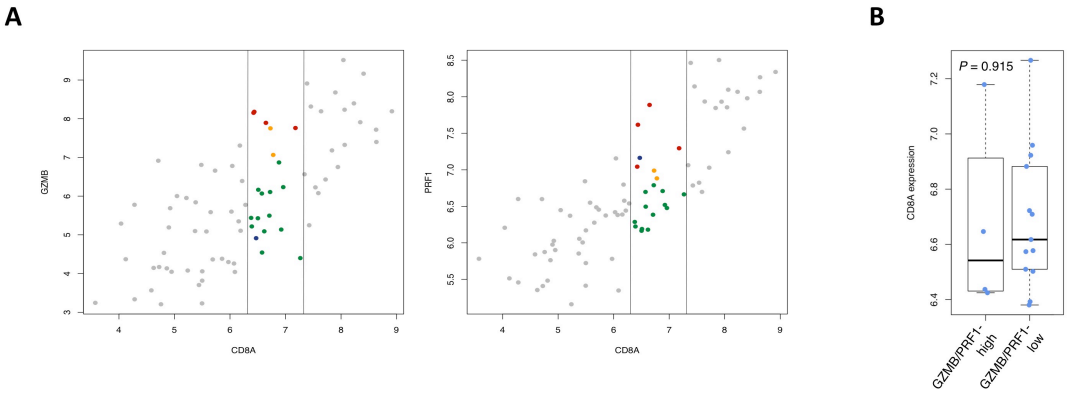
eFigure 2. Machine learning approach. Selecting the final markers of the MeTIL signature for quantification of TILs from the list of differential methylated markers between breast cells and T-lymphocytes by applying a random forest machine learning approach on cohort 1 (n = 105) for which PaTIL were available.



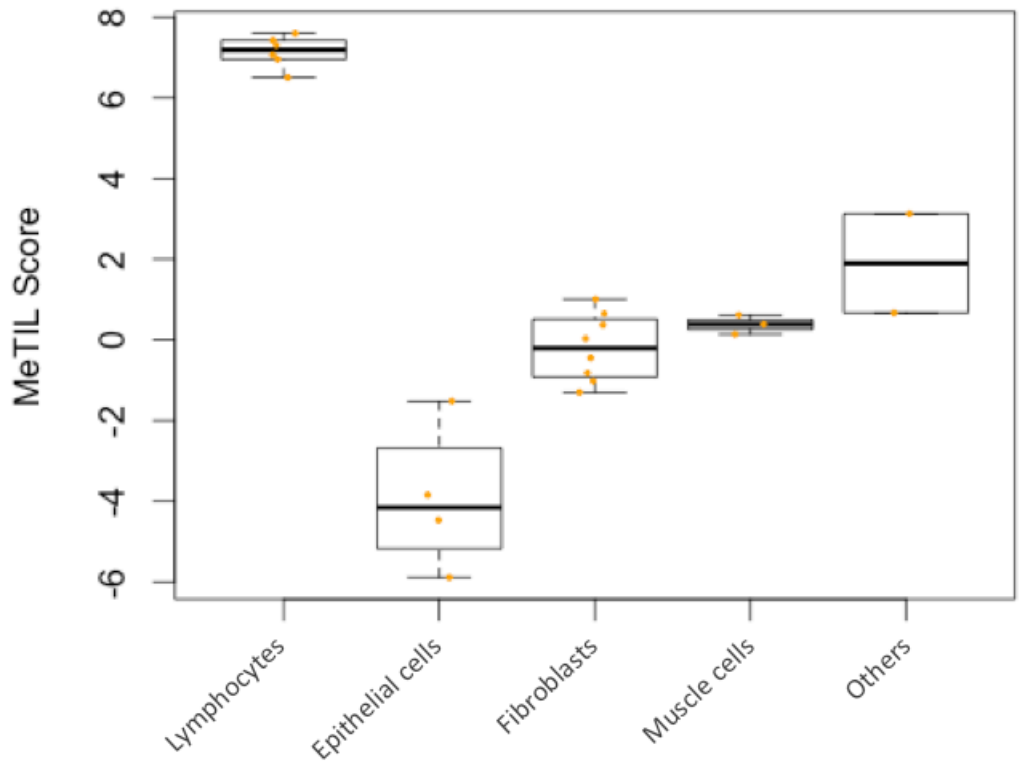
eFigure 3. MeTIL score correlates with IHC-based TIL counts. The MeTIL signature was transformed into a score and correlated with % of CD45+ (total leucocytes), CD3+ (total T-cells) and CD20+ (total B-cells) cells obtained by IHC staining in 61 samples of cohort 1. The MeTIL score values are plotted on the y-axis and % of CD45+, CD3+ and CD20+ cells on the x-axis. The correlation was assessed with a Spearman's rank correlation test. The Spearman's rank correlation coefficient (rho) and its p value are displayed for each plot.



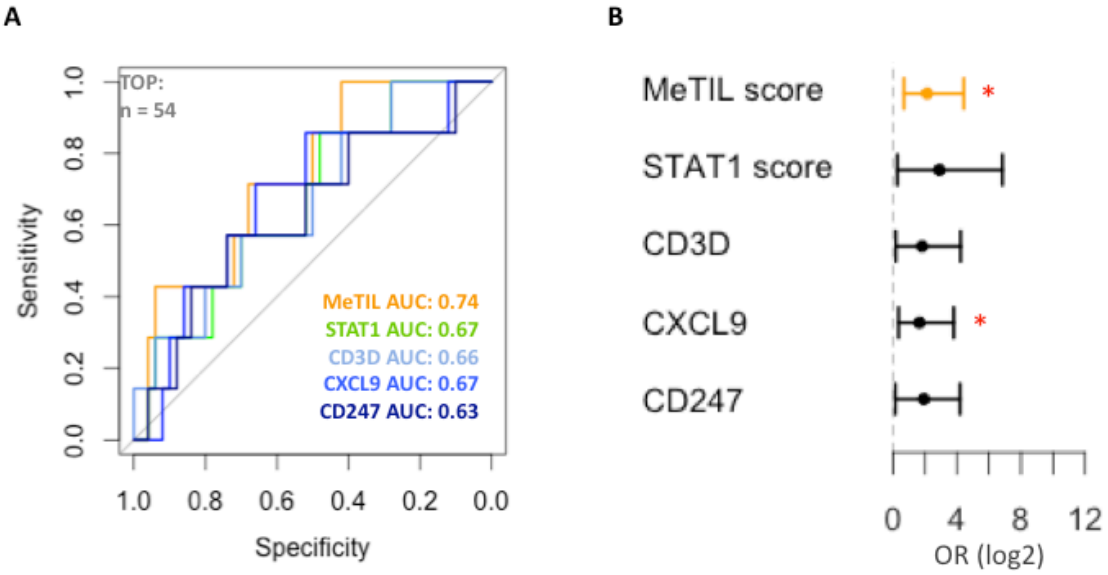
eFigure 4. MeTIL score values in T-cell subsets. MeTIL scores were computed in CD4+ or CD8+ T-cells sorted from eight whole blood samples. Differences between the MeTIL score of groups were assessed with a Student's t-test. P value is displayed in the lower, left corner of the plot.



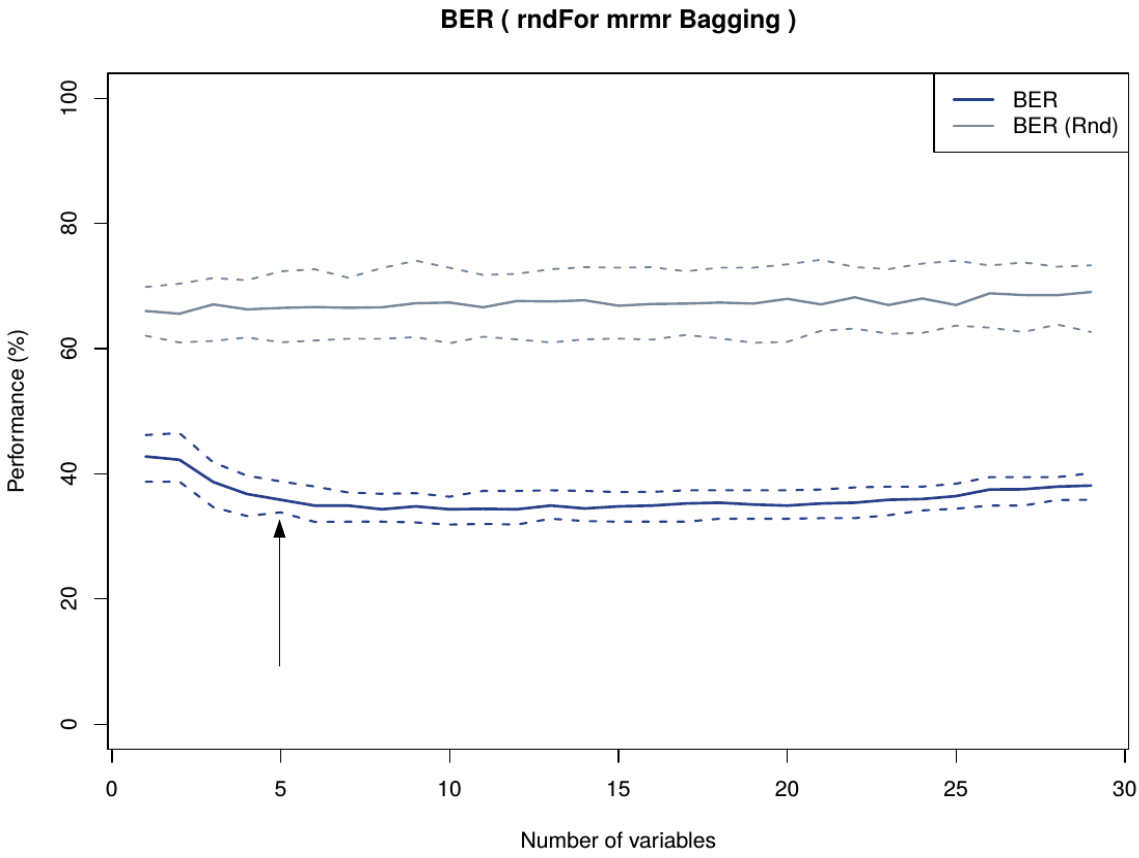
eFigure 5. MeTIL score reflects functionality of CTLs. A) Scatter plots displaying expression of *GZMB* and *CD8A* (left plot) and *PRF1* and *CD8A* (right plot) in samples of cohort 1. Note the linear correlation between *GZMB* and *CD8A* and *PRF1* and *CD8A*. In order to select for tumors enriched for functional or non-functional CTLs, we included tumors with *CD8A* log2 RNA-seq expression levels between 6.3 and 7.3 to assure the selection of tumors with comparable high levels of CTLs. The *CD8A* expression range was kept narrow to define groups of CTL-enriched tumors with high or low expression of *GZMB* and *PRF1* that was not confounded by expression of *CD8A*. The *CD8A* expression window was chosen based on the following criteria: 1) sufficiently high *CD8A* level to select for tumors enriched in CTLs, 2) not more than 1 log of expression between the beginning and end of window border to assure that samples have comparable level of CTLs and 3) the window with the largest range in expression of *GRZMB* and *PRF1* expression was chosen to define functional and non-functional CTLs. Red dots show samples with high expression of both *GZMB* and *PRF1* and green dots show samples with low expression of both *GZMB* and *PRF1*. Blue and yellow dots show samples in which the expression of *GZMB* and *PRF1* differed. Red-labeled samples were selected for the 'GZMB/PRF1-high group' and green samples for the 'GZMB/PRF1-low' group. **B)** Expression of *CD8A* in the 'GZMB/PRF1-high' and the 'GZMB/PRF1-low' group. Differences of *CD8A* expression between groups were assessed with Student's t- test. P value is displayed in the upper, left corner of the plot.



eFigure 6. MeTIL score values differ between cell types of the tumor microenvironment. The MeTIL score was computed for various samples from the ENCODE Consortium that represent different cell types typically found in a breast tumor biopsy. The MeTIL score values are plotted on the y-axis and the groups of cell types on the x-axis. The group 'Others' contains one sample classified as 'endothelial cells' and one sample classified as 'adipocytes'. Note, the MeTIL score displays highest values in lymphocytes as compared to other cell types fortifying its specificity for TIL evaluation.



eFigure 7. Predictive values of the MeTIL score or gene expression-based immune markers for response to preoperative anthracycline treatment. A) ROC curves for prediction of response to neoadjuvant anthracycline treatment based on the MeTIL score (orange) or gene expression-based immune markers (other colors) in 54 patients of the TOP cohort**. B)** Forest plot showing the log2 value of the odds ratios (OR) and CI of the MeTIL score (orange) or gene expression-based immune markers (black) for prediction of response to neoadjuvant anthracycline treatment in the TOP cohort. A red star indicates statistical significance ($p < 0.05$).



eFigure 8. Balanced-Error Rate. The performance of the developed random forest machine learning models was assessed through 'Balanced-Error Rates' (BERs) in a three-fold cross validation utilizing signature sizes ranging from a single cytosine to the entire set of T-lymphocyte-specific cytosines. The signature size, for which no more improvement of the BER was observed (five features signature size), was selected as the final model.

4. SUPPLEMENTAL TABLES

eTable 1. Probe information for T-cell-associated markers.

Ilmn ID	Corresponding gene	Full name	mRNA ID RefSeq	Cytosine position hg19	Surrounding sequence
cg00777121	RASSF1	Homo sapiens Ras association (RalGDS/AF-6) domain family member 1	NM_007182; NM_170714	chr3:50378191	GCGCGATGCGCAGCGCGTTGG CACGCTCCAGCCGGGTGCGGC CCTTCCCAGCGCGCCCAG[CG] GGTGCCAGCTCCCGCAGCTCA ATGAGCTCAGGCTCCCCCGAC ATGGCCCGGTTGGGCCCG
cg04396791	SHANK2	Homo sapiens SH3 and multiple ankyrin repeat domains 2	NM_012309; NM_133266	chr11:70508180	GGCAAAGAAAGTAAGTGGCCC GAACGGCGGCGGCGGCAGCG GCGGCCTCAGCAGCGGCGG[C G]TCGGGAATGAGCTCCTCCGG CTTGTCCCGACCACACGCATCT TCCAGGGAAGCCCGGGAA
cg05596756	FAM113B	Homo sapiens PC-esterase domain containing 1B	NM_138371; NM_138371; NR_026544	chr12:47610220	GACTTACCATTTCATGTTTTCAC ACATGTCAAAGTCTTGAAGCTT CGGTCATTAGTCTTC[CG]CTTC CTAAAGCAAGAAACTAAATGCT GACTTGTGTCGTTTGCCCCAGA TGCCACCGGAAA
cg06151165	VSX1	Homo sapiens visual system homeobox 1	NM_014588; NM_199425	chr20:25062254	CCGCCTAGATCTGGGCCGCTG GGGCCAGGGGTGCGGTGGGG CGATGGTCTGTGACCCCTG[CG]CGGCTCAGAGCCTAGGGGACA GGGGCAGGAGCGGAAAGCGC GGGCCTGATTACCGGACGT
cg06825142	DRD4	Homo sapiens dopamine receptor D4	NM_000797	chr11:637170	CGTGCCCGTCTCTCCCTGCGC AAAATTCCAAGATGAGCAAATA CTGGGCTCACGGTGGAG[CG]C CGCGGGGGCCCCCCTGAGCC GGGGCGGGTCGGGGGCGGGA CCAGGGTCCGGCCGGGGCG
cg07380416	CD6	Homo sapiens CD6 molecule	NM_006725	chr11:60739172	TGCAGACCAAAACCACAAGCA GAACAAGCAGGCGTGAGACAC TCACAGGTTGGGTTTGAT[CG]C ATGCGTGTCGGAGAGGAGAGA GCAGAGAGAGACACAGGAACA AGAACAGCAAAGGGTAG
cg07498879	OSM	Homo sapiens oncostatin M	NM_020530	chr22:30663041	AATGGGCGCAGGGCGCGGTGT ATGCCCGCTCCTCCTCCTGTTT TCTTCGAATTCGTTCTT[CG]AG GTCAGCCCTACGCCCAAGGAT GATGTAAAACCGCAGCCTCAG CCTTCCTAGGGTGAGT

cg08047457	RASSF1	Homo sapiens Ras association (RalGDS/AF-6) domain family member 1	NM_007182; NM_170714	chr3:50378413	AAGGAGCTGAGGAGAGCCGCG CAATGGAAACCTGGGTGCAGG GACTGTGGGGCCCGAAGG[CG] GGGCTGGGCGCGCTCTCGCAG AGCCCCCCCCGCCTTGCCCTT CCTTCCCTCCTTCGTCCC
cg09902130	CD6	Homo sapiens CD6 molecule	NM_006725	chr11:60739178	CCAAAACCACAAGCAGAACAAG CAGGCGTGAGACACTCACAGG TTGGGTTTGATCGCATG[CG]TG TCGGAGAGGAGAGAGCAGAGA GAGACACAGGAACAAGAACAG CAAAGGGTAGAGCAGA
cg10590292	BIN2	Homo sapiens bridging integrator 2	NM_016293	chr12:51717674	TCTCTCCAATCTGCCAGTTTTC ATGGAAGCCCACTCCTCGCTC CCGTTTCCCTGGGAGAG[CG]G GAGACCCTTCTCGTCAGCCCA GACCCTTCGGGGCCCCTGAGC ACCCTGCCCGCCCCTCC
cg12069309	SEMA3B	Homo sapiens sema domain, immunoglobulin domain, secreted, (semaphorin) 3B domain (Ig), short basic	NM_00100591 4; NM_004636	chr3:50312913	GGGGTCGCGTGCACGCGCTTC CAGCCCAGTGCCAAGAGGTGG GCGGGGTCGGGGTTGGGC[CG] CCGGGAGGGAGGCGAAGGGT CTTTCACTGCCCGGGGCTGAA AGAAGGGCTCACAGAAGAT
cg13470920	VAV1	Homo sapiens vav 1 guanine nucleotide exchange factor	NM_005428	chr19:6772831	TGGTGGAGGCTGCGAGGGTGC ACGGCCGGCCCTGGGCAGGC GGTAGCCATGGAGCTGTGG[CG]CCAATGCACCCACTGGCTCAT CCAGTGCCGGGTGCTGCCGCC CAGCCACCGCGTGACCTG
cg14519350	OSM	Homo sapiens oncostatin M	NM_020530	chr22:30661949	ACCGCTCACCACCACCCCCAA CGGGGCACCGGCAGCCACGTG GTCGGAAGGGGCCTATGG[CG] CAATGGCCCGGAAGGCAGTGA CACCATCGTTCCCGTCCTAAGA ACCATCCGGCCCTTTGG
cg16158681	MT3	Homo sapiens metallothionein 3	NM_005954	chr16:56623109	CAGGGAAGAGCTGGGAAATAC GCAAAGCGCCTTTTTCTCCACT TTCGGAGATGGTACGTG[CG]C GCTTCCACGCAGTGGCGGCTG CTGCGGCGAGCACGTCCCCTG CGGGACCCACGCGGGGA
cg16509569	HEM1	Homo sapiens NCK-associated protein 1-like	NM_005337	chr12:54891634	ATGTCTTTGACATCTGCTTACC AGCATAAATTAGCAGAGAAGCT CACTATCCTGAATGAT[CG]CGG TCAGGGGGTTCTCATCCGTATG TATAACATCAAGAAGGTAAGCA TGAACAATGGGAC
cg17078393	LCK	Homo sapiens lymphocyte-specific protein tyrosine kinase	NM_005356	chr1:32717002	AATGGGGCCAGAGGGCTCCCG GGCTGGGCAGGTAAGGAGCGC TGGTATTGGGGGCGCAGG[CG] CCGGGGTGAGAGGCCTGATAG CAGACGGCTGCAGCTGTGCGG

					GCCCAGGCTCCCTAGGGA
cg17518965	EDG6	Homo sapiens sphingosine-1-phosphate receptor 4	NM_003775	chr19:3178955	GGCCGGGCGCGGGGGGCCGG AGGATGGCGGCCTGGGGGCC CTGCGGGGGCTGTCGGTGGC[CG]CCAGCTGCCTGGTGGTGCT GGAGAACTTGCTGGTGCTGGC GGCCATCACCAGCCACATGC
cg20425130	KLHL6	Homo sapiens kelch-like family member 6	NM_130446	chr3:183273245	TGGCAGGAGAATTCCTGAATGT CCACACACAAGATGACATCTGT CAGAGCGTTTTCCATT[CG]CAG GGTTTCCAGGCCATTCTGAAGA ATTAAGGAGAGTCCCGCGTCG TCAAATTTGACCTT
cg20622019	ADA	Homo sapiens adenosine deaminase	NM_000022	chr20:43279793	TTCTTCCCCGGGGCGCATCCA CAAACCTGAGCTCCTTGGCGCT CACCACCATAACGCTGG[CG]TA CAAAGCTGCTTCCATTCACTGA GCACTCAGTAGGCGCCGGGCA CAGCACTAGGTGCCT
cg20792833	PTPRCAP	Homo sapiens protein tyrosine phosphatase, receptor type, C-associated protein	NM_005608	chr11:67205195	GCTGTGTCGAGCGAGAAGTGA GCTCAGTGCTCGTCTGCAGTG AAGGGTGGCCCAGGCTTC[CG] CTTCCTGCCCACATACCCCACC TGCCCCTCCCTGCTGCAGGAC CCCTGGTCCACACCAGA
cg21554552	RASSF1	Homo sapiens Ras association (RalGDS/AF-6) domain family member 1	NM_007182; NM_170714	chr3:50378425	AGAGCCGCGCAATGGAAACCT GGGTGCAGGGACTGTGGGGCC CGAAGGCGGGGCTGGGCG[CG]CTCTCGCAGAGCCCCCCCCGC CTTGCCCTTCCTTCCCTCCTTC GTCCCCTCCTCACACCC
cg23093496	Cl6orf54	Homo sapiens chromosome 16 open reading frame 54	NM_175900	chr16:29757323	TAGAGCCACTCTTCCTGGTGCT GGACTAAGAGGTGCAGGCTTG GAGGGTGCAGGGCGGTC[CG]C CTCTCAGACGTAGAGGCCCGG CCTCGGATGAAGGCGGAAGGG AGGGCACCGCCTGTTGC
cg23642747	INA	Homo sapiens internexin neuronal intermediate filament protein, alpha	NM_032727	chr10:105036645	GGGATGCGACAGAGCTGTGTG GTTTCCGGATGGGAAACCTCA GTCGTTTAGGCACCCCTC[CG]C TCGAGTCACTTCCGAAGCAGTC GATTCTTGGGGAGAAGCGCTG CGGAAAGGGGCGACTC
cg23797100	LAT	Homo sapiens linker for activation of T cells	NM_001014989; NM_001014987; NM_014387; NM_001014988	chr16:28996053	GCGACTGAGATGGGGGACTGC TCTTCTCCGAGTGAGCTGGCTG AGTGTGACCGTGAGTCA[CG]C CCCTGCTTCCCTGGAGCCTGT CCCTTGCCTCACAGGCCTGGC TGAGGTGGGGGTGGGCA
cg24545967	SH2D3C	Homo sapiens SH2 domain containing 3C	NM_170600	chr9:130540941	CTGGTCTTCTTGGTCCCCTCTG TCATCTTGGCAAATTGTGTGAA GCCCTTGGCCAGCTTG[CG]AG TGGCCACTCAAGGCTCGGGAC

					ATCAAAGTTGAGGTGCCTCTCC CACCTCCCCTTCCCC
cg24841244	CD3D	Homo sapiens CD3d molecule, delta (CD3-TCR complex)	NM_001040651; NM_000732	chr11:118213330	TGCGAGAGAAGGGTAGCCAGT ACCAGGCCAGAGAGAAACGTG CTATGTTCCATCTCCCAG[CG]G AACTCATCCAGTAGATAAAGCC AGGTCACCGAACTATCAGCCTG GGTGAGAGCTGCCCT
cg26285698	Cl6orf54	Homo sapiens chromosome 16 open reading frame 54	NM_175900	chr16:29757334	TTCCTGGTGCTGGACTAAGAG GTGCAGGCTTGGAGGGTGCAG GGCGGTCCGCCTCTCAGA[CG] TAGAGGCCCGGCCTCGGATGA AGGCGGAAGGGAGGGCACCG CCTGTTGCTGGGCAACTGT
cg27377213	PPP1R16B	Homo sapiens protein phosphatase 1, regulatory subunit 16B	NM_015568	chr20:37433803	TACGACCACTTCCTTCCTCCAG CGCTTCACCTCCCTTCTTTCCA GAAAGAGGAAATGATG[CG]AGT CCACACCGCAGGCCACAGCCC CCATGTCTCGGCCCCAGCAGC TGGGGGACCTGGCCC
cg27652350	ALDH1A3	Homo sapiens aldehyde dehydrogenase 1 family, member A3	NM_000693	chr15:101420989	GCCTCCTCGCTCTGTGGGCTG GGAGGCTACAGCGTTGGCAGA ACGGCGAGATTACCTGCC[CG] GGGCAGAAAGGAATAAAAGAA GATGGTGAGGCCTCCGAGGCC TCTGAGCTGGGGCTGAGG

NOTE: For each marker the following information is provided: Illumina array ID (Ilmn ID), gene symbol of the corresponding gene (Corresponding gene), full name of the corresponding gene (Full name), mRNA ID of corresponding gene transcript based on RefSeq database (mRNA ID RefSeq), chromosomal position of the target cytosine based on hg19 (Cytosine position hg19) and the sequence surrounding of target cytosine (Surrounding sequence).

eTable 2. Clinical data for cohort 1.

Sample Name	Subtype IHC	iTU-Ly	str-Ly	Grade	Size cm	Nodal Status	ER IHC	HER2 IHC	Age Diagnosis	RFS event	RFS time	RFS event censored	RFS time censored	OS event	OS time
BC102	HER2	5	10	3	5	1	0	1	52,94	0	13,42	0	10	0	13,42
BC11	Basal	1	5	3	2,2	0	0	0	52,64	0	8,32	0	8,32	0	8,32
BC116	Basal	0	0	3	3,7	0	0	0	57,32	0	9,78	0	9,78	0	9,78
BC124	Basal	10	40	3	1,7	0	0	0	53,7	0	6,19	0	6,19	0	6,19
BC18	Basal	5	60	3	2.8+1.5	0	0	0	59,56	0	9,45	0	9,45	0	9,45
BC24	LumB	50	20	3	3,4	0	1	0	53,78	0	8,84	0	8,84	0	8,84
BC28	Basal	40	20	3	2,9	0	0	0	49,61	1	0,51	1	0,51	1	4,99
BC34	Basal	2	5	3	1,5	0	0	0	37,66	1	3,34	1	3,34	0	7,37
BC35	Basal	90	10	3	2,5	1	0	0	54,16	1	0	1	0	0	7,81
BC37	Basal	70	0	3	2,8	0	0	0	51,95	0	7,84	0	7,84	0	7,84
BC45	Basal	10	20	3	0,9	0	0	0	54,86	0	7,16	0	7,16	0	7,16
BC49	LumB	0	0	3	1,6	0	1	0	60,46	0	6,71	0	6,71	0	6,71
BC51	Basal	30	10	3	2	0	0	0	73,69	1	5,01	1	5,01	0	6,55
BC56	Basal	5	50	3	3,9	1	0	0	36,17	0	6,2	0	6,2	0	6,2
BC57	Basal	0	0	3	1,8	0	0	0	48,7	0	6,21	0	6,21	0	6,21
BC6	HER2	NA	NA	3	1,5	0	0	1	35,86	0	7,21	0	7,21	0	7,21

BC62	HER2	0	0	3	2,7	0	1	1	33,44	1	2,63	1	2,63	1	6,45
BC67	Basal	50	80	2	1,4	1	0	0	32,16	0	14,08	0	10	0	14,08
BC7	Basal	40	80	3	2,2	0	0	0	65,42	1	7,01	1	7,01	0	8,19
BC73	Basal	10	30	3	1,7	1	0	0	75,13	0	5,74	0	5,74	0	5,74
BC80	Basal	30	90	3	1,1	0	0	0	53,98	1	2,05	1	2,05	1	5,33
BC95	Basal	50	80	3	2,5	1	0	0	56,55	1	1,48	1	1,48	1	1,92
BC1	HER2	0	10	3	2,2	1	0	1	74,94	1	0,93	1	0,93	0	8,28
BC10	HER2	5	20	3	5,1	1	0	1	50,12	0	2,12	0	2,12	1	2,52
BC108	HER2	0	30	3	2,6	1	0	1	55,22	1	0,74	1	0,74	1	0,87
BC112	HER2	NA	NA	3	6	0	0	1	47,82	1	10,22	0	10	0	10,82
BC113	HER2	0	5	3	3	1	0	1	64,83	1	3,94	1	3,94	0	1,21
BC121	HER2	0	10	3	>4.5	1	1	1	52,35	1	2,65	1	2,65	0	5,58
BC122	HER2	0	10	3	3	0	0	1	69,25	1	0,64	1	0,64	1	3,55
BC2	HER2	0	0	3	4,8	1	0	1	59,79	1	1,22	1	1,22	1	3,39
BC27	HER2	20	30	3	2,1	1	1	1	53,66	0	1,85	0	1,85	0	1,85
BC3	HER2	0	70	3	1,8	1	0	1	52,58	0	8,64	0	8,64	0	8,64
BC39	HER2	20	10	3	10	1	1	1	65,32	1	3,38	1	3,38	1	5,28
BC42	HER2	0	0	3	3,5	1	0	1	47,75	1	0	1	0	0	7,63
BC50	HER2	NA	NA	3	2,5	0	1	1	70,34	0	6,36	0	6,36	0	6,36
BC66	HER2	0	0	3	2,1	1	1	1	49,18	0	0,56	0	0,56	1	1,97

BC68	HER2	NA	NA	3	1,6	0	1	1	66,72	0	6,17	0	6,17	0	6,17
BC78	HER2	50	30	3	1,6	1	0	1	50,29	1	1,08	1	1,08	1	1,58
BC79	HER2	10	90	3	2,2	1	0	1	45,98	0	5,43	0	5,43	0	5,43
BC85	HER2	50	80	2	2,5	1	NA	1	63,61	0	13,96	0	10	0	13,96
BC86	HER2	0	10	3	3	0	0	1	46,95	0	13,39	0	10	0	13,39
BC92	HER2	0	0	2	3	1	0	1	76,43	1	2,09	1	2,09	1	5,89
BC94	HER2	0	5	3	3,5	1	0	1	43,96	1	8,67	1	8,67	1	11,32
BC101	LumA	0	0	1	1,5	0	1	0	67,57	0	7,9	0	7,9	0	7,9
BC103	Basal	0	5	2	3	1	0	0	46,83	0	13,27	0	10	0	13,27
BC106	LumA	0	0	1	2,4	0	1	0	51,81	0	5,36	0	5,36	0	5,36
BC107	LumB	0	0	3	>2	0	1	0	53,8	1	9,19	1	9,19	0	12,41
BC109	LumA	0	0	1	3	1	1	0	67,77	1	6,48	1	6,48	1	7,92
BC120	LumA	0	0	1	3	1	1	0	57,7	0	9,99	0	9,99	0	9,99
BC123	LumA	0	0	1	1,3	0	1	0	66,14	1	3,75	1	3,75	1	5,18
BC125	LumA	0	0	1	2,3	0	1	0	67,41	0	6,11	0	6,11	0	6,11
BC126	LumA	0	20	1	3,2	1	1	0	71,68	1	0,01	1	0,01	0	6,1
BC19	LumB	0	0	3	2.6+0.6	1	1	0	59,1	1	4,28	1	4,28	0	9,32
BC20	LumA	0	0	1	2,7	1	1	0	68,03	0	6,07	0	6,07	0	6,07
BC25	LumB	5	10	3	1,8	0	1	0	49,04	0	8,03	0	8,03	0	8,03
BC32	LumB	5	10	3	2,7	0	1	0	51,61	0	4,48	0	4,48	0	4,48

BC55	HER2	0	0	3	3	0	0	1	34,14	1	3,37	1	3,37	1	4,98
BC58	Basal	5	20	3	3,5	0	0	0	68,48	1	1,47	1	1,47	1	5,56
BC60	LumA	0	0	1	1,7	1	1	0	48,64	0	6,83	0	6,83	0	6,83
BC61	LumB	0	0	3	2,5	0	1	0	61,7	1	4,58	1	4,58	0	7,08
BC65	Basal	NA	NA	3	1.2-1.3	0	0	0	54,01	0	6,22	0	6,22	0	6,22
BC69	LumB	0	10	3	5,5	1	1	0	78,29	1	0,08	1	0,08	1	2,97
BC84	Basal	0	0	2	1,3	0	0	0	36,3	1	2,42	1	2,42	0	7,08
BC96	LumA	0	0	1	1,5	0	1	0	47,88	0	6,38	0	6,38	0	6,38
BC97	LumA	0	0	1	1,6	0	1	0	52,56	0	5,96	0	5,96	0	5,96
BC99	LumA	0	0	1	1,5	0	1	0	59,36	0	13,45	0	10	0	13,45
BC105	Basal	0	10	3	1,3	1	0	0	41,54	0	8,74	0	8,74	0	8,74
BC115	Basal	0	0	3	2,5	0	0	0	59,08	1	2,56	1	2,56	1	4,89
BC119	Basal	0	0	3	3,5	0	0	0	62,62	1	4,25	1	4,25	1	4,69
BC12	LumB	0	5	3	3,8	1	1	0	46,58	0	8,82	0	8,82	0	8,82
BC13	LumB	5	40	3	1,3	0	1	0	63,54	0	9,03	0	9,03	0	9,03
BC15	LumB	5	10	3	1,7	0	1	0	45,07	0	8,98	0	8,98	0	8,98
BC16	LumB	5	90	3	1.7+0.7 BIFOCAL	1	1	0	60,57	0	7,61	0	7,61	0	7,61
BC23	Basal	0	0	3	1,2	0	0	0	74,24	1	2,56	1	2,56	1	4,07
BC30	LumB	0	10	3	2,1	1	1	0	51,36	0	5,72	0	5,72	0	5,72

BC31	LumB	0	0	3	1,2	0	1	0	60,34	0	7,44	0	7,44	0	7,44
BC38	LumB	NA	NA	3	2,5	1	1	0	77,17	1	2,05	1	2,05	1	3,77
BC4	LumB	0	5	3	4,5	1	1	0	44,06	0	8,4	0	8,4	0	8,4
BC40	LumB	0	0	3	2,65	0	1	0	51,96	0	8,39	0	8,39	0	8,39
BC43	LumB	0	0	3	5,5	1	1	0	82,13	0	1,4	0	1,4	1	1,52
BC48	LumB	20	60	3	9,5	1	1	0	45,42	1	3,9	1	3,9	0	7,23
BC70	LumB	0	0	3	2,2	0	0	0	39,78	1	7,46	1	7,46	0	12,8
BC71	LumA	0	0	1	1,6	0	1	0	42,07	0	5,95	0	5,95	0	5,95
BC72	LumB	0	5	3	2,1	1	1	0	53,82	1	1,69	1	1,69	1	4,58
BC74	Basal	NA	NA	3	2,3	1	0	0	79,86	1	4,43	1	4,43	0	4,52
BC75	HER2	0	0	3	5,5	1	1	1	48,34	0	6,12	0	6,12	0	6,12
BC88	LumB	0	0	3	3	1	1	0	64,86	1	5,92	1	5,92	0	7,93
BC90	Basal	0	0	3	2,4	0	0	0	54,19	0	3,82	0	3,82	1	7,42
BC104	LumA	0	0	1	2,7	0	1	0	66,18	0	7,98	0	7,98	0	7,98
BC110	LumB	NA	NA	2	1.1+2.1	1	1	0	39,69	0	5,19	0	5,19	0	5,19
BC117	Basal	0	0	3	3,2	1	0	0	41,86	1	1,89	1	1,89	1	3,26
BC118	HER2	0	0	3	2,8	1	0	1	66,04	1	1,17	1	1,17	1	2,77
BC129	LumB	NA	NA	3	1,9	0	1	0	58,49	1	5,06	1	5,06	0	11,44
BC14	Basal	NA	NA	2	3,2	0	0	0	75,86	1	2,85	1	2,85	1	3,43
BC17	LumB	0	0	3	3,8	1	1	0	57,72	1	4,93	1	4,93	1	8,9

BC21	LumB	0	0	3	3,4	1	1	0	52,76	1	8,19	1	8,19	0	9,3
BC22	LumA	0	0	1	1,3	1	1	0	69,83	0	9,68	0	9,68	0	9,68
BC26	LumA	0	5	1	3,2	0	1	0	40,59	0	7,63	0	7,63	0	7,63
BC29	LumA	0	0	1	1,8	1	1	0	57,08	0	8,01	0	8,01	0	8,01
BC33	LumB	0	0	3	3,2	1	1	0	51,87	1	0,13	1	0,13	0	7,64
BC36	LumB	0	0	3	3,5	1	1	0	41,92	0	1,44	0	1,44	1	1,5
BC41	Basal	5	0	3	3	1	0	0	59,38	0	6,07	0	6,07	0	6,07
BC44	LumA	0	0	1	2,3	0	1	0	71,51	0	6,82	0	6,82	0	6,82
BC46	LumB	0	0	3	2,1	0	1	0	39,36	0	7,32	0	7,32	0	7,32
BC47	LumA	NA	NA	1	2,3	0	1	0	62,23	0	7,11	0	7,11	0	7,11
BC5	LumB	0	0	3	1,7	0	1	0	77,34	0	6,53	0	6,53	0	6,53
BC52	LumA	0	20	1	1,8	0	1	0	51,82	0	6,78	0	6,78	0	6,78
BC59	LumB	10	10	3	1,5	0	1	0	35,85	1	1,31	1	1,31	0	6,58
BC63	LumA	0	0	1	1	0	1	0	60,2	0	5,99	0	5,99	0	5,99
BC64	HER2	0	10	2	1,1	0	1	1	44,34	0	6,09	0	6,09	0	6,09
BC77	LumA	NA	NA	1	1,3	0	1	0	61,52	0	5,84	0	5,84	0	5,84
BC8	LumA	0	10	1	1,3	0	1	0	36,2	0	8,79	0	8,79	0	8,79
BC81	HER2	10	20	3	1,1	0	0	1	61	0	4,95	0	4,95	0	4,95
BC83	HER2	NA	NA	2	1,1	0	0	1	61,86	0	6,06	0	6,06	0	6,06
BC87	Basal	0	10	3	1,4	0	0	0	58,03	1	0,67	1	0,67	1	1,58

BC89	LumA	0	5	1	2	1	1	0	48,48	1	9,01	1	9,01	0	12,83
BC9	HER2	0	2	3	2,7	1	0	1	28,64	0	9,02	0	9,02	0	9,02
BC91	LumA	0	10	1	1,8	1	1	0	49,26	1	10,16	0	10	0	13,33
BC93	LumB	0	5	3	2,1	0	0	0	65,53	1	4,18	1	4,18	1	4,19

eTable 3. Histopathologic TIL evaluation methods.

Cohort	Histopathologic sample	Method	Reference
PNC1	Hematoxylin and eosin-stained full sections	Lymphocytic inilitration (intra-tumoral) was defined as the percentage of mononuclear cells within the epithelium of the invasive tumor cell nests. Histopathologic evaluation of TILs was performed independently by two well-trained pathologists; the mean value was used for the analyses presented.	Dedeurwaerder et al., EMBO Mol Med, 2011[1]; Denkert et al., J Clin Oncol, 2010[2]
	Immunohistochemistry (IHC)-stained full sections	Immunohistochemical staining of CD45 (Dako Denmark A/S, Glostrup, Denmark), CD3 (Dako Denmark A/S, Glostrup, Denmark) and CD20 (Dako Denmark A/S, Glostrup, Denmark) was performed according to standard procedures. TILs were scored as continuous variable on one full section per case and were reported as the percentage of tumor area infiltrated by lymphocytes. Histopathologic analysis of the TILs was performed independently by two well-trained pathologists.	Denkert et al., J Clin Oncol, 2010[2]
PNC2	Hematoxylin and eosin-stained full sections	Lymphocytic inilitration (intra-tumoral) was defined as the percentage of mononuclear cells within the epithelium of the invasive tumor cell nests. Histopathologic evaluation of TILs was performed independently by two well-trained pathologists; the mean value was used for the analyses presented.	Dedeurwaerder et al., EMBO Mol Med, 2011[1]; Denkert et al., J Clin Oncol, 2010[2]
TCGA	Hematoxylin and eosin-stained sections	A minimum of ten specimen fields were assessed to evaluate the percent of nuclei and necrosis. Then the total percent of the different cell types was evaluated. (Necrotic cells were not part of this calculation)	http://www.nationwidechildrens.org/tss-training-webinar[3]
TOP	---	---	---

NOTE: Methods for morphological TILs evaluation are described in detail in the following references: [1]Dedeurwaerder *et al*. (2011) DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol Med.* 3(12), 726-41. [2]Denkert *et al*. (2010) Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol*. 28(1), 105-13. [3]http://www.nationwidechildrens.org/tss-training-webinar.

eTable 4. Probe information for MeTIL signature markers.

Ilmn ID	Corresponding gene	Full name	mRNA ID RefSeq	Cytosine position hg19	Surrounding sequence
cg20792833	PTPRCAP	Homo sapiens protein tyrosine phosphatase, receptor type, C-associated protein	NM_005608	chr11:67205195	GCTGTGTCGAGCGAGAAGTGAGCT CAGTGCTCGTCTGCAGTGAAGGGT GGCCCAGGCTTC[CG]CTTCCTGCCC ACATACCCCACCTGCCCCTCCCTGC TGCAGGACCCCTGGTCCACACCAG A
cg20425130	KLHL6	Homo sapiens kelch-like family member 6	NM_130446	chr3:183273245	TGGCAGGAGAATTCCTGAATGTCCA CACACAAGATGACATCTGTCAGAGC GTTTTCCATT[CG]CAGGGTTTCCAG GCCATTCTGAAGAATTAAGGAGAGT CCCGCGTCGTCAAATTTGACCTT
cg23642747	INA	Homo sapiens internexin neuronal intermediate filament protein, alpha	NM_032727	chr10:105036645	GGGATGCGACAGAGCTGTGTGGTT TCCGGATGGGAAACCTCAGTCGTTT AGGCACCCCTC[CG]CTCGAGTCACT TCCGAAGCAGTCGATTCTTGGGGAG AAGCGCTGCGGAAAGGGGCGACTC
cg12069309	SEMA3B	Homo sapiens sema domain, immunoglobulin domain, secreted, (semaphorin) 3B domain (Ig), short basic	NM_001005914; NM_004636	chr3:50312913	GGGGTCGCGTGCACGCGCTTCCAG CCCAGTGCCAAGAGGTGGGCGGGG TCGGGGTTGGGC[CG]CCGGGAGGG AGGCGAAGGGTCTTTCACTGCCCG GGGCTGAAAGAAGGGCTCACAGAA GAT
cg21554552	RASSF1	Homo sapiens Ras association (RalGDS/AF-6) domain family member 1	NM_007182; NM_170714	chr3:50378425	AGAGCCGCGCAATGGAAACCTGGG TGCAGGGACTGTGGGGCCCGAAGG CGGGGCTGGGCG[CG]CTCTCGCAG AGCCCCCCCCGCCTTGCCCTTCCTT CCCTCCTTCGTCCCCTCCTCACACC C

eTable 5. Clinical data for cohort 2.

Sample Name	Age diagnosis	tumor size cm	nodes pos	grade	er status	pgr status	HER2 status	e ddfs	t ddfs	e os	t os	iTU.Ly	str.Ly	Subtype IHC
2-1	74.22	1.3	0	1	1	1	0	0	3.44	0	3.44	1	0	LumA
2-10	47.36	1.3	0	1	1	1	0	0	4.82	0	4.82	5	0	LumA
2-100	47.72	1.8	0	3	0	0	0	0	6.03	0	6.03	30	5	NA
2-101	68.97	4.0	0	3	0	0	0	0	0.96	0	0.96	15	2	Basal
2-102	72.99	0.4	0	3	0	0	0	0	4.89	0	4.89	5	0	Basal
2-103	56.54	5.5	1	3	0	0	0	0	5.14	0	5.14	20	5	Basal
2-104	65.38	3.0	0	3	0	0	0	0	4.65	0	4.65	40	60	Basal
2-105	55.48	1.4	0	3	0	0	0	1	1.09	0	4.75	10	1	Basal
2-106	60.21	0.8	0	3	0	0	0	0	3.81	0	3.81	45	5	Basal
2-107	36.96	1.6	0	3	0	0	0	0	4.44	0	4.44	5	0	Basal
2-108	72.29	2.9	0	3	0	0	0	1	0.82	1	1.01	60	5	Basal
2-109	46.73	2.2	3	3	0	0	0	0	1.96	0	1.96	25	1	Basal
2-11	49.95	0.9	0	1	1	1	0	0	5.06	0	5.06	5	0	LumA
2-110	41.40	1.5	2	3	1	1	0	0	4.32	0	4.32	5	0	LumB
2-111	65.58	1.8	2	3	0	0	0	0	3.42	0	3.42	15	0	Basal
2-112	35.15	1.8	1	3	0	0	0	1	1.90	1	3.01	10	2	Basal

2-113	60.16	1.7	1	3	0	0	0	0	4.99	0	4.99	5	0	Basal
2-114	60.96	1.1	0	3	0	0	0	0	3.06	0	3.06	30	1	Basal
2-115	55.99	2.2	0	3	0	0	0	0	3.61	0	3.61	60	15	Basal
2-116	23.38	1.4	0	3	0	1	0	0	3.18	0	3.18	20	1	Basal
2-117	38.92	4.2	0	3	0	0	0	0	3.53	0	3.53	NA	NA	Basal
2-118	48.90	2.2	37	3	0	0	0	0	2.07	0	2.07	40	5	Basal
2-119	50.99	1.9	0	3	0	0	0	0	2.06	0	2.06	NA	NA	Basal
2-12	72.08	1.1	0	1	1	1	0	0	7.06	0	7.06	5	1	LumA
2-120	41.24	2.1	0	3	0	0	0	0	2.18	0	2.18	30	5	Basal
2-13	91.43	1.6	NA	1	1	1	0	0	1.34	1	1.34	10	2	LumA
2-14	76.10	1.1	0	1	1	1	0	0	7.88	0	7.88	5	2	LumA
2-15	47.80	0.7	0	1	1	1	0	0	0.03	0	0.03	25	10	LumA
2-16	62.25	1.1	0	1	1	0	0	1	4.17	0	7.76	1	0	LumA
2-17	67.06	9.0	0	1	1	0	0	0	3.40	0	3.40	5	0	LumA
2-18	50.92	1.4	0	1	1	1	0	0	5.73	0	5.73	NA	1	LumA
2-19	83.44	2.3	NK	1	1	0	0	0	5.39	0	5.39	NA	NA	LumA
2-2	42.22	1.4	0	1	1	1	0	0	4.13	0	4.13	NA	NA	LumA
2-20	84.33	1.7	0	1	1	1	0	0	3.17	0	3.17	5	0	LumA
2-21	43.68	1.9	1	1	1	1	0	0	6.41	0	6.41	5	0	LumA
2-22	64.61	1.0	0	1	1	1	0	0	6.98	0	6.98	1	0	LumA

2-23	62.32	1.9	0	1	1	1	0	0	7.20	0	7.20	5	0	LumA
2-24	58.56	1.5	0	1	1	1	0	0	6.55	0	6.55	5	0	NA
2-25	47.25	1.6	0	1	1	1	0	0	7.33	0	7.33	5	0	LumA
2-26	56.18	1.3	0	1	1	1	0	0	4.85	0	4.85	5	0	LumA
2-27	48.68	1.8	0	1	1	1	0	0	2.40	0	2.40	1	0	LumA
2-28	58.54	1.7	0	1	1	1	0	0	2.48	0	2.48	20	5	LumA
2-29	31.95	2.0	1	3	1	0	0	0	12.73	0	12.73	NA	NA	LumB
2-3	61.28	1.4	0	1	1	1	0	0	4.40	0	4.40	NA	NA	LumA
2-30	47.50	3.0	NA	3	1	0	0	1	10.59	0	13.48	5	0	LumB
2-31	30.19	2.0	0	3	1	1	0	0	2.96	0	2.96	NA	NA	LumB
2-32	49.26	1.9	0	3	1	1	0	0	1.28	0	1.28	15	2	LumB
2-33	80.58	0.8	0	3	1	1	0	0	2.10	0	2.10	1	0	LumB
2-34	56.70	2.3	0	3	1	0	0	0	3.61	0	3.61	NA	NA	LumB
2-35	61.53	2.0	0	3	1	1	0	0	3.32	0	3.32	NA	NA	LumB
2-36	70.62	1.9	3	3	1	1	0	0	3.06	0	3.06	10	0	LumB
2-37	75.32	4.0	0	3	1	0	0	1	0.47	0	0.62	NA	NA	LumB
2-38	56.33	2.7	14	3	1	0	0	1	0.38	1	1.40	5	0	LumB
2-39	72.34	1.7	0	3	1	0	0	0	0.36	0	0.36	10	2	LumB
2-4	50.59	1.8	0	1	1	1	0	0	4.42	0	4.42	1	0	LumA
2-40	47.30	1.7	9	3	1	1	0	0	5.83	0	5.83	5	0	LumB

2-41	58.44	1.8	6	3	1	1	0	0	5.74	0	5.74	15	2	LumB
2-42	31.25	2.0	0	3	1	1	0	0	5.63	0	5.63	10	1	LumB
2-43	64.83	21.0	1	3	1	0	0	0	5.87	0	5.87	35	5	LumB
2-44	74.93	1.6	0	3	1	1	0	0	5.08	0	5.08	10	1	LumB
2-45	64.96	1.6	0	3	1	0	0	0	4.20	0	4.20	15	2	LumB
2-46	58.18	2.5	0	3	1	1	0	0	5.33	0	5.33	1	0	LumB
2-47	59.75	1.3	0	3	1	0	0	0	4.93	0	4.93	2	0	LumB
2-48	61.90	1.5	2	3	1	0	0	0	5.16	0	5.16	30	5	LumB
2-49	71.43	3.8	6	3	1	1	0	0	5.50	0	5.50	15	2	LumB
2-5	49.51	1.3	2	1	1	1	0	0	5.01	0	5.01	1	0	LumA
2-50	82.52	2.3	1	3	1	1	0	0	4.22	0	4.22	5	0	LumB
2-51	44.47	NA		NA	1	1	0	0	1.04	0	1.04	5	0	LumB
2-52	35.81	3.0	1	3	1	1	0	0	2.82	0	2.82	NA	NA	LumB
2-53	67.95	2.1	11	3	1	1	0	0	1.13	0	1.13	5	0	LumB
2-54	76.12	1.9	0	3	1	1	0	0	7.24	0	7.24	35	5	LumB
2-55	51.38	1.5	1	3	1	1	0	0	6.64	0	6.64	5	0	LumB
2-56	48.23	1.5	0	3	1	1	0	0	7.55	0	7.55	15	5	LumB
2-57	74.48	2.9	0	3	1	0	0	1	3.38	1	7.39	5	0	LumB
2-58	70.49	1.6	1	3	1	1	0	0	8.05	0	8.05	2	0	LumB
2-59	73.53	3.5	1	3	0	0	1	0	4.48	0	4.48	5	0	HER2

2-6	51.97	1.6	1	1	1	1	0	0	5.62	0	5.62	30	1	LumA
2-60	58.77	1.8	0	3	0	0	1	0	4.09	0	4.09	20	5	HER2
2-61	31.44	2.1	7	3	0	0	1	0	4.58	0	4.58	40	5	HER2
2-62	76.64	3.3	2	3	0	0	1	0	4.62	0	4.62	20	2	HER2
2-63	60.48	2.1	1	3	0	0	1	0	4.16	0	4.16	60	2	HER2
2-64	40.94	6.0	0	3	0	0	1	0	4.69	0	4.69	2	0	HER2
2-65	46.85	3.0	10	2	0	0	1	1	3.35	1	4.12	15	2	HER2
2-66	57.32	5.5	3	3	0	0	1	0	5.20	0	5.20	50	5	HER2
2-67	66.50	2.2	2	3	1	1	1	0	3.68	0	3.68	NA	NA	HER2
2-68	40.67	4.0	2	3	0	1	1	0	2.99	0	2.99	20	2	HER2
2-69	39.36	1.7	6	3	1	1	1	0	2.18	0	2.18	50	5	HER2
2-7	47.78	1.1	0	1	1	1	0	0	4.78	0	4.78	1	0	LumA
2-70	60.75	1.4	0	3	1	1	1	0	0.51	0	0.51	5	5	HER2
2-71	62.77	3.5	1	3	1	0	1	0	5.27	0	5.27	35	5	HER2
2-72	48.72	1.7	1	3	1	1	1	0	2.84	0	2.84	5	0	HER2
2-73	35.05	11.0	16	3	0	0	1	1	2.59	1	4.38	2	0	HER2
2-74	77.95	1.2	0	3	1	0	1	0	5.00	0	5.00	10	0	HER2
2-75	62.16	1.1	3	3	1	1	1	0	3.47	0	3.47	15	0	HER2
2-76	59.42	1.3	0	2	0	0	1	0	3.40	0	3.40	5	0	HER2
2-77	55.42	2.0	0	2	0	0	1	1	21.18	1	27.33	2	0	HER2

2-79	71.38	4.0	0	3	0	0	1	0	3.80	0	3.80	10	15	HER2
2-8	42.04	1.4	0	1	1	1	0	0	5.12	0	5.12	NA	NA	LumA
2-80	48.91	1.9	0	3	0	0	1	0	4.12	0	4.12	30	2	HER2
2-81	60.19	1.7	2	2	1	0	1	1	1.95	0	3.31	NA	NA	HER2
2-82	79.85	1.5	1	3	1	1	1	0	0.94	0	0.94	5	0	HER2
2-83	37.99	14.0	5	3	1	0*	1	1	13.62	0	20.25	NA	NA	HER2
2-84	82.83	3.4	1	3	1	1	1	0	3.71	0	3.71	NA	NA	HER2
2-85	59.77	1.6	0	3	0	0	0	0	7.31	0	7.31	40	5	Basal
2-86	78.39	4.3	0	3	0	0	0	0	3.75	0	3.75	70	5	Basal
2-87	57.60	1.6	1	3	0	0	0	0	7.20	0	7.20	100	5	Basal
2-88	35.27	2.8	0	3	0	0	0	0	5.00	0	5.00	15	0	Basal
2-89	53.92	1.8	0	3	0	0	0	0	5.70	0	5.70	10	0	Basal
2-9	68.30	1.3	0	1	1	1	0	0	0.95	0	0.95	5	0	LumA
2-90	66.91	3.0	0	3	0	0	0	0	3.07	0	3.07	20	30	Basal
2-91	33.14	3.0	9	3	0	0	0	0	2.34	1	2.34	10	0	Basal
2-92	67.61	3.0	4	3	0	0	0	0	2.38	1	2.38	15	0	Basal
2-93	63.80	1.9	1	3	0	0	0	0	7.65	0	7.65	35	2	Basal
2-94	80.70	0.9	0	3	0	0	0	0	2.42	0	2.42	20	50	Basal
2-95	41.93	1.8	1	3	0	0	0	0	7.22	0	7.22	10	0	Basal
2-96	82.12	5.5	12	3	0	0	0	1	1.94	1	3.50	1	0	Basal

2-97	39.32	1.6	0	3	0	0	0	1	1.44	0	1.70	35	15	Basal
2-98	50.56	1.1	0	3	0	0	0	0	0.02	0	0.02	25	10	Basal
2-99	56.84	0.7	3	2	0	0	0	1	0.73	1	1.25	30	5	Basal

eTable 6. Clinical data for TOP trial.

Sample Name	pCR	Age Bin	T	N	Grade	HER2FISHbin	ESR1bimod	ERBB2bimod	DMFS event	DMFS time	OS event	OS time
1109	0	1	T2	N1	3	0	0	0	0	693	0	693
123	0	1	T1	N0	3	1	0	0	NA	NA	NA	NA
1312	0	0	T1	N0	3	1	0	1	0	860	0	860
1366	0	0	T2	N0	3	0	0	0	0	1404	0	1404
1391	0	1	T2	N0	3	1	0	1	0	1082	0	1082
1536	1	0	T2	N0	3	1	0	1	0	620	0	620
1674	0	0	T2	N1	3	NA	0	0	0	1852	0	1852
1767	1	0	T2	N0	2	1	0	0	0	2011	0	2011
2090	0	0	T4	N1	3	0	0	0	1	NA	1	NA
2165	0	0	T2	N1	3	1	0	1	0	991	0	991
220	0	1	T2	N1	3	0	0	0	0	1358	0	1358
2325	0	1	T4	N1	3	1	0	0	0	675	0	675
2337	0	1	T2	N1	3	0	0	0	0	710	0	710
2744	0	0	T1	N3	3	NA	0	1	0	2069	0	2069
3276	0	1	T1	N0	3	0	0	0	0	1041	0	1041
3414	0	1	T2	N1	3	1	0	1	1	254	0	2187
3731	0	0	T2	N0	3	1	0	1	0	845	0	845

4082	0	1	T4	N1	3	0	0	0	1	314	1	314
4270	0	1	T2	N1	3	NA	0	0	1	63	1	359
4291	0	1	T2	N0	3	0	0	0	1	1414	1	1932
4511	0	1	T3	N0	3	0	0	0	0	1839	0	1839
4864	0	1	T1	N0	3	0	0	0	1	627	0	1211
4917	0	1	T2	N1	3	0	0	0	0	771	0	771
4938	0	1	T2	N0	3	NA	0	0	0	449	0	449
4984	0	0	T2	N3	3	1	0	1	0	1856	0	1856
5180	0	1	T2	N0	3	0	0	0	1	223	1	431
5328	1	1	T2	N0	3	0	0	0	0	1132	0	1132
5437	0	0	T1	N0	3	0	0	0	0	669	0	669
5622	0	0	T2	N1	3	1	0	1	0	794	0	794
5706	0	1	T2	N0	2	1	0	0	1	1072	0	1524
5751	0	0	T2	N1	3	NA	0	1	0	919	0	919
5783	1	1	T2	N0	3	0	1	NA	NA	NA	NA	NA
5920	0	1	T2	N2	3	0	0	0	0	NA	0	NA
6181	0	0	T2	N1	3	0	0	0	0	777	0	777
6198	0	1	T2	N1	2	0	0	0	0	1259	0	1259
6276	0	0	T2	N0	3	0	0	0	0	748	0	748
6754	0	1	T2	N0	2	0	0	0	0	1376	0	1376

6767	0	0	T4	N1	3	0	0	0	1	464	1	464
682	0	1	T4	N2	2	0	0	0	1	732	1	732
6839	0	1	T1	N1	3	0	0	0	0	1390	0	1390
6951	0	1	T4	N1	2	0	0	0	0	725	0	725
7173	1	1	T2	N2	3	1	0	1	0	1130	0	1130
7182	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
7469	0	0	T3	N1	3	0	0	0	1	291	1	858
7500	0	0	T3	N1	3	NA	0	0	1	558	0	726
7634	1	0	T2	N1	3	0	0	0	0	1208	0	1208
8319	0	0	T4	N1	2	0	0	0	0	1772	0	1772
845	0	1	T2	N0	3	1	0	0	0	1647	0	1647
8660	0	0	T2	N0	3	0	0	0	0	NA	0	NA
8739	0	1	T2	N1	3	0	0	0	NA	NA	NA	NA
8999	0	0	T2	N0	3	NA	0	0	1	249	1	547
9049	0	0	T2	N1	1	0	0	0	0	599	0	599
9064	0	0	T2	N1	3	0	0	0	0	1408	0	1408
9171	0	0	T2	N1	3	1	0	1	0	1719	0	1719
9386	0	0	T2	N0	2	0	0	0	0	956	0	956
9434	0	1	T4	N1	3	NA	0	0	1	160	0	737
9576	1	0	T4	N1	3	0	0	0	0	1748	0	1748

9665	0	0	T2	N0	3	1	0	1	0	1008	0	1008

eTable 7. Baseline characteristics of breast cancer patient cohorts.

Characteristic	Cohort 1 (n = 118)	Cohort 2 (n = 119)	TOP (n = 58)
Age, mean	56	58	50
Age			
≤ mean	64	57	28
> mean	54	62	29
Missing	/	/	1
Tumor size			
T1-T2	/	/	45
T3-T4	/	/	12
Missing	118	119	1
≤ 2	44	76	/
> 2	74	42	/
Missing	/	1	58
Histologic grade			
G1-G2	34	33	9
G3	84	85	48
Missing	/	1	1
Nodal status			
N0	64	69	24
N1-N3	54	48	33
Missing	/	2	1
Ki67, %			
≤ 25	/	63	/
> 25	/	55	/
Missing	118	1	58
ER status by IHC			
Positive	64	70	1

Negative	53	49	56
Missing	1	/	1
HER2 status by IHC			
Positive	30	25	13
Negative	88	94	43
Missing	/	/	2

Abbreviations: HER2, human epidermal growth factor receptor 2 positive subtype; ER, estrogen receptor; PR, progesterone receptor; IHC, immunohistochemistry.

eTable 8. Correlation between the MeTIL score or PaTIL and median survival of breast cancer patients by breast cancer subtype in various cohorts (univariate COX proportional hazards regression).

			MeTIL Score			PaTIL		
Subtype	**Cohort**	**Patients(Events)**	**HR**	**95 % CI**	***P***	**HR**	**95% CI**	***P***
TN	Cohort 1	28(14)	1.05	0.78 – 1.41	0.757	1.01	0.99 – 1.04	0.213
	Cohort 2	32(6)	**0.42**	**0.20 – 0.91**	**0.027**	0.92	0.75 – 1.14	0.463
	TOP	38(8)	**0.59**	**0.39 – 0.89**	**0.012**	/	/	/
LUM	Cohort 1	52(18)	**0.65**	**0.43 – 0.98**	**0.041**	1	0.93 – 1.06	0.894
	Cohort 2	47(3)	**0.25**	**0.06 – 0.96**	**0.043**	0.87	0.39 – 1.94	0.731
HER2	Cohort 1	25(13)	0.88	0.65 – 1.19	0.395	0.99	0.94 – 1.04	0.607
	Cohort 2	21(3)	0.61	0.23 – 1.63	0.327	0.79	0.34 – 1.79	0.565

NOTE: Breast cancer subtypes were defined based on status of ER, PR and HER2 evaluated with IHC or FISH, respectively.
Abbreviations: TILs, tumor-infiltrating lymphocytes; PaTIL, pathological assessment of TILs on H&E stained tumor sections; HR, hazard ratio; CI, confidence interval; P, p value; TN, triple negative; LUM, luminal subtype; HER2, human epidermal growth factor receptor 2 positive subtype; ER, estrogen receptor; PR, progesterone receptor; IHC, immunohistochemistry; FISH, fluorescent in situ hybridization.

eTable 9. Correlation between the MeTIL score and other clinical and pathological variables and response to neoadjuvant anthracycline therapy in TOP cohort (logit regression).

Characteristic	OR	95% CI	*P*
Age (≤ 50 vs > 50)	0,34	0.03 - 2.84	0,329
Tumor size (T1-T2 vs T3-T4)	0,88	0.03 - 15.3	0,932
Nodal status (N0 vs others)	0,70	0.06 - 83.7	0,762
Grade (G1 – G2 vs G3)	2,54	0.19 - 91.5	0,523
HER2 (negative vs positive*)	12,90	0.93 - 382	0,083
MeTIL score**	**4,38**	**1.62 - 21.5**	**0,020**

NOTE: *The HER2 status was defined based on FISH (fluorescent in situ hybridization) results. **The MeTIL score was treated as continuous predictor variable. Abbreviations: OR, odds ratio; CI, confidence interval; P, p value; HER2, human epidermal growth factor receptor 2.

eTable 10. Correlation between the MeTIL score or gene expression-based immune markers and response to neoadjuvant anthracycline therapy in TOP cohort (logit regression).

Characteristic	OR	95% CI	*P*
MeTIL score	**4,38**	**1.62 – 21.5**	**0,020**
STAT1 score	7,52	1.21 – 115	0,067
CD3D	3,54	1.13 – 18.8	0,063
CXCL9	**3,15**	**1.28 – 13.9**	**0,046**
CD247	3,83	1.11 – 18.3	0,051

NOTE: All markers were treated as continuous predictor variables. Abbreviations: OR, odds ratio; CI, confidence interval; P, p value.

eTable 11. Baseline characteristics of patients in different cancer cohorts from TCGA.

Cancer type	Characteristic	Patients (Events)
BLCA	years_to_birth	
	≤ 67	191 (64)
	> 67	204 (93)
	neoplasm_diseasestage	
	S0, SI, SII	136 (31)
	SIII, SIV	259 (126)
BRCA	years_to_birth	
	≤ 58	529 (56)
	> 58	505 (72)
	neoplasm_diseasestage	
	S0, SI, SII	780 (81)
	SIII, SIV	254 (47)
	pathology_N_stage	
	N0	506 (48)
	NI, NII, NIII	528 (80)
	radiations_radiation_regimenindication	
	No	292 (15)
	Yes	742 (113)
CESC	weight_kg_at_diagnosis	
	≤ 73.26	183 (45)
	> 73.26	121 (24)
COREAD	years_to_birth	
	≤ 64	79 (11)
	> 64	104 (28)
	neoplasm_diseasestage	
	S0, SI, SII	98 (15)
	SIII, SIV	85 (24)
	pathology_T_stage	

		T0, TI, TII	28 (1)
		TIII, TIV	155 (38)
	pathology_N_stage		
		N0	100 (13)
		NI, NII, NIII	83 (26)
ESCA	pathology_N_stage		
		N0	88 (28)
		NI, NII, NIII	84 (42)
	gender		
		Female	24 (6)
		Male	148 (64)
HNSC	years_to_birth		
		≤ 61	163 (46)
		> 61	148 (67)
	radiations_radiation_regimenindication		
		No	15 (3)
		Yes	296 (110)
KIRP	neoplasm_diseasestage		
		S0, SI	137 (6)
		SII, SIII, SIV	51 (14)
LGG	years_to_birth		
		≤ 43	136 (12)
		> 43	116 (21)
	tumor_grade		
		G2	126 (12)
		G3	126 (21)
LIHC	pathology_T_stage		
		T0, TI	182 (48)
		TII, TIII, TIV	175 (59)
	ethnicity		
		Caucassian	157 (66)

		Not-Caucassian	200 (41)
LUAD	neoplasm_diseasestage		
		S0, SI	151 (34)
		SII, SIII, SIV	111 (42)
LUSC	pathology_T_stage		
		T0, TI, TII	198 (71)
		TIII, TIV	46 (22)
OV	primary_site_of_disease		
		Ovary	103 (49)
		Peritoneum	2 (2)
PAAD	years_to_birth		
		≤ 64	67 (28)
		> 64	76 (41)
	pathology_T_stage		
		T0, TI, TII	28 (7)
		TIII, TIV	115 (62)
	number_of_lymph_nodes		
		0	36 (11)
		1 or +	107 (58)
	completeness_of_resection		
		r0	96 (42)
		r1, r2	47 (27)
	neoplasm_diseasestage		
		S0, SI	19 (4)
		SII, SIII, SIV	124 (65)
	gender		
		Female	65 (35)
		Male	78 (34)
PCPG	gender		
		Female	101 (2)
		Male	78 (4)

PRAD	psa_value		
		≤ 1.16	377 (4)
		> 1.16	38 (3)
	days_to_psa		
		≤ 714	263 (3)
		> 714	152 (4)
SARC	years_to_birth		
		≤ 61	125 (36)
		> 61	124 (47)
SKCM	years_to_birth		
		≤ 56	175 (85)
		> 56	175 (96)
	pathology_M_stage		
		0	330 (173)
		1	20 (8)
	days_to_submitted_specimen_dx		
		≤ 1308	232 (117)
		> 1308	118 (64)
	melanoma_primary_known		
		No	44 (15)
		Yes	306 (166)
STAD	years_to_birth		
		≤ 65	114 (34)
		> 65	142 (62)
	pathology_M_stage		
		0	240 (87)
		1	16 (9)
	number_of_lymph_nodes		
		0	65 (15)
		1 or +	191 (81)
	completeness_of_resection		

	r0	238 (81)
	r1, r2	18 (15)
	pathology_N_stage	
	N0, NI	140 (44)
	NII, NIII	116 (52)
TGCT	ethnicity	
	Caucassian	112 (3)
	Not-Caucassian	21 (1)
THCA	neoplasm_diseasestage	
	S0, SI	284 (1)
	SII, SIII, SIV	216 (13)
	multifocality	
	Multifocal	226 (1)
	Unifocal	274 (13)
THYM	gender	
	Female	59 (4)
	Male	62 (2)
	years_to_birth	
	≤ 58	57 (1)
	> 58	64 (5)
UCEC	years_to_birth	
	≤ 63	262 (28)
	> 63	277 (41)
	radiations_radiation_regimenindication	
	No	145 (12)
	Yes	394 (57)

Abbreviations: BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; COREAD, colon and rectum adenocarcinoma; ESCA, esophageal carcinoma; HNSC, head and neck squamous cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma.

eTable 12. Correlation between MeTIL score or PaTIL and median survival of patients in various cancer cohorts (univariate COX proportional hazards regression).

		MeTIL score			PaTIL		
Cancer Type	**Patients(Events)**	**HR**	**95% CI**	**P**	**HR**	**95% CI**	**P**
BLCA	395(157)	0.95	0.85 - 1.05	0.284	1	0.98 - 1.02	0.687
BRCA	1034(128)	0.93	0.84 - 1.03	0.171	1	0.98 - 1.01	0.567
CESC	304(69)	0.93	0.81 - 1.08	0.359	1.01	1 - 1.02	0.245
COREAD	183(39)	0.79	0.59 - 1.05	0.108	0.96	0.9 - 1.03	0.228
ESCA	172(70)	0.98	0.85 - 1.14	0.793	1.01	0.99 - 1.02	0.217
HNSC	**311(113)**	**0.85**	**0.74 - 0.98**	**0.025**	0.99	0.97 - 1.01	0.248
KIRP	188(20)	0.89	0.64 - 1.24	0.486	0.95	0.79 - 1.13	0.536
LGG	252(33)	1.19	0.95 - 1.49	0.136	0.32	0.02 - 6.74	0.468
LIHC	357(107)	0.93	0.83 - 1.04	0.18	1	0.96 - 1.04	0.9
LUAD	262(76)	0.91	0.78 - 1.08	0.279	0.99	0.97 - 1.01	0.374
LUSC	244(93)	0.86	0.74 - 1.01	0.0733	1.01	0.99 - 1.03	0.248
OV	105(51)	1.13	0.86 - 1.47	0.386	1	0.99 - 1	0.346
PAAD	143(69)	0.97	0.84 - 1.11	0.628	1	0.99 - 1.01	0.746
PCPG	**179(6)**	**0.46**	**0.24 - 0.88**	**0.0195**	1.03	0.99 - 1.08	0.146
PRAD	415(7)	1.18	0.69 - 2.03	0.544	0.94	0.8 - 1.11	0.457
SARC	249(83)	0.94	0.84 - 1.06	0.335	0.86	0.69 - 1.07	0.181
SKCM	**350(181)**	**0.91**	**0.85 - 0.98**	**0.0103**	0.96	0.92 - 1.01	0.105
STAD	256(96)	1.05	0.94 - 1.18	0.381	0.98	0.97 - 1	0.103
TGCT	133(4)	1.17	0.63 - 2.19	0.614	1.02	0.98 - 1.05	0.395
THCA	**500(14)**	**0.6**	**0.36 - 0.98**	**0.043**	0.47	0.18 - 1.19	0.111
THYM	**121(6)**	**0.66**	**0.44 - 0.99**	**0.0425**	0.69	0.3 - 1.56	0.372
UCEC	539(69)	0.91	0.76 - 1.1	0.332	1	0.99 - 1.02	0.778

Abbreviations: TILs, tumor-infiltrating lymphocytes; PaTIL, pathological assessment of TILs on H&E stained tumor sections; HR, hazard ratio; CI, confidence interval; P, p value; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; COREAD, colon and rectum adenocarcinoma; ESCA, esophageal carcinoma; HNSC, head and neck squamous cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma;

STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma

eTable 13. Correlation between the MeTIL score or PaTIL and median survival of cancer patients in context of other prognostic clinical and pathological variables in various cancer cohorts (multivariate COX proportional hazards regression).

		MeTIL score				PaTIL			
Cancer Type	**Patients (Events)**	**Variable**	**HR**	**95% CI**	**P**	**Variable**	**HR**	**95% CI**	**P**
BLCA	395(157)	MeTIL	0.92	0.83 - 1.02	0.13	PaTIL	1	0.97 - 1.02	0.68
		years_to_birth	1.45	1.06 - 2	0.0219	years_to_birth	1.44	1.05 - 1.98	0.0242
		neoplasm_diseasestage	2.39	1.61 - 3.54	1.58e-05	neoplasm_diseasestage	2.33	1.57 - 3.46	2.45e-05
BRCA	1034(128)	MeTIL	0.99	0.89 - 1.1	0.81	PaTIL	1	0.99 - 1.01	0.914
		years_to_birth	1.7	1.18 - 2.45	0.0041	years_to_birth	1.72	1.21 - 2.45	0.00253
		neoplasm_diseasestage	1.65	1.05 - 2.57	0.0285	neoplasm_diseasestage	1.64	1.05 - 2.56	0.0304
		pathology_N_stage	1.47	0.94 - 2.29	0.0883	pathology_N_stage	1.47	0.95 - 2.3	0.0857
		radiations_radiation_regimenindication	3.28	1.9 - 5.65	2,00E-05	radiations_radiation_regimenindication	3.28	1.9 - 5.66	1.89e-05
CESC	304(69)	MeTIL	0.93	0.8 - 1.07	0.322	PaTIL	1.01	1 - 1.02	0.259
		weight_kg_at_diagnosis	0.74	0.45 - 1.21	0.23	weight_kg_at_diagnosis	0.75	0.46 - 1.24	0.266
COREAD	183(39)	MeTIL	0.85	0.62 - 1.16	0.301	PaTIL	0.97	0.91 - 1.04	0.445
		years_to_birth	2.09	1.02 - 4.25	0.0427	years_to_birth	2.1	1.03 - 4.28	0.0423
		neoplasm_diseasestage	0.26	0.06 - 1.07	0.0616	neoplasm_diseasestage	0.23	0.06 - 0.96	0.0439
		pathology_T_stage	4.43	0.59 - 33.12	0.147	pathology_T_stage	3.99	0.54 - 29.63	0.177
		pathology_N_stage	10.61	2.38 - 47.31	0.00195	pathology_N_stage	12	2.72 - 52.86	0.00102

ESCA	172(70)	MeTIL	1.01	0.86 - 1.17	0.942	PaTIL	1.01	0.99 - 1.02	0.285
		pathology_N_stage	1.77	1.08 - 2.91	0.0232	pathology_N_stage	1.76	1.07 - 2.88	0.0253
		gender	1.86	0.79 - 4.34	0.154	gender	1.81	0.78 - 4.23	0.17
HNSC	**311(113)**	**MeTIL**	**0.84**	**0.73 - 0.97**	**0.0152**	PaTIL	0.99	0.97 - 1.01	0.228
		years_to_birth	1.54	1.05 - 2.25	0.0268	years_to_birth	1.51	1.03 - 2.2	0.0346
		radiations_radiation_regimenindication	2.26	0.72 - 7.13	0.165	radiations_radiation_regimenindication	2.14	0.68 - 6.77	0.194
KIRP	188(20)	MeTIL	0.93	0.67 - 1.29	0.67	PaTIL	0.98	0.84 - 1.15	0.835
		neoplasm_diseasestage	5.23	1.97 - 13.93	0.000923	neoplasm_diseasestage	5.22	1.94 - 14.01	0.00104
LGG	252(33)	MeTIL	1.15	0.92 - 1.44	0.207	PaTIL	0.48	0.05 - 4.48	0.517
		years_to_birth	2.27	1.06 - 4.84	0.0345	years_to_birth	2.41	1.13 - 5.15	0.0228
		tumor_grade	2.3	1.08 - 4.91	0.0311	tumor_grade	2.18	1.03 - 4.65	0.0428
LIHC	357(107)	MeTIL	0.93	0.84 - 1.04	0.231	PaTIL	1	0.96 - 1.03	0.865
		pathology_T_stage	1.47	1 - 2.18	0.0507	pathology_T_stage	1.52	1.03 - 2.24	0.0328
		ethnicity	0.63	0.42 - 0.93	0.0214	ethnicity	0.63	0.43 - 0.94	0.0246
LUAD	262(76)	MeTIL	0.91	0.78 - 1.07	0.255	PaTIL	0.99	0.97 - 1.02	0.508
		neoplasm_diseasestage	2.12	1.34 - 3.34	0.00124	neoplasm_diseasestage	2.08	1.32 - 3.28	0.00166
LUSC	**244(93)**	**MeTIL**	**0.85**	**0.72 - 0.99**	**0.0416**	PaTIL	1.01	0.99 - 1.03	0.262
		pathology_T_stage	1.6	0.97 - 2.63	0.0647	pathology_T_stage	1.46	0.89 - 2.39	0.134
OV	105(51)	MeTIL	1.12	0.85 - 1.46	0.418	PaTIL	1	0.99 - 1.01	0.377
		primary_site_of_disease	1.46	0.35 - 6.17	0.605	primary_site_of_disease	1.44	0.34 - 6.07	0.622

PAAD	143(69)	MeTIL	0.91	0.79 - 1.06	0.227	PaTIL	1	0.99 - 1.01	0.745
		years_to_birth	1.68	1.03 - 2.75	0.0389	years_to_birth	1.83	1.11 - 3.04	0.0187
		pathology_T_stage	2.53	1.1 - 5.79	0.0281	neoplasm_diseasestage	3.75	1.32 - 10.66	0.013
		number_of_lymph_nodes	1.95	0.99 - 3.84	0.0543	gender	0.64	0.39 - 1.05	0.0775
		completeness_of_resection	1.5	0.9 - 2.48	0.116	completeness_of_resection	1.73	1.05 - 2.86	0.0327
PCPG	**179(6)**	**MeTIL**	**0.35**	**0.16 - 0.79**	**0.0108**	PaTIL	1.03	0.99 - 1.08	0.131
		gender	5.56	0.76 - 40.8	0.0917	gender	3.28	0.54 - 19.85	0.195
PRAD	415(7)	MeTIL	1.35	0.73 - 2.5	0.344	PaTIL	0.95	0.82 - 1.1	0.462
		psa_value	15.04	2.84 - 79.48	0.00142	psa_value	12.02	2.29 - 63.11	0.00329
		days_to_psa	0.23	0.04 - 1.36	0.104				
SARC	249(83)	MeTIL	0.94	0.84 - 1.05	0.283	PaTIL	0.87	0.7 - 1.08	0.217
		years_to_birth	1.47	0.95 - 2.28	0.0844	years_to_birth	1.42	0.92 - 2.2	0.117
SKCM	**350(181)**	**MeTIL**	**0.85**	**0.78 - 0.91**	**8.71e-06**	PaTIL	0.96	0.91 - 1.01	0.152
		years_to_birth	1.45	1.06 - 1.97	0.0194	years_to_birth	1.42	1.04 - 1.93	0.027
		pathology_M_stage	1.99	0.96 - 4.14	0.0662	days_to_submitted_specimen_dx	0.25	0.17 - 0.35	5.12e-14
		days_to_submitted_specimen_dx	0.2	0.14 - 0.3	2.22e-16	melanoma_primary_known	1.74	1.01 - 3	0.0468
		melanoma_primary_known	1.51	0.87 - 2.62	0.141				
STAD	256(96)	MeTIL	0.99	0.89 - 1.11	0.93	PaTIL	0.98	0.96 - 1	0.0836
		years_to_birth	2.04	1.29 - 3.22	0.00224	years_to_birth	1.89	1.19 - 3	0.00666
		pathology_M_stage	2.26	1.07 - 4.78	0.0328	pathology_M_stage	2.52	1.18 - 5.38	0.0167

		number_of_lymph_nodes	2.29	1.29 - 4.04	0.00451	number_of_lymph_nodes	1.86	0.98 - 3.53	0.0576
		completeness_of_resection	2.05	1.11 - 3.8	0.0224	completeness_of_resection	2.06	1.13 - 3.76	0.0188
						pathology_N_stage	1.43	0.89 - 2.29	0.137
TGCT	133(4)	MeTIL	1.15	0.61 - 2.15	0.662	PaTIL	1.01	0.98 - 1.05	0.459
		ethnicity	3.49	0.31 - 39.04	0.31	ethnicity	3.29	0.29 - 36.93	0.334
THCA	500(14)	MeTIL	0.74	0.46 - 1.18	0.207	PaTIL	0.49	0.18 - 1.32	0.157
		neoplasm_diseasestage	15.71	2.04 - 120.89	0.00817	neoplasm_diseasestage	15.04	1.96 - 115.26	0.00908
		multifocality	5.07	0.65 - 39.69	0.122	multifocality	5.26	0.68 - 40.54	0.111
THYM	**121(6)**	**MeTIL**	**0.6**	**0.38 - 0.93**	**0.0242**	PaTIL	0.75	0.33 - 1.7	0.492
		gender	0.26	0.04 - 1.68	0.158	years_to_birth	5.13	0.59 - 44.52	0.138
UCEC	539(69)	MeTIL	0.89	0.74 - 1.07	0.216	PaTIL	1	0.99 - 1.02	0.761
		years_to_birth	1.57	0.96 - 2.56	0.0717	years_to_birth	1.48	0.92 - 2.4	0.11
		radiations_radiation_regimenindication	2.01	1.08 - 3.76	0.0286	radiations_radiation_regimenindication	2.07	1.11 - 3.86	0.0228

NOTE: Optimal multivariate models for each cancer subtype were determined by applying a forward and backward variable selection based on the Akaike's information criterion.
Abbreviations: TILs, tumor-infiltrating lymphocytes; PaTIL, pathological assessment of TILs on H&E stained tumor sections; HR, hazard ratio; CI, confidence interval; P, p value; BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; CESC, cervical squamous cell carcinoma and endocervical adenocarcinoma; COREAD, colon and rectum adenocarcinoma; ESCA, esophageal carcinoma; HNSC, head and neck squamous cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian serous cystadenocarcinoma; PAAD, pancreatic adenocarcinoma; PCPG, pheochromocytoma and paraganglioma; PRAD, prostate adenocarcinoma; SARC, sarcoma; SKCM, skin cutaneous melanoma; STAD, stomach adenocarcinoma; TGCT, testicular germ cell tumors; THCA, thyroid carcinoma; THYM, thymoma; UCEC, uterine corpus endometrial carcinoma.

eTable 14. Primer information for bisulfite pyrosequencing.

MeTIL marker	Sense primer	Antisense primer	Sequencing primer	Product size
RASSF1	[biot]GTGGTTAAGGTTAGGGATTAGTTGT	AACCATATCCAAAAAACCTAAACTCATTAA	ATTAAACTACCAAAACTAACAC	160 bp
SEMA3B	[biot]GGGGTAGGGGTTGGGT	ACATCCCATCTTCTATAAACCCTTCTTTC	ACCCTTCCCCTCCCTC	85 bp
KLHL6	GGTGGTAGGAGAATTTTTGAATGTTTATA	[biot]ACTCTCCTTAATTCTTCAAAATAACCT	TTGTTAGAGAGTTTTTTATT	101 bp
PTPRCAP	AAGGGTGGTTTAGGT	biot]ATCTAATATAAACCAAAAATCCTACAACA	AAGGGTGGTTTAGGT	88 bp
INA	AAGAAGAGGGGATAGGGTTAA	[biot]AAACAAAACCATCTACATC	AATTTTAGTAGTTTAGGTATTTTTT	163 bp

NOTE: PyroMark Assay Design 2.0 and PyroMark Q24 software (Qiagen) were used for the design of PCR and sequencing primers. PCR primers for each marker of the MeTIL signature were designed around the target cytosine from the Infinium probe. Thereby, the product size was kept as minimal as possible to anticipate amplification difficulties of fragmented bisulfite-treated FFPE DNA.

eTable 15. PCA parameters.

MeTIL marker	cg20792833	cg23642747	cg12069309	cg20425130	cg21554552
v1	-0,5102517	-0,4350007	-0,4651394	-0,4524386	-0,3596698
s1	0,1329287	0,1750277	0,1220225	0,1525058	0,2404768
c1	0,7135694	0,4625343	0,4007192	0,6471155	0,4601555

NOTE: Breast cancer subtypes were defined based on status of ER, PR and HER2 evaluated with IHC or FISH, respectively. Optimal multivariate models for each subtype were determined in the discovery cohort (cohort 1) by applying a forward and backward variable selection based on the Akaike's information criterion. Abbreviations: TILs, tumor-infiltrating lymphocytes; PaTIL, pathological assessment of TILs on H&E stained tumor sections; HR, hazard ratio; CI, confidence interval; P, p value; TN, triple negative; LUM, luminal subtype; HER2, human epidermal growth factor receptor 2 positive subtype; ER, estrogen receptor; PR, progesterone receptor; IHC, immunohistochemistry; FISH, fluorescent in situ hybridization.

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	3
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	5, 6
	4	Study objectives and hypotheses	7
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	21
Participants	**6**	Eligibility criteria	21
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	21
	8	Where and when potentially eligible participants were identified (setting, location and dates)	21
	9	Whether participants formed a consecutive, random or convenience series	NA
Test methods	**10a**	Index test, in sufficient detail to allow replication	24 & Suppl. methods
	10b	Reference standard, in sufficient detail to allow replication	21, 22
	11	Rationale for choosing the reference standard (if alternatives exist)	5, 16
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	NA
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	NA
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	21, 22
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	NA
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	25
	15	How indeterminate index test or reference standard results were handled	NA
	16	How missing data on the index test and reference standard were handled	Suppl. methods
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	NA
	18	Intended sample size and how it was determined	NA
RESULTS			
Participants	**19**	Flow of participants, using a diagram	NA
	20	Baseline demographic and clinical characteristics of participants	eTable7
	21a	Distribution of severity of disease in those with the target condition	NA
	21b	Distribution of alternative diagnoses in those without the target condition	NA
	22	Time interval and any clinical interventions between index test and reference standard	no
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	Figure 2 and Table 1
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	12, 13
	25	Any adverse events from performing the index test or the reference standard	no
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	17
	27	Implications for practice, including the intended use and clinical role of the index test	18, 20
OTHER INFORMATION			
	28	Registration number and name of registry	NA
	29	Where the full study protocol can be accessed	NA
	30	Sources of funding and other support; role of funders	4

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.