Method for predicting the efficacy of an antibody-drug-conjugate (ADC) using multiplex fluorescence in situ hybridization (FISH) and multiplex sequencing

Cross-reference to related applications

[001] The present application claims the right of priority of the European patent application 23213776 filed with the European Patent Office on 1 December 2023, the entire content of which is incorporated herein for all purposes.

Technical field

[002] The present invention relates to methods for predicting the efficacy of an antibody-drug-conjugate (ADC), said method comprising: (a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of ADC-associated biomarker/s are present in a sample obtained from a subject; (b) determining by multiplex sequencing whether or not mRNA species of ADC-associated biomarkers are present in said sample. The present invention further relates to methods for predicting the efficacy of an antibody-drug-conjugate (ADC), said method comprising: (a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of ADC-associated biomarker/s are present in a sample obtained from a subject; preferably said ADC-associated biomarker/s are selected from the group consisting of: ESR1, PGR, ERBB2 and MKI67; (b) determining by multiplex sequencing whether or not mRNA species of ADC-associated biomarkers are present in said sample; and (c) wherein a part of said sample in which said mRNA species of ADC-associated biomarkers was determined in step (a) is subjected to a laser capturing prior to performing step (b). The present invention further relates to methods for predicting the efficacy of an antibody-drug-conjugate (ADC) as described herein, wherein the ADC-associated biomarker/s from step (a) are same or different from the ADC-associated biomarkers from step (b), preferably wherein the ADC-associated biomarker/s from step (a) are different from the ADC-associated biomarkers from step (b), further preferably wherein in step (a) said ADC-associated biomarker/s are cancer-associated biomarkers, which are preferably used for visualization purposes. The present invention also relates to kits for performing the methods for predicting the efficacy of an antibody-drug-conjugate (ADC) as described herein and provided herein as well as use of such kits for performing the methods as described and provided herein.

Background of the invention

[003] Breast cancer (BCa) is a heterogeneous complex of diseases, a spectrum of many subtypes with distinct biological features that lead to differences in response to various treatment modalities and clinical outcomes. The diagnosis of BCa still suffers from imprecise diagnostic tests often relying on a single technology, and consequently failing in delivering accurate data for selecting the most effective treatment.

[004] BCa diagnosis depends on an expert clinical-pathological review of a diagnostic bio-specimen. As such, it can be subjective, equivocal and even result in misclassification. In this regard, increasing evidence suggests that a second opinion regarding tumor's stage, grade and histology can help to increase the accuracy of the classification. Tumor staging evaluates the extent to which a cancer has developed by growing and spreading. In the scenario of BCa, the nodal status is of particular importance, even influencing therapy selection. Grading reflects the microscopic cell appearance and its deviation from the tissue of origin, ranging from well-differentiated low-grade tumors with good prognosis to the most aggressive undifferentiated ones. From the histological point of view, over 70% of BCa tumors are classified as infiltrating duct carcinomas, no special type (IDC-NST), indicating that histologic typing represents a broad categorization rather than a detailed classification. It may not be sufficient to establish the underlying complex genetic alterations and the biological events involved in cancer development and progression. Thus, tumors of the same histologic type may show vastly different biological behavior and a sole histological assessment may not capture the varied clinical course of individual BCa cases.

[005] Importantly, the joint limitations of the clinical-pathological framework have led to overtreatment of an important proportion of women suffering from a certain type of BCa (early hormone receptor-positive cases with nodal involvement at diagnosis). For these patients, cytotoxic chemotherapy is oftentimes still indicated, even though most of these women have an excellent survival prognosis when treated with adjuvant endocrine therapy alone. In the current era of personalized medicine, a better understanding and classification is warranted.

[006] In this regard, breast tumors can be classified into molecularly distinct subgroups, based on similarities in gene expression profiles using microarrays (e.g., Sorlie, T. et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100, 8418–8423 (2003)). These intrinsic molecular subgroups were termed: luminal A, luminal B, HER2-overexpressing, basal-like BCa (BLBC) and normal breast-like tumors. This classification was successfully translated into the clinic and current BCa molecular subtyping is performed by immunohistochemistry (IHC), assessing expression of hormone receptors (estrogen receptor [ER], progesterone receptor [PR]), human epidermal growth factor receptor 2 (HER2), and proliferation regulator KI67. The actual subtypes are defined using combinations of these IHC markers as follows: luminal A-like (ER positive and/or PR positive and HER2 negative), luminal B-like (ER positive and/or PR positive/PR negative

and HER2 positive), HER2 enriched type (ER negative, PR negative, HER2 positive), and triple negative breast cancer ([TNBC], ER, PR, and HER2 negative). The latter often molecularly overlaps with BLBC. The intrinsic subtypes, which differ markedly in terms of prognosis, therapeutic treatment responsiveness, clinical outcomes, and both overall survival (OS) and relapse-free survival, have been reproduced by several studies using varying numbers of genes in the signature and provide valuable information to guide clinical decision making (e.g., Sorlie, T. et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100, 8418–8423 (2003)).

[007] The classification of BCa into distinct molecular subtypes together with advances in genomic profiling techniques like RT-qPCR, microarray, DNA hybridization, and next generation sequencing (NGS), paved the way for the development of a genomic paradigm to further characterize BCa subtypes and their associated clinical parameters. Some of these multigene assays have now been incorporated into guidelines in the USA (National Comprehensive Cancer Network (NCCN), American Society of Clinical Oncology (ASCO)) and Europe (European Society of Medical Oncology (ESMO)) for recommending adjuvant chemotherapy primarily for early ER-positive breast cancer.

[008] The first multigene molecular diagnostic test that was developed to inform clinical decisions is Oncotype DX, an RT-PCR based assay for the analysis of 21 genes. This assay, specifically developed for patients with ER+ disease at an early clinical stage, consists of 5 housekeeping genes used for normalization and 16 cancer-related genes that fall into several groups: proliferation, invasion, HER2, estrogen, and several miscellaneous genes. A formula is used to calculate a recurrence score between 0 and 100, which is further trichotomized into low, medium, or high risk (e.g., Paik, S. et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351, 2817–2826 (2004)). Patients with high risk show significantly increased risk of distant metastasis and the shortest OS, potentially indicating they will benefit from using chemotherapy in the treatment plan.

[009] MammaPrint, another multigene test, measures the expression of 70 genes using a microarray platform and reports a binary result (low risk or high risk) for recurrence without adjuvant chemotherapy at 10 years. The 70 genes included in the assay were selected by supervised classification of genes whose expression differed in good-prognosis and poor-prognosis tumors, among a group of 78 sporadic tumors. To perform the assay, a sample's gene expression profile is compared to the average good-prognosis profile, and a correlation coefficient is calculated. This coefficient is compared to a validated threshold to place the tumor into good or poor prognosis group. This threshold was optimized to minimize misclassifications, and in particular to minimize misclassification of poor-prognosis tumors into the good-prognosis group (e.g., van't Veer, L. J. et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–536 (2002)). MammaPrint can be

combined with BluePrint, an 80-gene test which stratifies patients into intrinsic molecular subgroups (luminal A, luminal B, HER2-type, basal-type). In addition to allocating effective treatment to appropriate patients, this test combination helps to correctly interpret the lack of neoadjuvant treatment response and tailor subsequent patient management (e.g., Whitworth, P. et al. Chemosensitivity and Endocrine Sensitivity in Clinical Luminal Breast Cancer Patients in the Prospective Neoadjuvant Breast Registry Symphony Trial (NBRST) Predicted by Molecular Subtyping. Ann Surg Oncol 24, 669–675 (2017).

[0010] The PAM50 (i.e., Prediction Analysis of Microarray using 50 classifier genes plus 5 reference genes) assay has been developed as a standardized method that categorizes breast cancers into luminal A, luminal B, HER2-enriched, and BLBC (e.g., Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 27, 1160–1167 (2009)). Based on a modified version of PAM50, Prosigna, a prognostic test approved by the US Food and Drug Administration, was developed for postmenopausal hormonal receptor positive BCa patients (e.g., Dowsett, M. et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. J Clin Oncol 31, 2783–2790 (2013); Gnant, M. et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. Ann Oncol 25, 339–345 (2014)). The test utilizes the NanoString nCounter technology based on two sequence-specific probes ("capture" and "reporter") that jointly hybridize to a target RNA. The reporter probe contains a molecular barcode consisting of six fluorophores whose linear order, determined by high-resolution imaging, creates a unique color code that reveals the identity of each captured target. Prosigna shows predictive value for distant recurrence, benefit of adjuvant chemotherapy, and in predicting response to neoadjuvant therapy and occurrence of late recurrence (e.g., Prat, A. et al. Prediction of Response to Neoadjuvant Chemotherapy Using Core Needle Biopsy Samples with the Prosigna Assay. Clin Cancer Res 22, 560–566 (2016); Filipits, M. et al. The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. Clin Cancer Res 20, 1298–1305 (2014)).

[0011] However, the assignment of the individual tumor to any subtype shows only moderate reproducibility depending on the array platform used, the composition of the entire tumor population, and setting of gene expression threshold (e.g., Mackay, A. et al. Microarray-Based Class Discovery for Molecular Classification of Breast Cancer: Analysis of Interobserver Agreement. JNCI: Journal of the National Cancer Institute 103, 662–673 (2011); Weigelt, B. et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. Lancet Oncol 11, 339–349 (2010).). For example, only ~80% of TNBC belonged to intrinsic

BLBC subtype, while 65% of HER2-positive tumors belonged to intrinsic HER2-overexpressing subtype. In general, BLBC tumors are the most reproducible, whereas luminal B and HER2 subtypes are the least reproducible. Similarly, there is an important disagreement among quantitative diagnostic tests using numerous genes in the signature (there is almost no overlap in the genes included in each of these tests), and some BCa subtypes remain neglected by efforts for treatment personalization (e.g., Vieira, A. F. & Schmitt, F. An Update on Breast Cancer Multigene Prognostic Tests-Emergent Clinical Biomarkers. Front Med (Lausanne) 5, 248 (2018)). Moreover, the cost and technical complexities limit the application of gene expression profiling in daily clinical practice even nowadays.

[0012] Therefore, better diagnostic and prognostic methods including methods for predicting the efficacy of an antibody-drug-conjugate (ADC) are urgently needed.

Summary of the invention

[0013] These and further disadvantages need to be overcome. The present invention therefore addresses these needs and technical objectives and provides a solution as described herein and as defined in the claims.

[0014] The present invention relates to a method for predicting the efficacy of an antibody-drug-conjugate (ADC), said method comprising:

(a) determining by multiplex fluorescence in situ hybridization (e.g., single cell multiplexed fluorescence *in situ* hybridization (FISH)) whether or not mRNA species of ADC-associated biomarker/s are present in a sample (e.g., single cell) obtained from a subject;

(b) determining by multiplex sequencing (e.g., single cell multiplex sequencing) whether or not mRNA species of ADC-associated biomarkers are present in said sample.

[0015] The present invention further relates to methods for predicting the efficacy of an antibody-drug-conjugate (ADC), said method comprising:

(a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of ADC-associated biomarker/s are present in a sample obtained from a subject; preferably said ADC-associated biomarker/s are selected from the group consisting of: ESR1, PGR, ERBB2 and MKI67;

(b) determining by multiplex sequencing whether or not mRNA species of ADC-associated biomarkers are present in said sample; and

(c) wherein a part of said sample in which said mRNA species of ADC-associated biomarkers was determined in step (a) is subjected to a laser capturing prior to performing step (b).

[0016] The present invention further relates to methods for stratification of a subject for treatment (e.g., cancer treatment, e.g., breast cancer treatment) with an ADC, comprising:

(a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of ADC-associated biomarker/s are present in a sample obtained from a subject; preferably said ADC-associated biomarker/s are selected from the group consisting of: ESR1, PGR, ERBB2 and MKI67;

(b) determining by multiplex sequencing whether or not mRNA species of ADC-associated biomarkers are present in said sample; and

(c) wherein a part of said sample in which said mRNA species of ADC-associated biomarkers was determined in step (a) is subjected to a laser capturing prior to performing step (b); wherein the presence of one or more of said ADC-associated biomarker is indicative that said ADC may be effective, and/or wherein the absence of one or more of said ADC-associated biomarker is indicative that said ADC may not be effective.

[0017] The present invention further relates to methods for improving performance of a multi-gene signature test (e.g. PAM50, AIMS), comprising subjecting a part of a sample obtained from a subject to a laser capturing prior to carrying out said test, preferably said multi-gene signature test is a cancer (e.g., breast cancer) multi-gene signature test.

[0018] The present invention further relates to methods as described herein, wherein said laser capturing is carried out by the means of a laser capture microdissection, preferably said method is carried out by the means of multiplexed RNA-FISH-guided laser capture microdissection.

[0019] The present invention further relates to methods as described herein, wherein the ADC-associated biomarker/s from step (a) are same or different from the ADC-associated biomarkers from step (b), preferably wherein the ADC-associated biomarker/s from step (a) are different from the ADC-associated biomarkers from step (b), further preferably wherein in step (a) said ADC-associated biomarker/s are cancer-associated biomarkers, which are optionally used for visualization purposes.

[0020] The present invention also relates to kits for performing the methods for predicting the efficacy of an antibody-drug-conjugate (ADC) as described herein and provided herein as well as use of such kits for performing the methods as described and provided herein.

[0021] In summary, the present invention also *inter alia* relates to the following advantages provided by the means and methods as described herein over known means and methods: (i) multiplex RNA-FISH provides spatial localization and/or visualization of mRNA biomarkers, e.g., allowing visualization of biomarker expression within specific cellular contexts (e.g., tumor cells and/or tissues); (ii) multiplex sequencing on visualized and laser captured cells/tissues providing a complementary layer of quantitative and quantitative analysis, which is capable of

verifying and quantifying biomarker expression with high precision and accuracy, because laser capturing (e.g., LCM) isolates specific regions of interest (e.g., tumor cells and/or tissues) from the biological sample localized and visualized by RNA-FISH. In other words, means and methods of the present invention synergistically combine spatial and quantitative techniques to ensure a robust and accurate assessment of biomarkers as described herein (e.g., resulting in minimizing false positives and negatives).

Brief description of the drawings

[0022] Figure 1: **Study design, analytical validity, and molecular subtyping.** (**a**) The retrospective cohort consisted of 1,082 formalin-fixed paraffin-embedded (FFPE) breast samples, which underwent multiplexed RNA-FISH-guided laser capture microdissection (LCM) coupled with RNA-sequencing. Annotation of the tumor area on an H&E section and the biomarker expression derived from multiplexed RNA-FISH were used to select regions of interest (ROIs) for LCM from cresyl violet sections. These tumor enriched samples were then sequenced to characterize gene expression signatures to provide diagnostic, prognostic, and predictive inferences from the cohort clinical data. (**b**) Analytical validity of mFISHseq compared to immunohistochemistry assessed by receiver operating characteristic (ROC) curves in 70:30 training and test datasets. Table shows biomarker thresholds defined in the training set and applied to the test set. (**c**) Consensus clustering of 293 genes used for molecular subtyping in 1,254 samples. The phenobars above the gene expression heatmap depict patient metadata (from top to bottom: biobank origin, clinical parameters, LCM or no LCM, subtyping approaches, prognostic signatures, and IHC and RNA-FISH results).

[0023] Figure 2: **Constructing consensus subtyping yields intrinsic molecular subtypes associated with survival and development of a T-DM1 response-predictive classifier** (**a**,**b**) Survival according to consensus subtyping with respect to their IHC-surrogate subtype and node negative (**a**) and positive (**b**) status. (**c**) Sankey diagram shows the IHC-surrogate subtypes and the proportion of samples reclassified by mFISHseq consensus subtyping. (**d**) Survival of consensus molecular subtypes. Univariate (**e**) and multivariate (**f**) Cox proportional hazards models (CPHM) on 19 ADC targets and their association with overall survival in all samples and stratified by subtype. (**g**) Univariate analysis of 70 genes/gene signatures and their association with pathologic complete response (pCR) in the T-DM1 arm of the I-SPY2 trial. (**h**) The 19 genes/gene signatures selected in the multivariate analysis on the training dataset. Green bars denote signatures associated with pCR; red bars indicate signatures associated with no pCR. (**i**) Performance of the T-DM1_pred classifier in the test set relative to *ERBB2* alone. AUC = area under the curve. (**j**) Scatter plots showing the distribution of

selected genes/gene signature scores of the T-DM1_pred classifier in patient samples according to pCR.

[0024] Figure 3: RNA-FISH image analysis pipeline using digital pathology/machine learning. Representative images depicting the RNA-FISH image analysis pipeline for tissue and cell segmentation. The tissue was stained for DAPI, ERBB2 (Opal520), MKI67 (Opal570), ESR1 (Opal620), and PGR (Opal690) and scanned at 20x. **(a)** The image analysis workflow started with raw images of multiple merged fluorescent channels. **(b)** The same image was split into individual separate fluorescent channels - DAPI, Opal520, Opal570, Opal620, Opal690, and autofluorescence, according to the spectral library for each fluorescent probe, and the autofluorescence was isolated from the image. **(c)** Using a machine learning algorithm, the tissue was segmented based on the selected training regions (red = tumor; green = stroma) with the training considering the fluorescent signal intensity of DAPI, ERBB2, MKI67, ESR1, and PGR. **(d)** Then the trained algorithm was applied to the whole image. **(e)** Individual cells were identified by the shape and intensity of DAPI staining, **(f)** then segmented, and fluorescent intensity values were normalized to exposure time and obtained for both tumor and stroma.

[0025] Figure 4: Gene signatures associated with survival and prediction of treatment response. (**a**) Gene expression heatmap illustrating the expression of 92 genes and 110 gene signatures in respect to molecular subtype and clinical parameters. Univariate (**b**) and multivariate (**c**) Cox proportional hazards models (CPHM) on ADC processing-related genes and gene signatures and their association with overall survival in all samples and stratified by subtype.

[0026] Figure 5: Expression of ADC targets in healthy versus invasive breast cancer tissue. Scatter dot plots illustrate the expression of ADC antigen targets assessed by RNA sequencing in healthy tissue, invasive breast cancer tissue, and individual clinical subtypes. *p < 0.05, **p < 0.01 ***p < 0.001, ****p < 0.0001.

[0027] Figure 6: Comparison of RNA-FISH/-SEQ with biobank IHC results. Graphs show the comparison of RNA-FISH and RNA-SEQ to biobank IHC results for PR (**a,b,e,f**), ER (**c,d,g,f**), HER2 (**i,j,m,n**), and Ki67 (**k,l,o,p**). Data are presented as a comparison to either IRS score for PR/ER, H-Score for HER2, percent of positive cells for Ki67, or as negative or positive IHC expression. Dotted lines in the violin plots show the median and quartiles. *p < 0.05, **p < 0.01 ***p < 0.001, ****p < 0.0001.

[0028] Figure 7: Expression of biomarkers as quantified by RNA-FISH and RNA-SEQ in IHC surrogate subtypes. Violin plots show *PGR* (**a,b**), *ESR1* (**c,d**), *ERBB2* (**e,f**), and *MKI67* (**g,h**) expression in each IHC surrogate molecular subtype, as determined by IHC biobank

results according to the 2013 St. Gallen classification of intrinsic surrogate subtypes, quantified by either RNA-FISH (**a,c,e,g**) or RNA-SEQ (**b,d,f,h**). Each dot represents an individual sample, the red dotted line denotes the median, and the black dotted lines denote the median and quartiles (25% and 75%).

[0029] Figure 8: Diagnostic performance of RNA-SEQ and RNA-FISH compared to IHC. PR (**a**, **b**, **e**, **f**) and ROC (**c**, **d**) curves from the 70:30 training (**a**, **c**, **e**) and test (**b**, **d**, **f**) datasets illustrate the performance of RNA-SEQ (**a**, **b**) and RNA-FISH (**c-f**) at classifying biomarker expression for *PGR*, *ESR1*, *ERBB2*, and *MKI67*. (**g**) The table shows the RNA-FISH signal intensity (SI) thresholds for each marker derived from the training set by identifying the threshold value that maximized agreement (Cohen's K) between IHC results. These threshold values were then prespecified for use in the test dataset. AUC, area under the curve; *ERBB2*, Erb-B2 Receptor Tyrosine Kinase 2 gene (HER2); *ESR1*, estrogen receptor 1 gene; *MKI67*, marker of proliferation Ki-67 gene; *PGR*, progesterone receptor gene. For Ki-67, ≤5% and ≥30% were used as cutoff values for negativity and positivity, respectively.

[0030] Figure 9: Comparison of molecular subtyping schemes in terms of survival and single sample concordance. (**a**) Graphs show the top 5 differentially expressed genes (log2 fold change) that are either upregulated or downregulated between each molecular subtype. The data are arranged such that the direction of the log2 fold change relates to the first subtype listed in the graphs title (i.e., LumA vs LumB comparison shows genes that are upregulated and downregulated in Luminal A samples relative to Luminal B samples). (**b**) Proportion of samples classified by each classifier into the respective IHC surrogate or intrinsic molecular subtypes. (**c**) Kaplan-Meier survival curves comparing overall survival by molecular subtype according to the IHC surrogate, mFISHseq, AIMS, or PAM50 subtyping schemes. (**d**) Cohen's kappa concordance between IHC surrogate, mFISHseq, AIMS, PAM50, consensus subtyping schemes for all samples and samples stratified by subtype. (**e**) Survival curves illustrating probability of overall survival stratified by IHC surrogate subtype for samples that showed perfect concordance between all four classifiers (4/4, IHC surrogate, mFISHseq, AIMS, and PAM50) as well as discordant samples that showed concordance between three out of four classifiers (3/4, IHC surrogate and any two multigene classifiers), two out of four classifiers (2/4, IHC surrogate and any one multigene classifier), or only the IHC surrogate classifier (1/4). (**f**) Metrics of instability as assessed by correlation to the corresponding PAM50 subtype centroid for samples with perfect concordance (4/4) or samples showing discordance in one or more classifiers (3/3, 2/4, and 1/4) stratified by IHC surrogate subtype.

[0031] Figure 10: Single sample discordance and reclassification by consensus subtyping. Overall survival of samples uniquely classified by only one subtyping approach

(IHC surrogate, mFISHseq, AIMS, and PAM50) as luminal A (**a**), luminal B (**c**), Her2 (**e**), and TNBC/basal-like (**g**). The proportion of samples with low, intermediate, or high genomic risk assessed by OncotypeDX, GENE70, and ROR-S as well as clinical risk in samples uniquely classified as luminal A (**b**), luminal B (**d**), Her2 (**f**), and TNBC/basal-like (**h**). The uniquely classified samples that were reclassified to a new consensus molecular subtype are depicted in the bottom right bar charts of **b**, **d**, **f**, and **h**.

[0032] Figure 11: Single sample discordance for the normal-like subtype and reclassification by consensus subtyping. (**a**) Overall survival of samples uniquely classified as normal-like by only one subtyping approach (AIMS, and PAM50). (**b**) The proportion of normal-like samples with low, intermediate, or high genomic risk assessed by OncotypeDX, GENE70, and ROR-S as well as clinical risk. (b, bottom right panel) The uniquely classified normal-like samples that were reclassified to a new consensus molecular subtype.

[0033] Figure 12: Differential gene expression analysis of *ERBB2* positive, low, and negative tumors. (**a**) Overview of samples stratified into *ESR1* positive and negative groups in reference to the original IHC classifications from the biobanks. Each *ESR1* group was further divided into ERRB2 positive, low, and negative subgroups according to TPM interquartile ranges. Each bar shows the proportion of samples in each subgroup according to the original IHC classifications from the biobanks. **(b)** Bar graphs show differentially expressed genes between *ERBB2* expression groups. **(c)** Line plots compare gene expression between *ERBB2* expression groups across 115 genes located near the 17q11-21 locus (e.g., extended Her2 amplicon). **(d)** Distribution of Her2 amplicon gene signature scores between *ERBB2* expression groups ranked from highest to lowest. Inset violin plots compare mean Her2 amplicon scores between *ERBB2* expression groups. Kruskal-Wallis test followed by Dunn's multiple comparisons test, ****P<0.0001.

[0034] Figure 13: Consensus prognostic risk categories show clinically relevant differences in survival (**a**) Proportion of risk categories (high, intermediate, and low) among each of the five prognostic classifiers, including Clinical risk, OncotypeDX, Risk of Recurrence by Subtype (ROR-S), GENE70, and Genomic Grade Index (GGI) on 567 ER+/HER2– with 0-3 positive lymph nodes. (**b**) Progression free survival (PFS) of each of the five prognostic classifiers. (**c**) Concordance of each classifier when comparing all risk categories (high, intermediate, and low; left table) or after consolidating into two categories (high and intermediate/low combined; right table). (**d**) Concordance of each classifier for high and low risk samples as illustrated by the number of concordant classifiers and (**e**) after consolidating into majority (agreement in ≥3 prognostic classifiers) and minority (agreement in 1-2 prognostic classifiers) categories. (**f**) Distribution of mRNA expression for estrogen receptor (*ESR1*,

yellow, left panel), progesterone receptor (*PGR*, green, left/middle panel), HER2 receptor (*ERBB2*, red, right/middle panel), and Ki67 marker of proliferation (*MKI67*, orange, right panel) in patients that were classified as high risk by a particular number of concordant classifiers (i.e., 1-5 concordant classifiers). (**g**) Kaplan Meier plots show PFS and overall survival (OS) for each consensus prognostic risk category (high, low, and ultra-low).

[0035] Figure 14: Development of consensus prognostic risk categories (**a**) Proportion of uniquely classified patients (i.e., patients classified in a risk category by only one classifier) in high, intermediate, and low risk categories. (**b**) Kaplan-Meier (KM) progression free survival (PFS) curves showing patients uniquely classified by an individual classifier into high (top panel), intermediate (middle panel), and low (bottom panel) risk categories. (**c**) The same PFS curves in (**b**) after consolidating unique classifications from OncotypeDX, ROR-S, GENE70, and GGI into a single Multigene category for each risk category. (**d**) Exploratory decision tree describing criteria for consensus classification of patients into high, low, and ultra-low risk categories. (**e**) KM curves for high (top panel), low (middle panel), and ultra-low (bottom panel) consensus risk categories depicting PFS after stratifying patients into treatment categories.

[0036] Figure 15: Genes and gene signatures associated with overall survival in the entire cohort. Forest plot illustrating significant Hazard ratios (HR) obtained from univariate Cox Proportional Hazard (Cox-PH) analyses, using a continuous signature scoring method calculated for the whole dataset (n=1013). Actual numbers in each analysis are lower due to missing follow up data and are denoted in the figure.

[0037] Figure 16: Gene signatures associated with overall survival in specific clinical subtypes. A Forest plot illustrating the significant Hazard ratios (HR) obtained from univariate Cox Proportional Hazard (Cox-PH) analyses, using continuous gene signature scoring separately calculated for each biobank clinical subtype (n=1015). Actual numbers in each analysis are lower due to missing follow up data and are denoted in the figure.

[0038] Figure 17: Comparison of laser capture microdissection with bulk processing on biomarker expression, molecular subtyping, and prognostic classifiers. (**a**) Photomicrographs depict examples of hematoxylin and eosin-stained resected tumor specimens with low, intermediate, or high tumor content represented in shaded annotations. Scale bars represents 2mm length. (**b,c**) Change in gene expression of *PGR*, *ESR1*, *ERBB2*, and *MKI67* in specimens that were classified as IHC positive (**b**) or IHC negative (**c**). (**d**) Dot plots show the dynamic range of gene expression for each biomarker in LCM vs no LCM matched samples. Dotted lines represent the median. (**e**) Expression of cell-type specific markers in LCM vs no LCM samples containing either low, intermediate, or high tumor content. (**f,g**) Sankey diagrams illustrate change in mFISHseq consensus subtypes (**f**) and in PAM50

risk of recurrence by subtype (ROR-S) classification (**g**) for LCM samples and their paired undissected scrolls.

[0039] Figure 18: Clinical and molecular parameters associated with discordant risk assignment. (a) Venn diagram depicting the number of overlapping genes among each of the four multigene prognostic classifiers. (b) Dot plot of clinical parameters significantly associated with high and low risk. (c-h) Graphs illustrate the proportion of selected clinical parameters associated with discordance in either high or low risk, including tumor size (c), node status (d), tumor grade (e), histological subtypes (f), clinical risk as described in the MINDACT trial (g), and treatment with chemotherapy (h). The percentages at the top of each bar denote the proportion of patients with pT2-pT3 tumors (c), node positive status (d), grade 1 (G1) tumors (e), invasive lobular carcinoma (f), low clinical risk (g), and adjuvant chemotherapy treatment (h). The heatmaps (i) show the significant genes/gene signatures associated with discordance in 1 or 2 prognostic classifiers relative to patients with unanimous agreement (group labeled as 0) for either high (left heatmap) or low (right heatmap) risk. The legends refer to the row metadata for group of gene/gene signature, -log10 (adjusted p-value) for all significant Kruskal-Wallis tests (adjusted p-value < 0.05), results from Dunn's multiple comparison tests for each pairwise comparison (gray box, adjusted p-value <0.05; white box, adjusted p-value >0.05), and the z-score normalized expression. This figure demonstrates the clinical and molecular factors that drive discordance in risk assignment and provides supporting evidence for consensus prognostic risk groups.

[0040] Figure 19: Cell types and states associated with discordance and immunotherapy response. (a,b) The heatmaps show the significantly different cell types/states in patients that have one or two (groups labeled as 1-2) prognostic classifiers that are discordant with patients that are unanimously classified (group labeled as 0) as either high (a) or low (b) risk as the reference group. The legends refer to the row metadata for group (cell type), -log10 (FDR) with red values denoting a significant Kruskal-Wallis test (FDR < 0.05), results from Dunn's multiple comparison tests for each pairwise comparison (gray box, FDR <0.05; white box, FDR >0.05; note that post hoc comparisons are depicted even if the Kruskal-Wallis test was not significant), and the cube root transformed cell type/state abundances. (**c-h**) Bar graphs illustrate the exemplary cell types/states that are differentially expressed when comparing high and low risk samples (top panel graphs) as well as patients that have 1 or 2 discordant classifiers for high risk relative to those who have unanimous agreement (bottom panels). Cell types/states include CD8 T cells – Exhausted / effector memory (SO3) (**c**), CD4 T cells – Exhausted / effector memory / Treg (SO1) (**d**), Dendritic cells – Mature immunogenic (S03) (**e**), Monocytes/Macrophages – M2-like proliferative (S07) (**f**), Epithelial cells – Pro-inflammatory (SO4) (**g**), Fibroblasts – Pro-migratory-like (SO8) (**h**). Statistical comparisons are

only shown for bottom panels. *p < 0.05, **p < 0.01 ***p < 0.001, ****p < 0.0001. (**i**). Bar graphs show the percentage of patients who achieved pathological complete response following neoadjuvant treatment with paclitaxel combined with pembrolizumab in the I-SPY2 trial stratified by biomarker subgroups. This figure demonstrates the clinical and molecular factors that drive discordance in risk assignment and provides supporting evidence for consensus prognostic risk groups and shows a cell type specific signature that predicts response to immunotherapy.

[0041] Figure 20: Development and validation of a classifier for T-DM1 sensitivity. (a) Univariate logistic regression analysis of 71 prespecified ADC-related genes/gene signatures and their association with pathologic complete response (pCR) in the T-DM1 arm of the I-SPY2 trial (n=52). (b) The 19 genes/gene signatures selected in the multivariate logistic regression with elastic net modeling using 10-fold cross validation and their association with pCR. Green bars denote signatures associated with pCR; red bars indicate signatures associated with no pCR; black bars depict signatures not associated with either pCR or no pCR. (c) ROC curves showing performance of two T-DM1_pred classifiers in the test set relative to *ERBB2* mRNA alone in the T-DM1 arm. Univariate T-DM1_pred is a single score derived from all 19 features, while multivariate T-DM1_pred includes all 19 features in a multivariate regression model. (d) ROC curves showing performance of the T-DM1_pred classifier in the test set relative to *ERBB2* mRNA alone in both the pertuzumab (n=44) and taxane/anthracycline (n=31) control arms. AUC = area under the curve. (e) Scatter plots showing the distribution of selected genes/gene signature scores of the T-DM1_pred classifier in patient samples according to pCR.

[0042] Figure 21: Deployment of mFISHseq as a research-use only test. (a) Table outlining the number of genes and gene signatures and their relevant drug targets/pathways that were used for RUO testing. (b) Proportion of 48 patients according to treatment setting (left bar), molecular subtype (middle bar, includes 58 samples total to account for 10 patients that had 2 subtypes collected by LCM), and TNBC subtype (right bar). (c) Frequency of therapies that were recommended as the top 3 according to expression of genes and gene signatures tailored to each patient. (d-h) Expression of 20 ADC antigen targets (d) as well as targets relevant to payloads for topoisomerase and microtubule inhibitors (e), endocytosis (f), lysosome activity (g), and resistance (h). Panels b (middle bar) and d-h include 58 samples total comprised of 48 patients with 10 patients who had two subtypes collected by LCM). Panel b (right bar) contains 36 samples total comprised of 31 patients with 4 patients who had two subtypes collected by LCM and one patient that had both a primary and metastatic tumor analyzed. (i) Examples of patients belonging to putative ADC treatment-responsive groups based on expression of ADC relevant biomarkers. Patients 1, 18, 15, and 23 are predicted to

respond to SG, T-Dxd, both SG and T-Dxd, and neither SG and T-Dxd, respectively. This figure showing a proof-of-concept foundation for using ADC markers for patient stratification.

[0043] Figure 22: Benchmarking reclassified samples with subtype-specific markers. The dotted lines in the Sankey diagram show samples that were reclassified from either IHC surrogate LumA to LumB (a-c) or IHC surrogate LumB to LumA by consensus subtyping (d-f). The box and whisker plots show the expression of several genes and gene signatures for proliferation (a,d), estrogen (b,e), and Her2 (c,f) pathways. (a-c) Plots compare the samples reclassified from IHC surrogate LumA to LumB (gray bars) to the ground truth samples for LumA (consensus LumA, green bars) and LumB (consensus LumB, blue bars). (d-f) Plots compare the samples reclassified from IHC surrogate LumB to LumA (gray bars) to the ground truth samples for LumB (consensus LumB, blue bars) and LumA (consensus LumA, green bars). Boxes denote the interquartile range with the line plotted as the median, whiskers show the 5 and 95 percentiles, and dots are individual samples outside this range. *$p < 0.05$, **$p < 0.01$ ***$p < 0.001$, ****$p < 0.0001$.

[0044] Figure 23: Consensus subtyping on the METABRIC cohort yields intrinsic molecular subtypes associated with survival. (a-d) Overall survival according to consensus subtyping with respect to the IHC surrogate subtype according to IHC results, including Luminal A (a, n=746), Luminal B (b, n=672), Her2+ (c, n=130), and TNBC (d, n=293) stratified by nodal status (left and right panels depict node negative and positive, respectively). (e) Sankey diagram shows the IHC surrogate subtypes and the proportion (%) of samples reclassified by consensus subtyping. (f) Overall survival of consensus molecular subtypes. (g) Forest plots showing univariate and multivariate Cox proportional hazards models comparing prognostic utility of IHC surrogate vs Consensus molecular subtypes. Multivariate models included both tumor size (pT1 vs pT2-pT4) and node status (pN0 vs pN1-pN3). Hazard ratios show the overall survival estimates with 95% CIs.

[0045] Figure 24: Consensus subtyping on the TCGA cohort yields intrinsic molecular subtypes associated with survival. (a-d) Overall survival according to consensus subtyping with respect to the IHC surrogate subtype according to IHC results, including Luminal A (a, n=382), Luminal B (b, n=313), Her2+ (c, n=66), and TNBC (d, n=151) stratified by nodal status (left and right panels depict node negative and positive, respectively). (e) Sankey diagram shows the IHC surrogate subtypes and the proportion (%) of samples reclassified by consensus subtyping. (f) Overall survival of consensus molecular subtypes. (g) Forest plots showing univariate and multivariate Cox proportional hazards models comparing prognostic utility of IHC surrogate vs Consensus molecular subtypes. Multivariate models included both

tumor size (pT1 vs pT2-pT4) and node status (pN0 vs pN1-pN3). Hazard ratios show the overall survival estimates with 95% CIs.

[0046] Figure 25: Consensus prognostic risk categories show clinically relevant differences in survival in METABRIC and TCGA cohorts. External validation of consensus prognostic risk categories in METABRIC (a,c,e) and TCGA (b,d,f) cohorts. (a,b) Concordance of each classifier for high and low risk samples as illustrated by the number of concordant classifiers and (c,d) after consolidating into majority (agreement in ≥3 prognostic classifiers) and minority (agreement in 1-2 prognostic classifiers) categories. (e,f) Kaplan Meier plots show relapse free status (e, top panel) or progression free survival (PFS; f, top panel)) and overall survival (OS; e,f, bottom panels) for each consensus prognostic risk category (high, low, and ultra-low). All analyses contain 1,172 (METABRIC) or 465 (TCGA) patient samples that would be eligible for prognostic multigene tests in a real-world clinical setting (ER or PR+, HER2−, and 0-3 positive lymph nodes.

[0047] Figure 26: Single cell distribution and specificity of genes included in each multigene classifier. Stacked bar graphs show the number of genes classified as Cell Enriched, Cell Enhanced, Cell Group Enriched, Cell Group Enhanced, Low Specificity, and Not detected using slightly modified criteria from The Human Protein Atlas (see Supplementary Methods). The distribution and specificity overview of the metadata from Supplementary Figures 26-31 is provided for each classifier, including (a) mFISHseq, (b) AIMS, (c) PAM50, (d) OncotypeDX, (e) GENE70, and (f) GGI. Bars are colored based on the major cell type defined by The Human Protein Atlas portal.

[0048] Figure 27: Single cell expression of ADC markers and impact of LCM on their expression. (a) The left heatmap shows the unsupervised hierarchically clustered max-normalized nTPM expression of 69 ADC marker genes (2 genes undetected) in 20 different cell types from healthy breast tissue obtained from the Human Protein Atlas. Note that the single cell expression data from multiple cell types (e.g., breast glandular cells.1-6) are also consolidated into cell groups (e.g., breast glandular cells) that precede the individual cell types. The row (gene) metadata includes log10 (median nTPM +1) normalized expression as well as a summary of cell type/group enrichment, enhancement, specificity, and whether the gene was detected in the healthy single cell dataset. The right dot plot heatmap shows 0-1 normalized mean-scaled gene expression profiles in cell types obtained from breast cancer tissue from the Broad Institute Single Cell Portal. The rightmost dot plot heatmap shows the same breast cancer data from Cancer Epithelial cells stratified by different breast cancer subtypes (HER2+, ER+, and TNBC). The gene order is the same as the healthy tissue clustering. The dot size represents the percentage of cells expressing each gene. The row (gene) metadata includes

the color-coded ADC marker group that each individual gene is categorized. CAFs, cancer-associated fibroblasts; PVL, perivascular-like cells. (b) Stacked bar graphs show the number of ADC marker genes classified as Cell Enriched, Cell Enhanced, Cell Group Enriched, Cell Group Enhanced, Low Specificity, and Not detected using slightly modified criteria from The Human Protein Atlas. Bars are colored based on the major cell type defined by The Human Protein Atlas portal. The additional pie charts show the specific genes in different ADC classes that are cell type/group enriched or lack specificity. (c-h) Expression of selected ADC antigen targets (c-d), payload targets (e), endocytosis markers (f), lysosome markers (g), and resistance markers (h) in paired LCM vs no LCM samples containing either low, intermediate, or high tumor content. This figure shows single-cell RNA expression of ADC markers and the impact of LCM on their expression.

[0049] Figure 28: An exemplary vignette of a patient treated with the ADC, sacituzumab govitecan. (a) Patient 1 has TNBC and was unresponsive to 7 cycles of neoadjuvant taxane/carboplatin leading to a mastectomy. The resected tumor was classified as invasive ductal carcinoma (pT3, pN1, G3) and ER/PR/HER2− with 60% Ki67. After surgery, the patient received postoperative radiation therapy and capecitabine and then had a locoregional relapse. In the advanced/metastatic setting the patient received the VEGF inhibitor, bevacizumab, with a documented response, and then after disease progression is currently on sacituzumab govitecan. (b) We conducted mFISHseq on both the initial biopsy of the primary tumor and the resected tumor from mastectomy. The histopathology revealed two distinct regions of the resected tumor that were characterized either with or without infiltrating lymphocytes. (c) After dissecting both regions (Sample A/B; shown in red/yellow annotations), mFISHseq confirmed the IHC results and assigned both ROIs as basal-like; however, the two regions were classified as different TNBC subtypes: mesenchymal (M) and immunomodulatory (IM). The basal-like subtype and TNBC M subtype matched the subtyping of the primary tumor. (d) To identify treatment predictive signatures common between both ROIs of the resected tumor, we investigated a tailored list of markers relevant for metastatic TNBC, revealing that both ROIs contained high scores for an angiogenesis/hypoxia/inflammation signature that may predict response to bevacizumab (Avastin) and high expression for *TACSTD2* and *TOP1* (i.e., antigen and payload targets for sacituzumab govitecan). This provided two viable treatments in the first and second lines for metastatic TNBC. This evidence concurred with follow up data documenting a durable clinical response following bevacizumab. Interestingly, analysis of the primary tumor also provides biological rationale for the lack of response to neoadjuvant taxane/carboplatin, which would be predicted given the M TNBC subtype and low signature scores for taxanes (MAPs_Mitotic_kinases_neoadj_chemo17) and carboplatin (VCpred_TN).

[0050] Figure 29: An exemplary vignette of a patient treated with the ADC, T-DM1. (a) Patient 15 has Luminal B and was neoadjuvantly treated with anthracycline, taxane, trastuzumab and pertuzumab. Biopsy from tumor surgery was classified as invasive carcinoma of no special type (pT1c, pN1a, G2) with 90% ER, 10% PR, 40% Ki67 and Her2 status 3+ assessed by IHC. After surgery, the patient had residual disease, received radiation therapy, and is scheduled to receive T-DM1 (14 cycles). (b) We conducted mFISHseq on the tumor biopsy and based on homogeneous histopathology and biomarker expression, we dissected a single ROI (shown in red annotations). (c) mFISHseq confirmed previous IHC results, and consensus subtyping classified the patient as luminal B. (d) Elevated levels of *ERBB2* and its related Her2 amplicon signature, along with low to moderate expression of resistance markers (pSTAT3-GS and *NRG1*), indicate that Her2-related therapies such as trastuzumab, pertuzumab, and trastuzumab emtansine may be beneficial. Notably, this patient had a moderate T-DM1_pred score (score = 0.29; 57th percentile) that was above the threshold for predicting response to T-DM1. While high MAPs signatures predict response to neoadjuvant taxane chemotherapy, there were also markers of resistance like a high Residual_disease_ER.Pos signature, which was clinically confirmed. High expression of ERBB2 and TACSTD2 also suggest the novel ADCs, trastuzumab deruxtecan and sacituzumab govitecan, may be beneficial in the advanced setting, if necessary, which is also supported by high TOP1, the target of the cytotoxic payloads of both ADCs. Currently, we are awaiting follow up data on the patient.

[0051] Figure 30: An exemplary vignette of a patient treated with the ADC, T-DM1. (a) Patient 19 has Luminal B breast cancer and was treated with neoadjuvant endocrine therapy which was followed by a mastectomy. The resected tumor was classified as invasive ductal carcinoma (pT4b, pN1a, G2) and ER+/PR−/HER2+ (IHC 2+ with positive amplification by DNA-FISH) with 15% Ki67. After surgery, the patient received adjuvant taxol and trastuzumab and endocrine treatment with letrozole. After 12 months of endocrine treatment, the patient stopped responding and tumor cell dissemination was detected. In the advanced/metastatic setting, the first line treatment with T-DM1 was applied with a duration of response of 6 months. (b) We conducted mFISHseq on the resected tumor from the mastectomy and dissected two heterogeneous ROIs of the tumor based on *ERBB2* expression (high *ERBB2* (Sample A), low *ERBB2* (Sample B)). (c) mFISHseq results for both ROIs (Sample A/B) agreed with IHC results except for Sample B, which was classified as *ERBB2* low (d) High expression of *ERBB2* and related signatures (Her2 amplicon_MDX and Module7_*ERBB2*), especially in Sample A, suggest sensitivity to anti-Her2 targeted therapies. Interestingly, only Sample A had a T-DM1_pred score (score = 0.03; 48th percentile) that was above the threshold for predicting response to T-DM1. Sample B, which had lower *ERBB2* expression had a lower T-DM1_pred

score (score = −01.20; 14th percentile), which was predicted to be insensitive to T-DM1. This evidence concurred with follow up data documenting a short-term clinical response to T-DM1 and is also in agreement with a phase II clinical trial that showed lower responses to T-DM1 in patients with Her2 (ERBB2) heterogeneity. High scores for *ERBB2*, *TACSTD2*, and *TOP1* (i.e., antigen and payload targets for trastuzumab deruxtecan and sacituzumab govitecan), providing two viable treatment options. Trastuzumab deruxtecan is currently planned for administration as the second line of treatment in this setting.

[0052] Figure 31: An exemplary vignette of a patient treated with the ADC, trastuzumab deruxtecan. (a) Patient 25 was diagnosed with breast cancer and de novo metastasis to the liver, lumbar vertebra (L2), and left iliac bone. She underwent a metastasectomy of segment 8 (S8) of the liver, left mastectomy, and sentinel lymph node biopsy (SNB) revealing an invasive ductal carcinoma (pT1c, cpN1, pM1, G2) that was ER+/PR+/HER2− with high (30%) Ki67. The patient received adjuvant cyclophosphamide, epirubicin, and 5-fluorouracil regimen for only two cycles due to poor tolerance followed by a year of tamoxifen. After discovering multiple liver metastases, the patient received first line palliative paclitaxel chemotherapy and the VEGFA-inhibitor, bevacizumab, for eight cycles, resulting in partial remission (PR). This was followed by letrozole (8 months), which was changed to exemestane (10 months) due to poor tolerance. An ultrasound and CT revealed disease progression (PD) and then the patient was administered disulfiram in a clinical trial. Another CT detected liver progression, which was treated with percutaneous radiofrequency ablation (RFA). In the second line, the patient received capecitabine (10 cycles) followed by fulvestrant (17 doses); however, disease progression was detected by liver MRI. The patient then received a repeat exposure to capecitabine (18 cycles, 4th line), metronomic capecitabine and cyclophosphamide chemotherapy (5th line), and, after further disease progression, two rounds of microwave ablation. A liver biopsy was performed, revealing infiltration by invasive ductal carcinoma of the breast (ER+/PR−/HER2− with 25% Ki67. This was the specimen we processed using Multiplex8+. After this biopsy, bone scintigraphy revealed a solitary metastasis in the L2 vertebra that was subsequently treated with stereotactic body radiation therapy (SBRT). Another MRI revealed new metastatic lesions in the liver, L2 vertebra, and left iliac bone. The patient received letrozole and a CDK4/6 inhibitor and then after 5 month a liver MRI showed progression. The patient then continued with docetaxel (6th line) and trastuzumab deruxtecan (7th line) before passing away in 2024. (b-c) Multplex8+ was performed on the liver biopsy, which confirmed the IHC results, although *ERBB2* was classified as low, and yielded a luminal B molecular subtype. (d) The low luminal signatures and low E2F4 signature were correct in predicting the lack of response to endocrine therapies and CDK4/6 inhibitors, respectively. Due to *ERBB2* low and high levels of TOP1, the payload target of trastuzumab deruxtecan (T-Dxd),

this therapy was predicted to be beneficial. However, after a secondary analysis, the lack of a durable response to T-Dxd could be explained by low signatures for other ADC markers, including CTSB (the cathepsin responsible for cleaving the tetrapeptide (GFLG) linker of T-Dxd), signatures comprising other cathepsins and lysosome function, and markers of endocytosis like CAV1, which is key for internalizing the HER2 receptor. Overall, this challenging case highlights the potential to use more than a single biomarker (e.g., HER2 or *ERBB2*) to determine response to novel ADCs like T-Dxd.

Detailed description of the invention

[0053] Definitions:

[0054] The term "ADC-associated biomarkers" or "Antibody-Drug-Conjugate-associated biomarkers" as used herein may refer to specific biological molecules (e.g., mRNA species) or biomarkers, including genes, mRNAs, proteins and/or measurable characteristics, that can be used for predicting the efficacy responsiveness, suitability of antibody-drug conjugates (ADCs) in therapeutic contexts, preferably in cancer treatment, and/or stratification a subject for treatment with ADCs and may include ADC-processing-related genes products. Such ADC-associated biomarkers may include one or more of the following: targets for the ADC antibody (e.g., HER2, CD30), biomolecules involved in ADC processing (e.g., endocytosis biomarkers, lysosomal activity biomarkers), pathways associated with ADC-resistance, payload interaction sites (e.g., tubulin-binding proteins for microtubule-disrupting payloads). Preferably, ADC-associated biomarkers of the present invention may include, but are not limited to: (i) ADC-processing-related genes and products; (ii) ADC-antigen targets; (iii) endocytosis; (iv) lysosomal function; (v) payload targets; (vi) resistance pathways; and/or (vii) linker; as decribed herein below (e.g., in Examples 1-6 and/or Figures 1-31 herein). Further preferably, ADC-associated biomarkers of the present invention may be selected from the group consisting of: Estrogen receptor (ESR1), Progesterone receptor (PGR), Receptor tyrosine-protein kinase erbB-2 (ERBB2) and Proliferation marker protein Ki-67 (MKI67). Because ADCs are primarily developed for cancer therapy, the term ADC-associated biomarkers as used herein may overlap with the term cancer-associated biomarkers as used herein. Preferably, ADC-associated biomarkers as used herein are not cancer-associated biomarkers. Other ADC-associated biomarkers are known in the art and available through resources known in the art, e.g., ADC Review (https://www.adcreview.com); and DrugBank (https://go.drugbank.com).

[0055] The term "cancer-associated biomarkers" as used herein may refer to specific biological molecules and/or biomarkers (e.g., mRNA species) that can be used in indicating the presence/diagnosis, survival, assessing status, risk and/or prognosis of cancer. Such cancer-associated biomarkers may include but are not imited to tumor antigens (e.g., HER2, EGFR), genetic mutations associated with cancer (e.g., TP53, BRCA1/2), pathway activation

markers (e.g., PI3K/AKT signaling). Preferably, "cancer-associated biomarkers" of the present invention are cancer-associated biomarkers as described herein below (e.g., in Examples 1-6 and/or Figures 1-31 herein). Other cancer-associated biomarkers are known in the art and available through resources known in the art, e.g., "The Cancer Genome Atlas (TCGA) (https://www.cancer.gov/ccg/research/genome-sequencing/tcga); "OncoKB" (https://www.oncokb.org/); COSMIC (Catalogue of Somatic Mutations in Cancer) (https://cancer.sanger.ac.uk/cosmic) etc. Accordingly, such platforms can be searched for ADC-associated biomarkers, specific ADCs, cancer-associated biomarkers, cancer types and/or biomarkers.

[0056] The term "biomarker" as used herein may refer to a biological molecule/s (e.g., RNA species), gene/s and/or measurable characteristic/s (e.g., gene expression signature/s) that can be used as an indicator of a biological state, condition (e.g., disease, e.g., cancer) and/or process. Examples of cancer-associated biomarkers include but are not limited to ERBB2, PD-L1 and TP53. Such biomarker/s may serve as a diagnostic, prognostic, predictive and/or therapeutic target in the context of diseases, e.g., in cancer. Preferably, biomarkers of the present invention are ADC-associated biomarkers and cancer-associated biomarkers biomarkers as described herein (e.g., in Examples 1-6 and/or Figures 1-31 herein).

[0057] The term "method for predicting the efficacy of an antibody-drug-conjugate (ADC)" as used herein may refer to a method evaluating the presence and expression levels of specific ADC-associated biomarkers in a biological sample obtained from a subject. Such method may combine advanced molecular techniques to determine the likelihood of a positive therapeutic response to an ADC based on the molecular profile of the tumor or disease-related tissue.

[0058] The term "method of cancer subtyping" as used herein may refer to a method of selecting/dividing (e.g., classifying) cancers into distinct subgroups based on shared characteristics, which may include genetic, molecular, cellular, histological and/or clinical features. Examples of cancer subtyping may include but is not limited to, subtyping based on hormone receptor status and HER2 expression (e.g., "Luminal A-like BCa", "Luminal B-like BCa", "HER2-enriched BCa", and "Triple-Negative BCa").

[0059] The term "method for stratification of a subject for treatment" as used herein may refer to a method of selecting/dividing (e.g., classifying) patients into subgroups or strata based on specific biological, molecular, genetic, clinical and/or demographic characteristics. Such stratification may enable the selection of the most appropriate therapeutic approach (e.g., ADC treatment) for the subject while ensuring optimal efficacy and/or minimizing possible adverse effects (e.g., of an ADC treatment).

[0060] The term "method for improving performance of a multi-gene signature test" as used herein may refer to a method of improving the accuracy, precision, reliability and/or predictive power of a multi-gene signature test, such as PAM50 or AIMS, by optimizing pre-analytical

and/or analytical steps. Preferably, such methods as described herein specifically focuse on incorporating advanced sample preparation techniques and/or methods of the present invention and/or integrating predictive models for assessing the efficacy of a targeted therapy, such as an antibody-drug-conjugate (ADC), in a specific clinical context (e.g., breast cancer).

[0061] The term "multi-gene signature test" as used herein may refer to a diagnostic and/or prognostic tool that assesses the expression levels of multiple genes within a sample, e.g., in order to generate a molecular profile associated with a specific cancer type.

[0062] The term "calculating an average overall score of said multi-gene signature test" as used herein may refer to a method/process of aggregating the expression levels and/or weighted contributions of individual genes included in a multi-gene signature panel to derive a single representative score. Such score may reflect the combined molecular profile of the biological sample and provide clinically significant information, such as tumor classification, prognosis, and/or prediction of therapeutic response.

[0063] The term "method for predicting survival and/or assessing status, risk and/or prognosis of cancer" as used herein may refer to a method/process combining molecular analysis methods to evaluate the expression of specific cancer-associated biomarkers in a biological sample obtained from a subject such that results may provide clinically information regarding the likelihood of survival, current state of the disease, potential risks and/or overall prognosis, particularly in cancers such as breast cancer.

[0064] The term "multiplex fluorescence in situ hybridization" as used herein may refer to a molecular diagnostic method for detecting and visualizing multiple nucleic acid sequences (e.g., RNA species, preferably mRNA) simultaneously in a biological sample. Such method may involve: hybridizing fluorescently labeled probes specific to the target nucleic acid sequences (e.g., RNA species, preferably mRNA) with the sample; using distinct fluorophores or combinations thereof to label the probes, enabling simultaneous detection of multiple targets; and visualizing the fluorescent signals through a fluorescence microscope or imaging system, e.g., where each signal corresponds to a unique target sequence.

[0065] The term "multiplex fluorescence in situ hybridization" as used herein may refer to a high-throughput sequencing method that may enable the simultaneous analysis of multiple nucleic acid sequences (e.g., RNA species, preferably mRNA) from one or more samples in a single sequencing run. For example, by utilizing unique barcodes or indexes for each target or sample this method may allow for efficient, parallel processing and identification of multiple molecular targets.

[0066] The term "laser capturing" as used herein may refer to a precise and targeted technique for isolating specific cells or regions of interest (e.g., cancer cells/tissues localized nd visualized by RNA-FISH) from a biological sample using a focused laser beam.

[0067] In the course of the present invention the analytical validity of a test using multiplexed RNA-FISH-guided laser capture microdissection coupled with RNA-sequencing (mFISHseq) with 93% accuracy compared to immunohistochemistry was demonstrated on a retrospective cohort of 1,082 breast samples (e.g., Example 1 herein).

[0068] In some aspects of the present invention, by combining three molecular classifiers for consensus subtyping, mFISHseq alleviated single sample discordance, provided near perfect concordance with other classifiers ($\kappa > 0.85$), and reclassified 30% of samples into different subtypes with prognostic implications.

[0069] In some aspects of the present invention, in multivariate Cox analyses on a curated list of 102 ADC processing-related genes/gene signatures, many markers were independent predictors of survival and showed subtype-specificity, thus highlighting potential ADC treatment-responsive subgroups. In some aspects of the pending invention, a classifier that predicted response to trastuzumab emtansine (AUC = 0.96) was better than *ERBB2* alone (e.g., Example 1 herein).

[0070] In other aspects, in the course of the present invention RNA sequencing was used for assessing ERBB2 expression levels. First, the cohort was separated into ESR1 positive and negative groups using a cutoff of 6.2 TPM (e.g., Example 2 herein), which resulted in 98.3% (674/686 samples) of samples also being IHC positive. In the ESR1 negative group, 78.5% (259/330) samples were IHC negative, while the majority (53.5%) of IHC positive samples contained low ER expression. Then, the samples in both ESR1 groups were stratified by using the interquartile range of ERBB2 TPM (e.g., Example 2 herein), resulting in ERBB2 positive (above 75th percentile), ERBB2 low (25th-75th percentile), and ERBB2 negative (below 25th percentile).

[0071] In some aspects, it was found that the genes that best differentiated *ERRB2* positive samples from *ERBB2* low and negative, irrespective of *ESR1* status, were located along the core of the Her2 amplicon (i.e., *ERBB2*, *PNMT*, *GRB7*, *TCAP*, *PGAP3*, *STARD3*, *MIEN1*) and are frequently co-amplified with *ERRB2*.

[0072] In some aspects, the Her2 amplicon was further explored and it was found that *ERRB2* low samples showed intermediate expression levels for many genes near the 17q12 locus when compared to either *ERRB2* positive or negative samples, a finding that was consistent across *ESR1* status (e.g., Example 2 herein). To see if expression of the Her2 amplicon could segregate *ERBB2* low from positive and negative samples, a Her2 amplicon module score was calculated and plotted each sample from highest to lowest score, which resulted in good separation of *ERBB2* positive, low, and negative samples (e.g., Example 2 herein).

[0073] In some aspects, these data suggest that ERBB2 expression, select differentially expressed genes, and Her2 amplicon signatures could be used as an aid to stratify tumors into

ERBB2 positive, low, and negative, ultimately resolving the ambiguity of IHC for low Her2 expression.

[0074] In some aspects the present invention relates to methods for predicting survival and/or assessing status, risk and/or prognosis of cancer (e.g., breast cancer), said method comprising: (a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of cancer-associated biomarker/s (e.g., as disclosed herein, e.g., in Example 3 herein) are present in a sample obtained from a subject; and (b) determining by multiplex sequencing whether or not mRNA species of cancer--associated biomarkers (e.g., as disclosed herein, e.g., in Example 3 herein) are present in said sample; (c) wherein a part of said sample in which said mRNA species of cancer-associated biomarkers was determined in step (a) is subjected to a laser capturing (e.g., as disclosed herein, e.g., as in Example 4 herein) prior to performing step (b); wherein the presence of one or more of said cancer-associated biomarker/s is indicative that the survival, risk and/or prognosis of cancer is low, and/or wherein the absence of one or more of said cancer-associated biomarker/s may be indicative that the survival, risk and/or prognosis of cancer is high.

[0075] In some further aspects, risk classifiers (e.g., 5 risk classifiers) are identified by the means of the methods disclosed herein.

[0076] In some aspects, while all 5 risk classifiers have prognostic utility at the population level, they show discordance when looking at individual patient samples.

[0077] In some aspects, patients with discordant results are identified with methods disclosed herein, wherein those patients tend to have survival that differs from those samples that are classified as a particular risk category by all classifiers.

[0078] In some aspects, such discordance at the patient level can be mitigated by constructing consensus categories identified by the methods disclosed herein that take into consideration results from all five classifiers.

[0079] In some aspects, patients are identified by the means of the methods disclosed herein that will benefit from chemoendocrine therapy (e.g., high risk) versus endocrine therapy alone (low risk) while also providing more accurate and precise results than any single test.

[0080] In some aspects, patients are identified by the means of the methods disclosed herein that are ultra-low risk, which are patients that have excellent outcomes and are prime candidates for de-escalating adjuvant treatment altogether – they may only need surgery to remove the tumor and can be spared both endocrine therapy and chemotherapy.

[0081] In some aspects, laser capturing was carried out as described herein. Laser capturing applied in the methods of the present invention is preferably laser capture microdissection (e.g., as disclosed in Example 4 herein).

[0082] Laser microdissection allows the isolation of, e.g., distinct cells from samples, in particular tissue section samples based on visual selection. Visual selection may be computer

and software aided. Visual selection includes, *inter alia*, slide scanning, RNA visualization, RNA biomarker quantification, spectral imaging, de-mixing, multiplexed visualization, whole-slide scanning, image analysis.

[0083] Means and methods for laser capture microdissection are, for example, provided by Arcturus or Leica (see, e.g. Hipp et al., J Path Inform 2018; 9, 45).

[0084] Preferably, signals, such as fluorescence signals, immunohistochemical signals or signals from hematoxylin and/or eosin may be detected by microscopy when multiplex FISH is performed. Such microscopy may be automated, computer and/or software aided.

[0085] The present inventors observed that laser capturing significantly improved the results obtained by the methods of the present invention. Particularly, they observed that laser capture microdissection reduces bias introduced by sampling heterogeneous tissues and ensures concordance between RNA FISH and RNA sequencing. Accordingly, it is preferred in the context of the methods of the present invention that, after having performed multiplex RNA FISH, that part(s) of the sample in which mRNA species of interest were determined by RNA FISH to be present is/are subject to laser capturing, in particular laser capture microdissection prior to performing multiplex sequencing of RNA. By that, those parts of the sample showing signals in RNA FISH for one or more ADC-associated biomarker/s are, so to say, concentrated and subjected to a potential confirmation step by performing multiplex RNA sequencing.

[0086] In some aspects of the present invention, it can be determined whether RNAs which was/were detected by multiplex RNA FISH may (again) be detected by multiplex RNA sequencing, thereby confirming the results obtained by multiplex RNA FISH. This two-step procedure avoids, e.g., false-positive and/or false-negative results obtained by RNA FISH and/or RNA sequencing, since RNAs which are detected by multiplex RNA FISH are required to be (re)detected by multiplex RNA sequencing.

[0087] In the context of performing the methods of the present invention, in particular laser capture microdissection, it is preferred that a sample, particularly a tissue section sample is placed on glass slides, preferably positively charged glass slides, frame slides, frame slides with PET membrane, or glass slides with membrane,

[0088] Laser capture microdissection is preferably done as described in the appended examples.

[0089] A preferred sample used in the methods of the present invention is a tissue section sample. The tissue section sample may be formaldehyde fixed, flash frozen (FFFF) or formaldehyde fixed (FF) and/or paraffin embedded (PE), preferably formaldehyde fixed, paraffin embedded (FFPE). Formaldehyde fixation is preferably done by formalin which is a 37% aqueous solution of formaldehyde in water.

[0090] Preferably, a sample when used in the context of the present invention is obtained by microtome sectioning of a paraffin embedded sample which is preferably additionally

formaldehyde fixed. Ideally and preferably, said sample obtained by microtome sectioning is after obtainment incubated at a temperature between 25°C to 0°C.

[0091] Indeed, the present inventors found that it is advantageous to place a sample, preferably obtained by microtome sectioning after its obtainment, ideally immediately, at a temperature between 25°C and 0°C, e.g., 25°C, 24°C, 23°C, 22°C, 21°C, 20°C, 19°C, 18°C, 17°C, 16°C, 15°C, 14°C, 13°C, 12°C, 11°C, 10°C, 9°C, 8°C, 7°C, 6°C, 5°C, 4°C, 3°C, 2°C, 1°C, or 0°C, more preferably, 10°C, 9°C, 8°C, 7°C, 6°C, 5°C, 4°C, 3°C, 2°C, 1°C, or 0°C.

[0092] Preferably, an ice-water bath may be used for incubating the sample at such a temperature. Microtome sectioning is preferably done as described in the appended examples or known in the art.

[0093] As described herein, when placing a sample, preferably a tissue section sample, preferably formaldehyde fixed, paraffin embedded which is preferably obtained by microtome sectioning, after obtainment at a temperature between 25°C to 0°C, achieved by an ice-cold water bath, diffusion of RNA, in particular small RNA, e.g., miRNA out of cells comprised by the sample is prevented, thereby ensuring high concentrations of RNA, in particular, small RNAs. A sample which is preferably processed as described herein can be placed after its obtainment through microtome sectioning at a temperature between 25°C and 0°C, while samples prepared in accordance with the teaching of the prior art were not placed at such a cold temperature. In fact, the prior art does not teach placing a sample obtained through microtome sectioning at such a temperature. Rather, prior art teaches to immediately place a sample at a temperature of about 45°C to 42°C to flatten the microtome sectioned sample.

[0094] If a sample which may be paraffin embedded is to be diagnosed by the methods of the present invention, it is preferably deparaffinized prior to performing step(a) and/or step(b) of the methods as described herein. Deparaffinization is preferably achieved as described herein, in particular in the appended examples.

[0095] The present invention offers novel personalized methods/tests relating to but not limited to cancer (e.g., breast cancer (BCa)) which assesses the expression level of multiple biomarkers (e.g., mRNA). It combines visualization techniques (FISH) and subsequent multi-sample sequencing technologies in one single test with a high level of specificity and sensitivity. This reduces analysis biases and allows for simultaneous assessment of multiple RNA biomarkers (for example for BCa, e.g., 9 up to 300 or more biomarkers) on a single patient's sample. The method of the present invention combines two independent qualitative/quantitative technologies, allowing cross-validation of biomarkers analysis, and eliminating or drastically reducing diagnostic errors. A specific personalised treatment and therapy length may then be suggested, as well as the probability of recurrence, which will help oncologists design a personalized and effective treatment. The inventive method described

and provided herein may be applied for different diseases, particularly cancer, preferably any solid tumor (e.g., breast cancer) and any stage of cancer.

[0096] As used herein and as known in the art, the term “mRNA” is used interchangeably with “messenger RNA” and means RNA molecules that convey genetic information from DNA to the ribosome, where they specify the amino acid sequence of the protein products of gene expression. mRNAs are unique for different genes and may exhibit specific expression patterns in different kinds of tissue such as, e.g., tumor tissue.

[0097] Generally, as used herein, the terms „polynucleotide“, „nucleic acid“ or „nucleic acid molecule“ are to be construed synonymously. Generally, nucleic acid molecules may comprise *inter alia* DNA molecules (including cDNA, complementary DNA), RNA molecules (e.g., miRNA, mRNA, rRNA, tRNA, snRNA, siRNA, scRNA, snoRNA, and others as known in the art), LNA (locked nucleic acid) molecules (see, e.g., Kaur et al., Biochem (2006), 45(23): 7347-7355), oligonucleotide thiophosphates, substituted ribo-oligonucleotides or PNA (peptide nucleic acid) molecules. Furthermore, the term "nucleic acid molecule" may refer to DNA or RNA or hybrids thereof or any modification thereof that is known in the art (see, e.g., US 5525711, US 471 1955, US 5792608 or EP 302175 for examples of modifications). The polynucleotide sequence may be single- or double- stranded, linear or circular, natural or synthetic, and without any size limitation. For instance, the polynucleotide sequence may be genomic DNA, cDNA, mitochondrial DNA, mRNA, antisense RNA, ribosomal RNA or a DNA encoding such RNAs or chimeroplasts (Gamper, Nucleic Acids Research, 2000, 28, 4332 - 4339). Said polynucleotide sequence may be in the form of a vector, plasmid or of viral DNA or RNA. Also described herein are nucleic acid molecules, which are complementary to the nucleic acid molecules described above and nucleic acid molecules, which are able to hybridize to nucleic acid molecules described herein. A nucleic acid molecule described herein may also be a fragment of the nucleic acid molecules in context of the present invention. Particularly, such a fragment is a functional fragment. Examples for such functional fragments are nucleic acid molecules, which can serve as primers.

[0098] As used herein, nucleic acid molecules may comprise different types of nucleotides, comprising naturally occurring nucleotides, modified nucleotides, and artificial nucleotides. Nucleotides as used herein generally comprise nucleosides, naturally occurring nucleosides, modified nucleosides, and artificial nucleosides. As known in the art, naturally occurring nucleosides comprise purine bases or pyrimidine bases. Examples for naturally occurring nucleosides comprise (deoxy)adenosine, (deoxy)guanosine, (deoxy)uridine, thymidine, and (deoxy)cytidine. Nucleosides as part of nucleotides (and, thus, nucleic acid molecules) as described herein may generally encompass structures comprising any purine or pyrimidine nucleoside and derivatives or analogues thereof. That is, “purine nucleoside” or “pyrimidine nucleoside” as used in context with the present invention generally comprises any kind of

purine or pyrimidine as well as derivatives or analogues thereof as described herein respectively, as well as a sugar, e.g., a pentose. In one embodiment of the present invention, the purine nucleoside may be selected from the group consisting of (deoxy)adenosine, inosine, and (deoxy)guanosine and derivatives or analogues thereof. A derivative may be, e.g., a nucleoside with a purine selected from the group consisting of a deazapurine, an azidopurine, an alkylpurine, a thiopurine, a bromopurine, an O-alkylpurine, and an isopurine, for example a deazapurine such as, e.g., 7-deazapurine. That is, in one aspect of the present invention, the purine nucleoside may be a nucleoside with a purine selected from the group consisting of a deazapurine, an azidopurine, an alkylpurine, a thiopurine, a bromopurine, an O-alkylpurine, and an isopurine, for example a deazapurine such as, e.g., 7-deazapurine. In another aspect of the present invention, the purine nucleoside may be selected from the group consisting of 1-methyl(deoxy)adenosine, 2-methyl-(deoxy)adenosine, N^6-methyl(deoxy)adenosine, N^6,N^6-dimethyl(deoxy)adenosine, 7-deaza(deoxy)adenosine, 7-deaza-8-aza(deoxy)adenosine, 7-deaza-7-bromo(deoxy)adenosine, 7-deaza-7-iodo(deoxy)adenosine, 8-azido(deoxy)adenosine, 8-bromo(deoxy)adenosine, 8-iodo(deoxy)adenosine, 8-bromo-2'-deoxy(deoxy)adenosine, 2'-O-methyladenosin, inosin, 1-methylinosin, 2'-O-methylinosin, 1-methyl(deoxy)guanosine, 7-methyl(deoxy)guanosine, N^2-methyl(deoxy)guanosine, N^2,N^2-dimethyl-guanosine, isoguanosine, 7-deaza(deoxy)guanosine, 7-deaza-8-aza(deoxy)guanosine, 7-deaza-7-bromo(deoxy)guanosine, 7-deaza-7-iodo(deoxy)guanosine, 6-thio(deoxy)guanosine, O^6-methyl(deoxy)guanosine, 8-azido(deoxy)guanosine, 8-bromo(deoxy)guanosine, 8-iodo(deoxy)guanosine, 2'-O-methylguanosine, 8-azidoinosine, 7-azainosine, 8-bromoinosine, 8-iodoinosine, 1-methylinosine, and 4-methylinosine. In a further aspect of the present invention, the purine nucleosides may be selected from the group consisting of a queuosine, an archaeosine, a wyosine and a N^6-threonylcarbamoyladenosine. In one aspect of the present invention, the pyrimidine nucleoside may be selected from the group consisting of (deoxy)cytidine, (deoxy)thymidine, (deoxy)ribothymidine, (deoxy)uridine, and derivatives thereof. A derivative may be, e.g., a nucleoside with a pyrimidine selected from the group consisting of an alkylpyrimidine, a thiopyrimidine, a bromopyrimidine, an O-alkylpyrimidine, an isopyrimidine, an acetylpyrimidine hydropyrimidine, and a pseudopyrimidine. That is, in one aspect of the present invention, the pyrimidine nucleoside may be a nucleoside with a pyrimidine selected from the group consisting of an alkylpyrimidine, a thiopyrimidine, a bromopyrimidine, an O-alkylpyrimidine, an isopyrimidine, an acetylpyrimidine hydropyrimidine, and a pseudopyrimidine. In another aspect of the present invention, the pyrimidine nucleoside may be selected from the group consisting of 3-methyl-(deoxy)cytidine, N^4-methyl(deoxy)cytidine, N^4,N^4-dimethyl(deoxy)cytidine, iso(deoxy)cytidine, pseudo(deoxy)cytidine, pseudoiso(deoxy)cytidine, 2-thio(deoxy)cytidine, N^4-acetyl(deoxy)cytidine, 3-methyl(deoxy)uridine, pseudo(deoxy)uridine, 1-methyl-

pseudo(deoxy)uridine, 5,6-dihydro(deoxy)uridine, 2-thio(deoxy)uridine, 4-thio(deoxy)uridine, 5-bromodeoxy(deoxy)uridine, 2'-deoxyuridine, 4-thio(deoxy)thymidine, 5,6-dihydro(deoxy)thymidine, O^4-methylthymidine, difluortoluene, and other nucleobase surrogates. As mentioned, the nucleosides as described and provided herein generally comprise a purine or pyrimidine or derivative or analogue thereof as described herein as well as sugar moiety such as, e.g., a pentose. Generally, the pentose as part of the purine or pyrimidine nucleoside or derivative or analogue thereof as described herein may be, *inter alia*, ribose, deoxyribose, arabinose, or methylribose (2-*O*-methyribose), for example, a ribose or a deoxyribose. That is, the nucleoside may be, e.g., a (ribosyl)nucleoside, a desoxy(ribosyl)nucleoside, an arabinosylnucleoside or an (methylribosyl)nucleoside, for example a (ribosyl)nucleoside or a deoxy(ribosyl)nucleoside.

[0099] As used herein, the terms "desoxy" and "deoxy" as prefixes of molecule terms are used synonymously and indicate the absence of an oxygen atom or a hydroxyl-group, e.g., in a given pentose such as ribose or others.

[00100] In context with the present invention, the presence of ADC-associated biomarker/s in step(a) and step(b) is preferably indicative of the efficacy of said ADC.

[00101] In a further embodiment of the present invention, the method described and provided herein allows predicting the efficacy of an Antibody-Drug-Conjugate (ADC) based on the presence or absence of said ADC-associated biomarker/s.

[00102] Generally, in context with the present invention, the efficacy to be predicted with the method described and provided herein may be any disease which is connected to the presence of specific biomarkers detectable with FISH and which can be sequenced. Accordingly, in a preferred embodiment of the present invention, such biomarkers are nucleic acid molecules, preferably miRNA or mRNA, most preferably mRNA molecules specific for the respective disease.

[00103] In one embodiment of the present invention, the disease is cancer. In this context, the cancer may be any type of cancer, for example it is a solid cancer. Specifically, if the disease is cancer, it may be, e.g., solid cancer (e.g., on-small cell lung cancer, breast cancer, colorectal cancer, pancreatic cancer, ovarian cancer, skin cancer, prostate cancer, cancer of the brain or nervous system, head and neck cancer, testicular cancer, lung cancer, liver cancer, kidney cancer, bladder cancer, gastrointestinal cancer, bone cancer, cancer of the endocrine system, cancer of the lymphatic system, fibrosarcoma, neurectodermal tumor, mesothelioma, epidermoid carcinoma, or Kaposi's sarcoma), and blood cancer (e.g., leukemias). In a specific embodiment of the present invention, the disease is breast cancer (BCa).

[00104] In accordance with the present invention, the mRNA species to be detected by the method described and provided herein may be any mRNA whose elevated or reduced

occurrence or expression is specific for a certain disease to be diagnosed as described herein. Its elevated or reduced occurrence or expression may also be typical for specific diseased tissue, e.g., tumor tissue (for cancer).

[00105] For example, for breast cancer, typical mRNA species (e.g., breast cancer-associated biomarkers) to be detected in accordance with the method of the present invention may include those specific for HER2 (HER2, GRB7), Proliferation (Ki-67, STK15, Survivin, Cyclin B1, MYBL2), Oestrogen (ER, PR, Bcl2, SCUBE2), Invasion (Stromelysin 3, Cathepsin L2), and others (GSTM1, BAG1, CD68), where HER2, ER and PR represent the most relevant markers in accordance with the present invention. Further mRNA markers whose presence may be determined in accordance with the method of the present invention may include: HER2, ER, PR (the most relevant breast cancer diagnosis/prognosis transcripts), NUPR1 (marker for taxol resistance), and CSF1 (marker for invasion in triple negative breast cancer cases). Accordingly, in one embodiment of the present invention, the mRNA species whose presence may be determined in accordance with the present invention may be selected from the group consisting of HER2, ER, PR, NUPR1, CSF1, GRB7, Ki-67, STK15, Survivin, Cyclin B1, MYBL2, ER, PR, Bcl2, SCUBE2, Stromelysin 3, Cathepsin L2, GSTM1, BAG1, and CD68, preferably including HER2, ER, and PR. Of course, in the methods of the present invention any mRNA which is an ADC-associated biomarker/s (e.g., disease-associated biomarker/s) can be assayed for its elevated or reduced occurrence as described herein.

[00106] Furthermore, in accordance with the present invention, the presence of control markers may be determined by multiplex FISH in step(a). In one embodiment of the present invention, the presence of additionally one or more RNAs selected from the group consisting of 28S rRNA, poly(A) RNA, Beta-actin, GAPDH, RPLPO, GUS, and TFRC may be determined by FISH as control.

[00107] In context with the present invention, in a first step of the method described and provided herein, multiplex fluorescence in situ hybridization (FISH) is used to determine whether or not mRNA species of ADC-associated biomarker (e.g., disease-associated biomarkers) are present in a sample obtained from a subject. As used herein, “multiplex FISH” is a FISH assay as known to the person skilled in the art (cf., e.g., Lee et al., RNA (2011), 17(6): 1076-1089) where the presence of an mRNA and/or at least one, two, three, four, five, six, seven, eight, nine, ten, or more than ten different miRNA species is determined. In one embodiment of the present invention, the presence of several mRNAs, scRNAs, snRNAs, rRNAs, other RNAs, and/or at least one, two, three, four, five, six, seven, eight, nine, ten, or more than ten miRNA species is determined by “multiplex” FISH assay simultaneously. In accordance with the present invention, simultaneous FISH measurement of different RNA species may result in multi-color FISH as different RNA species may be detected using different dye labels.

[00108] Preferably, multiplex RNA-FISH is done as described in the appended examples.

[00109] For example, particularly where the ADC-response to be predicted by the inventive method is breast cancer, in one embodiment of the present invention, the presence of at least one, two, three, four, five, six, seven, eight, nine, ten, or more than ten mRNAs (HER2, ER, PR) is determined simultaneously.

[00110] Generally, in context with the present invention, the methods described and provided herein is performed starting from a patient's tissue (e.g., cancer tissue, such as BCa tissue) removed during biopsy or surgery. Tissues are first analysed by multiplex FISH assay allowing the simultaneous visualization and quantification of several (up to 9-300 in BCa) representative biomarkers. The assay is based on fluorescently labelled nucleic-acid probes (available from, e.g., MultiplexDX®) which recognize selected mRNA or miRNA targets. The hybridization of FISH probes to RNAs allows the visualization of the biomarkers within single cells by common fluorescent microscopy. In one embodiment of the present invention, the present invention may use "click labelling" to increase the binding affinity of FISH probes based on the chemical fusion of two closely neighbouring probes on a specific RNA target. Accordingly, it is preferred that probes used for fluorescence in situ hybridization (FISH) of RNA may be conjugated by click chemistry. Click chemistry cross-linking of probes is preferably done as described in the appended examples. In accordance with the present invention, the specificity of FISH probes allows the detection of multiple biomarkers (multi-color FISH) simultaneously (multiplex FISH) and precise biomarkers' quantification derived from FISH signal intensity.

[00111] As known to the person skilled in the art, in FISH, any relatively stable mis-hybridization to abundant RNAs will result in false-positive signals which reduce specificity. Although probe hybridization in FISH assays cannot be directly compared to hybridization of primers or siRNAs in PCR or RNAi, respectively, it is clear that during PCR, the end of primers, whose perfect binding is necessary for starting the enzymatic chain reaction, is the most sensitive for any mismatch or bulge within primer-target duplex resulting in an aberrant PCR reaction. While 1% mis-hybridization of siRNAs to abundant RNAs such as ribosomal RNA (rRNA) is irrelevant, such a level of mis-hybridization to abundant off-target transcripts in RNA FISH would result predominantly in rRNA signal (since rRNA is at least 1000-100,000 times more abundant than mRNAs, 1% mis-hybridization to rRNA would be up to 1,000 times higher than target mRNA signal). Thus, probe mis-hybridization remains a large technological barrier.

[00112] Probes to be employed in the inventive method described herein may be of any kind suitable for application in FISH as known in the art. Preferably, the probe is based on a nucleic acid molecule. In one embodiment of the present invention, the probe is an LNA molecule. The probe can be of any length, dependent on the RNA molecule whose presence

is to be determined according to the method of the present invention. For example, in one embodiment of the present invention, the probe may be up to 40 nucleotides in length, preferably up to 35, 30 or 20 nucleotides. In an additional embodiment of the present invention, the probe is at least 7 nucleotides in length, preferably at least 8, 9 or 10 nucleotides.

[00113] Probes for RNA detection in FISH can be designed as commonly known in the art. For example, in accordance with the present invention, the following parameters may be pre-set: 1. A target sequence (RNA of interest), 2. Hybridization temperature of RNA FISH protocol, and, 3. The desired number of FISH probes per target. A list of sequences that can be subsequently inserted into the synthesis program of a DNA synthesizer may then be generated by a program provided by MultiplexDX® or other commercially available programs. Based on predicted hybridization behaviors and the least cross-hybridization to the most abundant RNAs (e.g. rRNAs, ncRNAs, tRNAs, mt-RNAs, scRNAs, etc.), an appropriate sequence for the probe may be selected. In accordance with the present invention, optimal probes may fulfil the following characteristics: (a) minimized cross-hybridization to the most abundant RNAs, and (b) equalized LNA/DNA probe melting temperatures (T_Ms)/binding affinities by varying probe lengths, exact positioning of probes, and LNA content per probe. Based on the chosen hybridization conditions, RNA FISH probes may be, e.g., 11- to 15-nt long locked nucleic acid (LNA)-modified DNA probes (LNA/DNA) with ~3 to 5 LNA modifications per probe. T_M prediction may be based on algorithms currently used by, e.g., IDT DNA oligo analyzer for DNA/DNA, DNA/RNA, and LNA/DNA duplexes, taking target type (DNA or RNA), oligo concentration, and Na^+ concentration as the main input parameters. Optionally, this may be modified in accordance with the present invention for 50% Formamide, which may be used in RNA FISH protocols as a denaturant.

[00114] For the multiplex FISH as to be employed in context with the method of the present invention, for example short, fluorescent, multi-labeled LNA/DNA probes (about 11-15 nt long, about 4-8 fluorophores per probe, about 20-50 probes per target) may be used for visualization of any RNA target (up to 5 kb). The probes may be, e.g., synthesized, deprotected, and desalted on an automated Dr. Oligo 48 (cf. http://www.biolytic.com/t-dr-oligo-48-dna-rna-oligo-synthesizer.aspx) or Dr. Oligo 768 (cf. https://www.biolytic.com/t-dna-rna-oligo-synthesizer.aspx) DNA/RNA synthesizer or other means known in the art, each with several desired modifications if required. Each fluorescent multi-labeled probe (e.g., about 11-15 nt) may be prepared with an elongation segment containing multiple (4-8) 5-Octadiynyl-dU or Abasic-Alkyne or other suitable dye, each separated by, e.g., 5-nt segment or C18 spacer (GlenResearch). In accordance with this invention, this allows very powerful, selective and efficient click chemistry labelling with azide-labeled fluorophores or any fluorophore of interest (e.g., ATTO-, Alexa-, Cy, BODIPY-, DyLight-, Cyto-, Seta-, RadiantDy- and CF-dyes selecting some of the brightest, the most photo- and water-soluble, avoiding unspecific protein binding).

Click chemistry, due to its robust specificity and quantitative efficiency when compared to NHS-ester chemistry, is in accordance with the present invention a suitable approach to enable multiple labelling of the probes with 5 or more fluorophores per probe. Comparing click chemistry labeling to NHS-ester labeling, even though fluorophore-azides are 33-50% more expensive than fluorophore-NHS esters, in accordance with the present invention usually only 1.2X excess of azide may be used, but about, e.g., 10X excess of NHS-ester in labeling reaction with oligo, making click chemistry more cost-efficient.

[00115] In accordance with the present invention, the multiplex FISH as to be employed in the inventive method described and provided herein, may be for example performed as follows. In one embodiment, tissue cells may be fixed prior to detection of biomarkers as described herein. Suitable methods for cell and tissue fixation and subsequent biomarker detection are known in the art and also described in, e.g., US 9359636, US 9005893 and US 8394588. In accordance with the present invention, by *in situ* click ligation of two closely neighboring probes (ca. 15-nt long), a more stable 30-nt long double-probe may be formed, followed by mishybridization-targeting stringent wash, which will wash-off mishybridized probes completely and reduce background while preserving perfect specific binding of all probes (as depicted in Examples herein). This allows to achieve specific RNA FISH signals of even low abundant RNA targets.

[00116] As described above, it is preferred that a sample is fixed prior to performing step(a) of the methods as described herein. Fixation may preferably be achieved by treatment with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) or 5-ethylthio-1H-tetrazole (ETT). That said, fixation is preferably done as described herein, in particular in the appended examples.

[00117] As regards click-labelling as described in context with the present invention, for example, one probe may be labeled with azide (through 3'-amino-modified LNA/DNA oligonucleotide), and another may be tagged with alkyne on the 5'-end, directly labeled in oligonucleotide synthesis. Each probe may preferably be labelled with the same fluorophore. Then, click chemistry may be performed at room temperature to conjugate the probes. This significantly increases binding affinity of the formed double-probe. This in turn may enable use of mishybridization-targeting stringent wash, which washes-off mishybridization coming from 15-nt long probes completely by increasing temperature of the wash above previous hybridization temperature (e.g., if hybridization temperature is 40 ºC, then the stringent wash may be 45-50ºC). Increased stability of the double-probe following conjugation and the stringent wash may preserve perfect and specific binding of all probes. In accordance of the present invention, multiplex FISH approach may be performed as follows: 1. RNA FISH using 15nt long LNA/DNA probes at 40 °C; 2. Cool down and click-chemistry performed at room

temperature; 3. Stringent wash at 45-50 °C (5-10 °C above previous hybridization temperature).

[00118] As set forth herein, the multiplex FISH to be employed in context with the inventive method described and provided herein allows simultaneous determination of the presence of several RNA species as described herein. For example, in accordance with the present invention, an 8-color multiplex FISH may be applied as follows.

[00119] Using spectral imaging with linear unmixing, it may be examined whether all individual signals within the entire spectrum (sum of all spectra) must be near-equal, which appears to be one of the most relevant limitations of current methods.

[00120] The term "hybridization" or "hybridizes" as used herein in context with nucleic acid molecules/ sequences incl. miRs, mRNAs, DNAs, cDNAs, LNAs, and others as described herein, or a portion or fragment thereof may relate to hybridizations under stringent, low stringent or non-stringent conditions. In one embodiment, the conditions are preferably stringent. Said hybridization conditions may be established according to conventional protocols described, for example, in Sambrook, Russell "Molecular Cloning, A Laboratory Manual", Cold Spring Harbor Laboratory, N. Y. (2001); Current Protocols in Molecular Biology, Update May 9, 2012, Print ISSN: 1934-3639, Online ISSN: 1934-3647; Ausubel, "Current Protocols in Molecular Biology", Green Publishing Associates and Wiley Interscience, N. Y. (1989), or Higgins and Hames (Eds.), "Nucleic acid hybridization, a practical approach", IRL Press Oxford, Washington DC, (1985). The setting of conditions is well within the skill of the artisan and can be determined according to protocols described in the art. Thus, the detection of only specifically hybridizing sequences will usually require stringent hybridization and washing conditions such as 0.1 x SSC, 0.1% SDS at 65 °C ("stringent conditions" as used herein). Non-stringent hybridization conditions for the detection of homologous or not exactly complementary sequences may be set at 6 x SSC, 1% SDS at 65 °C ("non-stringent conditions" as used herein). As is well known, the length of the probe and the composition of the nucleic acid to be determined constitute further parameters of the hybridization conditions. Variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility. In accordance to the invention described herein, low stringent hybridization conditions for the detection of homologous or not exactly complementary sequences may, for example, be set at 6 x SSC, 0.5% SDS at 65 °C ("low stringent conditions" as used herein). As is well known, the length of the probe and the

composition of the nucleic acid to be determined constitute further parameters of the hybridization conditions.

[00121] Hybridizing nucleic acid molecules also comprise fragments of the above described molecules. Such fragments may represent nucleic acid molecules serving as inhibitors as described herein or a functional fragment thereof. Furthermore, nucleic acid molecules, which hybridize with any of the aforementioned nucleic acid molecules, also include complementary fragments, derivatives and variants of these molecules. Additionally, a hybridization complex refers to a complex between two nucleic acid sequences (e.g., cDNA/LNA) by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T (or U for RNA as known to the skilled person) bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (e.g., Cot or Rot analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., membranes, filters, chips, pins or glass slides to which, e.g., cells have been fixed). The terms “complementary” or “complementarity” refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-U" binds to the complementary sequence "U-C-A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between single-stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands.

[00122] Having performed RNA FISH as described herein, it is preferred that a sample is stained with hematoxylin and eosin after FISH of RNA was performed.

[00123] As described herein, in the context of the methods of the present invention, *inter alia*, 28S rRNA and poly(A) RNA are determined by FISH as control. This control, i.e., 28S rRNA and poly(A) RNA are of particular interest as control, since the ratio between poly(A) RNA and 28S rRNA is indicative of the potential degradation of RNA in a sample. Accordingly, it is preferred that prior to performing step(a) and/or step(b) of said method, said sample is checked for the ratio between poly(A) (m)RNA and 28S rRNA.

[00124] Indeed, the present inventors observed that a quality control of a sample as described herein prior to performing multiplex RNA FISH and/or multiplex RNA sequencing may be helpful. Surprisingly, the present inventors found that effects on RNA degradation are reflected by the ratio of PolyA (m)RNA/28S rRNA. Preferably, the ratio (fluorescence activity/units) of PolyA (m) RNA/28S rRNA can be used to assess RNA degradation in

samples, e.g., formalin-fixed paraffin-embedded (FFPE) tissue section samples, with ratios below 1 signifying degraded RNA (as a perhaps ideal ratio 1.15).

[00125] Accordingly, it is preferred that prior to performing multiplex RNA FISH (step (a) as referred to herein) and/or multiplex RNA sequencing (step(b) as referred to herein) in the context of the methods of the present invention, a sample as described herein is checked whether the ratio between polyA (m)RNA and 28S rRNA is 1:1 or greater than 1:1, e.g. 1.05 : 1, 1.1 : 1, 1.15 : 1, 1.2 : 1, 1.3 : 1, 1.4 : 1, 1.5: 1, 2 : 1 or 3: 1.

[00126] However, in the alternative, it is preferred that prior to performing multiplex RNA FISH (step (a) as referred to herein) and/or multiplex RNA sequencing (step(b) as referred to herein) in the context of the methods of the present invention, a sample as described herein is checked whether the ratio between polyA (m)RNA and 28S rRNA is 1:1 or lower than 1:1, e.g. 1 : 1.05, 1 : 1.1, 1 : 1.15, 1 : 1.2, 1 : 1.3, 1 : 1.4, 1 : 1.5, 1 : 2, or 1 : 3.

[00127] As described herein, according to the method provided by the present invention, following the multiplex FISH step, a multiplex sequencing step follows (step(b)).

[00128] Accordingly, after the FISH step, a sequencing step can be performed in order to confirm the result of step(a) (multiplex FISH) and/or to determine the presence of additional biomarkers indicative for the respective disease.

[00129] That is, in context with the present invention, the same biomarkers may be sequenced in step(b) whose presence has been determined in step(a), or other or additional biomarkers than those of step(a) may be sequenced in step(b). In one embodiment of the present invention, the same biomarkers may be sequenced in step(b) whose presence has been determined in step(a). The sequencing of the biomarkers may be performed by methods known in the art and as described and exemplified herein. For example, in accordance with the present invention, after tissue has been stained in step(a) during the FISH assay, and where the disease to be diagnosed is cancer (e.g., breast cancer), the tumor may be selectively isolated from normal tissue and immune cells by, e.g., laser micro-dissection. From the homogenous population of cancer cells, total RNA may then be extracted and subjected to RNAseq (cf. Cánovas et al., Sci Reports (2014), 4 (art. no. 5297), doi: 10.1038/srep05297) to confirm quantification of the first analysis (visualization in multiplex FISH of step(a)) and to reveal the presence and quantity of the same or even a much broader panel of specific biomarkers. Accordingly, in the context of the methods of the present invention, it is preferred that RNA is extracted from a sample when multiplex sequencing of RNA is performed. Preferably, RNA extraction is done as described in the appended examples. The sequencing of different RNA markers as described may be based on suitable cDNA libraries (DNA synthetized from RNA population used as template for sequencing) known in the art (e.g., Illumina TrueSeq cDNA library, or ClontechSMART cDNA library as described, e.g., in https://www.takarabio.com/products/cdna-synthesis/cdna-synthesis-kits/library-construction-

kits). Sequencing may generally be done by RNA sequencing methods known in the art and as *inter alia* described in, e.g., Illumina TrueSeq cDNA library preparation protocol (https://support.illumina.com/content/dam/illumine-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqrna/truseq-rna-sample-prep-v2-guide-15026495-f.pdf). Preferably, RNA sequencing is done as described in the appended examples.

[00130] As described above, in the context of multiplex sequencing of RNA, a cDNA library may be prepared. Accordingly, it is preferred that for the preparation of a cDNA library 3'-adapter molecules comprising a barcode are used. Preferably, these 3'-adapter molecules comprising a barcode are preadenylated. It is also preferred that these 3'-adapter molecules comprising a barcode are a collection. These 3'-adapter molecules comprising a barcode, which may preferably be preadenylated, are either pooled or unpooled, with pooled being preferred. To be more precise, the collection of these 3'-adapter molecules comprising a barcode, which may preferably be preadenylated, may preferably be pooled or unpooled, with a pooled collection being preferred. Such a collection comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or more 3'-adapter molecules comprising a barcode and which may preferably be preadenylated.

[00131] Indeed, much to their surprise, the present inventors observed when using 3'-adapter molecules for the preparation of a cDNA library in the context of performing multiplex RNA sequencing that 3'-adapter molecules comprising a barcode reduce ligation and technical biases in RNA sequencing. In fact, in contrast to the present invention, the preparation of cDNA libraries for RNA sequencing in the prior art is made by using 3'-adapter molecules not having a barcode. For cDNA library preparation as done in the prior art, the barcode is only ligated later during the preparation of the cDNA library.

[00132] cDNA libraries can be generated using standard 3'-adapters (sample I-1 – I-3) when compared to cDNA libraries using 3'-adapters comprising a barcode, e.g., pooled MDX barcoded pre-adenylated 3'-adapters) (MDX-4a-c) unpooled MDX barcoded pre-adenylated 3'-adapters (MDX-7 – MDX-9) do not reduce ligation bias as shown by density estimates that are closer to the expected read counts (denoted as dotted line at 0), whereas 3'-adapters used in the context of the methods of the present invention, which comprise a barcode and which may preferably be preadenylated, reduce ligation bias as shown by density estimates that are closer to the expected read counts (denoted as dotted line at 0). Hence, the 3' adapters of the present invention are advantageous for the preparation of a cDNA library when performing multiplex RNA sequencing. This advantage could not have been expected.

[00133] The present invention further relates to a kit for performing the method as described and provided herein.

[00134] Preferably, the kit comprises glass slides, preferably positively charged glass slides, frame slides, frame slides with PET membrane, or glass slides with membrane. It is also preferred that the kit further comprises formaldehyde and/or paraffin.

[00135] The kit may preferably also comprise EDC or ETT and/or hematoxylin and/or eosin.

[00136] The kit may preferably also comprise siliconized nuclease-free containers.

[00137] Moreover, the kit preferably further comprises probes for fluorescence in situ hybridization of RNA and/or 3'-adapter molecules comprising a barcode for preparing a cDNA library from RNA. These 3'-adapter molecules comprising a barcode which may be comprised by the kit are preferably preadenylated.

[00138] The probes for RNA FISH preferably comprised by the kit may preferably be specific for HER2, ER, PR, NUPR1, CSF1, GRB7, Ki-67, STK15, Survivin, Cyclin B1, MYBL2, ER, PR, Bcl2, SCUBE2, Stromelysin 3, Cathepsin L2, GSTM1, BAG1, and CD68.

[00139] Also, the probes for RNA FISH preferably further comprised by the kit may preferably be specific for miR-21, miR-29a, miR-221, and miR-375.

[00140] Similarly, the probes for RNA FISH preferably further comprised by the kit may preferably be specific for U2 snRNA and/or 7SL scRNA.

[00141] Preferably, as control, the kit may also comprise probes which are specific for 28S rRNA, poly(A) RNA, Beta-actin, GAPDH, RPLPO, GUS, TFRC.

[00142] It is also preferred that the 3'-adapter molecules comprising a barcode which may be comprised by the kit are a collection of 3'-adapter molecules comprising a barcode. Preferably, the 3'-adapter molecules comprising a barcode comprised by the kit are pooled or unpooled, with unooled being preferred.

[00143] The present invention further relates to the use of such a kit.

[00144] The embodiments, which characterize the present invention, are described herein, shown in the Figures, illustrated in the Examples, and reflected in the claims.

[00145] It must be noted that as used herein, the singular forms "a", "an", and "the", include plural references unless the context clearly indicates otherwise. Thus, for example, reference to "a reagent" includes one or more of such different reagents and reference to "the method" includes reference to equivalent steps and methods known to those of ordinary skill in the art that could be modified or substituted for the methods described herein.

[00146] Unless otherwise indicated, the term "at least" preceding a series of elements is to be understood to refer to every element in the series. Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the present invention.

[00147] The term "and/or", wherever used herein, includes the meaning of "and", "or" and "all or any other combination of the elements connected by said term".

[00148] The term "about" or "approximately" as used herein means within 20%, preferably within 10%, more preferably within 5%, and most preferably within 3% of a given value or range.

[00149] Throughout this specification and the claims which follow, unless the context requires otherwise, the word “comprise”, and variations such as “comprises” and “comprising”, will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integer or step. When used herein the term “comprising” can be substituted with the term “containing” or “including” or sometimes when used herein with the term “having”.

[00150] When used herein “consisting of" excludes any element, step, or ingredient not specified in the claim element. When used herein, "consisting essentially of" does not exclude materials or steps that do not materially affect the basic and novel characteristics of the claim.

[00151] Unless specifically stated otherwise, in each instance herein any of the terms "comprising", "consisting essentially of" and "consisting of" may be replaced with either of the other two terms. For example, where a given feature, compound or range is indicted as “comprised by” a respective broader term, such broader term may also “consist of” such feature, compound or range.

[00152] It should be understood that this invention is not limited to the particular methodology, protocols, and reagents, etc., described herein and as such can vary. The terminology used herein is for the purpose of describing particular embodiments only and is not intended to limit the scope of the present invention, which is defined solely by the claims.

[00153] All publications and patents cited throughout the text of this specification (including all patents, patent applications, scientific publications, manufacturer’s specifications, instructions, etc.), whether supra or infra, are hereby incorporated by reference in their entirety. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention. To the extent the material incorporated by reference contradicts or is inconsistent with this specification, the specification will supersede any such material.

[00154] As described herein references are made to UniProtKB Accession Numbers (http://www.uniprot.org/, e.g., as available in UniProtKB UniProt release 2023_05, Released on November 08, 2023, unless indicated otherwise or otherwise inherent).

[00155] **The present invention may also be characterized by the following items:**

1. A method for predicting the efficacy of an antibody-drug-conjugate (ADC, e.g., Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan), said method comprising:

a) determining by multiplex fluorescence in situ hybridization (e.g., multiplexed fluorescence *in situ* hybridization (FISH) based on a single cell or based on plurality of cells) whether or not mRNA species of ADC-associated biomarker/s (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 and/or 3-6 herein) are present in a sample obtained from a subject; preferably said ADC-associated biomarker/s are selected from the group consisting of: *ESR1* (e.g., having Uniprot Accession Number: P03372), *PGR* (e.g., having Uniprot Accession Number: P06401), *ERBB2* (e.g., having Uniprot Accession Number: P04626) and *MKI67* (e.g., having Uniprot Accession Number: P46013); and

b) determining by multiplex sequencing (e.g., multiplex sequencing based on a single cell or based on the plurality of cells) whether or not mRNA species of ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 and/or 3-6 herein) are present in said sample.

2. The method of any one of the preceding items, wherein (a) and (b) are carried out on a single cell or on a plurality of cells.

3. The method of any one of the preceding items, wherein said method further comprising: (c) wherein a part of said sample in which said mRNA species of ADC-associated biomarkers was determined in step (a) is subjected to a laser capturing (e.g., as disclosed herein, e.g., as in Example 4 herein) prior to performing step (b).

4. The method of any one of the preceding items, wherein said method comprising:

a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of ADC-associated biomarker/s (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) are present in a sample obtained from a subject; preferably said ADC-associated biomarker/s are selected from the group consisting of: *ESR1*, *PGR*, *ERBB2* and *MKI67*; and

b) determining by multiplex sequencing whether or not mRNA species of ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 and/or 3-6 herein) are present in said sample;

wherein a part of said sample in which said mRNA species of ADC-associated biomarkers was determined in step (a) is subjected to a laser capturing (e.g., as disclosed herein, e.g., as in Example 4 herein) prior to performing step (b).

5. The method of any one of the preceding items, wherein said determining comprises comparing levels (e.g., expression levels) of said mRNA species of said ADC-associated biomarker obtained from said subject to corresponding levels (e.g., expression levels) of said mRNA species obtained from a heathy subject/s and/or cohort/s.

6. The method of any one of the preceding items, wherein the ADC-associated biomarker/s from step (a) are same or different from the ADC-associated biomarkers from step (b), preferably wherein the ADC-associated biomarker/s from step (a) are different from the ADC-associated biomarkers from step (b), further preferably wherein in step (a) said ADC-associated biomarker/s are cancer-associated biomarkers (e.g., used for visualization purposes).

7. The method of any one of the preceding items, wherein the presence and/or upregulation of ADC-associated biomarker/s in step (a) and step (b) is indicative of the efficacy of said ADC, wherein said efficacy preferably comprises treatment responsiveness to said ADC.

8. The method of any one of the preceding items, wherein said laser capturing is carried out by the means of a laser capture microdissection, preferably said method is carried out by the means of multiplexed RNA-FISH-guided laser capture microdissection (e.g., as disclosed herein, e.g., as in Example 4 herein).

9. The method of any one of the preceding items, wherein said multiplexed RNA-FISH-guided laser capture microdissection followed by RNA sequencing predicts a presence or absence of a therapeutic response to said ADC.

10. The method of any one of the preceding items, wherein said method further comprises selecting and/or monitoring a therapy and/or therapeutic regimen based on said presence or absence of said therapeutic response to said ADC and/or said method further comprises administering said ADC to said subject.

11. The method of any one of the preceding items, wherein said method is a method of cancer subtyping, preferably said method is a method of breast cancer subtyping, further preferably said breast cancer subtyping is selecting a breast cancer subtype from the group consisting of:

a) luminal A-like breast cancer (e.g., ER-positive, e.g., wherein said ER having Uniprot Accession Number: P03372; and/or PR-positive, e.g., wherein said PR having Uniprot Accession Number: P06401; and HER2-negative, e.g., wherein said HER2 having Uniprot Accession Number: P04626;

b) luminal B-like breast cancer (e.g., ER-positive, e.g., wherein said ER having Uniprot Accession Number: P03372; and/or PR-positive/PR-negative, e.g., wherein said PR having Uniprot Accession Number: P06401; and HER2-positive, e.g., wherein said HER2 having Uniprot Accession Number: P04626);

c) HER2 enriched type (e.g., ER-negative, e.g., wherein said ER having Uniprot Accession Number: P03372; PR-negative, e.g., wherein said PR having Uniprot Accession Number: P06401; or HER2-positive, e.g., wherein said HER2 having Uniprot Accession Number: P04626); and

d) triple negative (TNBC), (e.g., ER-, PR- and HER2-negative, e.g., Uniprot Accession Numbers as in (a)).

12. The method of any one of the preceding items, wherein said ADC-associated biomarkers are ADC-associated cancer biomarkers, preferably said ADC-associated cancer biomarkers are ADC-associated breast cancer biomarkers.

13. The method of any one of the preceding items, wherein said ADC-associated biomarkers (e.g., from step (a) and/or step (b)) are one or more of the following:

a) said ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) comprising or consisting of: ADC-associated biomarkers related to or associated with: ADC antigen targets, endocytosis, lysosomal function, payload targets, resistance pathways, and/or linker;

b) said ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) comprising or consisting of: ESR1, PGR and/or MKI67, optionally additionally ERBB2;

c) said ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) comprising or consisting of: estrogen receptors (ERs); progesterone receptors (PRs), HER2 and/or Ki67;

d) said ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) comprising or consisting of: ADC-associated biomarkers related to or associated with one or more of the following: estrogen (e.g., ESR1 and/or PGR); Her2 (e.g., ERBB2 and/or NRG1); CDK4/6 (e.g., CCNB1); Immune function

(e.g., Chemokine12 and/or CD274); chemotherapy (e.g., SLC29A1, SLC19A1 and/or MAPs); DNA repair; Topoisomerase inhibitor/s (e.g., UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1 and/or PROM1); ABC transport (e.g., ABCB1, ABCC1, ABCC3 and/or MTUS1); endocytosis (e.g., RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2 and/or FLOT1 to FLOT2 ratio); enzymatic cleavage (e.g., CTSB, GUSB and/or GLB1); TME (e.g., VEGFA); lysosome (e.g., Cathepsins, LAMP1, LAMP2, SLC46A3 and/or SLX4); proliferation (e.g., Module11_Prolif); ADC target (e.g., ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75 and/or ROR1); molecular subtyping (e.g., Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA); microtubule genes (e.g., TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and/or TUBA1B);

e) said ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) are selected from the group consisting of: ESR1, PGR, ERBB2, NRG1, CCNB1, Chemokine12, CD274, SLC29A1, SLC19A1, MAPs, UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1, PROM1, ABCB1, ABCC1, ABCC3, MTUS1, RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2, FLOT1 to FLOT2 ratio, CTSB, GUSB, GLB1, VEGFA, Cathepsins, LAMP1, LAMP2, SLC46A3, SLX4, Module11_Prolif, ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75, ROR1, Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA), TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and TUBA1B; and/or

f) said ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) are one or more (e.g., all) of the following: ESR1, PGR, ERBB2, NRG1, CCNB1, Chemokine12, CD274, SLC29A1, SLC19A1, MAPs, UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1, PROM1, ABCB1, ABCC1, ABCC3, MTUS1, RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2, FLOT1 to FLOT2 ratio, CTSB, GUSB, GLB1, VEGFA, Cathepsins, LAMP1, LAMP2, SLC46A3, SLX4, Module11_Prolif, ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75, ROR1, Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA), TUBB, TUBE1,

TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and TUBA1B.

14. The method of any one of the preceding items, wherein said biomarker is: (i) one or more of the biomarkers listed in Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and/or 17 and/or 18-31 herein; and/or (ii) one or more genes and/or gene signatures as listed in Table 4, 5 and/or 6 herein.

Table 4: List of signature genes and corresponding groups (e.g., https://pubmed.ncbi.nlm.nih.gov/):

Group	Signature	PMID	Calculation
Estrogen	ESR1	Internal GE data	Single gene
Estrogen	PGR	Internal GE data	Single gene
Estrogen	ESR1_PGR_ave	Internal GE data	1) Mean center by gene over all samples, 2) average over genes
Estrogen	GSEA_HALLMARK_ESTROGEN_RESPONSE_EARLY	GSEA	Median expression
Estrogen	GSEA_HALLMARK_ESTROGEN_RESPONSE_LATE	GSEA	Median expression
Estrogen	TAMR13_scores	PMID: 26607490, PMID: 18498629	Genefu
Estrogen	E2F4_scores	PMID: 29581135	An E2F4 activation z-score was developed by adding values across all genes for each tumor to generate an un-escalated E2F4 score. The un-escalated E2F4 score was then standardized to a z-score by subtracting from each patient's score the mean score in the cohort, and then dividing it by the scores' standard deviation.
Estrogen	MOTERA_sig_ESR1_mutations_Cancer_Res_2021	PMID: 34711608	1) Mean center by gene over all samples, 2) average over genes
Transcription factor	TFEB	PMID: 35882754	Single gene
Her2	EGFR	N/A	Single gene

Her2	Module7_ERBB2_sig_ave_weighted	PMID: 24516633	1) Mean center by gene over all samples, 2) weighted average as published
Her2	CDK12	PMID: 31468695	Single gene
Her2	CUL7	PMID: 31461670	Single gene
Her2	ARID1A	PMID: 27172896	Single gene
Her2	ANXA1	PMID: 27172896	Single gene
Her2	MUC4	PMID: 27698002	Single gene
Her2	NRG1	PMID: 24097864, PMID: 34385295	Single gene
Her2	Her2_amplicon_MDX	N/A	1) Mean center by gene over all samples, 2) average over genes
Her2	pSTAT3_Sig	PMID: 26234940	The pSTAT3-GS score (Sigscore) is derived by computing the sum of the products of the gene coefficient in the module (Si) by the corresponding gene expression value (ei) according to the following formula: Sigscore=∑i=1nei*Si; Missing genes: LOC100507494, LOC100506071, LOC100996578
CDK4/6	CCNE1	PMID: 30807234	Single gene
CDK4/6	CCND3	PMID: 28270497	Single gene
CDK4/6	CCNB1	PMID: 28821558	Single gene
CDK4/6	CDKN2D	PMID: 28270497	Single gene
CDK4/6	CDK6	PMID: 27748766	Single gene
CDK4/6	CDK4	N/A	Single gene
CDK4/6	Retinoblastoma_pathway_loss_of_function_Rbsig	PMID: 27634906	Median expression
CDK4/6	RB1_PTEN_loss	PMID: 35005994	Median expression
CDK4/6	cMET_FAK_CDK2_axis	PMID: 35005994	Median expression
CDK4/6	CDK2	PMID: 35005994	Single gene
CDK4/6	cmyc	PMID: 35005994	Single gene

CDK4/6	FAT1	PMID: 35005994	Single gene
CDK4/6	AKT1	PMID: 35005994	Single gene
CDK4/6	RB1	PMID: 35005994	Single gene
CDK4/6	KRAS_HRAS_NRAS	PMID: 35005994	Median expression
CDK4/6	FGFR2_ERBB2	PMID: 35005994	Median expression
CDK4/6	TK1	PMID: 35005994	Single gene
NKTR fusions	NTRK1	N/A	Single gene
NKTR fusions	NTRK2	N/A	Single gene
NKTR fusions	NTRK3	N/A	Single gene
PI3K	PIK3CA-GS	PMID: 26607490	GeneFu
PI3K	FOXM1	PMID: 32976773	Single gene
PARP	BIOCARTA_ATRBRCA_PATHWAY20	GSEA	Median Expression
PARP	PMID_17493263_basal_epithelial	PMID:17493263	Median expression
PARP	PMID_17493263_luminal_epithelial	PMID:17493263	Median expression
PARP	PMID_17493263_second_luminal	PMID:17493263	Median expression
PARP	PMID_17493263_proliferation	PMID:17493263	Median expression
PARP	PMID_17493263_interferon	PMID:17493263	Median expression
PARP	PMID_17493263_fibroblast	PMID:17493263	Median expression
PARP	PMID_17493263_Kras2	PMID:17493263	Median expression

PARP	PARP_sensitivity_MDACC_NPJ	PMID:28649435	Median expression
PARP	PARP_Sensitivity_Signature_POSITIVE	PMID:28439535	Median Expression
PARP	PARP_Sensitivity_Signature_NEGATIVE	PMID:28439535	Median Expression
PARP	olaparib_Resistance	PMID:22875744	Median Expression
PARP	olaparib_Sensitivity	PMID:22875744	Median Expression
PARP	VCpred	PMID: 35623341	1) mean center by gene over all samples 2) calculate weighted average = (13.60*CXCL13 -6.48*BRCA1 + 6.41*APEX1 + 5.32*FEN1 + 4.85*CD8A - 4.84*SEM1 + 4.78*APEX2 - 4.60*RNMT + 4.51*CCR7 + 3.99*H2AFX + 3.88*POLD3 - 3.49*PRKDC + 3.48*C1QA + 3.33*CLIC5 - 3.24*RAD51 + 3.10 *DDB2 - 2.83*SPP1 - 2.80 *POLD2 - 2.80*POLB + 2.72*LIG1 - 2.67*GTF2H5 – 2.63*PMS2 + 2.60*LY9 -2.34*SHPRH + 6.27*ARAF), 3) Z-score. SEM1 is also known as SHFM1.
PARP	PARP1	N/A	Single gene
Immune	Module5_TcellBcell	PMID: 24516633	1) Mean center by gene over all samples, 2) average over genes
Immune	ICS5	PMID: 24172169	1) Mean center by gene over all samples, 2) average over genes
Immune	B_cells	PMID:28239471	1) Average over genes, 2) mean center
Immune	Dendritic_cells	PMID:28239471	1) Average over genes, 2) mean center
Immune	Mast_cells	PMID:28239471	1) Average over genes, 2) mean center
Immune	STAT1	PMID:19272155	1) Mean center by gene over all samples, 2) average over genes
Immune	Chemokine12	PMID: 21703392	1) Mean center by gene over all samples, 2) average over genes
Immune	Module3_IFN	PMID: 24516633	1) Mean center by gene over all samples, 2) average over genes
Immune	PDCD1	N/A	Single gene
Immune	CD274	N/A	Single gene
Immune	CTLA4	N/A	Single gene
Chemotherapies	Early_Relapse_ER.Pos	PMID:21558518	Median expression
Chemotherapies	Early_Relapse_ER.Neg_median_MDX	PMID:21558518	Median expression

Chemotherapies	Pathologic_response_ER.Pos	PMID:21558518	Median expression
Chemotherapies	Pathologic_response_ER.Neg	PMID:21558518	Median expression
Chemotherapies	Residual_disease_ER.Pos	PMID:21558518	Median expression
Chemotherapies	Residual_disease_ER.Neg	PMID:21558518	Median expression
Chemotherapies	HS_Red1	PMID:21214954	Median expression
Chemotherapies	HS_Red8	PMID:21214954	Median expression
Chemotherapies	Scorr_PTEN_Present_Correlation_PNAS_correlated	PMID:17452630	Median expression
Chemotherapies	Scorr_PTEN_Present_Correlation_PNAS_anticorrelated	PMID:17452630	Median expression
Chemotherapies	HS_Green15	PMID:21214954	Median expression
Chemotherapies	TOP2A	PMID: 25897160	Single gene
Chemotherapies	TYMS	N/A	Single gene
Chemotherapies	DHFR	N/A	Single gene
Chemotherapies	SLC29A1	PMID: 11801546	Single gene
Chemotherapies	SLC19A1	PMID: 33479665	Single gene
Chemotherapies	MAPs_Mitotic_kinases_neoadj_chemo118	PMID: 31685623	Median expression
Chemotherapies	MAPs_Mitotic_kinases_neoadj_chemo17	PMID: 31685623	Median expression

Prognostic risk scores	RISK_OncotypeDX_score	PMID: 26607490	Genefu
Prognostic risk scores	RISK_GENE70_score	PMID: 26607490	Genefu
Prognostic risk scores	RISK_PAM50_ROR_score	PMID: 26607490	Genefu
Prognostic risk scores	RISK_GGI_score	PMID: 26607490	Genefu
DNA repair	Original_mammaprint	PMID: 11823860	Median expression
DNA repair	HRD_Race_Classifier	PMID: 36129803	Median expression
DNA repair	MPG	PMID: 22801474	Single gene
DNA repair	TP53	PMID: 22801474	Single gene
DNA repair	MGMT	PMID: 35582025	Single gene
DNA repair	HALLMARK_DNA_REPAIR	GSEA	Median expression
DNA repair	BIOCARTA_DNAFRAGMENT_PATHWAY	GSEA	Median expression
Topoisomerase inhibitors	UGT1A1	N/A	Single gene
Topoisomerase inhibitors	CYP3A4	PMID: 30881507	Single gene
Topoisomerase inhibitors	CYP3A5	PMID: 30881507	Single gene
Topoisomerase inhibitors	TOP1	PMID: 37108395	Single gene
Topoisomerase inhibitors	RAD51	PMID: 33196706	Single gene

Topoisomerase inhibitors	ERCC1	PMID: 33196706	Single gene
Topoisomerase inhibitors	PROM1	PMID: 37108395	Single gene
Topoisomerase inhibitors	GOMF_DNA_TOPOISOMERASE_ACTIVITY	GSEA	Median expression
	ABCB1	PMID: 35534238	Single gene
	ABCC1	PMID: 25646013	Single gene
	ABCC3	N/A	Single gene
	MTUS1	PMID: 23396587, PMID: 31685623	Single gene
Endocytosis	PMCA2_ATP2B2	PMID: 27148852	Single gene
Endocytosis	SH3GL1	PMID: 28974266	Single gene
Endocytosis	CAV1	PMID: 29054985, PMID: 34992350	Single gene
Endocytosis	RAB5A	PMID: 34741021	Single gene
Endocytosis	GOCC_CAVEOLA	GSEA	Median expression
Endocytosis	KEGG_ENDOCYTOSIS	GSEA	Median expression
Endocytosis	GOBP_RECEPTOR_MEDIATED_ENDOCYTOSIS	GSEA	Median expression
Endocytosis	GOBP_REGULATION_OF_CLATHRIN_DEPENDENT_ENDOCYTOSIS	GSEA	Median expression

Endocytosis	GOBP_MACROPINOCYTOSIS	GSEA	Median expression
Endocytosis	RAB5B	PMID: 25646013	Single gene
Endocytosis	ATG9A	PMID: 25646013	Single gene
Endocytosis	HTT	PMID: 25646013	Single gene
Endocytosis	Endocytic_coat	PMID: 29410531	1) Mean center by gene over all samples, 2) average over genes
Endocytosis	Endocytic_actin	PMID: 29410531	1) Mean center by gene over all samples, 2) average over genes
Endocytosis	Endocytic_scission	PMID: 29410531	1) Mean center by gene over all samples, 2) average over genes
Endocytosis	Endocytic_uncoat	PMID: 29410531	1) Mean center by gene over all samples, 2) average over genes
Endocytosis	GSEA_GOBP_MEMBRANE_RAFT_ASSEMBLY	GSEA	1) Mean center by gene over all samples, 2) average over genes
Endocytosis	ATP2B2	PMID: 27148852	Single gene
Endocytosis	HSP90AB1	PMID: 16352662	Single gene
Endocytosis	STUB1	PMID: 12239347, PMID: 12574167	Single gene
Endocytosis	CUL5	PMID: 25102001	Single gene
Endocytosis	FLOT1	PMID: 22869152	Single gene
Endocytosis	FLOT2	PMID: 22869152	Single gene
Endocytosis	FLOT1_FLOT2_ratio	Manual curation (see PMID: 22869152)	Ratio
Lysosome	Cathepsins	PMID: 16990854	1) Mean center by gene over all samples, 2) average over genes
Enzymatic cleavage	CTSB	N/A	Single gene
Enzymatic cleavage	GUSB	N/A	Single gene
Enzymatic cleavage	GLB1	PMID: 28818506	Single gene

TME	Glycolysis_Signature	PMID:19291283	Median expression
TME	VEGF_Signature	PMID:19291283	Median expression
TME	SEMENZA_HIF1_TARGETS	GSEA	Median expression
TME	VEGFA	N/A	Single gene
TME	Hypoxia_Angiogenesis_Inflammatory_MDX	N/A	1) Mean center by gene over all samples, 2) average over genes
TME	Metabolic_sig_Trop2_TNBC_Survival	PMID: 34711841	Average expression of 5 genes
Lysosome	LAMP1	PMID: 29409507	Single gene
Lysosome	LAMP2	PMID: 22874552	Single gene
Lysosome	SLC46A3	PMID: 30131388	Single gene
Lysosome	GSEA_LYSOSOME_M13845	GSEA	Median expression
Lysosome	GSEA_LYSOSOMAL_MEMBRANE_M14296	GSEA	Median expression
Lysosome	GSEA_LYSOSOME_ORGANIZATION_AND_BIOGENESIS_M5971	GSEA	Median expression
Lysosome	GSEA_KEGG_LYSOSOME_M11266	GSEA	Median expression
Lysosome	GSEA_GOMF_LYSOZYME_ACTIVITY_M18340	GSEA	Median expression
Lysosome	SLX4	DOI:https://doi.org/10.1016/j.annonc.2022.03.277	Single gene

Lysosome	GSEA_KEGG_ABC_TRANSPORTERS_M11911	GSEA	Median expression
Cell type marker	B_cells_memory	PMID:25822800	Median expression
Cell type marker	B_cells_naive	PMID:25822800	Median expression
Cell type marker	Dendritic_cells_activated	PMID:25822800	Median expression
Cell type marker	Dendritic_cells_resting	PMID:25822800	Median expression
Cell type marker	Eosinophils	PMID:25822800	Median expression
Cell type marker	Macrophages_M0	PMID:25822800	Median expression
Cell type marker	Macrophages_M1	PMID:25822800	Median expression
Cell type marker	Macrophages_M2	PMID:25822800	Median expression
Cell type marker	Mast_cells_resting	PMID:25822800	Median expression
Cell type marker	Mast_cells_activated	PMID:25822800	Median expression
Cell type marker	Monocytes_median	PMID:25822800	Median expression
Cell type marker	Neutrophils	PMID:25822800	Median expression
Cell type marker	NK_cells_resting	PMID:25822800	Median expression
Cell type marker	NK_cells_activated	PMID:25822800	Median expression
Cell type marker	Plasma_cells	PMID:25822800	Median expression
Cell type marker	T_cells_CD4_memory_activated	PMID:25822800	Median expression
Cell type marker	T_cells_CD4_memory_resting	PMID:25822800	Median expression
Cell type marker	T_cells_CD4_naive	PMID:25822800	Median expression
Cell type marker	T_cells_CD8	PMID:25822800	Median expression
Cell type marker	T_cells_follicular_helper	PMID:25822800	Median expression

Cell type marker	T_cells_gamma_delta	PMID:25822800	Median expression
Cell type marker	T_cells_regulatory_Tregs	PMID:25822800	Median expression
Proliferation	AURKA	N/A	Single gene
Proliferation	MKI67	N/A	Single gene
Proliferation	Proliferation_MDX	N/A	1) Mean center by gene over all samples, 2) average over genes
Proliferation	Module11_Prolif	PMID: 24516633	1) Mean center by gene over all samples, 2) average over genes
ADC target	ERBB2	N/A	Single gene
ADC target	NECTIN4	N/A	Single gene
ADC target	TACSTD2	PMID: 35882754	Single gene
ADC target	ERBB3	N/A	Single gene
ADC target	FOLR1	N/A	Single gene
ADC target	MSLN	N/A	Single gene
ADC target	F3	N/A	Single gene
ADC target	MET	N/A	Single gene
ADC target	SLC39A6	N/A	Single gene
ADC target	TPBG	N/A	Single gene
ADC target	ROR2	N/A	Single gene
ADC target	CD276	N/A	Single gene
ADC target	VTCN1	N/A	Single gene
ADC target	CEACAM5	N/A	Single gene
ADC target	CCR2	N/A	Single gene
ADC target	STING1	N/A	Single gene
ADC target	FOLH1	N/A	Single gene
ADC target	KAAG1	N/A	Single gene

ADC target	LY75	N/A	Single gene
ADC target	ROR1	N/A	Single gene
Molecular subtyping	Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA	PMID: 26607490	Correlation to centroid Genefu
Molecular subtyping	Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumB	PMID: 26607490	Correlation to centroid Genefu
Molecular subtyping	Genefu_Luminal_index_Correlation_to_centroid_PAM50_Her2	PMID: 26607490	Correlation to centroid Genefu
Molecular subtyping	Genefu_Correlation_to_centroid_PAM50_basal	PMID: 26607490	Correlation to centroid Genefu
Molecular subtyping	Genefu_Correlation_to_centroid_PAM50_normal	PMID: 26607490	Correlation to centroid Genefu
Microtubule genes	TUBB	N/A	Single gene
Microtubule genes	TUBE1	N/A	Single gene
Microtubule genes	TUBB2A	N/A	Single gene
Microtubule genes	TUBB3	N/A	Single gene
Microtubule genes	TUBA1A	N/A	Single gene
Microtubule genes	TUBB4B	N/A	Single gene

Microtubule genes	TUBB6	N/A	Single gene
Microtubule genes	TUBD1	N/A	Single gene
Microtubule genes	TUBA4A	N/A	Single gene
Microtubule genes	TUBG1	N/A	Single gene
Microtubule genes	TUBA1C	N/A	Single gene
Microtubule genes	TUBG2	N/A	Single gene
Microtubule genes	TUBA1B	N/A	Single gene

Table 5. Extended list of signature biomarkers with additional information (e.g., https://www.ensembl.org/)

Gene_id	Gene_name	Gene_id	Gene_name
ENSG00000141736	ERBB2	ENSG00000256650	RERG-IT1
ENSG00000161395	PGAP3	ENSG00000183833	CFAP91
ENSG00000141738	GRB7	ENSG00000086696	HSD17B2
ENSG00000141741	MIEN1	ENSG00000144354	CDCA7
ENSG00000265178	MIR4728	ENSG00000131203	IDO1
ENSG00000131748	STARD3	ENSG00000286480	
ENSG00000173991	TCAP	ENSG00000130487	KLHDC7B
ENSG00000141744	PNMT	ENSG00000244676	
ENSG00000167258	CDK12	ENSG00000262406	MMP12
ENSG00000273576		ENSG00000026559	KCNG1
ENSG00000277363	SRCIN1	ENSG00000253838	
ENSG00000106541	AGR2	ENSG00000204019	CT83
ENSG00000184709	LRRC26	ENSG00000198930	CSAG1
ENSG00000214546		ENSG00000210741	MIR196A1
ENSG00000279286		ENSG00000287064	
ENSG00000129514	FOXA1	ENSG00000221867	MAGEA3
ENSG00000139865	TTC6	ENSG00000163207	IVL
ENSG00000266469		ENSG00000124102	PI3
ENSG00000009765	IYD	ENSG00000186075	ZPBP2
ENSG00000176532	PRR15	ENSG00000229314	ORM1
ENSG00000174469	CNTNAP2	ENSG00000166535	A2ML1

ENSG00000065371	ROPN1	ENSG00000280411	IGHV1-69D
ENSG00000228639	ROCR	ENSG00000285822	
ENSG00000151715	TMEM45B	ENSG00000205420	KRT6A
ENSG00000172057	ORMDL3	ENSG00000205426	KRT81
ENSG00000198807	PAX9	ENSG00000197172	MAGEA6
ENSG00000134762	DSC3	ENSG00000167656	LY6D
ENSG00000258616	LINC02303	ENSG00000239311	
ENSG00000073605	GSDMB	ENSG00000228278	ORM2
ENSG00000288605		ENSG00000184330	S100A7A
ENSG00000185634	SHC4	ENSG00000170523	KRT83
ENSG00000125686	MED1	ENSG00000213401	MAGEA12
ENSG00000114547	ROPN1B	ENSG00000169469	SPRR1B
ENSG00000109846	CRYAB	ENSG00000148346	LCN2
ENSG00000162949	CAPN13	ENSG00000287448	
ENSG00000100146	SOX10	ENSG00000203688	LINC02487
ENSG00000172478	MAB21L4	ENSG00000156219	ART3
ENSG00000277972	CISD3	ENSG00000143556	S100A7
ENSG00000234918		ENSG00000198729	PPP1R14C
ENSG00000277258	PCGF2	ENSG00000258910	LINC01956
ENSG00000131771	PPP1R1B	ENSG00000224853	LINC00393
ENSG00000160307	S100B	ENSG00000275216	
ENSG00000086548	CEACAM6	ENSG00000186832	KRT16
ENSG00000108846	ABCC3	ENSG00000164434	FABP7
ENSG00000267259	ERVE-1	ENSG00000169474	SPRR1A
ENSG00000182253	SYNM	ENSG00000137673	MMP7
ENSG00000138685	FGF2	ENSG00000244474	UGT1A4
ENSG00000066468	FGFR2	ENSG00000181617	FDCSP
ENSG00000171532	NEUROD2	ENSG00000268756	
ENSG00000259087		ENSG00000224271	
ENSG00000091831	ESR1	ENSG00000060718	COL11A1
ENSG00000248360	LINC00504	ENSG00000105141	CASP14
ENSG00000107485	GATA3	ENSG00000164199	ADGRV1
ENSG00000109436	TBC1D9	ENSG00000189377	CXCL17
ENSG00000187720	THSD4	ENSG00000124107	SLPI
ENSG00000074410	CA12	ENSG00000151882	CCL28
ENSG00000260586		ENSG00000277693	
ENSG00000173467	AGR3	ENSG00000179455	MKRN3
ENSG00000168350	DEGS2	ENSG00000253301	LINC01606
ENSG00000134830	C5AR2	ENSG00000248810	LINC02432
ENSG00000120262	CCDC170	ENSG00000173809	TDRD12
ENSG00000135069	PSAT1	ENSG00000275757	
ENSG00000196405	EVL	ENSG00000138829	FBN2
ENSG00000100219	XBP1	ENSG00000151632	AKR1C2
ENSG00000134352	IL6ST	ENSG00000165272	AQP3
ENSG00000186868	MAPT	ENSG00000145824	CXCL14

ENSG00000163879	DNALI1	ENSG00000224846	NQO2-AS1
ENSG00000082175	PGR	ENSG00000260239	LINC02533
ENSG00000178935	ZNF552	ENSG00000212609	RNU1-139P
ENSG00000178568	ERBB4	ENSG00000204832	ST8SIA6-AS1
ENSG00000158258	CLSTN2	ENSG00000116299	ELAPOR1
ENSG00000134755	DSC2	ENSG00000230937	MIR205HG
ENSG00000160182	TFF1	ENSG00000133124	IRS4
ENSG00000237080	EHMT2-AS1	ENSG00000167244	IGF2
ENSG00000203697	CAPN8	ENSG00000049540	ELN
ENSG00000207155	RNY1P14	ENSG00000178372	CALML5
ENSG00000141337	ARSG	ENSG00000160862	AZGP1
ENSG00000183044	ABAT	ENSG00000156284	CLDN8
ENSG00000196177	ACADSB	ENSG00000127249	ATP13A4
ENSG00000278570	NR2E3	ENSG00000103534	TMC5
ENSG00000108176	DNAJC12	ENSG00000227471	AKR1B15
ENSG00000196208	GREB1	ENSG00000164120	HPGD
ENSG00000075275	CELSR1	ENSG00000236699	ARHGEF38
ENSG00000203709	ENSG00000132746	ALDH3B2	
ENSG00000151892	GFRA1	ENSG00000166743	ACSM1
ENSG00000144218	AFF3	ENSG00000135226	UGT2B28
ENSG00000279685	MAPT-IT1	ENSG00000213759	UGT2B11
ENSG00000157303	SUSD3	ENSG00000196228	SULT1C3
ENSG00000259459	LINC02568	ENSG00000183569	SERHL2
ENSG00000166106	ADAMTS15	ENSG00000179593	ALOX15B
ENSG00000175356	SCUBE2	ENSG00000100867	DHRS2
ENSG00000092068	SLC7A8	ENSG00000109906	ZBTB16
ENSG00000143412	ANXA9	ENSG00000156968	MPV17L
ENSG00000200385		ENSG00000129988	LBP
ENSG00000115648	MLPH	ENSG00000144452	ABCA12
ENSG00000185133	INPP5J	ENSG00000115255	REEP6
ENSG00000286208		ENSG00000134240	HMGCS2
ENSG00000054598	FOXC1	ENSG00000140798	ABCC12
ENSG00000253125		ENSG00000169083	AR
ENSG00000146476	ARMT1	ENSG00000137975	CLCA2
ENSG00000160180	TFF3	ENSG00000124664	SPDEF
ENSG00000150667	FSIP1	ENSG00000148513	ANKRD30A
ENSG00000224559	LINC01087	ENSG00000008196	TFAP2B
ENSG00000104332	SFRP1	ENSG00000189058	APOD
ENSG00000050438	SLC4A8	ENSG00000172551	MUCL1
ENSG00000108932	SLC16A6	ENSG00000121270	ABCC11
ENSG00000171428	NAT1	ENSG00000244468	
ENSG00000164626	KCNK5	ENSG00000159763	PIP
ENSG00000141424	SLC39A6	ENSG00000196091	MYBPC1
ENSG00000057593	F7	ENSG00000171345	KRT19
ENSG00000253509		ENSG00000141750	STAC2

ENSG00000183779	ZNF703	ENSG00000120885	CLU
ENSG00000099812	MISP	ENSG00000132465	JCHAIN
ENSG00000119866	BCL11A	ENSG00000211890	IGHA2
ENSG00000102243	VGLL1	ENSG00000211895	IGHA1
ENSG00000140443	IGF1R	ENSG00000162896	PIGR
ENSG00000134533	RERG	ENSG00000012223	LTF
ENSG00000176887	SOX11	ENSG00000019169	MARCO
ENSG00000171791	BCL2	ENSG00000089685	BIRC5
ENSG00000173894	CBX2	ENSG00000101057	MYBL2
ENSG00000170500	LONRF2	ENSG00000119547	ONECUT2
ENSG00000251034		ENSG00000130167	TSPAN16
ENSG00000251141	MRPS30-DT	ENSG00000137261	KIAA0319
ENSG00000278022		ENSG00000137648	TMPRSS4
ENSG00000112414	ADGRG6	ENSG00000167104	BPIFB6
ENSG00000143546	S100A8	ENSG00000167653	PSCA
ENSG00000164309	CMYA5	ENSG00000170703	TTLL6
ENSG00000113739	STC2	ENSG00000179363	TMEM31
ENSG00000171124	FUT3	ENSG00000204710	SPDYC
ENSG00000163362	INAVA	ENSG00000235281	
ENSG00000128510	CPA4	ENSG00000236193	LINC01953
ENSG00000163220	S100A9	ENSG00000241636	LINC01323
ENSG00000130045	NXNL2	ENSG00000248431	
ENSG00000101447	FAM83D	ENSG00000250874	
ENSG00000165125	TRPV6	ENSG00000254574	
ENSG00000250643		ENSG00000254721	
ENSG00000244265	SIAH2-AS1	ENSG00000257883	
ENSG00000152076	CCDC74B	ENSG00000258815	LINC02820
ENSG00000163040	CCDC74A	ENSG00000259306	
ENSG00000264480	MIR4714	ENSG00000260532	
ENSG00000124171	PARD6B	ENSG00000261763	
ENSG00000286042	LCAL1	ENSG00000262152	LINC00514
ENSG00000285108		ENSG00000278041	
ENSG00000156689	GLYATL2	ENSG00000278704	
ENSG00000178750	STX19	ENSG00000283439	SPEM3
ENSG00000187527	ATP13A5	ENSG00000285800	
ENSG00000140479	PCSK6		

15. The method of any one of the preceding items, wherein said ADC is Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan, optionally said subject in step (a) is suffering from a cancer (e.g., breast cancer).

16. A method for stratification of a subject for treatment (e.g., cancer treatment, e.g., breast cancer treatment) with an ADC (e.g., Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan), comprising:

a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of ADC-associated biomarker/s (e.g., as disclosed herein, e.g., in Examples1 and/or 2 and/or 3-6 herein) are present in a sample obtained from a subject; preferably said ADC-associated biomarker/s are selected from the group consisting of: *ESR1* (e.g., having Uniprot Accession Number: P03372), *PGR* (e.g., having Uniprot Accession Number: P06401), *ERBB2* (e.g., having Uniprot Accession Number: P04626) and *MKI67* (e.g., having Uniprot Accession Number: P46013); and

b) determining by multiplex sequencing whether or not mRNA species of ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 and/or 3-6 herein) are present in said sample;

wherein a part of said sample in which said mRNA species of ADC-associated biomarkers was determined in step (a) is subjected to a laser capturing prior to performing step (b); wherein the presence of one or more of said ADC-associated biomarker is indicative that said ADC may be effective, and/or wherein the absence of one or more of said ADC-associated biomarker is indicative that said ADC may not be effective;

preferably wherein said method comprising: (a') carrying out the method for predicting the efficacy of an antibody-drug-conjugate (ADC) according to any one of the preceding items, wherein: (b') the presence of one or more of said ADC-associated biomarkers is indicative that said ADC may be effective; and/or (c') wherein the absence of one or more of said ADC-associated biomarkers is indicative that said ADC may not be effective; preferably said method is an *ex vivo* or *in vitro* method.

17. The method of any one of the preceding items, wherein the ADC-associated biomarker/s from step (a) are same or different from the ADC-associated biomarkers from step (b), preferably wherein the ADC-associated biomarker/s from step (a) are different from the ADC-associated biomarkers from step (b), further preferably wherein in step (a) said ADC-associated biomarker/s are cancer-associated biomarkers (e.g., used for visualization purposes).

18. The method of any one of the preceding items, wherein said ADC is Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan, optionally said subject in step (a) is suffering from a cancer (e.g., breast cancer).

19. A method for improving performance of a multi-gene signature test (e.g., PAM50 (Prediction Analysis of Microarray using 50 classifier genes) and/or AIMS (Absolute Intrinsic Molecular Subtyping), comprising subjecting a part of a sample obtained from a subject to a laser capturing prior to carrying out said test, preferably said multi-gene signature test is a cancer (e.g., breast cancer) multi-gene signature test.

20. The method of any one of the preceding items, further comprising calculating an average overall score of said multi-gene signature test.

21. A method for predicting survival and/or assessing status, risk and/or prognosis of cancer (e.g., breast cancer), said method comprising:

a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of cancer-associated biomarker/s (e.g., as disclosed herein, e.g., in Example 3 herein, e.g., Table 6) are present in a sample obtained from a subject; and

b) determining by multiplex sequencing whether or not mRNA species of cancer--associated biomarkers (e.g., as disclosed herein, e.g., in Example 3 herein, e.g., Table 6) are present in said sample;

wherein a part of said sample in which said mRNA species of cancer-associated biomarkers was determined in step (a) is subjected to a laser capturing (e.g., as disclosed herein, e.g., as in Example 4 herein) prior to performing step (b); wherein the presence of one or more of said cancer-associated biomarker/s is indicative that the survival, risk and/or prognosis of cancer is low, and/or wherein the absence of one or more of said cancer-associated biomarker/s may be indicative that the survival, risk and/or prognosis of cancer is high.

22. The method of any one of the preceding items, wherein the cancer-associated biomarker/s from step (a) are same or different from the cancer-associated biomarkers from step (b), preferably wherein the cancer-associated biomarker/s from step (a) are different from the cancer-associated biomarkers from step (b).

23. An antibody-drug-conjugate (ADC, e.g., Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan) for use in the treatment of a subject

suffering from cancer (e.g., breast cancer), wherein said subject is stratified in accordance with the method of any one of the preceding items, optionally said ADC is Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan.

24. A kit for performing the method of any one of the preceding items (e.g., as depicted in Examples 1-6 herein).

25. The kit of any one of the preceding items, comprising glass slides, preferably positively charged glass slides, frame slides, frame slides with PET membrane, or glass slides with membrane.

26. The kit of any one of the preceding items, comprising formaldehyde and/or paraffin.

27. The kit of any one of the preceding items, comprising probes for fluorescence in situ hybridization of RNA.

28. The kit of any one of the preceding items, comprising probes for FISH of RNA specific for one or more of the following:

a) ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) related to or associated with: ADC antigen targets, endocytosis, lysosomal function, payload targets, resistance pathways and/or linker;

b) *ESR1* (e.g., having Uniprot Accession Number: P03372), *PGR* (e.g., having Uniprot Accession Number: P06401), and/or *MKI67* (e.g., having Uniprot Accession Number: P46013), optionally additionally *ERBB2* (e.g., having Uniprot Accession Number: P04626);

c) estrogen receptors (ERs); progesterone receptors (PRs), HER2 and/or Ki67;

d) ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) are related to or associated with one or more of the following: estrogen (e.g., ESR1 and/or PGR); Her2 (e.g., ERBB2 and/or NRG1); CDK4/6 (e.g., CCNB1); Immune function (e.g., Chemokine12 and/or CD274); chemotherapy (e.g., SLC29A1, SLC19A1 and/or MAPs); DNA repair; Topoisomerase inhibitor/s (e.g., UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1 and/or PROM1); ABC transport (e.g., ABCB1, ABCC1, ABCC3 and/or MTUS1); endocytosis (e.g., RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2 and/or FLOT1 to FLOT2 ratio); enzymatic cleavage (e.g., CTSB, GUSB and/or GLB1); TME (e.g., VEGFA); lysosome (e.g., Cathepsins, LAMP1, LAMP2,

SLC46A3 and/or SLX4); proliferation (e.g., Module11_Prolif); ADC target (e.g., ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75 and/or ROR1); molecular subtyping (e.g., Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA); microtubule genes (e.g., TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and/or TUBA1B);

e) said ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) are selected from the group consisting of: ESR1, PGR, ERBB2, NRG1, CCNB1, Chemokine12, CD274, SLC29A1, SLC19A1, MAPs, UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1, PROM1, ABCB1, ABCC1, ABCC3, MTUS1, RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2, FLOT1 to FLOT2 ratio, CTSB, GUSB, GLB1, VEGFA, Cathepsins, LAMP1, LAMP2, SLC46A3, SLX4, Module11_Prolif, ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75, ROR1, Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA), TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and TUBA1B; and/or

f) said ADC-associated biomarkers (e.g., as disclosed herein, e.g., in Examples 1 and/or 2 herein) are one or more of the following: ESR1, PGR, ERBB2, NRG1, CCNB1, Chemokine12, CD274, SLC29A1, SLC19A1, MAPs, UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1, PROM1, ABCB1, ABCC1, ABCC3, MTUS1, RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2, FLOT1 to FLOT2 ratio, CTSB, GUSB, GLB1, VEGFA, Cathepsins, LAMP1, LAMP2, SLC46A3, SLX4, Module11_Prolif, ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75, ROR1, Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA), TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and TUBA1B.

29. The kit of any one of the preceding items, comprising one or more probes for: (i) one or more of the biomarkers listed in Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and/or 17 and/or 18-31 herein; and/or (ii) one or more genes and/or gene signatures as listed in Table 4, 5 and/or 6 herein.

30. Use of a kit of any one of the preceding items, for performing the methods of any one of the preceding items.

31. Use of laser capturing for improving the prediction of the efficacy of an Antibody-Drug-Conjugate (ADC, e.g., Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan), stratification of a subject for treatment (e.g., cancer treatment) with an ADC or performance of a multi-gene signature test (e.g., in cancer), optionally according to any one of the preceding items.

32. The method for improving the prediction of the efficacy of an antibody-drug-conjugate (ADC, e.g., Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan), stratification of a subject for treatment (e.g., cancer treatment) with an ADC (e.g., Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan) and/or performance of a multi-gene signature test (e.g., in cancer), said method comprising: carrying out a laser capture on said sample, preferably carrying out a laser capture microdissection (LCM), further preferably carrying out a multiplexed RNA-FISH-guided laser capture microdissection.

33. The method, kit, use or ADC according to any one of the preceding items, wherein said ADC is Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan and/or Trastuzumab-drug-conjugate (Trastuzumab-DC), optionally said subject is suffering from a cancer (e.g., breast cancer); further optionally said subject is suffering from a cancer and is Her2-positive (e.g., as in Example 2 herein).

34. The method or use according to any one of the preceding items, wherein said method or use is an *in vitro* or *ex vivo* method or use.

35. The method, kit, use or ADC according to any one of the preceding items, wherein said ADC-associated biomarker/s are biomarkers (e.g., genes, mutation/s, polymorphism/s, mRNA/s, protein/s, enzyme/s, their activities and/or post-translational modification/s (e.g., phosphorylation status) and/or pathways) associated with an ADC treatment, preferably said association comprises one or more of the following: alteration (e.g., up-regulation, down-regulation and/or any modification) of an expression level (e.g., of RNA/s and/or protein/s), and/or alteration of activity and/or genetic status (e.g.., mutation/s and/or polymorphism/s) of said biomarker/s in response to said ADC-treatment (e.g., in cancer treatment).

36. The method, kit, use or ADC according to any one of the preceding items, wherein said biomarker/s (e.g., ADC-associated) comprising or consisting of cancer-associated

biomarker/s, optionally said ADC-associated biomarker/s are cancer-associated biomarker/s (e.g., as disclosed herein, e.g., in Table 4, 5 and/or 6 herein).

37. The method, kit, use or ADC according to any one of the preceding items, wherein said cancer is breast cancer.

38. The method, kit, use or ADC according to any one of the preceding items, carried out as described herein, preferably as described in Example 1, Example 2, Example 3 and/or Example 4 and/or 5-6 herein.

39. The method, kit, use or ADC according to any one of the preceding items, wherein said ADC biomarker/s are one or many (or selected from the group consisting of) as described herein, preferably as described in Example 1, Example 2 and/or Example 3 and/or 4-6 herein, further preferably said ADC biomarker/s are one or many (e.g., all) genes and/or gene signatures as described herein.

***

[00156] The invention is further illustrated by the following examples, however, without being limited to the example or by any specific embodiment of the examples.

[00157] Examples of the invention

[00158] Example 1: Multiplexed RNA-FISH-guided Laser Capture Microdissection RNA Sequencing Improves Breast Cancer Molecular Subtyping and Predicts Response to Antibody Drug Conjugates

[00159] Material and methods

[00160] Study Design

[00161] We conducted a retrospective clinical validation of our mFISHseq BCa diagnostic test using 1,082 archived FFPE BCa samples collected from two European biobanks (Biobank Graz, PATH Biobank), one hospital (Malaga, Spain), and two commercial companies (AMS Bio and Precision for Medicine). Clinicopathological information associated with each sample (age, receptor histological status, tumor grade, therapy history, survival data, etc.) was accrued in collaboration with the biobanks and follows the Bio-specimen Reporting for Improved Study Quality (BRISQ) criteria and used to perform association analyses with molecular data. Inclusion criteria for the study consisted of females with histologically confirmed invasive BCa, availability of anonymized data regarding pathological diagnosis (IHC status, TNM staging), therapy (hormone/targeted/chemo- or radiotherapy), and survival (progression-free survival, overall survival), as well as signed and dated informed consent. The

only exclusion criteria were pre-existing conditions or concurrent diagnosis of a cancer other than breast cancer or other disease that may influence the interpretation of the study results. The project was approved by the Ethics Committee of the Bratislava Self-Governing Region (Ref. No. 05320/2020/HF).

[00162] Patient specimens were processed in batches (see **Supplementary Methods herein below**) using a stratified randomization approach to ensure that each batch contains a representative sampling of the IHC-surrogate subtypes (i.e., luminal A, luminal B, HER2+, TNBC). Researchers who processed batches and conducted the data processing and analysis were blind to IHC biomarker status (e.g., ER, PR, HER2, and KI67) and other clinical information.

[00163] To assess the analytical validity of mFISHseq, the dataset was divided into a training and test set (70:30 split) using a stratified randomization approach to ensure similar proportions of positive and negative biomarkers (as defined by IHC) and sufficient patient outcomes. Other analyses like consensus subtyping and characterization of genes/gene signatures utilized the full dataset.

[00164] **Tissue processing and H&E staining**

[00165] We obtained at least eight 5 µm sections from FFPE BCa specimens using a Leica Histocore Multicut. Adjacent sections were collected in the following order: 1 section on a glass slide for H&E, 1 section on a functionalized PEN membrane slide for LCM, 1 section on a glass slide for RNA-FISH, and 1 section taken as a scroll and frozen at –20 °C for later RNA extraction and RNA sequencing (see **Supplementary Methods herein below**). To ensure proper identification of invasive breast cancer, we stained one section using H&E, cover slipped, and then obtained a whole-slide scan. The resulting image was annotated by a trained researcher to identify the invasive breast cancer component for later microdissection. If necessary, a board-certified pathologist either annotated or reviewed challenging cases. Importantly, the H&E-stained section was adjacent to the PEN membrane slide used for laser capture microdissection to ensure comparable anatomical morphology between the slide that was annotated and the slide that was microdissected.

[00166] **Multiplexed RNA-FISH**

[00167] For RNA-FISH we used the Advanced Cell Diagnostics RNAscope™ Multiplex Fluorescent V2 Assay to detect *PGR*, *ESR1*, *ERBB2*, and *MKI67* according to the manufacturer's instructions. These markers were detected using Akoya Opal 690, 620, 520, and 570 fluorophores, respectively. Visualization of these markers allowed us to capture the heterogeneity of the tumor tissue and isolate key regions of interest using LCM to obtain tumor-specific regions of interest, while eliminating otherwise healthy tissue, stroma, and adipose cells that may "mask" true gene expression differences.

[00168] Whole-slide imaging, image annotation, and image analysis

[00169] Following H&E staining and RNA-FISH, we used the Akoya Vectra Polaris to obtain brightfield and fluorescent whole slide scans (20x objective) that could be further analyzed and annotated for microdissection. The H&E whole slide scans were annotated by a trained researcher using the open-source program QuPath (https://qupath.github.io/). Annotations were color-coded to identify invasive breast cancer (segregated by histological subtype if more than one is present in a specimen), ductal carcinoma in situ (DCIS), and healthy/normal tissue. If necessary, a board-certified pathologist either annotated or reviewed challenging cases. The RNA-FISH whole slide scans were annotated by a trained researcher using Akoya's Phenochart software (Vectra Polaris). The RNA-FISH annotation consisted of a qualitative overview of the intensity and distribution of fluorescent signals from ER, PR, HER2, and KI67. Based on the annotated H&E and RNA-FISH images, specific regions of interest were selected for LCM with an emphasis on regions that displayed the expression of biomarkers of interest (e.g., hotspots), areas identified as invasive by a trained researcher/pathologist, and the margins of invasive tumors. Moreover, in the case of specimens that displayed histologic or biomarker expression heterogeneity in the form of different molecular expression patterns (e.g., ER/PR+ and HER2- regions versus ER/PR- and HER2+ regions) or histological subtypes (e.g., invasive ductal vs invasive lobular carcinoma) distinct regions of interest were annotated and separately subjected to LCM and downstream analyses (see **Supplementary Methods herein below**).

[00170] For RNA-FISH image analysis, regions of interest that were dissected by laser capture microdissection were stamped on the digital whole slide scans for further processing of biomarker signals using Akoya's InForm software (**Figure 9**). At least 1-3 stamped regions, depending on the size of the area dissected by LCM, were analyzed (at 20x objective). The analysis consisted of the following steps: 1) spectral unmixing and autofluorescence isolation using a synthetic spectral library; 2) using machine learning algorithms to segment the tissue into different regions (tumor versus stroma) as well as to segment individual cells into nuclear and cytoplasmic components; and 3) scoring the expression of each biomarker. The average fluorescence intensity for each marker was assessed specifically in the tumor segment of the image and the researcher conducting the analysis was blinded to the known IHC results and the clinicopathological data.

[00171] Laser capture microdissection

[00172] We followed established protocols from Leica for conducting LCM in a manner that maintained RNA integrity. This included conducting a rapid, cresyl violet stain, limiting dissection times to under 1 hour per sample, and taking precautionary measures to ensure RNA integrity. Regions selected for dissection were identified by comparing the annotated H&E

and RNA-FISH images with the adjacent cresyl violet stained section. We aimed to dissect approximately 10-20 mm^2 of tissue per sample to ensure an adequate amount of material for RNA extraction. For samples with less tumor area, we conducted LCM on multiple PEN membrane slides to obtain sufficient tissue.

[00173] RNA isolation and quality control

[00174] The Macherey Nagel NucleoSpin totalRNA FFPE XS kit was used for RNA isolation (see **Supplementary Methods herein below**). After RNA isolation, RNA quantity was measured using the Qubit RNA HS (High Sensitivity) Assay Kit with a Qubit 4 Fluorometer and RNA quality using the Agilent High Sensitivity RNA ScreenTape with an Agilent 4150 TapeStation. The DV_{200} value of the sample (i.e., the percentage of fragments ≥ 200 bases in length) was calculated as recommended by Illumina. Samples with DV_{200} values > 15% were considered as viable samples for library preparation.

[00175] RNA library preparation and sequencing

[00176] We used the Takara SMARTer Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian kit to prepare total RNA-SEQ libraries following the manufacturer's instructions. To control for batch library preparation effects, we included a single natural positive control sample in each library preparation batch and a synthetic spike-in control in each sample (see **Supplementary Methods herein below**). Following library preparation, the quantity and fragment size range of the library were assessed using both the Qubit dsDNA HS kit (Qubit 4 Fluorometer) and the Agilent High Sensitivity DNA ScreenTape kit (Agilent 4150 TapeStation). Successfully prepared libraries contained sufficient library (≥ 4ng/µl) to pool on an Illumina NovaSeq 6000 sequencing instrument and fragment range spanning 200 – 1,000 bp, with a local maximum ~250 – 350 bp. Individual sequencing libraries were pooled and sequenced on an Illumina NovaSeq 6000 using SP, S1, S2, or S4 flow cells depending on pool size). Pooled libraries were spiked with 10% PhiX as recommended by both Illumina and Takara for low-complexity libraries sequenced on patterned flow cells. Paired-end sequencing (2 x 100 bp) was conducted with the aim of obtaining approximately 100 million reads per sample. The bioinformatics pipeline for RNA sequencing is described in **Supplementary Methods herein below**).

[00177] Statistical Methods

[00178] All statistical tests were conducted using GraphPad Prism 9 and R studio packages. Unless otherwise stated, the level of significance was set at $P < 0.05$ for both adjusted (q-value) and unadjusted P-values. For determining the significance of continuous variables, paired or unpaired t-tests were used for normally distributed data for two groups, while linear mixed models or ANOVA were used for normal data for three or more groups. Non-

normal data for two groups were analyzed using Mann-Whitney *U* tests (independent samples), while non-normal data for three or more groups were analyzed using Kruskal-Wallis tests. Appropriate corrections for multiple comparisons were conducted using Tukey's test (parametric), Dunn's test (non-parametric), or Benjamini-Hochberg (or Benjamini, Krieger, and Yekutieli) FDR procedures (non-parametric) to adjust P values. ROC and precision-recall curves were constructed using either GraphPad Prism 9 or the R package pROC 1.3.1. Diagnostic performance metrics were calculated using the R package caret 6.0-94.

[00179] **Cohort descriptive statistics** – Standard descriptive statistics (presented as either median or mean ± standard error of the mean (SEM)) were used to summarize sample characteristics (e.g., data for biospecimen information, demographics, pathological, therapy type, survival) for all specimens in the study and were segregated by source and IHC-surrogate subtype.

[00180] **RNA-FISH** – For each RNA-FISH target (*ESR1*, *PGR*, *ERBB2*, *MKI67*), fluorescent intensity (normalized to exposure time and segregated into the cytoplasm, nucleus, and total cell) was assessed. Standard descriptive statistics and violin plots (presented as either Median or Mean ± SEM or interquartile range) were used to summarize results. To compare between RNA-FISH and known IHC results within individual specimens, continuous data (e.g., fluorescent intensity), were analyzed using Pearson correlations for continuous data (e.g., fluorescent intensity) and Spearman's correlations for ordinal data.

[00181] **Survival / Outcome analyses** – Kaplan-Meier analysis and/or Cox Proportional Hazard models were used to quantify associations made between specific dependent variables and/or genes and gene signature predictors with known clinical outcome data (overall survival, progression-free survival). Both univariate and multivariate analyses with clinical parameters (tumor size and node status) were conducted using the R package survival (v3.5-7).

[00182] **Supplementary Methods**

[00183] **Study Design**

[00184] Given the large number of samples, we divided the workflow into various batches of at least 8 samples (8-24 samples depending on the laboratory technique) that were nested into three layers of batches that followed a sequential order: 1) LCM (6-16 samples per batch, including samples with more than one sample dissected by LCM), 2) RNA isolation (8-24 samples), and 3) RNA-SEQ library preparation (8-46 samples). Batches earlier in the workflow were maintained in later batches. For example, two LCM batches could be combined into one RNA isolation batch and then two RNA isolation batches could be combined into one RNA-SEQ library preparation batch. No samples from a given batch were split into separate

batches in the downstream workflow, unless a sample had to be redone due to technical issues.

[00185] Tissue processing

[00186] For tissues provided by PATH Biobank, sectioning was performed at MultiplexDX where several serial sections were obtained in the following order: 2 sections on Epredia Superfrost plus slides (one slide used for RNA-FISH and the second slide used as a backup for failed RNA-FISH runs), 1 section on a Leica PEN membrane slide (2 µm thick membrane and certified RNase-free) that was used for a rapid cresyl violet stain followed by laser capture microdissection, 1 section mounted on an Epredia Superfrost plus slide that was used for H&E staining and reviewed by a board certified pathologist, and 1 section was taken as a scroll, placed into a 1.5 mL microcentrifuge tube (Eppendorf DNA LoBind Tubes, Cat. No. 0030108051), and frozen at –20 °C for later RNA extraction and RNA sequencing. To ensure enough sections in case of technical issues that led to failed assays or for follow up analyses, we took an additional second set of serial sections in the same order as described above. Sectioning was conducted in an RNase-free manner using a slightly modified protocol recommended by Leica to ensure RNA integrity for laser capture microdissection and further molecular analysis of RNA species. Sections were stored in standard microscope slide boxes either at room temperature or the refrigerator until further analyses were performed. For tissues obtained from Biobank Graz, the tissue sectioning was performed at the Center for Medical Research at the Medical University of Graz, where 8 serial sections were obtained in the following order: 1st section on Epredia Superfrost plus slide, 2nd section on Epredia Superfrost plus slide, 3rd section on a Leica PEN membrane slide, 4th section on Epredia Superfrost plus slide, 5th section on Epredia Superfrost plus slide, 6th section on a Leica PEN membrane slide, 7th section on Epredia Superfrost plus, 8th section taken as a scroll/curl, placed into a 1.5 mL microcentrifuge tube (Eppendorf DNA LoBind® Tubes, Cat. No. 0030108051 or equivalent), and frozen at –20 °C for later RNA extraction and RNA-SEQ.

[00187] Whole-slide imaging, image annotation, and image analysis

[00188] The Vectra Polaris is the ideal setup for RNA-FISH images because the instrument contains proprietary filters that allow spectral unmixing of fluorophore signals as well as the ability to isolate unwanted autofluorescence. This alleviates two limitations of RNA-FISH: crosstalk/bleed-through fluorophore emission spectra (especially with high target expression) as well as background noise due to inherent autofluorescence (e.g., lipofuscin, collagen, blood cells) and autofluorescence from tissue preparation (e.g., formalin fixation).

[00189] During our selection process of the appropriate OPAL dyes for each target, we found that RNA-FISH biomarkers that were detected using Opal 520, a fluorescent dye that emits in the green wavelength (525 nm emission maxima), showed the poorest concordance

with RNA-SEQ data. This is likely due to the high levels of background and endogenous tissue autofluorescence that may arise from tissue processing. Because of this, we selected the highest expressed target *ERBB2* to be detected with Opal 520 to maximize the signal to noise.

[00190] To select regions of interest for LCM, we considered both the H&E-stained specimen and RNA-FISH biomarker expression to identify regions of interest (ROI) that displayed heterogeneity. This included specimens that contained more than one histological subtype (ductal carcinoma, lobular carcinoma, mucinous, etc.), the presence of DCIS/LCIS mixed with invasive BCa, or biomarker heterogeneity. Biomarker heterogeneity consisted of ROIs that were positive for a particular marker in one area, but negative for the same marker in another area on the same tissue section (i.e., all or none heterogeneity). Alternatively, we also considered heterogeneity in expression levels where a biomarker is highly expressed in one area but lowly expressed in another area on the same tissue section (i.e., high or low heterogeneity). In some instances, we also conducted LCM on both invasive BCa and adjacent healthy breast epithelial cells.

[00191] **RNA isolation and quality control**

[00192] The Macherey Nagel NucleoSpin totalRNA FFPE XS kit was used for RNA isolation using the manufacturer´s protocol with the following deviations as suggested by the user instructions: 1) omission of the deparaffinization step since this already occurs before cresyl violet staining; 2) lengthening proteinase K digestion to 90 min at 56 °C; 3) inclusion of the optional on-column DNase digestion step to remove residual DNA. Laser capture microdissected samples were processed in batches of 8-16 samples with a single person conducting the isolation procedure for a single batch.

[00193] **RNA library preparation and sequencing**

[00194] We followed several of the manufacturer's recommended changes for using highly degraded FFPE specimens with the Takara SMARTer Stranded Total RNA-Seq Kit v3 - Pico Input Mammalian, including 1) total RNA input range from 0.5-15 ng, although we found the best performance to be from 1 to 20 ng with PCR cycles adjusted accordingly; 2) inclusion of the DNase pretreatment step 3) omission of the fragmentation step; 4) 10 PCR cycles for first PCR amplification step; and 5) 10-13 PCR cycles for the 2nd PCR amplification step depending on starting input. For 1 ng input, we used 13 PCR cycles and for 10 ng and 20 ng inputs, we used 10 PCR cycles. Libraries were prepared in batches of 8-48 samples by an individual scientist, while maintaining the batch organization of prior procedures like laser capture microdissection and RNA isolation.

[00195] The positive control sample consisted of a mixture of total RNA extracted from 3 biological replicates of each IHC-surrogate subtype (e.g., luminal A and B, HER2+, and TNBC), including both node-negative and node-positive specimens (24 specimens total). This

cocktail positive control was aliquoted into single-use aliquots and one sample was prepared for each library preparation batch alongside all other samples. This positive control had a similar quality, quantity, and gene expression relative to our experimental samples and aided in identifying technical batch effects. A second control consisted of a panel of 92 non-overlapping synthetic spike-in RNA controls characterized by the External RNA Control Consortium (https://www.nist.gov/programs-projects/external-rna-controls-consortium) [2] and a mixture of 69 isoform transcripts (SIRV-Set 3; Lexogen, Vienna, Austria) that could be used to identify processing artifacts in individual samples and monitor the performance of library preparation (e.g., dynamic range, technical variability, the limit of detection, etc.). These synthetic spike-ins were introduced into every sample before library preparation at an attomolar input concentration that resulted in an estimated 1% of mRNA reads by using Lexogen's calculator worksheet.

[00196] The bioinformatics pipeline for RNA sequencing is outlined below. Details about statistical analyses for each pipeline can be found in their respective user manuals so it is only briefly summarized here:

[00197] 1) FASTQC (v0.11.9) – used for initial quality check of raw sequencing data to provide general descriptive stats about sequence quality, identify duplicates, potential biases for downstream processing, sample outliers, and over-represented sequences/contaminants.

[00198] 2) Cutadapt (v3.4) – Removal of adapters and filtering for sequence quality and length.

[00199] 3) STAR (v2.7.7a) – alignment tool used to map reads to the human genome, providing metrics about mapped/unmapped reads, low-quality reads, and ambiguous reads.

[00200] 4) SAMtools (v1.16.1) – used for indexing and sorting bam files.

[00201] 5) UMI-tools (v1.0.0) – used for deduplicating bam files based on unique molecular identifiers introduced during library preparation.

[00202] 6) FeatureCounts (v1.5.2) – this was used to identify gene and transcript abundance across various RNA biotypes (e.g., mRNA, snRNA, lincRNA, rRNA, etc.). Read counts were normalized to transcripts per million (TPM) for an overview of biotypes and transcript abundance.

[00203] 7) DESeq2 (v.1.36.0) and Wilcoxon rank sum tests – Raw integer count matrices were imported into DeSeq2, a tool used to assess differential gene expression using a generalized linear model under consideration of a negative binomial distribution (Poisson-gamma model). To estimate differentially expressed genes, the shrinkage estimator apeglm was used. Wilcoxon rank sum tests were also used to test for differential expression in a non-parametric approach. To account for multiple hypothesis testing, FDR adjusted P-values ($P < 0.01$) were used to assess significance.

[00204] Generation of molecular subtyping gene list

[00205] To generate a list of genes for molecular subtyping of samples into luminal A, luminal B, HER2, and basal-like subtypes, we conducted differential gene expression (DE) analysis using DESeq2 and the Wilcoxon rank-sum test, a nonparametric approach that better controls the false discovery rate (FDR) when analyzing large sample sizes. We assessed DE between each subtype combination (e.g., luminal A vs luminal B, luminal A vs Her2, etc.) and took the union of the top 50 significant DE genes (FDR adjusted $p < 0.05$ with fold change > 4), which yielded 196 genes (pseudogenes were manually removed). We generated a second list of genes to better differentiate high vs low OncotypeDX luminal samples using the following approach: First, we filtered luminal A and luminal B samples, which did not agree between the biobank surrogate subtypes based on St. Gallen criteria (e.g., Ki-67 < 30 for luminal A or Grade 3 for luminal B) and our classification based on 196 genes. Then we classified discordant luminal samples to either high or low risk based on their OncotypeDX score. Next, we performed DE analysis between luminal high and low risk samples by DESeq2 and Wilcoxon, took the intersected results from these two methods, and selected 30 significant DE genes (FDR adjusted $p < 0.05$ with fold change > 4; pseudogenes were manually removed). We also conducted unsupervised clustering and selected 78 genes that were enriched in the luminal androgen receptor (LAR) TNBC subtype. We then took the union of our 196 subtype DE genes, 30 luminal high/low risk genes, and 78 LAR TNBC subtype genes, to generate a final list of 293 genes. To classify individual samples, we used consensus clustering (kmeans = 4; ConsensusClusterPlus, v.1.60.0 package in R) with this 293-gene list. Heatmaps were generated by the ComplexHeatmap package in R (v.2.12.1).

[00206] Benchmarking subtyping classifiers and development of consensus subtyping scheme

[00207] The mFISHseq 293-gene molecular subtype classifier was benchmarked against the IHC-surrogate subtypes (using the definitions proposed by the 2013 St. Gallen International Breast Cancer Conference Panel), research based PAM50 and AIMS implemented in the GeneFu package, as well as the six and four Lehmann's TNBC subtypes using TNBCtype. To assess risk of single samples belonging to a particular subtype, we again used GeneFu to derive scores for several research based prognostic signatures, including the 21-gene OncotypeDX, GENE70 (research version of Mammaprint), the PAM50 Risk of Recurrence by Subtype only (ROR-S), and Genomic Grade Index (GGI). For clinical risk, we used a modified Adjuvant! Online criteria outlined in the TAILORx and MINDACT trials to classify samples as low or high risk.

[00208] To derive a consensus subtype, we used a simple voting scheme between multigene subtyping tools (mFISHseq, PAM50, and AIMS) where the consensus subtype was

determined by the majority. In seven cases (Luminal A: n=3; Luminal B: n=2; TNBC: n=2), the three multigene subtyping tools assigned a different subtype to each case, so we used the IHC-surrogate subtype as the tiebreaker. Other comparison metrics included concordance between all samples and stratified by subtype using Cohen's kappa statistics, survival using KM plots and log rank tests, and stability assessed by correlation to PAM50 centroids.

[00209] Univariate and multivariate logistic regression using I-SPY2 trial data

[00210] We downloaded data from the T–DM1, control, and pertuzumab arms of the I-SPY2 trial from NCBI's Gene Expression Omnibus (GEO) under accession code GSE181574. The T-DM1 arm contains 52 patients (ER+/HER2+: n=35, ER–/HER2+: n=17) treated with T-DM1 and pertuzumab with 30 patients achieving pathological complete response (pCR). The control arm contains 31 patients (ER+/HER2+: n=19, ER–/HER2+: n=12) treated with paclitaxel and trastuzumab with 8 patients achieving pCR. The pertuzumab arm (used as a control arm for pertuzumab in the T-DM1 arm) contains 44 patients (ER+/HER2+: n=29, ER–/HER2+: n=15) treated with paclitaxel, pertuzumab, and trastuzumab with 26 patients achieving pCR. The T-DM1 dataset was stratified into training and test (50:50) data based on pCR and hormonal status. To associate the gene/gene signature with pCR, we performed univariate logistic regression in R through lmtest (v09-40) with the likelihood ratio test assessing significance. We further combined genes and gene signatures into the multivariate logistic model with elastic net regularization using tidymodels (v 1.1.1) with hyperparameter tuning by 10-fold cross validation. The final 19-feature classifier was then applied to the T-DM1 test set and compared with *ERBB2* alone. The same 19-feature classifier and ERBB2 alone was also tested in the control and pertuzumab arms to assess the specificity in predictive utility.

[00211] METABRIC microarray and clinical data for 2,509 patients was downloaded from the cBioPortal for Cancer Genomics (https://www.cbioportal.org/study/summary?id=brca_metabric). TCGA-BRCA data was downloaded from several different sources: 1. TCGA RNA-sequencing data was downloaded from the NIH GDC Data Portal database (https://portal.gdc.cancer.gov/), filtering for TCGA-BRCA, transcriptomic profiling, RNA-seq and Gene Expression Quantification to get 1231 samples; 2. TCGA clinical data was downloaded from cBioPortal for Cancer Genomics from both the TCGA, GDC dataset (n=1,103) and the TCGA, PanCancer Atlas (n=1,084); 3. TCGA PAM50 Ground truth subtypes were downloaded from cBioPortal for Cancer Genomics from the flagship dataset (TCGA, Nature 2012, n=825 with 521 samples having PAM50 subtype calls) and a more updated analysis from Perou and colleagues molecular analysis of TCGA breast cancer histologic types (PMID: 35465400).

[00212] Results

[00213] Breast cancer (BCa) intrinsic molecular subgroups (e.g., luminal A/B (LumA/LumB), HER2-overexpressing (OE), basal- and normal-like) differ in prognosis, response to therapy, and clinical outcomes. Multigene assays aid therapeutic decision-making; however, their single sample reproducibility is only moderate and potentially exacerbated by their use of bulk processing/crude macrodissection, which loses spatial context and might lead to gene expression inaccuracies from non-tumor elements. Current tools are limited in predicting response to antibody-drug conjugates (ADCs), which demonstrate efficacy despite negative/low antigen expression by immunohistochemistry (IHC). Enhancing predictive accuracy may require assessment of biomarkers, related not only to the antigen but also intracellular ADC processing and cytotoxic payload targets.

[00214] We retrospectively validated mFISHseq, which uses multiplexed fluorescent in situ hybridization (RNA-FISH) of ESR1, PGR, ERBB2 and MKI67 to define regions of interest (ROIs) used for spatially-resolved tumor enrichment through laser capture microdissection (LCM) and RNA-sequencing (SEQ, Figure 1a). We demonstrate its analytical validity against IHC, use in a consensus subtyping approach that mitigates limitations of individual classifiers, and value for ADC response prediction.

[00215] By analyzing 1,082 formalin-fixed paraffin-embedded (FFPE) BCa samples (Extended Table 1), we compared RNA-FISH/-SEQ to IHC for estrogen (ER)/progesterone (PR) receptors, HER2, and Ki67. We observed similar expression patterns for each marker and when stratified into IHC-surrogate subtypes (**Figures 6 and 7**). To determine biomarker thresholds for mFISHseq, we used training and test (70:30) cohorts and found excellent concordance between RNA-SEQ and IHC for all biomarkers (ROC and precision-recall AUCs ≥ 0.93) except ERBB2 had slightly lower precision-recall (AUC ≥ 0.86; **Figure 1b; Figure 8a,b**; Supplementary Table 1). Although RNA-FISH was primarily used to guide LCM, RNA-FISH quantification by digital pathology and machine learning (**Figure 1**) revealed very good concordance with IHC (ROC and precision-recall AUCs ≥ 0.80; **Figure 8c-g**), supporting diagnostic utility for both orthogonal methodologies.

[00216] **Extended Table 1. Cohort descriptive statistics. One specimen classified as G2&G3. [2]Two specimens classified as G2&G3. [3]Two specimens classified as G4.**

Characteristic	Luminal A			Luminal B			Her2			TNBC		
Subtype sample #	*432*			*313*			*87*			*181*		
Biobank	PATH	GRAZ	Mal + AMS	PATH	GRAZ	Mal + AMS	PATH	GRAZ	Mal + AMS	PATH	GRAZ	Mal + AMS
Sample #	205 (47%)	205 (47%)	22 (5%)	185 (59%)	125 (40%)	3 (1%)	61 (70%)	23 (26%)	3 (3%)	141 (78%)	37 (20%)	3 (2%)

Sex												
Female	205 (100%)	202 (99%)	-	185 (100%)	121 (97%)	-	61 (100%)	23 (100%)	-	141 (100%)	37 (100%)	-
Male	0 (0%)	3 (1%)	-	0 (0%)	4 (3%)	-	0 (0%)	0 (0%)	-	0 (0%)	0 (0%)	-
Median age at surgery (years)	62	57	62	62	59	62	62	52	62	57	52	52
Menopausal status												
Premenopausal	35 (17%)	-	-	42 (23%)	-	-	7 (11%)	-	-	29 (21%)	-	-
Perimenopausal	6 (3%)	-	-	3 (2%)	-	-	1 (2%)	-	-	8 (6%)	-	-
Postmenopausal	159 (78%)	-	-	137 (74%)	-	-	43 (70%)	-	-	93 (66%)	-	-
Unavailable	5 (2%)	-	-	3 (2%)	-	-	10 (16%)	-	-	11 (8%)	-	-
Tumor size (T)												
pT1	128 (62%)	150 (73%)	7 (32%)	79 (43%)	71 (56%)	1 (33%)	26 (43%)	14 (61%)	1 (33%)	68 (48%)	22 (59%)	0 (0%)
pT2	69 (34%)	45(22%)	11 (50%)	94 (51%)	48 (38%)	1 (33%)	31 (51%)	8 (35%)	1 (33%)	65 (46%)	14 (38%)	2 (67%)
pT3 + pT4	8 (4%)	10 (5%)	4 (18%)	12 (6%)	6 (5%)	1 (33%)	4 (7%)	1 (4%)	1 (33%)	8 (6%)	1 (3%)	1 (33%)
Node status (N)												
pN0	98 (48%)	127 (62%)	5 (23%)	93 (50%)	70 (56%)	0 (0%)	30 (49%)	10 (43%)	0 (0%)	99 (70%)	28 (76%)	0 (0%)
pN1	88 (43%)	57 (28%)	11 (50%)	65 (35%)	36 (29%)	1 (33%)	21 (34%)	6 (26%)	0 (0%)	29 (21%)	8 (22%)	0 (0%)
pN2	13 (6%)	10 (5%)	4 (18%)	18 (10%)	10 (8%)	0 (0%)	4 (7%)	3 (13%)	2 (67%)	6 (4%)	0 (0%)	1 (33%)
pN3	6 (3%)	8 (4%)	1 (5%)	9 (5%)	8 (6%)	1 (33%)	5 (8%)	2 (9%)	0 (0%)	7 (5%)	1 (3%)	1 (33%)
Unavailable	0 (0%)	3 (1%)	1 (5%)	0 (0%)	1 (1%)	1 (33%)	1 (2%)	2 (9%)	1 (33%)	0 (0%)	0 (0%)	1 (33%)
Grade (G)												

G1	61 (30 %)	60 (29%)	3 (14%)	1 (1%)	4 (3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)	1 (3%)	0 (0%)
G2	144 (70 %)	140[1] (68%)	10 (45%)	30 (16%)	31[2] (25%)	3 (100%)	17 (28 %)	3 (13 %)	1 (33%)	32 (23%)	7 (19%)	1 (33%)
G3	0 (0%)	4 (2%)	9 (41%)	154 (83%)	89 (71%)	0 (0%)	44 (72 %)	20 (87 %)	2 67%)	108 (77%)	28[3] (76%)	2 (67%)
Unavailabl e	0 (0%)	1 (<1%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (3%)	0 (0%)
Lymphatic invasion (L)												
L0	137 (67 %)	-	15 (68%)	105 (57%)	-	2 (67%)	31 (51 %)	-	2 (67%)	99 (70%)	-	0 (0%)
L1	38 (19 %)	-	6 (27%)	55 (30%)	-	0 (0%)	24 (39 %)	-	0 (0%)	35 (25%)	-	2 (67%)
Unavailable	30 (15 %)	-	1 (5%)	25 (14%)	-	1 (33%)	6 (10 %)	-	1 (33%)	7 (5%)	-	1 (33%)
Residual tumor (PATH only)												
R0	204 (10 0%)	-	19 (86%)	180 (97%)	-	2 (67%)	60 (98 %)	-	2 (67%)	138 (98%)	-	2 (67%)
R1	0 (0%)	-	2 (9%)	4 (2%)	-	0 (0%)	1 (2%)	-	0 (0%)	2 (1%)	-	0 (0%)
Unavailabl e	1 (<1 %)	-	1 (5%)	1 (1%)	-	1 (33%)	0 (0%)	-	1 (33%)	1 (1%)	-	1 (33%)
Histopath ological tumor types												
Invasive ductal carcinoma (IDC)	142 (69 %)	135 (66%)	17 (77%)	154 (83%)	103 (82%)	3 (100%)	52 (85 %)	21 (91 %)	3 (100 %)	125 (89%)	26 (70%)	3 (100 %)
IDC mixed	3 (1%)	10 (5%)	2 (9%)	10 (5%)	3 (2%)	0 (0%)	6 (10 %)	0 (0%)	0 (0%)	7 (5%)	2 (5%)	0 (0%)

Invasive lobular carcinoma (ILC)	38 (19 %)	42 (20%)	3 (14%)	16 (9%)	14 (11%)	0 (0%)	1 (2%)	1 (4%)	0 (0%)	1 (1%)	1 (3%)	0 (0%)
ILC mixed	4 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Metaplasti c carcinoma	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (2%)	5 (14%)	0 (0%)
Mucinous carcinoma	9 (4%)	8 (4%)	0 (0%)	1 (1%)	2 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Other rare histopatho logical subtypes	8 (4%)	9 (4%)	0 (0%)	4 (2%)	3 (2%)	0 (0%)	2 (3%)	0 (0%)	0 (0%)	5 (4%)	2 (5%)	0 (0%)
Unavailabl e	1 (<1 %)	1 (<1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (4%)	0 (0%)	0 (0%)	1 (3%)	0 (0%)
Estrogen receptor positive (%)	205 (10 0%)	-	-	185 (100 %)	-	-	0 (0%)	-	-	0 (0%)	-	-
IRS 0	0 (0%)	-	-	0 (0%)	-	-	41 (67 %)	-	-	97 (69%)	-	-
IRS 1-2	1 (<1 %)	-	-	23 (12%)	-	-	2 (3%)	-	-	4 (3%)	-	-
IRS 3-4	4 (2%)	-	-	20 (11%)	-	-	0 (0%)	-	-	0 (0%)	-	-
IRS 6-8	27 (13 %)	-	-	25 (14%)	-	-	0 (0%)	-	-	0 (0%)	-	-
IRS 9	18 (9%)	-	-	20 (11%)	-	-	0 (0%)	-	-	0 (0%)	-	-
IRS 12	155 (76 %)	-	-	95 (51%)	-	-	0 (0%)	-	-	0 (0%)	-	-
Unavailabl e	0 (0%)	-	-	2 (1%)	-	-	18 (30 %)	-	-	40 (28%)	-	-
Progester one receptor positive (%)	181 (88 %)	-	-	130 (70%)	-	-	0 (0%)	-	-	0 (0%)	-	-
IRS 0	20 (10 %)	-	-	43 (23%)	-	-	43 (70 %)	-	-	97 (69%)	-	-
IRS 1-2	12 (6%)	-	-	29 (16%)	-	-	0 (0%)	-	-	5 (4%)	-	-

IRS 3-4	24 (12 %)	-	-	23 (12%)	-	-	0 (0%)	-	-	0 (0%)	-	-
IRS 6-8	37 (18 %)	-	-	34 (18%)	-	-	0 (0%)	-	-	0 (0%)	-	-
IRS 9	34 (17 %)	-	-	21 (11%)	-	-	0 (0%)	-	-	0 (0%)	-	-
IRS 12	78 (38 %)	-	-	33 (18%)	-	-	0 (0%)	-	-	0 (0%)	-	-
Unavailabl e	0 (0%)	-	-	2 (1%)	-	-	18 (30 %)	-	-	39 (28%)	-	-
HER2 positive	0 (0%)	-	-	51 (28%)	-	-	61 (10 0%)	-	-	0 (0%)	-	-
HER2 0	104 (51 %)	-	-	65 (35%)	-	-	0 (0%)	-	-	105 (74%)	-	-
HER2 1+	90 (44 %)	-	-	58 (31%)	-	-	0 (0%)	-	-	32 (23%)	-	-
HER2 2+ Neg.	10 (5%)	-	-	8 (4%)	-	-	0 (0%)	-	-	1 (1%)	-	-
HER2 2+ Pos.	0 (0%)	-	-	8 (4%)	-	-	5 (8%)	-	-	0 (0%)	-	-
HER2 3+	0 (0%)	-	-	40 (22%)	-	-	53 (87 %)	-	-	0 (0%)	-	-
Unavailabl e	1 (<1 %)	-	-	6 (3%)	-	-	3 (5%)	-	-	3 (2%)	-	-
KI67 positive												
≤5%	13 (6%)	46 (22%)	-	1 (1%)	3 (2%)	-	0 (0%)	0 (0%)	-	1 (1%)	0 (0%)	-
>5 to <30%	63 (31 %)	108 (52%)	-	22 (12%)	19 (15%)	-	1 (2%)	2 (9%)	-	3 (2%)	2 (5%)	-
≥30%	3 (1%)	0 (0%)	-	35 (19%)	70 (56%)	-	4 (7%)	11 (48 %)	-	21 (15%)	22 (59%)	-
Unavailabl e	126 (61 %)	51 (25%)	-	127 (69%)	33 (26%)	-	56 (92 %)	10 (43 %)	-	116 (82%)	13 (35%)	-
Treatment type												
Chemothe rapy												

Yes	51 (25 %)	41 (20%)	-	136 (74%)	73 (58%)	-	54 (89 %)	19 (83 %)	-	125 (89%)	35 (95%)	-
No	152 (74 %)	164 (80%)	-	48 (26%)	52 (41%)	-	5 (8%)	4 (17 %)	-	14 (10%)	2 (5%)	-
Unavailabl e	2 (1%)	0 (0%)	-	1 (1%)	0 (0%)	-	2 (3%)	0 (0%)	-	2 (1%)	0 (0%)	-
Anti-hormonal therapy												
Yes	194 (95 %)	100 (49%)	-	163 (88%)	61 (48%)	-	10 (16 %)	0 (0%)	-	9 (6%)	0 (0%)	-
No	9 (4%)	106 (51%)	-	21 (11%)	65 (52%)	-	49 (80 %)	23 (10 0%)	-	131 (93%)	37 (100 %)	-
Unavailabl e	2 (1%)	0 (0%)	-	1 (1%)	0 (0%)	-	2 (3%)	0 (0%)	-	1 (1%)	0 (0%)	-
AI inhibitor therapy												
Yes	129 (63 %)	101 (49%)	-	123 (66%)	54 (43%)	-	9 (15 %)	0 (0%)	-	5 (4%)	0 (0%)	-
No	72 (35 %)	104 (51%)	-	59 (32%)	71 (57%)	-	52 (85 %)	23 (10 0%)	-	136 (96%)	37 (100 %)	-
Unavailabl e	4 (2%)	0 (0%)	-	3 (2%)	0 (0%)	-	0 (0%)	0 (0%)	-	0 (0%)	0 (0%)	-
Anti-Her2 therapy												
Yes	0 (0%)	0 (0%)	-	42 (23%)	20 (16%)	-	50 (82 %)	22 (96 %)	-	1 (1%)	0 (0%)	-
No	203 (99 %)	205 (100 %)	-	140 (76%)	105 (83%)	-	9 (15 %)	1 (4%)	-	139 (99%)	37 (100 %)	-
Unavailabl e	2 (1%)	0 (0%)	-	3 (2%)	0 (0%)	-	2 (3%)	0 (0%)	-	1 (1%)	0 (0%)	-
Radiotherapy												
IORT	6 (3%)	0 (0%)	-	2 (1%)	0 (0%)	-	0 (0%)	0 (0%)	-	1 (1%)	0 (0%)	-
N	18 (9%)	40 (19%)	-	14 (8%)	21 (17%)	-	4 (7%)	5 (22 %)	-	15 (11%)	5 (14%)	-
Y	147 (72 %)	165 (80%)	-	131 (71%)	104 (83%)	-	51 (84 %)	18 (78 %)	-	114 (81%)	32 (86%)	-

Y + IORT	31 (15 %)	0 (0%)	-	37 (20%)	0 (0%)	-	5 (8%)	0 (0%)	-	10 (7%)	0 (0%)	-
Unavailabl e	3 (1%)	0 (0%)	-	1 (1%)	0 (0%)	-	1 (2%)	0 (0%)	-	1 (1%)	0 (0%)	-
Metastasi s												
Yes	8 (4%)	35 (17%)	-	12 (6%)	41 (33%)	-	3 (5%)	7 (30 %)	-	18	5	-
No	173 (84 %)	170 (83%)	-	120 (65%)	84 (67%)	-	48 (79 %)	16 (70 %)	-	93	32	-
Unavailabl e/Lost to follow-up	24 (12 %)	0 (0%)	-	53 (29%)	0 (0%)	-	10 (16 %)	0 (0%)	-	30	0 (0%)	-
Progressio n free survival (Median)	70 mo nth s	90 mon ths	-	64.5 mont hs	94 mon ths	-	98. 5 mo nth s	89 mo nth s	-	84.5 mont hs	123 mon ths	-
Unavailabl e/ lost to follow-up	8 (4%)	1 (<1%)	-	23 (12%)	2 (2%)	-	3 (5%)	0 (0%)	-	19	0 (0%)	-
Overall survival (Median)	72 mo nth s	97 mon ths	-	66 mont hs	103 mon ths	-	102 mo nth s	101 mo nth s	-	85.5 mont hs	123 mon ths	-
Unavailabl e/ lost to follow-up	11 (5%)	0 (0%)	-	22 (12%)	0 (0%)	-	1 (2%)	0 (0%)	-	15	0 (0%)	-
Deceased												
Yes	40 (20 %)	25 (12%)	-	61 (33%)	19 (15%)	-	17 (28 %)	6 (26 %)	-	50 (35%)	3 (8%)	-
No	165 (80 %)	0 (0%)	-	124 (67%)	0 (0%)	-	44 (72 %)	0 (0%)	-	91 (65%)	0 (0%)	-
Unavailabl e	0 (0%)	180 (88%)	-	0 (0%)	106 (85%)	-	0 (0%)	17 (74 %)	-	0 (0%)	34 (92%)	-

[00217] Supplementary Table 1. Diagnostic performance of RNA-seq training and test data set:

Marker	PR		ER		HER2		Ki-67	
Train/Test data	Train	Test	Train	Test	Train	Test	Train	Test
TP	393	165	486	212	89	40	106	45
FP	25	11	16	11	6	6	1	1
TN	238	104	171	71	578	258	49	14
FN	36	28	29	18	28	8	9	6
Accuracy	0.9118	0.8734	0.9359	0.9071	0.9515	0.9551	0.9394	0.8939
95% CI	0.8882	0.831	0.9152	0.8692	0.9329	0.9259	0.8914	0.7936
	0.9319	0.9084	0.9529	0.9369	0.9662	0.9753	0.9706	0.9563
No Information Rate	0.6199	0.6266	0.7336	0.7372	0.8331	0.8462	0.697	0.7727
P-Value [Acc >NIR]	<2e-16	<2e-16	<2e-16	4.35E-14	<2.2e-16	9.42E-10	1.19E-14	9.51E-03
Kappa	0.8144	0.7372	0.8395	0.7665	0.8114	0.8247	0.8627	0.7298
Mcnemar's Test P-Value	0.2004	0.01041	0.07364	0.2652	0.0003164	0.7893	0.02686	0.13057
Sensitivity	0.9161	0.8549	0.9437	0.9217	0.7607	0.8333	0.9217	0.8824
Specificity	0.9049	0.9043	0.9144	0.8659	0.9897	0.9773	0.98	0.9333
Pos Pred Value	0.9402	0.9375	0.9681	0.9507	0.9368	0.8696	0.9907	0.9783
Neg Pred Value	0.8686	0.7879	0.855	0.7978	0.9538	0.9699	0.8448	0.7
Precision	0.9402	0.9375	0.9681	0.9507	0.9368	0.8696	0.9907	0.9783
Recall	0.9161	0.8549	0.9437	0.9217	0.7607	0.8333	0.9217	0.8824
F1	0.928	0.8943	0.9558	0.936	0.8396	0.8511	0.955	0.9278
Prevalence	0.6199	0.6266	0.7336	0.7372	0.1669	0.1538	0.697	0.7727
Detection Rate	0.5679	0.5357	0.6923	0.6795	0.127	0.1282	0.6424	0.6818
Detection Prevalence	0.604	0.5714	0.7151	0.7147	0.1355	0.1474	0.6485	0.697
Balanced Accuracy	0.9105	0.8796	0.9291	0.8938	0.8752	0.9053	0.9509	0.9078

[00218] To classify BCa samples into molecular subtypes, we conducted differential gene expression analysis that revealed 293 genes (Extended Table 2,3 and Figure 9a). Semi-supervised consensus clustering on these 293 genes was benchmarked against IHC-surrogate, PAM50, and Absolute Intrinsic Molecular Subtyping (AIMS) classifications (Figure 1c). Multigene methods classified fewer LumA and more HER2-OE/basal-like samples than IHC-surrogate subtyping (Figure 9b). While subtyping methods had comparable survival (Figure 9c), they had only moderate concordance between each multigene classifier and IHC-surrogate subtypes and substantial concordance among multigene classifiers (Figure 9d). Indeed, 45% (459/1013) of IHC-surrogate samples showed discordance in ≥1 classifier and these samples had divergent survival (Figure 9e). Discordant samples showed instability with progressively lower correlations to PAM50 centroids (Figure 9f).

[00219] Extended Table 2. Overview of differential gene expression analysis among molecular subtypes:

Groups	DESeq2 (LFC shrinkage)	Wilcoxon rank-sum test	Intersection
Her2onl vs TNBC	1138	1372	1006
Downregulated	*590*	*704*	*523*
Upregulated	*548*	*668*	*483*
Her2onl vs LumA	3636	4784	3364
Downregulated	*1188*	*1312*	*1094*
Upregulated	*2448*	*3472*	*2270*
Her2onl vs LumB	1577	2181	1338
Downregulated	*621*	*734*	*547*
Upregulated	*956*	*1447*	791
LumA vs TNBC	4950	5568	4661
Downregulated	*3330*	*3902*	*3149*
Upregulated	*1620*	*1666*	*1512*
LumB vs TNBC	2635	2954	2356
Downregulated	*1564*	*1785*	*1360*
Upregulated	*1071*	*1169*	*996*
LumA vs LumB	746	802	525
Downregulated	*621*	*686*	443
Upregulated	*125*	*116*	82

[00220] To improve sample level agreement, we implemented a consensus intrinsic subtype using a simple voting scheme for mFISHseq, PAM50, and AIMS and found differences between IHC-surrogate and consensus subtypes with prognostic implications. For example, 24% (102/432) of IHC-surrogate LumA samples were reclassified as LumB, showing poorer survival for node negative samples relative to those unanimously classified as LumA by all methods (Figure 2a,b). Approximately 15% (46/313) of IHC-surrogate LumB samples were reclassified as basal-like and 21% (65/313) as LumA, with poorer and better survival, respectively, compared to samples unanimously classified as LumB (Figure 2a,b). Consensus subtyping reclassified 4% (7/181) of TNBC IHC-surrogate samples into LumA/B and normal-like subtypes and 13% (23/181) into HER2-OE. These reclassified samples showed more favorable and poorer survival, respectively, when compared to consensus basal-like samples (Figure 2a,b).

[00221] We ascertained discordance at the single sample level by exploring the 41% (415/1013) of samples uniquely classified into a specific molecular subtype by a single

classifier (Figures 10-11). Uniquely classified LumA showed shorter survival and higher genomic/clinical risk relative to samples classified as LumA by all tools (LumA-All; Figure 10a,b). In contrast, mFISHseq and PAM50 uniquely classified LumB had more favorable survival and enriched intermediate/low risk scores compared to samples classified as LumB by all tools (LumB-All; Figure 10c,d). Consensus subtyping reclassified most of these samples to a subtype that better matched their risk and survival: 83% (n=91/109) of unique LumA and 97% (n=69/71) of mFISHseq and PAM50 uniquely classified LumB were reclassified as LumB and LumA, respectively (Figure 10b,d). IHC-surrogate only LumB samples, which showed similar survival to LumB-All, were reclassified by consensus subtyping into LumA (n=29)/normal-like (n=6) and HER2-OE (n=24)/basal-like (n=46; Figure 10d), which showed either more favorable or poorer survival, respectively (Figure 10c, inset).

[00222] Samples uniquely classified as HER2-OE showed disparate findings depending on the classifier (Figure 10e,f). Relative to HER2-All samples, IHC-surrogate HER2+ samples had equivalent survival, higher genomic/clinical risk, and were reclassified as basal-like by consensus subtyping; AIMS HER2-OE samples had poorer survival, higher genomic/clinical risk, and most samples were reclassified as basal-like (n=15/24) or LumB (n=8/24); mFISHseq and PAM50 HER2-OE samples had better survival and a subset of these samples (n=9) with lower GENE70 and/or clinical risk experienced no deaths and were predominantly reclassified into LumA (n=7/9).

[00223] While IHC-surrogate only TNBC samples showed poorer survival compared to those classified as TNBC/basal-like by all methods (Basal-like-All), the samples classified as basal-like by multigene classifiers had better prognosis (Figure 10g). IHC-surrogate only TNBC samples encompassed 13 samples with low GENE70 or clinical risk and favorable survival, and 15 samples with high risk and poor survival (Figure 10g, inset,5f). Consensus subtyping stratified the samples into clinically meaningful subgroups with deaths occurring in 10/22 samples reclassified as HER2-OE/basal-like and only 1/5 samples reclassified as LumA/B (Figure 10f). AIMS and PAM50 independently classified normal-like samples were reclassified by consensus subtyping to basal-like (n=18), LumA (n=14), and HER2-OE (n=3), providing more clinically relevant subtyping (Figure 11a,b).

[00224] Consensus subtyping reclassified 21% (214/1013) of uniquely classified IHC-surrogate samples into intrinsic subtypes that better fit survival data. Similarly, samples uniquely classified by mFISHseq (n=55), AIMS (n=83), and PAM50 (n=63) were reclassified by a consensus of the other two multigene classifiers to a subtype that matched survival. Overall, 30% (305/1013) of samples were reclassified from their IHC-surrogate subtype into a different consensus subtype showing prognostic utility (Figure 2c,d). This improved concordance between the consensus subtypes and other multigene classifiers to near perfect agreement, even for more challenging luminal samples (Figure 9d).

[00225] Biomarkers predicting response to ADCs are an unmet need. Gene expression analysis of 20 targets of FDA/EMA-approved ADC (or in clinical trials) revealed high variability, differential expression between normal and tumor tissue, and subtype-specific enrichment (Extended Figure 3). Expression of various ADC targets were associated with survival in a subtype-specific manner and occasionally in opposite ways. TACSTD2 (TROP2), targeted by sacituzumab govitecan, was associated with favorable survival in LumA but poor survival in TNBC. ERBB2, ERBB3, and TPBG predicted favorable survival in LumB but poor survival in TNBC (Figure 2e,f).

[00226] Because response to ADCs likely encompasses broad ADC-processing pathways, we investigated 70 genes and 32 gene signatures related to endocytosis, lysosomal function, payload targets, and resistance pathways, positing that combining markers may elucidate ADC-responsive subgroups. These ADC-processing genes/gene signatures displayed marked variability (Extended Figure 2a) and were independent predictors of survival (Extended Figure 2b,c). This suggests patients could be stratified into subgroups based on expression of ADC-relevant markers.

[00227] Given our retrospective cohort did not include ADC-treated patients, we re-analyzed the trastuzumab emtansine (T-DM1) arm of the I-SPY2 trial to determine whether combinations of ADC-processing markers would predict response to T-DM1. Univariate logistic regression on 71 pre-specified ADC-relevant markers in a training/test set (50:50) revealed various ADC-processing markers predicted pathologic complete response (pCR; Figure 2g). Elastic net multivariate regression (10-fold cross validation) yielded a 19-feature classifier that displayed superior predictive utility than ERBB2 alone (ROC AUC of 0.96 vs 0.87, Figure 2h,i). The dominant features in the signature predicting T-DM1 efficacy were related to the antigen, endocytosis, lysosome function, maytansine payload, and resistance markers (Figure 2i,j) This highlights the clinical utility of integrating ADC-processing markers into predictive models for effective patient selection.

[00228] In summary, on a retrospective cohort of 1,082 samples, mFISHseq showed excellent concordance with IHC biomarkers, a finding that is consistent with others and underscores its analytical validity. Although molecular subtyping has clinical utility, there is discordance at the single-sample level. Indeed, the OPTIMA study showed marked disagreement in 40.7% of patients' commercial prognostic tests, despite being conducted by the vendors. Our consensus subtyping alleviated this discordance and reclassified 30% of samples into subtypes that better corresponded to their overall survival and prognostic risk, thus resolving a problem that has impeded adoption of multigene subtyping in clinical practice. Changes in the gene list has the greatest influence on intra-/ inter-predictor concordance, providing support to our consensus subtyping approach that combines several multigene classifiers with different gene lists for subtype predictions. The substantial inter-predictor

concordance we observed without consensus subtyping was higher than prior reports, potentially due to using LCM. Bulk/macrodissected specimens may introduce erroneous gene expression from non-tumor tissue, leading to misclassification of BCa samples, which is exemplified by the refinement of the six TNBC subtypes into four following LCM use.

[00229] Finally, through a comprehensive analyses of ADC processing-related genes/gene signatures, we characterized potential ADC treatment-response subgroup and then validated this concept using the T-DM1 arm of the I-SPY2 trial, where we identified a classifier with 19 features related to ADC processing that predicted response. This classifier outperformed ERBB2 expression alone, supporting the notion that using more than antigen target levels, patients can be stratified into ADC-responsive subgroups.

[00230] **Example 2: Unraveling the tumor biology of Her2-low**

[00231] The recent approval of the antibody drug conjugate trastuzumab deruxtecan for treatment of individuals with low levels of Her2 in the advanced/metastatic setting has spurred lively debate on whether Her2 constitutes a unique biological entity[53]. Rather than assessing Her2 low in the context of IHC results, which are broadly acknowledged to poorly separate low levels of Her2 from the absence of Her2 [54], we took an alternative approach and instead used RNA sequencing as the ground truth for *ERBB2* expression levels. First, we separated the cohort into *ESR1* positive and negative groups using a cutoff of 6.2 TPM (**Figure 12a**), which resulted in 98.3% (674/686 samples) of samples also being IHC positive. In the *ESR1* negative group, 78.5% (259/330) samples were IHC negative, while the majority (53.5%) of IHC positive samples contained low ER expression. Then we stratified the samples in both *ESR1* groups by using the interquartile range of *ERBB2* TPM (**Figure 12a**), resulting in *ERBB2* positive (above 75th percentile), *ERBB2* low (25th-75th percentile), and *ERBB2* negative (below 25th percentile).

[00232] As shown in **Table 3**, differential expression analysis of these *ERRB2* expression subgroups revealed a minimal number of upregulated genes in *ERBB2* low relative to *ERBB2* negative, with the *ESR1* positive group showing the least upregulated genes (**Figure 12b**).

[00233] **Table 3. Overview of differential gene expression analysis among *ERBB2* positive, low, and negative samples stratified by *ESR1* status.**

	Groups	**DESeq2 (LFC shrinkage)**	**Wilcoxon rank-sum test**	**Intersection**
ERneg	Her2low vs HER2neg	851	682	330
	Downregulated	*785*	*601*	*318*
		66	*81*	*12*

	Upregulated			
	Her2low vs HER2pos *Downregulated* *Upregulated*	1051 *240* *811*	1999 *868* *1131*	*727* *207* *520*
	Her2neg vs HER2pos *Downregulated* *Upregulated*	2427 *412* *2015*	2310 *923* *1387*	1073 *355* *718*
ERpos	Her2low vs HER2neg *Downregulated* *Upregulated*	5436 *5428* *8*	4779 *4759* *20*	3180 *3178* *2*
	Her2low vs HER2pos *Downregulated* *Upregulated*	131 *62* *69*	141 *85* *56*	37 *29* *8*
	Her2neg vs HER2pos *Downregulated* *Upregulated*	7497 *84* *7413*	603 *76* *527*	396 *48* *348*

[00234] For the *ESR1* positive subgroup, only two genes were found to be upregulated in *ERBB2* low by both DeSeq2 and Wilcoxon, including *CRISPLD1*, a secretory protein that is part of the cysteine-rich secretory proteins (CAP) superfamily implicated in cancer and immunity, and the *UGT2A3* pseudogene. Expanding the analysis to upregulated DEGs by either DeSeq2 or Wilcoxon revealed several upregulated genes, including *KRT1*, *FLG*, *FLG2*, and *OLFM4*, with shared common functions as structural components of skin, cell adhesion processes, and protein binding ability. In the ESR1 negative subgroup, twelve genes were found to be upregulated in *ERBB2* low by both DeSeq2 and Wilcoxon. Most of these genes (*KLK10*, SLURP1, LINC02571, *AARD*, *SERHL2*, *SYT8*, *TNNI2*, and *ANGPT1*) were also found to be upregulated in the basal-like subtype relative to luminal and Her2 subtypes. Nevertheless, it is intriguing that the *ERBB2* low subgroup displayed overexpression of this markers relative to *ERBB2* negative, since the latter is more exemplary of a true basal-like phenotype. Notably ERBB2 was also upregulated in *ERBB2* low and some other genes located at the 17q12 locus (i.e., Her2 amplicon) were upregulated in the DeSeq2 (*PNMT*) and Wilcoxon

(*GRB7*, *TBC1D3L*) analyses. In contrast, to the handful of upregulated genes, *ERBB2* low specimens, irrespective of *ESR1* status, showed broad downregulation of regulatory non-coding RNAs such as snoRNAs suggesting *ERBB2* low may have altered ribosomal RNA processing and post-transcriptional regulation (**Figure 12b**, top right panels for ER positive and ER negative).

[00235] As expected, the genes that best differentiated *ERRB2* positive samples from *ERBB2* low and negative, irrespective of *ESR1* status, were located along the core of the Her2 amplicon (i.e., *ERBB2*, *PNMT*, *GRB7*, *TCAP*, *PGAP3*, *STARD3*, *MIEN1*) and are frequently co-amplified with *ERRB2*. We explored the Her2 amplicon further and found *ERRB2* low samples showed intermediate expression levels for many genes near the 17q12 locus when compared to either *ERRB2* positive or negative samples, a finding that was consistent across *ESR1* status (**Figure 12c**). To see if expression of the Her2 amplicon could segregate *ERBB2* low from positive and negative samples, we calculated a Her2 amplicon module score and plotted each sample from highest to lowest score, which resulted in good separation of *ERBB2* positive, low, and negative samples (**Figure 12d**). Altogether, these data suggest that *ERBB2* expression, select differentially expressed genes, and Her2 amplicon signatures could be used as an aid to stratify tumors into *ERBB2* positive, low, and negative, ultimately resolving the ambiguity of IHC for low Her2 expression.

[00236] **Disscussion**

[00237] Results from the Destiny breast trials have created a paradigm shift in our thinking of Her2 expression as a simple binary positive or negative classification, but rather as a continuum of Her2 expression. Since nearly 60% of all tumors may express low levels of Her2 and consequently may respond to trastuzumab deruxtecan as well as other anti-Her2 drug conjugates shown efficacy in Her2-low patients (e.g., trastuzumab duocarmazine, disitamab vedotin, and MRG002), this has led to the acknowledgement that current IHC diagnostics are insufficient to quantify Her2 at a broad dynamic range. Thus, novel quantitative assays that can accurately define Her2 expression are critical to identify patients eligible for emerging anti-Her2 drug conjugates. Many researchers have approached this challenge by exploring biological and clinicopathological differences of Her2-low tumors albeit with disparate results. To address this, we surmised that RNA sequencing expression data of *ERBB2* may provide a more suitable ground truth compared to IHC, which is known to suffer from poor concordance when classifying Her2-low (IHC +1/+2 unamplified by DNA ISH) and negative (IHC 0) specimens. Therefore, we used our RNA sequencing data to stratify *ERBB2* expression in respect to ER status and conducted robust differential expression analyses using two approaches to mitigate false positive results. Few upregulated genes were observed in *ERBB2* low, relative to *ERBB2* negative. In contrast, downregulation of genes accounted for

most of the differences between *ERBB2* low and negative, with *ERBB2*-low samples showing reduced expression of a broad panel of snoRNAs and other regulatory RNAs, irrespective of ER status. This is a novel finding with important implications for Her2-low given the broad regulatory roles of snoRNA in processing of ribosomal RNA and post-transcriptional regulation together with evidence of their dysregulation in breast cancer. This suggests that rather than being a distinct biological subtype akin to luminal A/B, Her2-enriched, and basal-like/TNBC, Her2-low specimens are mainly characterized by subtle changes in ribosomal RNA processing and post-transcriptional regulatory programs on top of their main molecular background (e.g., luminal vs basal). We also demonstrate several novel approaches to quantify Her2 in low expression ranges by using RNA expression assays like RNA-FISH and RNA-seq to characterize *ERBB2* levels as well as genes located near the 17q12 Her2 amplicon locus. These approaches could supplement current IHC diagnostics to provide additional information about Her2 expression in gray zone cases.

[00238] Example 3: Interrogation of genes and gene signatures to identify prognostic subgroups that may predict prognosis and therapeutic response

[00239] We also investigated the performance and concordance of several multigene prognostic risk assays since these tests can provide information on the effectiveness of adjuvant chemotherapy in select ER+/Her2– patients. We compared clinical risk, as assessed using criteria in the MINDACT trial, to research-based versions of OncotypeDX, the PAM50 Risk of Recurrence by Sample (ROR-S), GENE70 (i.e., MammaPrint), and the Genomic Grade Index (GGI). This analysis was restricted to 567 patients that were ER+/Her2– by IHC and either node negative (pN0) or node positive (1-3 positive nodes, pN1), encompassing the indicated population of patients eligible for the commercial tests. Relative to clinical risk, all multigene assays classified fewer patients as high risk with OncotypeDX and GGI having the highest and lowest proportion of high-risk patients, respectively. GENE70 and GGI, which use a 2-category (high and low) risk assessment scale, classified more patients as low risk, relative to clinical risk, OncotypeDX, and ROR-S (**Figure 13a**).

[00240] All risk classifiers had comparable prognostic utility with low- and high-risk patients showing favorable and poor progression-free survival (PFS), respectively (**Figure 3b**). Agreement between multigene classifiers and clinical risk was only fair (κ: GGI=0.367, GENE70=0.364, ROR-S=0.343, OncotypeDX=0.288), while agreement among multigene classifiers was moderate to substantial (κ range: 0.431 - 0.724), with GENE70 and GGI showing the highest agreement and GENE70 and OncotypeDX having the lowest (**Figure 13c**). Since OncotypeDX and ROR-S use a 3-category (high, intermediate, low) risk scale and patients classified as intermediate by these two classifiers showed only slight agreement

(κ=0.112) we dichotomized the 3-categories into 2 categories (high vs intermediate/low) improved concordance, especially among multigene classifiers (κ range: 0.653 - 0.796; **Figure 13c**). Only 39.2% of patients (n=222/567) were unanimously classified by all five approaches as either high (n=162) or low (n=60) risk leaving 61.8% of patients (n=345) with a discordant result in at least one classifier. Within these discordant patients, another 27.5% of patients (n=156/567) were categorized as either high (n=52) or low (n=104) by four classifiers and 23.5% of patients (n=133/567) were categorized as either high (n=33) or low (n=100) by three classifiers. Thus, the five prognostic classifiers reached a majority consensus (i.e., at least 3 classifiers in agreement) to stratify 43.6% (n=247/567) and 47.1% (n=267/567) of patients into high and low risk, respectively.

[00241] Notably, patients differed markedly in terms of outcome depending on the number of concordant classifiers for a particular risk category (**Figure 13d**). Patients with at least four classifiers in agreement for either high or low risk showed the poorest and best probability of PFS at 10 years (59.2 % for 4/5 classifiers - 65.1% for 5/5 classifiers in agreement for high-risk vs 89.0% for 4/5 classifiers - 88.2% for 5/5 classifiers in agreement for low-risk; **Figure 13d**). When separating patients based on whether a majority (3 or more classifiers in agreement) or minority (1 or 2 classifiers in agreement) of the five classifiers predicted the same risk category, we observed that the minority group displayed PFS that was intermediate compared to the majority as well as patients classified unanimously as high or low risk (**Figure 13e**). Patientss classified as high risk by the majority of classifiers had lower expression of *PGR* and higher *MKI67* mRNA relative to patients classified as high risk by the minority of classifiers, while there was no difference in *ESR1* expression (**Figure 13f**). Patients unanimously classified as high risk also had higher *ERBB2* expression (e.g., ERBB2 low) relative to patients that were not classified as high risk by any classifier (**Figure 13f**).

[00242] Each classifier independently categorized a proportion of patients into high (n=120), intermediate (n=186), and low (n=65) risk while no other classifier agreed with these unique classifications (**Figure 14a**). Clinical risk had the most unique classifications (n= 118) followed by ROR-S (n=104), OncotypeDX (n=103), GGI (n=25), and GENE70 (n=21). Uniquely classified intermediate and high-risk patients had better progression free survival relative to patients classified as intermediate risk by both OncotypeDX and ROR-S and high risk by all classifiers, respectively (**Figure 14b**). The better survival for the multigene assays was more evident when each classifier was combined because of the low patients per classifier (**Figure 14c**). Patients independently categorized as low risk by GGI or GENE70 had poorer survival and those uniquely classified as low risk by Clinical risk were similar relative to patients classified as low risk by all classifiers (**Figure 14b,c**).

[00243] Given the discordance observed at the single-patient level for risk categorization, we constructed consensus prognostic risk categories by combining the results for all five classifiers into 3-risk categories: high risk (If ≥3 prognostic classifiers agree on intermediate/high risk), low risk (If ≥3 prognostic classifiers agree on low risk), and ultra-low risk (If all 5 prognostic classifiers agree on low risk and the patient is node negative (pN0) and PR+) as outlined in the decision tree in **Figure 14d**. Consensus high risk patients (n=300) showed poor outcomes (PFS and OS) with 49 relapses (distant and local) and 36 deaths within 5-years and another 24 relapses and 17 deaths from 5-10 years. Consensus low risk patients (n=214) had better outcomes with 13 relapses and 10 deaths within 5-years and another 10 relapses and 7 deaths from 5-10 years. Consensus ultra-low risk patients (n=53) showed the best outcomes with 1 relapse and 0 deaths within 5-years and another 2 relapses and 1 death from 5-10 years (**Figure 13g**). When stratifying the consensus outcome groups based on treatment (**Figure 14e**), we observed that high risk patients benefited most from chemoendocrine therapy, relative to endocrine or chemotherapy alone, while low/ultra-low patients did not benefit from chemoendocrine therapy, highlighted the clinical utility of our consensus prognostic risk categories in de-escalating overtreatment in low/ultra-low risk patients.

[00244] To investigate potential markers of prognosis and response to treatment, we first ran univariate Cox analysis on 30,693 transcripts in our RNA-SEQ dataset on all samples as well as samples stratified into IHC surrogate subtypes. This analysis yielded 9,433 genes associated with survival ($p<0.05$) across all samples, 5463 genes for LumA, 4,154 for LumB, 2,055 for Her2, and 2,726 for TNBC samples. We then took the top 100 genes in each comparison, and used this list to identify high probability prognostic markers as well as combinations of prognostic markers that act as surrogate markers for cancer pathways such as **proliferation, luminal pathways, invasion/metastasis, immune function, DNA damage/repair, Her2 pathways, AKT/mTOR, tumor suppression (e.g., PTEN, Retinoblastoma, TP53), hypoxia/glycolysis/metabolism, endocytosis/lysosome**, etc (see Tables 4, 5). For these signatures, we also considered Her2 amplicon genes (see Example 2 herein) and our list of 293 molecular subtyping genes to build gene signatures. As an example of the utility of this list in predicting survival, we constructed a gene signature for proliferation (called Proliferation_MDX) and a hypoxic tumor microenvironment (called Hypoxia_Angiogenesis_Inflammation_MDX). We ran these two signatures and a curated a list of **92 genes and 110 gene signatures** (including the two abovementioned, see Table 6) and assessed their expression in our cohort (n=1013) in relation to BCa IHC surrogate subtype and overall survival.

Table 6: Prognostic genes and gene signatures (890 prognostic genes) associated with survival (e.g., https://www.ensembl.org/)

Gene_id	Gene_name	Gene_id	Gene_name
ENSG00000105141	CASP14	ENSG00000225057	NA
ENSG00000169814	BTD	ENSG00000148459	PDSS1
ENSG00000148926	ADM	ENSG00000286504	NA
ENSG00000171314	PGAM1	ENSG00000287765	NA
ENSG00000228288	PCAT6	ENSG00000165805	C12orf50
ENSG00000187742	SECISBP2	ENSG00000130368	MAS1
ENSG00000156239	N6AMT1	ENSG00000260750	NA
ENSG00000279066	HEXD-IT1	ENSG00000101306	MYLK2
ENSG00000196275	GTF2IRD2	ENSG00000238099	LINC01625
ENSG00000189001	SBSN	ENSG00000259134	LINC00924
ENSG00000136932	TRMO	ENSG00000244128	LINC01322
ENSG00000150433	TMEM218	ENSG00000251003	ZFPM2-AS1
ENSG00000114026	OGG1	ENSG00000273523	NA
ENSG00000102144	PGK1	ENSG00000263089	NA
ENSG00000286042	LCAL1	ENSG00000230749	MEIS1-AS2
ENSG00000224660	SH3BP5-AS1	ENSG00000231329	NA
ENSG00000213139	CRYGS	ENSG00000163673	DCLK3
ENSG00000237280	NA	ENSG00000174502	SLC26A9
ENSG00000105877	DNAH11	ENSG00000155719	OTOA
ENSG00000169032	MAP2K1	ENSG00000172731	LRRC20
ENSG00000277224	H2BC7	ENSG00000285572	NA
ENSG00000249685	NA	ENSG00000081721	DUSP12
ENSG00000095059	DHPS	ENSG00000111339	ART4
ENSG00000203650	LINC01285	ENSG00000288074	NA
ENSG00000167977	KCTD5	ENSG00000138152	BTBD16
ENSG00000127252	PLAAT1	ENSG00000232712	KIZ-AS1
ENSG00000247095	MIR210HG	ENSG00000228709	LINC02575
ENSG00000259196	HMBOX1-IT1	ENSG00000142405	NLRP12
ENSG00000185164	NOMO2	ENSG00000260588	NA
ENSG00000090861	AARS1	ENSG00000261460	NA
ENSG00000272812	NA	ENSG00000254266	PKIA-AS1
ENSG00000167553	TUBA1C	ENSG00000174145	NWD2
ENSG00000157303	SUSD3	ENSG00000133069	TMCC2
ENSG00000176171	BNIP3	ENSG00000125864	BFSP1
ENSG00000153774	CFDP1	ENSG00000030304	MUSK
ENSG00000231298	MANCR	ENSG00000112039	FANCE
ENSG00000273350	NA	ENSG00000028528	SNX1
ENSG00000134333	LDHA	ENSG00000187783	TMEM72
ENSG00000182199	SHMT2	ENSG00000228543	NA

ENSG00000277386	NA	ENSG00000157064	NMNAT2
ENSG00000108474	PIGL	ENSG00000040608	RTN4R
ENSG00000187840	EIF4EBP1	ENSG00000229766	NA
ENSG00000186812	ZNF397	ENSG00000196242	OR2C3
ENSG00000157796	WDR19	ENSG00000175262	C1orf127
ENSG00000011677	GABRA3	ENSG00000197587	DMBX1
ENSG00000259336	NA	ENSG00000224063	NA
ENSG00000109107	ALDOC	ENSG00000283240	NA
ENSG00000112984	KIF20A	ENSG00000248455	LINC02217
ENSG00000188312	CENPP	ENSG00000149488	TMC2
ENSG00000158402	CDC25C	ENSG00000143858	SYT2
ENSG00000061656	SPAG4	ENSG00000284196	NA
ENSG00000178567	EPM2AIP1	ENSG00000258926	NA
ENSG00000175063	UBE2C	ENSG00000187808	SOWAHD
ENSG00000237289	CKMT1B	ENSG00000224086	PPM1F-AS1
ENSG00000129173	E2F8	ENSG00000253658	LINC01592
ENSG00000172893	DHCR7	ENSG00000170703	TTLL6
ENSG00000103257	SLC7A5	ENSG00000232624	LINC01517
ENSG00000266965	NA	ENSG00000256193	LINC00507
ENSG00000274737	NA	ENSG00000259664	LINC02254
ENSG00000205336	ADGRG1	ENSG00000147570	DNAJC5B
ENSG00000139734	DIAPH3	ENSG00000243284	VSIG8
ENSG00000164362	TERT	ENSG00000230623	NA
ENSG00000198431	TXNRD1	ENSG00000226741	LINC02554
ENSG00000197930	ERO1A	ENSG00000254656	RTL1
ENSG00000101057	MYBL2	ENSG00000226913	BSN-DT
ENSG00000215912	TTC34	ENSG00000231204	NA
ENSG00000035499	DEPDC1B	ENSG00000258957	NA
ENSG00000077152	UBE2T	ENSG00000225706	PTPRD-AS1
ENSG00000248323	LUCAT1	ENSG00000237222	LINC01968
ENSG00000148154	UGCG	ENSG00000198711	SSBP3-AS1
ENSG00000181652	ATG9B	ENSG00000183747	ACSM2A
ENSG00000111203	ITFG2	ENSG00000230773	NA
ENSG00000051180	RAD51	ENSG00000177369	FLJ40194
ENSG00000229127	KANSL1L-AS1	ENSG00000238959	RNU7-19P
ENSG00000067225	PKM	ENSG00000241560	ZBTB20-AS1
ENSG00000104321	TRPA1	ENSG00000116560	SFPQ
ENSG00000121152	NCAPH	ENSG00000162747	FCGR3B
ENSG00000111640	GAPDH	ENSG00000147059	SPIN2A
ENSG00000127564	PKMYT1	ENSG00000125414	MYH2
ENSG00000174652	ZNF266	ENSG00000115474	KCNJ13
ENSG00000134057	CCNB1	ENSG00000114854	TNNC1

ENSG00000258441	LINC00641	ENSG00000122482	ZNF644
ENSG00000172270	BSG	ENSG00000015592	STMN4
ENSG00000267766	NA	ENSG00000226125	LINC01907
ENSG00000126267	COX6B1	ENSG00000167914	GSDMA
ENSG00000100075	SLC25A1	ENSG00000260664	NA
ENSG00000261326	LINC01355	ENSG00000148584	A1CF
ENSG00000242182	RN7SL745P	ENSG00000277481	PKD1L3
ENSG00000181830	SLC35C1	ENSG00000232406	NA
ENSG00000040275	SPDL1	ENSG00000143257	NR1I3
ENSG00000163808	KIF15	ENSG00000188693	CYP51A1-AS1
ENSG00000174903	RAB1B	ENSG00000227906	SNAP25-AS1
ENSG00000029993	HMGB3	ENSG00000237126	NA
ENSG00000266602	NA	ENSG00000249628	LINC00942
ENSG00000078674	PCM1	ENSG00000117501	MROH9
ENSG00000267013	LINC01929	ENSG00000223812	NA
ENSG00000111669	TPI1	ENSG00000256422	LINC02552
ENSG00000072571	HMMR	ENSG00000108839	ALOX12
ENSG00000214021	TTLL3	ENSG00000274895	NA
ENSG00000188321	ZNF559	ENSG00000231170	NA
ENSG00000184840	TMED9	ENSG00000180537	RNF182
ENSG00000107159	CA9	ENSG00000254578	NA
ENSG00000199038	MIR210	ENSG00000287543	NA
ENSG00000105497	ZNF175	ENSG00000287544	NA
ENSG00000143252	SDHC	ENSG00000151704	KCNJ1
ENSG00000251023	NA	ENSG00000279467	NA
ENSG00000233191	NA	ENSG00000126070	AGO3
ENSG00000157456	CCNB2	ENSG00000226167	AP4B1-AS1
ENSG00000011638	TMEM159	ENSG00000286311	NA
ENSG00000164087	POC1A	ENSG00000134489	HRH4
ENSG00000171124	FUT3	ENSG00000285847	NA
ENSG00000254996	ANKHD1-EIF4EBP3	ENSG00000178409	BEND3
ENSG00000168701	TMEM208	ENSG00000166035	LIPC
ENSG00000269054	ZNF497-AS1	ENSG00000174371	EXO1
ENSG00000122884	P4HA1	ENSG00000249618	LINC02465
ENSG00000182473	EXOC7	ENSG00000269416	LINC01224
ENSG00000124731	TREM1	ENSG00000283213	NA
ENSG00000065833	ME1	ENSG00000234377	OBI1-AS1
ENSG00000279342	NA	ENSG00000286084	NA
ENSG00000134352	IL6ST	ENSG00000280809	LINC00836
ENSG00000024526	DEPDC1	ENSG00000228063	LYPLAL1-DT
ENSG00000213585	VDAC1	ENSG00000143520	FLG2

ENSG00000087586	AURKA	ENSG00000278090	LUNAR1
ENSG00000123416	TUBA1B	ENSG00000129195	PIMREG
ENSG00000196550	FAM72A	ENSG00000163467	TSACC
ENSG00000088325	TPX2	ENSG00000070019	GUCY2C
ENSG00000135972	MRPS9	ENSG00000133661	SFTPD
ENSG00000109854	HTATIP2	ENSG00000081248	CACNA1S
ENSG00000267274	NA	ENSG00000245864	MEF2C-AS2
ENSG00000163220	S100A9	ENSG00000250986	LINC02600
ENSG00000101447	FAM83D	ENSG00000163449	TMEM169
ENSG00000174547	MRPL11	ENSG00000230805	NA
ENSG00000087086	FTL	ENSG00000143994	ABHD1
ENSG00000165480	SKA3	ENSG00000074800	ENO1
ENSG00000123131	PRDX4	ENSG00000128944	KNSTRN
ENSG00000105699	LSR	ENSG00000116685	KIAA2013
ENSG00000121417	ZNF211	ENSG00000145293	ENOPH1
ENSG00000167578	RAB4B	ENSG00000115866	DARS1
ENSG00000198901	PRC1	ENSG00000158373	H2BC5
ENSG00000101470	TNNC2	ENSG00000278112	NA
ENSG00000142945	KIF2C	ENSG00000134153	EMC7
ENSG00000170027	YWHAG	ENSG00000258667	HIF1A-AS3
ENSG00000273284	NA	ENSG00000100003	SEC14L2
ENSG00000176383	B3GNT4	ENSG00000164611	PTTG1
ENSG00000166123	GPT2	ENSG00000238832	NA
ENSG00000011426	ANLN	ENSG00000150316	CWC15
ENSG00000247828	TMEM161B-AS1	ENSG00000142657	PGD
ENSG00000269044	NA	ENSG00000104147	OIP5
ENSG00000131375	CAPN7	ENSG00000231721	LINC-PINT
ENSG00000267298	NA	ENSG00000080986	NDC80
ENSG00000158710	TAGLN2	ENSG00000127663	KDM4B
ENSG00000093009	CDC45	ENSG00000119203	CPSF3
ENSG00000185917	SETD4	ENSG00000162972	MAIP1
ENSG00000115307	AUP1	ENSG00000124380	SNRNP27
ENSG00000145494	NDUFS6	ENSG00000269578	NA
ENSG00000241404	EGFL8	ENSG00000090889	KIF4A
ENSG00000184270	H2AC21	ENSG00000201564	RN7SKP50
ENSG00000274300	MIR6864	ENSG00000180596	H2BC4
ENSG00000137804	NUSAP1	ENSG00000269343	ZNF587B
ENSG00000169679	BUB1	ENSG00000167528	ZNF641
ENSG00000023608	SNAPC1	ENSG00000006611	USH1C
ENSG00000287643	NA	ENSG00000164109	MAD2L1
ENSG00000119004	CYP20A1	ENSG00000221947	XKR9
ENSG00000147872	PLIN2	ENSG00000224975	INE1

ENSG00000137812	KNL1	ENSG00000269968	NA
ENSG00000196757	ZNF700	ENSG00000121310	ECHDC2
ENSG00000197892	KIF13B	ENSG00000092853	CLSPN
ENSG00000102287	GABRE	ENSG00000274641	H2BC17
ENSG00000117650	NEK2	ENSG00000188610	FAM72B
ENSG00000228742	LINC02577	ENSG00000081181	ARG2
ENSG00000140534	TICRR	ENSG00000182117	NOP10
ENSG00000231407	GORAB-AS1	ENSG00000270562	NA
ENSG00000177380	PPFIA3	ENSG00000201435	RNU4-24P
ENSG00000163728	TTC14	ENSG00000110218	PANX1
ENSG00000168159	RNF187	ENSG00000157110	RBPMS
ENSG00000198393	ZNF26	ENSG00000227486	NA
ENSG00000105143	SLC1A6	ENSG00000184260	H2AC20
ENSG00000221914	PPP2R2A	ENSG00000116030	SUMO1
ENSG00000134690	CDCA8	ENSG00000165733	BMS1
ENSG00000260583	NA	ENSG00000140280	LYSMD2
ENSG00000166851	PLK1	ENSG00000178741	COX5A
ENSG00000103035	PSMD7	ENSG00000197903	H2BC12
ENSG00000177156	TALDO1	ENSG00000158164	TMSB15A
ENSG00000103569	AQP9	ENSG00000253671	NA
ENSG00000112110	MRPL18	ENSG00000124635	H2BC11
ENSG00000109805	NCAPG	ENSG00000138778	CENPE
ENSG00000158406	H4C8	ENSG00000123485	HJURP
ENSG00000199933	RNY1P16	ENSG00000142731	PLK4
ENSG00000252498	RNU6-1016P	ENSG00000137807	KIF23
ENSG00000286037	NA	ENSG00000259172	NA
ENSG00000106268	NUDT1	ENSG00000276727	NA
ENSG00000142751	GPN2	ENSG00000233396	LINC01719
ENSG00000270179	NA	ENSG00000005469	CROT
ENSG00000075218	GTSE1	ENSG00000116670	MAD2L2
ENSG00000245849	RAD51-AS1	ENSG00000223751	NA
ENSG00000258673	LINC01397	ENSG00000175662	TOM1L2
ENSG00000105677	TMEM147	ENSG00000105258	POLR2I
ENSG00000178718	RPP25	ENSG00000168439	STIP1
ENSG00000171848	RRM2	ENSG00000138629	UBL7
ENSG00000156970	BUB1B	ENSG00000276757	RN7SL192P
ENSG00000214770	NA	ENSG00000187514	PTMA
ENSG00000159885	ZNF222	ENSG00000185130	H2BC13
ENSG00000232648	NA	ENSG00000108106	UBE2S
ENSG00000258944	NA	ENSG00000109618	SEPSECS
ENSG00000285278	TFAP2A-AS2	ENSG00000140525	FANCI
ENSG00000231871	IPO9-AS1	ENSG00000167513	CDT1

ENSG00000095303	PTGS1	ENSG00000164032	H2AZ1
ENSG00000159259	CHAF1B	ENSG00000115163	CENPA
ENSG00000241395	RN7SL344P	ENSG00000115053	NCL
ENSG00000102383	ZDHHC15	ENSG00000197299	BLM
ENSG00000186496	ZNF396	ENSG00000097021	ACOT7
ENSG00000272054	NA	ENSG00000143228	NUF2
ENSG00000102387	TAF7L	ENSG00000145354	CISD2
ENSG00000237491	LINC01409	ENSG00000105792	CFAP69
ENSG00000287016	NA	ENSG00000165637	VDAC2
ENSG00000176809	LRRC37A3	ENSG00000187741	FANCA
ENSG00000288533	NA	ENSG00000169230	PRELID1
ENSG00000171557	FGG	ENSG00000242256	RN7SL57P
ENSG00000288095	NA	ENSG00000099246	RAB18
ENSG00000092098	RNF31	ENSG00000249395	CASC9
ENSG00000187838	PLSCR3	ENSG00000140598	EFL1
ENSG00000260672	NA	ENSG00000149554	CHEK1
ENSG00000271335	CCNY-AS1	ENSG00000115484	CCT4
ENSG00000267855	NDUFA7	ENSG00000142396	ERVK3-1
ENSG00000288663	NA	ENSG00000270130	NA
ENSG00000118307	CFAP94	ENSG00000145824	CXCL14
ENSG00000268751	SCGB1B2P	ENSG00000119888	EPCAM
ENSG00000250410	NA	ENSG00000257660	ADCY6-DT
ENSG00000132423	COQ3	ENSG00000176244	ACBD7
ENSG00000278774	U2	ENSG00000130294	KIF1A
ENSG00000256771	ZNF253	ENSG00000147573	TRIM55
ENSG00000264188	NA	ENSG00000169684	CHRNA5
ENSG00000149932	TMEM219	ENSG00000257509	NA
ENSG00000240533	RN7SL69P	ENSG00000166226	CCT2
ENSG00000099194	SCD	ENSG00000288542	NA
ENSG00000228022	HCG20	ENSG00000196787	H2AC11
ENSG00000136492	BRIP1	ENSG00000184368	MAP7D2
ENSG00000280029	NA	ENSG00000206530	CFAP44
ENSG00000267216	ZNF8-ERVK3-1	ENSG00000169564	PCBP1
ENSG00000280303	ERICD	ENSG00000100591	AHSA1
ENSG00000223305	RN7SKP30	ENSG00000004961	HCCS
ENSG00000185798	WDR53	ENSG00000197046	SIGLEC15
ENSG00000256073	URB1-AS1	ENSG00000273979	Metazoa
ENSG00000183474	GTF2H2C	ENSG00000269559	NA
ENSG00000240731	NA	ENSG00000144354	CDCA7
ENSG00000219200	RNASEK	ENSG00000213551	DNAJC9
ENSG00000266618	MIR4742	ENSG00000234129	NA
ENSG00000096654	ZNF184	ENSG00000105854	PON2

ENSG00000239122	RNU2-11P	ENSG00000128271	ADORA2A
ENSG00000258620	NA	ENSG00000265241	RBM8A
ENSG00000272108	NA	ENSG00000065427	KARS1
ENSG00000136866	ZFP37	ENSG00000086061	DNAJA1
ENSG00000239149	SNORA59A	ENSG00000166803	PCLAF
ENSG00000231113	NA	ENSG00000198826	ARHGAP11A
ENSG00000231711	LINC00899	ENSG00000182768	NGRN
ENSG00000261775	NA	ENSG00000131153	GINS2
ENSG00000090857	PDPR	ENSG00000171241	SHCBP1
ENSG00000087087	SRRT	ENSG00000064763	FAR2
ENSG00000183281	PLGLB1	ENSG00000138180	CEP55
ENSG00000197279	ZNF165	ENSG00000169607	CKAP2L
ENSG00000120328	PCDHB12	ENSG00000135119	RNFT2
ENSG00000274023	NA	ENSG00000131876	SNRPA1
ENSG00000022976	ZNF839	ENSG00000174446	SNAPC5
ENSG00000121410	A1BG	ENSG00000135900	MRPL44
ENSG00000252756	RNU6-577P	ENSG00000141741	MIEN1
ENSG00000063854	HAGH	ENSG00000263325	NA
ENSG00000163159	VPS72	ENSG00000145740	SLC30A5
ENSG00000111880	RNGTT	ENSG00000008838	MED24
ENSG00000253263	NA	ENSG00000144712	CAND2
ENSG00000153485	TMEM251	ENSG00000197168	NEK5
ENSG00000054654	SYNE2	ENSG00000108349	CASC3
ENSG00000200999	NA	ENSG00000141736	ERBB2
ENSG00000251611	FAM160A1-DT	ENSG00000137478	FCHSD2
ENSG00000285693	NA	ENSG00000172057	ORMDL3
ENSG00000178796	RIIAD1	ENSG00000090554	FLT3LG
ENSG00000227896	NA	ENSG00000005379	TSPOAP1
ENSG00000150907	FOXO1	ENSG00000041353	RAB27B
ENSG00000166398	GARRE1	ENSG00000115355	CCDC88A
ENSG00000162434	JAK1	ENSG00000161395	PGAP3
ENSG00000135537	AFG1L	ENSG00000167608	TMC4
ENSG00000138668	HNRNPD	ENSG00000182150	ERCC6L2
ENSG00000204644	ZFP57	ENSG00000145495	MARCHF6
ENSG00000263624	NA	ENSG00000126247	CAPNS1
ENSG00000103510	KAT8	ENSG00000139517	LNX2
ENSG00000272024	NA	ENSG00000143578	CREB3L4
ENSG00000157540	DYRK1A	ENSG00000196975	ANXA4
ENSG00000244945	RUFY1-AS1	ENSG00000189362	NEMP2
ENSG00000130348	QRSL1	ENSG00000162836	ACP6
ENSG00000235100	NA	ENSG00000090020	SLC9A1
ENSG00000261211	NA	ENSG00000171475	WIPF2

ENSG00000270083	NA	ENSG00000249100	NA
ENSG00000274979	NA	ENSG00000116209	TMEM59
ENSG00000246090	NA	ENSG00000196505	GDAP2
ENSG00000211452	DIO1	ENSG00000137731	FXYD2
ENSG00000114446	IFT57	ENSG00000141738	GRB7
ENSG00000287642	NA	ENSG00000072310	SREBF1
ENSG00000198932	GPRASP1	ENSG00000084073	ZMPSTE24
ENSG00000177842	ZNF620	ENSG00000234912	SNHG20
ENSG00000257176	NA	ENSG00000149308	NPAT
ENSG00000109685	NSD2	ENSG00000136240	KDELR2
ENSG00000261916	NA	ENSG00000267795	SMIM22
ENSG00000170396	ZNF804A	ENSG00000284602	NA
ENSG00000228084	NA	ENSG00000106537	TSPAN13
ENSG00000187699	C2orf88	ENSG00000288642	CDR1
ENSG00000267474	NA	ENSG00000208018	MIR645
ENSG00000259416	LINC02883	ENSG00000237943	PRKCQ-AS1
ENSG00000267469	NA	ENSG00000155066	PROM2
ENSG00000176894	PXMP2	ENSG00000005001	PRSS22
ENSG00000267251	NA	ENSG00000232504	ST3GAL5-AS1
ENSG00000112309	B3GAT2	ENSG00000123388	HOXC11
ENSG00000241357	NA	ENSG00000138386	NAB1
ENSG00000286565	NA	ENSG00000170903	MSANTD4
ENSG00000164663	USP49	ENSG00000137726	FXYD6
ENSG00000163481	RNF25	ENSG00000086232	EIF2AK1
ENSG00000054392	HHAT	ENSG00000136238	RAC1
ENSG00000126773	PCNX4	ENSG00000013288	MAN2B2
ENSG00000203706	SERTAD4-AS1	ENSG00000117472	TSPAN1
ENSG00000234183	LINC01952	ENSG00000168528	SERINC2
ENSG00000149596	JPH2	ENSG00000071242	RPS6KA2
ENSG00000243005	RN7SL16P	ENSG00000186160	CYP4Z1
ENSG00000258136	NA	ENSG00000111716	LDHB
ENSG00000130021	PUDP	ENSG00000169193	CCDC126
ENSG00000112651	MRPL2	ENSG00000004142	POLDIP2
ENSG00000243279	PRAF2	ENSG00000115884	SDC1
ENSG00000047662	FAM184B	ENSG00000168938	PPIC
ENSG00000198934	MAGEE1	ENSG00000231528	FAM225A
ENSG00000204576	PRR3	ENSG00000176871	WSB2
ENSG00000204850	NA	ENSG00000072134	EPN2
ENSG00000245556	SCAMP1-AS1	ENSG00000278156	TSC22D1-AS1
ENSG00000238966	NA	ENSG00000123091	RNF11
ENSG00000169550	MUC15	ENSG00000125534	PPDPF
ENSG00000286825	NA	ENSG00000179889	PDXDC1

ENSG00000273311	DGCR11	ENSG00000197081	IGF2R
ENSG00000145331	TRMT10A	ENSG00000108423	TUBD1
ENSG00000018280	SLC11A1	ENSG00000162433	AK4
ENSG00000221838	AP4M1	ENSG00000075624	ACTB
ENSG00000225806	NA	ENSG00000196440	ARMCX4
ENSG00000287119	NA	ENSG00000003147	ICA1
ENSG00000258301	VASH1-AS1	ENSG00000123364	HOXC13
ENSG00000185043	CIB1	ENSG00000185920	PTCH1
ENSG00000042813	ZPBP	ENSG00000198125	MB
ENSG00000252933	RNU6-1223P	ENSG00000188064	WNT7B
ENSG00000178996	SNX18	ENSG00000212124	TAS2R19
ENSG00000266533	MIR3619	ENSG00000213977	TAX1BP3
ENSG00000184788	SATL1	ENSG00000201662	RNU6-60P
ENSG00000121988	ZRANB3	ENSG00000091542	ALKBH5
ENSG00000083097	DOP1A	ENSG00000124444	ZNF576
ENSG00000230648	NA	ENSG00000272931	LRRC8D-DT
ENSG00000103507	BCKDK	ENSG00000102547	CAB39L
ENSG00000185433	LINC00158	ENSG00000006432	MAP3K9
ENSG00000198040	ZNF84	ENSG00000042445	RETSAT
ENSG00000237836	PHKA2-AS1	ENSG00000134072	CAMK1
ENSG00000137944	KYAT3	ENSG00000166347	CYB5A
ENSG00000267285	NA	ENSG00000287377	NA
ENSG00000256116	NA	ENSG00000267259	ERVE-1
ENSG00000262049	NA	ENSG00000164535	DAGLB
ENSG00000168209	DDIT4	ENSG00000144668	ITGA9
ENSG00000105404	RABAC1	ENSG00000113719	ERGIC1
ENSG00000120324	PCDHB10	ENSG00000130684	ZNF337
ENSG00000231160	KLF3-AS1	ENSG00000099875	MKNK2
ENSG00000200107	RNU6-1249P	ENSG00000081923	ATP8B1
ENSG00000169612	RAMAC	ENSG00000287070	NA
ENSG00000236535	RC3H1-IT1	ENSG00000157014	TATDN2
ENSG00000264812	NA	ENSG00000265178	MIR4728
ENSG00000225623	AGBL4-IT1	ENSG00000205129	C4orf47
ENSG00000165029	ABCA1	ENSG00000117461	PIK3R3
ENSG00000234127	TRIM26	ENSG00000255062	NA
ENSG00000230873	STMND1	ENSG00000196696	PDXDC2P-NPIPB14P
ENSG00000152315	KCNK13	ENSG00000113732	ATP6V0E1
ENSG00000135341	MAP3K7	ENSG00000273306	NA
ENSG00000255129	NA	ENSG00000198382	UVRAG
ENSG00000182903	ZNF721	ENSG00000284657	NA
ENSG00000273891	NA	ENSG00000233355	CHRM3-AS2
ENSG00000081818	PCDHB4	ENSG00000143537	ADAM15

ENSG00000263400	TMEM220-AS1	ENSG00000213859	KCTD11
ENSG00000125965	GDF5	ENSG00000204619	PPP1R11
ENSG00000275332	NA	ENSG00000131236	CAP1
ENSG00000236289	GACAT3	ENSG00000110906	KCTD10
ENSG00000232474	NCKAP5-IT1	ENSG00000163191	S100A11
ENSG00000234171	RNASEH1-AS1	ENSG00000119403	PHF19
ENSG00000252213	SNORA74D	ENSG00000164244	PRRC1
ENSG00000286954	NA	ENSG00000089723	OTUB2
ENSG00000229431	MED8-AS1	ENSG00000006625	GGCT
ENSG00000030582	GRN	ENSG00000110484	SCGB2A2
ENSG00000287032	NA	ENSG00000187446	CHP1
ENSG00000177692	DNAJC28	ENSG00000111907	TPD52L1
ENSG00000163093	BBS5	ENSG00000206053	JPT2
ENSG00000146243	IRAK1BP1	ENSG00000133138	TBC1D8B
ENSG00000287275	NA	ENSG00000119397	CNTRL
ENSG00000135482	ZC3H10	ENSG00000230724	LINC01001
ENSG00000200550	RNU6-137P	ENSG00000186907	RTN4RL2
ENSG00000214773	NA	ENSG00000167778	SPRYD3
ENSG00000285230	RALY-AS1	ENSG00000085365	SCAMP1
ENSG00000267096	NA	ENSG00000110171	TRIM3
ENSG00000259153	NA	ENSG00000152939	MARVELD2
ENSG00000247373	TMED2-DT	ENSG00000254093	PINX1
ENSG00000230667	SETSIP	ENSG00000198185	ZNF334
ENSG00000254440	PBOV1	ENSG00000276573	NA
ENSG00000241388	HNF1A-AS1	ENSG00000144045	DQX1
ENSG00000163092	XIRP2	ENSG00000124356	STAMBP
ENSG00000133055	MYBPH	ENSG00000155366	RHOC
ENSG00000265142	MIR133A1HG	ENSG00000264895	NA
ENSG00000159409	CELF3	ENSG00000072803	FBXW11
ENSG00000136943	CTSV	ENSG00000106460	TMEM106B
ENSG00000257545	NA	ENSG00000101152	DNAJC5
ENSG00000277089	CCL3-AS1	ENSG00000072682	P4HA2
ENSG00000197893	NRAP	ENSG00000244165	P2RY11
ENSG00000287445	NA	ENSG00000260093	NA
ENSG00000066248	NGEF	ENSG00000181019	NQO1
ENSG00000126262	FFAR2	ENSG00000267390	KDSR-DT
ENSG00000137869	CYP19A1	ENSG00000260349	NA
ENSG00000283662	NA	ENSG00000249042	NA
ENSG00000235481	UBE2R2-AS1	ENSG00000162366	PDZK1IP1
ENSG00000114200	BCHE	ENSG00000117862	TXNDC12
ENSG00000234754	LINC02817	ENSG00000221817	PPP3CB-AS1
ENSG00000226856	THORLNC	ENSG00000255833	TIFAB

ENSG00000173705	SUSD5	ENSG00000183018	SPNS2
ENSG00000232860	SMG7-AS1	ENSG00000198576	ARC
ENSG00000130528	HRC	ENSG00000031081	ARHGAP31
ENSG00000119608	PROX2	ENSG00000261215	NA
ENSG00000237463	LRRC52-AS1	ENSG00000157734	SNX22
ENSG00000132677	RHBG	ENSG00000080298	RFX3
ENSG00000285734	NA	ENSG00000255449	NA
ENSG00000120337	TNFSF18	ENSG00000198858	R3HDM4
ENSG00000125851	PCSK2	ENSG00000235109	ZSCAN31
ENSG00000177354	C10orf71	ENSG00000169057	MECP2
ENSG00000233384	CNIH3-AS2	ENSG00000235079	ZRANB2-AS1
ENSG00000285570	NA	ENSG00000188738	FSIP2
ENSG00000132972	RNF17	ENSG00000231795	ITCH-IT1
ENSG00000168876	ANKRD49	ENSG00000134058	CDK7
ENSG00000286662	NA	ENSG00000140511	HAPLN3
ENSG00000273828	NA	ENSG00000268108	NA
ENSG00000109061	MYH1	ENSG00000113712	CSNK1A1
ENSG00000287466	NA	ENSG00000152404	CWF19L2
ENSG00000148702	HABP2	ENSG00000204334	ERICH2
ENSG00000254718	NA	ENSG00000205890	NA
ENSG00000158882	TOMM40L	ENSG00000197568	HHLA3
ENSG00000159173	TNNI1	ENSG00000177082	WDR73
ENSG00000253394	LINC00534	ENSG00000001617	SEMA3F
ENSG00000262152	LINC00514	ENSG00000006047	YBX2
ENSG00000205832	C16orf96	ENSG00000106105	GARS1
ENSG00000230550	ERLNC1	ENSG00000063601	MTMR1
ENSG00000232077	LINC01031	ENSG00000131791	PRKAB2
ENSG00000156395	SORCS3	ENSG00000173801	JUP
ENSG00000229021	NA	ENSG00000128602	SMO

[00245] A univariate Cox analysis on all samples, irrespective of IHC subtype, revealed the signatures associated with poor survival were the hypoxia/angiogenesis (Metabolic_sig_Trop2_TNBC_Survival, VEGF_Signature, Glycolysis_Signature, including our Hypoxia_Angiogenesis_Inflammatory_MDX signature), prognostic scores (PAM50 ROR-S, GENE70, OncotypeDX, and GGI), proliferation markers (AURKA, CCNB1, CCNE1, FOXM1, MKI67, TK1, PMID_17493263_proliferation, Module11_Prolif, and including our Proliferation_MDX signature), and DNA damage repair (RAD51, TK1), while signatures associated with favorable outcome included luminal (PGR, TAMR13_scores, MOTERA, Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA, and GSEA_HALLMARK_ESTROGEN_RESPONSE_EARLY) and immune (Mast_cells; **Figure**

15) pathways. The prognostic scores and some proliferation markers were also associated with poor survival in luminal A and B subtypes, but not Her2 or TNBC (**Figure 16**). The hypoxia/angiogenesis signatures predicted poor survival in luminal B and TNBC. Different immune signatures were prognostic for different subtypes. While dendritic cell and neutrophil signatures were associated with poor prognosis in luminal A and luminal B, signatures related to dendritic cells, monocytes/macrophages, and T and B cells were associated with a favorable prognosis in TNBC samples.

[00246] Overall, this demonstrates that various genes and gene signatures that encompass cancer hallmark pathways and/or ADC-processing-related pathways (see **Table 6**) can be used to identify subgroups of patients that have poor and favorable survival. These patient groups may be ideal targets for therapeutic intervention using novel targeted therapies (e.g., immunotherapies, ADCs) as well as de-escalation strategies to withhold potentially toxic treatment (e.g., chemotherapy) from individuals who would otherwise have excellent outcomes in the absence of treatment. Moreover, our curated gene lists (500 prognostic genes, **293 subtyping genes, and 92 genes and 110 gene signatures** related to cancer hallmark and ADC-processing pathways) can be a powerful tool to develop new prognostic and predictive signatures for patient stratification into breast cancer molecular subtypes, prognostic risk groups, and treatment-responsive subgroups. And combining multiple gene signatures may provide superior performance in terms of accuracy, precision, and mitigating misclassifications that arise when using a single gene list/classifier as we demonstrate in Example 1 for consensus subtyping and Example 3 for consensus prognostic risk groups.

[00247] Example 4: Laser capture microdissection enables tumor-specific gene expression and accurate multigene profiling

[00248] Breast tumors display considerable intra-tumoral heterogeneity spanning histological, morphological/cellular, genetic, and molecular features, which has important implications for patient diagnosis, treatment, and prognosis. Particularly important for multi-gene assays (e.g., PAM50, OncotypeDX, etc.) is the tissue and cellular composition of the tumor because specimens with low tumor content (or tumor cellularity) can lead to spurious gene expression results. To investigate the effects of tumor content on gene expression, we selected a panel of 44 samples with variable tumor content and compared the transcriptome profiles of LCM with adjacent sections that did not undergo LCM (**Figure 17a**). LCM samples showed enrichment for each marker (*PGR*, *ESR1*, *ERBB2*, and *MKI67*), but only for samples that were classified as IHC-positive for the respective marker (**Figure 17b**). Genes that were classified as IHC-negative showed reduced expression when comparing LCM samples to matched undissected sections, except for *ERBB2*, which was either unaltered or enriched (**Figure 17c**). Interestingly, this observation may be related to the low levels of *ERBB2*/Her2

that may be present in specimens classified as Her2 negative (IHC 0) or Her2-low as defined by an IHC score of +1 or +2 without DNA amplification. LCM, compared to no LCM, also resulted in a broader dynamic range for all markers (**Figure 17d**), presumably because more sequencing reads are distributed to transcripts derived from cancerous tissues rather than normal, healthy epithelial, connective, and adipose tissues. In support of this, LCM enriched Cadherin-1 (*CDH1*) expression, a cell-type marker of breast glandular epithelial cells (i.e., tumor cell marker) and reduced the expression of Vimentin (*VIM*) and Platelet and Endothelial Cell Adhesion Molecule 1 (*PECAM1*), markers for fibroblasts and endothelial cells, respectively (**Figure 17e**).

[00249] We further explored the impact of LCM on molecular subtyping and prognostic risk classifiers as many of these procedures rely on bulk processing or tumor content thresholds to ensure reliable results. Molecular subtyping was particularly susceptible to the presence of non-tumor tissue with 61% (25/41) of AIMS and 32% (13/41) of both mFISHseq and PAM50 samples switching molecular subtype when comparing paired LCM and no LCM samples. This also resulted in 41% (17/41) samples switching their consensus subtype (**Figure 17f**). Expectedly, the most common consensus subtype change was to the normal-like subtype (15/41) and then two samples changed from LumB to LumA and one sample from LumB to Her2 (**Figure 17f**). Bulk processing also influenced the prognostic classifiers to a lesser extent with 41% (17/41) of ROR-S, 22% (9/41) of GGI, 15% (6/41) of GENE70, and 10% (4/41) of OncotypeDX samples switching prognostic risk groups. All samples that switched prognostic risk groups were classified as a lower risk group (e.g., high to intermediate or high to low), which could have profound implications for treatment since low-risk individuals may forego receiving potentially beneficial chemotherapy. For ROR-S, 12 patients (6 high and 6 intermediate risk by LCM) switched to the low prognostic risk group (**Figure 17g**) and these individuals would be incorrectly recommended to not receive chemotherapy. Overall, this highlights the importance of LCM in enriching tumor specific gene expression to provide accurate assessment of the four main breast cancer biomarkers and multigene testing. Methodologies, including commercial tests, that fail to adequately eliminate non-tumor elements may lead to erroneous gene expression, misclassification of patients, and dire consequences for prescribing inappropriate treatment regimens.

[00250] **Discussion**

[00251] Multiplex8+ is a test that utilizes two independent methods, RNA-FISH and RNA-sequencing, to characterize breast tumor biology. Visualizing gene expression of the four main breast cancer biomarkers in a single multiplexed reaction and then using digital pathology tools for robust quantification, allows Multiplex8+ to assess a broad dynamic range of expression and characterize tumor heterogeneity. Users may then identify regions of interest

(ROI) based on biomarker expression and histology and then capture these spatially-defined ROIs for downstream NGS. Total RNA sequencing facilitates transcriptome-wide expression profiling to classify the molecular subtype of breast cancer and quantify predictive and prognostic gene signatures. Here, we report the results of a large-scale retrospective validation on a cohort of 1082 FFPE breast cancer specimens using the Multiplex8+ test.

[00252] A limitation of RNA-based expression assays such as RT-qPCR, microarray, and next generation sequencing is that processing a bulk tissue specimen not only loses the spatial information about biomarker expression but may introduce erroneous gene expression from healthy stroma/epithelia, ductal/lobular carcinoma in situ, and other non-tumor tissue. Indeed, by comparing paired undissected and LCM samples, we demonstrated that precise microdissection of tumor tissue is critical for enriching tumor-specific biomarker expression, providing broad dynamic range expression profiles and ensuring correct classification of molecular subtypes. Several studies have also demonstrated that contamination from non-tumor tissue can lead gene expression profiling tests to misclassify breast cancer samples into molecular subtypes and prognostic risk groups. For example, the “normal-like” subtype predicted by the PAM50 classifier is simply a byproduct of sample contamination from normal tissue and other non-tumor elements (e.g., blood/hemorrhage, necrosis, DCIS/LCIS) or lack of sufficient tumor content/cellularity. A similar issue arose with the characterization of the six molecular subtypes of TNBC described by Lehmann and colleagues. In a follow-up study, the authors used LCM to precisely isolate tumor tissue and demonstrated that the immunomodulatory and mesenchymal stem-like subtypes result from elevated infiltrating lymphocytes and tumor-associated mesenchymal cells, respectively, which led them to refine the TNBC classification system to four subtypes.

[00253] Most, if not all, commercial multigene tests for breast cancer attempt to mitigate this limitation by requiring specimens to either contain a specific tumor content/cellularity or undergo manual macrodissection (i.e., visual dissection with a scalper/razor blade) to enrich for tumor. While manual macrodissection is fast and inexpensive, it lacks single-cell resolution and consequently the precision necessary to eliminate non-tumor elements like DCIS/LCIS, infiltrating lymphocytes, and other intratumoral components, and thus carries the risk of leading to erroneous sample classification. This stresses the importance of incorporating laser capture microdissection into the Multiplex8+ workflow to minimize contamination from non-tumor elements and obtain spatially defined, tumor-enriched cell populations, ultimately facilitating precise and accurate biomarker quantification and molecular subtyping.

[00254] **Example 5: Additional analysis and data interrogation of Example 1 herein:**

[00255] 5.1 Clinical and molecular markers associated with discordant risk classification

[00256] To better understand the drivers of concordance/discordance among the four multigene prognostic signatures and clinical risk, we first compared their similarities. The multigene prognostic signatures had minimal overlap with no genes being present in all classifiers (**Figure 18a**). GGI and ROR-S shared the most genes (n=14), followed by OncotypeDX and ROR-S (n=11), GENE70 and GGI (n=8), OncotypeDX and GGI (n=5), GENE70 and ROR-S (n=3), and OncotypeDX and GENE70 with the least (n=1, *SCUBE2*), which partially explained the concordance between signatures. ROR-S shared the most genes with all three other classifiers (n=28) followed by GGI (n=27), OncotypeDX (n=17), and GENE70 (n=12). Several proliferation-related genes (*BIRC5*, *CCNB1*, *KNTC2*, *MELK*, *MKI67*, *MYBL2*) occured in three of the gene lists, highlighting that proliferation and cell-cycle related genes form a core component of these multigene prognostic signatures.

[00257] We observed several clinical parameters that were associated with discordance between high and low risk, including tumor grade, clinical risk (MINDACT criteria and AJCC stage), chemotherapy, tumor size, and node status (**Figure 18b**). When comparing discordance within a single risk category (i.e., high risk with all classifiers in agreement or only 1 or 2 in disagrement), we also observed changes in the same clinical parameters in patients with one or two discordant classifiers compared with those with unanimous concordance. High risk patients that had 1 or 2 discordant classifiers, compared to those with unanimous concordance for high risk, had greater proportions of T1 tumors, node negative cases, grade 1-2 tumors, low risk scores, and patients treated less frequently with adjuvant chemotherapy (**Figure 18c-h**). Low risk patients that had 1 or 2 discordant classifiers, compared to those with unanimous concordance for low risk, displayed the opposite pattern, with greater proportions of T2-3 tumors, node positive cases, grade 2-3 tumors, high risk scores, and patients treated more frequently with adjuvant chemotherapy (**Figure 18c-h**). Interesingly, both high and low risk patients that had discordance showed enrichment in the proportion of invasive lobular carcinomas (**Figure 18f**), suggesting histology is a factor that drives discordance.

[00258] Next, we interrogated a list of 92 genes and 110 gene signatures that span cancer hallmark pathways and treatment response to ascertain the tumor biology of high and low risk discordance. When comparing samples that were classified as high or low risk by the consensus of all five prognostic classifiers (i.e., those where the majority (≥3) of classifiers agreed), we found 121 genes/gene signatures significantly different between risk groups, when high risk was used as the reference group for discordance. Expectedly, these were largely dominated by proliferation (*FOXM1*, *AURKA*, *MKI67*, *CCNB1*, Module11_Prolif), cell cycle

(RB/PTEN loss, E2F4), DNA damage repair (*RAD51*, *PARP1*, *TYMS*, HALLMARK_DNA_REPAIR), luminal/estrogen pathways (*PGR*, Correlation_to_centroid_PAM50_LumA), and multigene prognostic signature scores. Many signatures involved in immune function (STAT1, Mast cells, Module3_IFN), metabolism/angiogenesis/hypoxia (Metabolic_sig_Trop2_TNBC_Survival), and ADC mechanisms of action were also altered between high and low risk patients, a novel finding considering many multigene prognostic signatures do not capture the biology of these pathways. Regarding ADCs, these spanned antigen targets for ADCs (*F3*, *ROR1/2*, *TPBG*, *FOLH1*, *ERBB2*, *TACSTD2*), endocytosis (*CAV1*, Endocytic_uncoat), and lysosome function (*LAMP2*, *CTSB*). As the number of discordant classifiers increased in reference to high risk, the expression patterns of genes and gene signatures steadily increased (or decreased), highlighting the continuum of gene expression that separates high and low risk patients.

[00259] When comparing discordance within a single risk category, where either all multigene prognostic classifiers were unanimous or had 1 or 2 discordant classifiers (but retained the assigned consensus risk category), the significantly altered genes/gene signatures markedly differed for discordant high and low risk samples (**Figure 18i**). Discordant low risk patients were distinguished by 23 genes/gene signatures composed of proliferation, cell cycle, and estrogen response markers (**Figure 18i**, right heatmap) In contrast, discordant high risk patients differed in 70 genes/gene signatures that spanned the same pathways observed in discordant low risk samples as well as pathways such as DNA damage repair and PARP sensitivity, metabolism/hypoxia/angiogenesis, and immune activation (**Figure 18i**, left heatmap), suggesting high risk tumor biology is more diverse than low risk.

[00260] 5.2 Cellular states associated with risk discordance and immunotherapy response

[00261] The differences in immune markers led us to use the EcoTyper/CIBERSORTx digital cytometry to elucidate the abundances of 11 cell types stratified into 71 cell states across high and low risk patients. High versus low-risk patients showed marked differences in cellular composition with 50 out of 71 cell states reaching statistical significance. This revealed unique patterns of immune states with high-risk patients displaying markers of active adaptive and innate immune response, while low risk patients were enriched in immunoregulatory and resting (normal) immune cells. Interestingly, when investigating cell type and state abundances in the context of discordance within a single risk category, discordant high-risk samples had significantly altered tumor immune microenvironments (TMEs) relative to discordant low risk samples, with the former and latter differing in 39% (28/71) versus 11% (8/71) of cell states, respectively (**Figure 19a,b**). High risk patients, relative to low risk, had active immune TMEs

comprised of exhausted/effector memory CD8/CD4 T cells, mature immunogenic dendritic cells, and M2-like proliferative macrophages (**Figure 19c-f,** top panels) as well as reduced composition of resting or normal immune cells. Non-immune cells like epithelial, endothelial, and fibroblasts were also differentially expresseed in high and low-risk patients, with high-risk patients having enriched pro-inflammatory epithelial cells and pro-migratory-like fibroblasts (**Figure 19f-i,** top panels). Interestingly, high risk patients with 1 or 2 discordant multigene prognostic signatures showed reduced active immune responses (**Figure 19c-f,** bottom panels), increased resting/normal immune cells, and epithelial/endothelial/fibroblast cells (**Figure 19g-h,** bottom panels) that resembled the "immune cold" TMEs of low-risk patients. Altogether, high and low risk patients (and patients with discordance in high and low risk assignment) have unique immune ecosystems, which could be a source of discordance since multigene prognostic signatures have minimal overlap on immune genes.

[00262] Given the promising results of neoadjuvant immunotherapy in hormone receptor positive breast cancer as shown in the prospective KEYNOTE-756 and I-SPY2 trials, we sought to determine whether these cell state signatures that are enriched in high-risk patients could serve as biomarkers for patient stratification. To test this hypothesis, we utilized the publicly available clinical and microarray data from the 40 HR+/HER2− patients in the paclitaxel combined with pembrolizumab (PEMBRO) arm of the multicenter, phase 2 prospective I-SPY2 trial, where 30% (12/40) of patients achieved the primary endpoint of pathological complete response (pCR). Using the online EcoTyper/CIBERSORTx tool, we deconvolved bulk microarray gene expression data into cell type/state signatures and dichotomized them into high and low subgroups based on median expression in the PEMBRO arm. Strikingly, PEMBRO arm patients with high cutoff signatures for Exhausted/effector memory CD8 T cells, mature immunogenic dendritic cells, and classical M1 monocytes/macrophages all experienced a higher proportion of pCRs (55%, n=11/20) relative to those with low signatures (5%, n=1/20), an absolute pCR gain of 25% compared to all PEMBRO arm patients (**Figure 19i**). Combining all three signatures yielded a subgroup with 65% (n=11/17) of patients achieving pCR (**Figure 19i**). These cell type/state signatures were among the top 20 differentially expressed signatures between high and low risk patients and the top 10 signatures between high risk discordant samples (**Figure 19a**), highlighting the notion that high consensus prognostic risk patients may be ideal candidates for neoadjuvant immunotherapy. Indeed, these signatures performed similarly to the MammaPrint High 2 (55%, n=6/11 patients with pCR) and response-predictive subtyping-5 (RPS-5, HER2−/Immune+; 69%, n=11/16 patients with pCR) biomarker groups characterized in I-SPY2 (**Figure 19i**), while covering a slightly larger eligible patient population.

[00263] 5.3 Development of a classifier for sensitivity to T-DM1

[00264] Because the sensitivity and resistance to ADCs likely encompasses broad cellular targets/pathways involved in ADC processing, we posited that combining markers may elucidate ADC-responsive subgroups that can be exploited for effective patient selection. Given that our retrospective cohort did not include ADC-treated patients, we re-analyzed the trastuzumab emtansine (T-DM1) arm and control arms of the multicenter, adaptive-randomized, phase 2 prospective I-SPY2 (Investigation of Serial studies to Predict Your Therapeutic Response with Imaging and Molecular AnaLysis 2, NCT01042379) clinical trial to determine whether combinations of ADC-processing markers would predict response to T-DM1. The T-DM1 arm contains 52 HER2+ patients (35 ER+ and 17 ER− treated with T-DM1 in the neoadjuvant setting with pathological complete response (pCR) as the primary endpoint (see e.g., **Methods** for control arm details). Univariate logistic regression on 71 pre-specified ADC-relevant markers revealed associations with pCR (**Figure 20a**). Elastic net multivariate logistic regression (10-fold cross validation) yielded a 19-feature classifier (**Figure 20b**) that displayed superior predictive utility than *ERBB2* mRNA alone (ROC AUC of 0.99 vs 0.87, **Figure 20c**). Importantly, the T-DM1 predictor only had moderate/low predictive utility in the trastuzumab/pertuzumab and chemotherapy control arms (ROC AUC of 0.78 and 0.62; **Figure 20d**), underlining its specificity to T-DM1 rather than anti-HER2 targeted therapies. The dominant features in the signature predicting T-DM1 efficacy were related to the antigen, endocytosis (*FLOT1/FLOT2* ratio, *RAB5A*), lysosome function (*GLB1*, *HTT*), maytansine payload (*TUBA1B*), and resistance markers (e.g., the multidrug resistance transporter, *ABCC3*; **Figure 20e**). To our knowledge, this constitutes the first evidence of integrating ADC-processing markers into predictive models for effective patient selection, paving the way for future prospective clinical validation.

[00265] **5.4 Real-world implementation of mFISHseq**

[00266] To demonstrate the feasibility of implementing mFISHseq in a real-world setting, we conducted an RUO version of the test on 48 patients, which included assigning consensus subtyping/prognostic risk groups and assessing 40 genes and 28 gene signatures spanning cancer pathways relevant for treatment and prognosis (**Figure 21a**). These patients comprised all clinical settings and molecular subtypes (**Figure 21b**). The most frequently recommended therapies were novel, targeted therapies such as ADCs, PARP inhibitors, and immunotherapies (**Figure 21c** and **Supplementary Methods** for details on how therapies are recommended). **Figures 28-31** show the RUO testing results in several patient vignettes that describe clinicopathological information, relevant mFISHseq results, and information about patient follow up data to determine potential efficacy of the predictive signatures. A detailed report of each patient can be found at https://multiplex8.com/medical-professional (Note that the patient ID numbers do not reveal any identifying information). The RUO version of

mFISHseq (called Multiplex8+) provided unique insights for each patient that could identify potential treatments (including subsequent lines of therapy) and helped to explain why prior treatments performed poorly. Like our retrospective study, there was exceptional concordance between IHC results and mFISHseq with agreement in 91% (n=164/181) of cases for all four biomarkers, with Her2 IHC and *ERBB2* mFISHseq showing 96% agreement. Most discordant results were observed for *ESR1* (n=7) followed by *MKI67* (n=5) and *PGR* (n=3). In two patients (#7 and #12), both ER and PR were in gray zone areas for IHC (i.e., ER/PR low ≤10%) and mFISHseq classified them as negative. Notably, both patients were classified as basal-like by consensus subtyping, mesenchymal (M) TNBC subtype, and had high expression of proliferation, immune, and/or angiogenesis suggesting the mFISHseq results reflected the underlying tumor biology better than IHC.

[00267] To further illustrate a hypothetical framework that could be used in a future prospective validation for identifying patients responsive to ADCs, we interrogated the expression of targets related to ADC antigens, cytotoxic payloads, endocytosis, lysosome function, and resistance (**Figures 21d-h**) in these RUO patients. Expression of ADC markers was compared with either all patient samples from our retrospective cohort or samples restricted to the relevant IHC-surrogate subtype to assign a percentile ranking score and then categorized into tertiles (high, intermediate, low). Based on the expression of these ADC relevant genes and gene signatures, patients could be stratified into putative groups that may be responsive or unresponsive to ADCs (**Figure 21i**), which could then serve as prespecified biomarker groups in future prospective studies.

[00268] 5.5 Molecular drivers of molecular subtyping discordance

[00269] While survival may be a suitable ground truth to indicate a better classification when comparing LumA patients to other more aggressive subtypes (and vice-versa), it is challenging to interpret for other subtype comparisons that have similar outcomes. Therefore, we used another ground truth metric to benchmark reclassified samples that included a panel of genes and gene signatures that are important for molecular subtyping (luminal, proliferation, Her2, and basal-like markers). Samples reclassified from IHC surrogate LumA to LumB showed increased expression of proliferation markers (*MKI67*, *AURKA*, *PCNA*, *TOP2A*, Module11_Prolif) compared to both ground truth LumA (consensus LumA) and LumB samples (consensus LumB without the reclassified LumB samples) but samples reclassified as LumB were statistically more like consensus LumB (**Figure 22a**). Samples reclassified as LumB also had reduced expression of luminal markers (*PGR* and the GSEA Estrogen Response Early signature) but only relative to consensus LumA (**Figure 22b**) and showed an intermediate

ERBB2 phenotype (**Figure 22c**). A similar finding was observed for samples reclassified from LumB to LumA, albeit with the opposite changes in gene expression (**Figure 22d-f**).

[00270] 5.6 Validation of consensus subtyping on METABRIC and TCGA cohorts

[00271] We used both The Cancer Genome Atlas breast cancer (TCGA-BRCA) and Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohorts to externally validate our consensus molecular subtyping and prognostic risk. For benchmarking subtype classifications, we obtained ground truth PAM50 classifications from the flagship METABRIC paper (de Bruijn, I. et al. Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. Cancer Res 83, 3861–3867 (2023)) and Perou and colleagues molecular analysis of TCGA breast cancer histologic types (Perou, C. M. et al. Molecular portraits of human breast tumours. Nature 406, 747–752 (2000)) (see **Supplementary Methods**). Individual classifiers showed more variability in the proportions of luminal subtypes with AIMS classifying the least number of luminal subtypes in all datasets. Normal-like classifications were 1.5-3x higher in METABRIC and TCGA datasets, relative to our retrospective MDX-BRCA cohort that underwent LCM, suggesting the greater presence of non-tumor elements, which influence the proportion of normal-like subtype calls. There were also 82 METABRIC and 24 TCGA samples with indeterminate subtypes (i.e., each molecular classifier assigned a different subtype), considerably higher than the 8 indeterminate samples found in our MDX-BRCA cohort

[00272] While all subtype classifiers showed prognostic utility in METABRIC, only mFISHseq and PAM50 ground truth did in TCGA. Concordance between IHC surrogate and molecular subtypes was only moderate in both datasets; while the concordance between molecular classifiers was moderate/substantial, having slightly lower concordance than our MDX-BRCA cohort, supporting our observation that LCM may improve concordance.

[00273] In METABRIC and TCGA, 57% (1109/1929) and 52% (490/951) of samples showed discordance in ≥1 classifier, respectively, and similar to our retrospective findings, some discordant samples had altered survival, especially in the case of LumB. As the number of discordant samples for a given subtype increased, the correlation with the PAM50 centroid decreased suggesting changes in gene expression underlie the discordance. Inspection of samples that were uniquely classified by a single approach revealed altered survival, especially in LumA patients who had poorer outcomes as well as changes in biomarker expression, relative to those unanimously classified as LumA. For example, uniquely classified LumA samples had higher *MKI67* and lower *PGR* expression relative to samples classified as LumA by all multigene classifiers. Conversely, uniquely classified LumB samples had lower

MKI67 and higher *PGR* expression relative to samples classified as LumB by all multigene classifiers.

[00274] Like our MDX-BRCA cohort, taking the consensus subtype call from the three multigene classifiers mitigated the discordance, ultimately assigning 27% of METABRIC and 32% of TCGA patients to subtypes that better fit their biology and outcome (**Figures 23 and 24**). In the case of METABRIC, these consensus reclassifications had clinically relevant differences in outcome when compared with their original IHC surrogate classifications, with Consensus LumA and Consensus LumB having better and poorer overall survival, respectively (**Figures 23a,b**). While the differences in clinical outcome were less evident in the TCGA cohort, in both TCGA and METABRIC, consensus subtyping had better prognostic utility in both univariate and multivariate Cox models when compared with IHC surrogate (**Figures 23g and 24g**). Our investigation of the correctness of the reclassified samples using a panel of genes and gene signatures that are important for molecular subtyping revealed nearly identical comparisons in TCGA, METABRIC, and MDX-BRCA cohorts. Samples reclassified by consensus subtyping displayed tumor markers that closely aligned with their new subtype rather than the original IHC surrogate or discordant multigene classification.

[00275] 5.7 Validation of consensus risk groups on METABRIC and TCGA cohorts

[00276] Compared to our MDX-BRCA cohort, in the METABRIC and TCGA cohorts, each prognostic classifier assigned risk groups in similar patterns (e.g., GENE70/GGI had the most low risk patients and Clinical risk/OncotypeDX had the most high risk patients, and fewer classifiers showed prognostic utility. Concordance among prognostic classifiers was also dramatically lower than our MDX-BRCA cohort, both between each multigene classifier and clinical risk and among multigene classifiers. Only 22.8% (n=188 high risk, n=11 low risk/872 eligible patients) of METABRIC and 26.2% (n=97 high risk, n=25 low risk/465 eligible patients) of TCGA patients were unanimously assigned to the same risk group by all five classifiers, highlighting the substantial single sample discordance. As the number of discordant classifiers increased for either high or low risk, patients differed markedly in terms of outcome (**Figures 25a,b**). Stratification based on whether a majority (≥3 classifiers) or minority (1-2 classifiers) of the classifiers predicted the same risk category revealed similar differences in outcome, with the largest differences observed in the 5-10 year range (**Figures 25c,d**).

[00277] We further explored single sample discordance by investigating the outcomes of patients classified into a risk group by a single classifier, which no other classifier agreed upon. All three cohorts (MDX, METABRIC, and TCGA) showed similar patterns of uniquely classified samples with clinical risk and OncotypeDX having the most unique high risk samples

and clinical risk and GGI having the most unique low risk samples. These classifiers were less frequently present in the majority consensus risk category than ROR-S and GENE70. Uniquely classified samples also showed clinically relevant differences in PFS when benchmarked against samples classified into a particular risk group by all classifiers: uniquely classified high risk samples had better PFS whereas uniquely classified low risk samples had poorer PFS relative to unanimously classified samples. Similar to our MDX cohort, taking the consensus of the five risk classifiers yielded high and low risk groups that had differences in PFS/RFS past 10 years in both METABRIC and TCGA and in OS past 20 years in METABRIC. The ultra-low risk group, although limited in patients due to low concordance for low risk, displayed excellent PFS/RFS, with only a single relapse across both cohorts **Figures 25e,f**). Overall, these results highlight specific biases for each prognostic classifier that spanned over three independent cohorts, thus underlining potential misclassifications that may arise by using a single risk classification scheme and providing a useful consensus approach to mitigate these limitations.

[00278] 5.8 Single-cell enrichment and specificity of multigene expression classifiers

[00279] Our observation that sample classification is dependent on the amount of tumor content, led us to determine if particular gene expression assays may be more susceptible to non-tumor tissue due to non-specific gene expression. We analyzed two single-cell transcriptomic datasets: a healthy breast tissue dataset (Tarhan, L. et al. Single Cell Portal: an interactive home for single-cell genomics data. 2023.07.13.548886 Preprint at https://doi.org/10.1101/2023.07.13.548886 (2023)) from the Human Protein Atlas (HPA) (Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* **53**, 1334–1347 (2021); Hatzis, C. et al. GENOMIC PREDICTOR OF RESPONSE AND SURVIVAL FOLLOWING TAXANE-ANTHRACYCLINE CHEMOTHERAPY FOR INVASIVE BREAST CANCER. JAMA 305, 1873–1881 (2011)) and a breast cancer dataset from the Broad Institute's Single Cell Portal (Bardia, A. et al. Biomarker analyses in the phase III ASCENT study of sacituzumab govitecan versus chemotherapy in patients with metastatic triple-negative breast cancer. Annals of Oncology 32, 1148–1156 (2021)).

[00280] When comparing molecular subtyping classifiers (**Figure 26a-c**), mFISHseq contained the highest proportion of Cell Enhanced genes (65%) in breast glandular epithelial cells followed by PAM50 (50%) and AIMS (43%). AIMS contained the highest proportion of Cell Enhanced genes in non-breast glandular cells (24%), including adipocytes (6%), mesenchymal cells (7%), and blood and immune cells (8%), whereas mFISHseq contained the least proportion of Cell Enhanced genes in non-breast glandular cells (12%), including

adipocytes (<1%), mesenchymal cells (4%), and blood and immune cells (6%). PAM50 had low proportions of Cell Enhanced genes in non-breast glandular cells (14%), which was dominated by mesenchymal cells (2%), and blood and immune cells (12%). Group Enriched and Group Enhanced genes displayed similar patterns of expression among the three classifiers. PAM50 contained the highest proportion of Low specificity genes (20%) followed by AIMS (14%) and mFISHseq (8%). Notably, seven (out of 17) of the low specificity PAM50 genes comprise the subset of 18 genes used to calculate the proliferation score used in the ROR algorithm, including *CCNE1*, *CDC20*, *CDC6*, *CENPF*, *ORC6*, *TYMS*, and *UBET2*. Many of these genes showed the highest expression in immune cells in healthy breast tissue, highlighting how non-tumor elements may influence subtype and risk classifications. These results may also explain why AIMS, which had the highest proportion of enhanced/enriched genes in non-glandular cell types (e.g., adipose and stromal cells) showed the highest susceptibility to non-tumor tissue when comparing samples that did or did not undergo LCM.

[00281] Compared to molecular subtyping, the prognostic risk classifiers showed lower proportions of gene enrichment/enhancement in breast glandular epithelial cells, greater enrichment/enhancement in immune cells, and more low specificity genes (**Figure 26d-f**). OncotypeDX contained the highest proportion of Cell Enhanced genes (48%) in breast glandular epithelial cells followed by GENE70 (24%) and GGI (16%). GENE70 contained the highest proportion of Cell Enhanced genes in non-breast glandular cells (32%), including blood and immune cells (16%), endothelial cells (6%), and mesenchymal cells (6%), whereas OncotypeDX contained the least proportion of Cell Enhanced genes in non-breast glandular cells (14%) with all being enhanced in blood and immune cells. GGI also contained a high proportion of Cell Enhanced genes in non-breast glandular cells (31%) with most enhanced in blood and immune cells (23%). Strikingly, over half of the genes in GGI were Low specificity (51%), which is likely due to this signature being developed by a differential gene expression analysis comparing low versus high histological grade tumors apparently without respect to tumor content/cellularity. Thus, many non-tumor related genes would be represented in this analysis. GENE70 and OncotypeDX had 44% and 38% of Low specificity genes, respectively, overall revealing considerable susceptibility of non-tumor elements in determining risk scores.

[00282] Out of all classifiers, mFISHseq had the most genes that were Not detected (14%), which could be due to the differences in RNAseq and scRNAseq library preparation and whether skin is present in the breast sample, as some of these genes are expressed in bulk RNAseq datasets from the HPA in either breast or skin tissue (e.g., *AKR1B15*, *A2ML1*, *CASP14*, *GREP1*, *IGHV1-69D*, *IVL*, *KRT83*, *KIAA0319*, *SULT1C3*).

[00283] When comparing cancer to healthy tissue, most genes were elevated in cancer epithelial cells, however, some genes showed higher expression in healthy tissue. *BAG1*, an important regulator of the oncogene BCL2, is a shared gene in OncotypeDX and PAM50, one of the most highly expressed in both multigene panels, and shows the highest expression in normal breast epithelial cells. In PAM50 and AIMS, the basal cytokeratins (*KRT5*, *14*, and *17*) that are commonly associated with the basal-like molecular subtype all showed higher expression and in a higher proportion of normal cells relative to cancer cells. Although there are more examples, the last to highlight is *GRB7*, an important gene that is often co-amplified in HER2+ breast cancers as part of the HER2 amplicon. This gene, which is included in mFISHseq, AIMS, PAM50, and OncotypeDX, had slightly higher expression in normal vs cancer epithelial cells. Taken together, many genes show substantial expression in normal tissue cells (epithelial, immune, etc) and could have profound influences on the final subtype or risk group call depending on tumor content/cellularity and the methodology used to enrich (e.g., LCM vs macrodissection) or not enrich for exclusively tumor cells.

[00284] **Example 6: Cell type specificity of ADC markers and impact of LCM**

[00285] By binding to the antigen target on non-tumor cells, an ADC may have on-target, off-tumor actions that may result in side effects due to damage to healthy tissues. To ascertain the specificity and distribution of ADC markers (**Figure 27a-b**), we utilized scRNA datasets as described earlier. While nearly half of the ADC antigen targets were enriched in breast glandular cells, the remaining showed low specificity (*STING1*, *LY75*, *F3*) or enriched expression in non-tumor cells (*MET* – adipocytes, *CCR2* – T-cells, *MSLN* – macrophages, *CD276*, *ROR1*, and *ROR2* – fibroblasts). Some antigen targets also showed higher expression in healthy relative to cancer epithelial cells, including *F3*, *TACSTD2*, *VTCN1*, *FOLH1*, and *FOLR1* (**Figure 27a**). This highlights how expression in non-tumor cells could influence ADC response, by increased on-target, off-tumor actions leading to potentially toxic side-effects, and reduced on-target, on-tumor actions, leading to suboptimal anti-tumor activity. Additionally, any bulk sampling methodology to identify biomarkers for ADC response could be confounded by gene expression coming from tumor-extrinsic factors. This notion is supported by our comparisons of ADC antigen target expression in paired samples that either did or did not undergo LCM, where we observed enrichment, irrespective of tumor content, in antigen markers that were highly expressed in breast glandular cells (e.g., *ERBB2*, *NECTIN4*, *TACSTD2*, *ERBB3*, *SLC39A6*, *TPBG*; **Figure 27c-d**). In contrast, most antigen targets with low cell type specificity or enrichment in non-glandular cells in the scRNA-seq data were reduced or showed no change in expression following LCM (**Figure 27c-d**).

[00286] Given that camptothecins constitute the most common payload class for ADCs and their molecular target is topoisomerase 1, it is notable that its protein-coding gene *TOP1* showed low specificity in healthy breast tissue with the highest expression in T-cells and, although cancer epithelial cells showed the highest expression, healthy epithelial cells showed comparable expression (**Figure 27a-b**). TOP1 also showed little to no enrichment following LCM, highlighting its low specificity (**Figure 27e**). This may have implications for ADCs with cleavable linkers or membrane permeable payloads that result in bystander effects (i.e., diffusion of the cleaved payload to other cells, both tumor and healthy). Moreover, if the ADC contains a target that is highly expressed in non-tumor cells, then off-tumor actions could be magnified. Other ADC markers generally showed low cell-type specificity and distribution including 59% (13/22) of payload targets, 75% (9/12) of endocytosis genes, and 60% (6/10) of lysosome/enzymatic cleavage genes (**Figure 27a-b**), perhaps unsurprisingly given that these organelles/markers are ubiquitously expressed. These ADC markers for endocytosis, lysosomes, and resistance showed little to know enrichment following LCM (**Figure 27f-h**). For example, several cathepsins (*CTSB*, *CTSS*, and *CTSK*) were enriched in fibroblasts/cancer-associated fibroblasts and macrophages/myeloid cells and *CTSB* showed no enrichment following LCM (**Figure 27a-b,g**). These findings are important since these enzymes are thought to be key for cleaving the protease-cleavable dipeptide valine-citrulline (VC) and tetrapeptide (Gly-Phe-Leu-Gly, GFLG) linkers found in ADCs like trastuzumab deruxtecan (HER2), trastuzumab duocarmazine (HER2), datopotamab deruxtecan (TROP2), patritumab deruxtecan (HER3), tisotumab vedotin (TF), and ozuriftamab vedotin (ROR2) among others.

[00287] Altogether, the scRNA analyses and impact of LCM show that many ADC markers are expressed in multiple cell types, including healthy cells and non-tumor stromal cells, and that LCM may help to enrich genes that are expressed highly in breast glandular cells. This raises the prospect that gene expression signatures may be confounded by methods that either bulk process or macrodissect samples and may be insufficient to stratify patients.

Claims

1. A method for predicting the efficacy of an antibody-drug-conjugate (ADC), said method comprising:

   a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of ADC-associated biomarkers are present in a sample obtained from a subject; preferably said ADC-associated biomarkers are selected from the group consisting of: *Estrogen receptor (ESR1), Progesterone receptor (PGR), Receptor tyrosine-protein kinase erbB-2 (ERBB2) and Proliferation marker protein Ki-67 (MKI67)*;

   b) determining by multiplex sequencing whether or not mRNA species of ADC-associated biomarkers are present in said sample; and

   (c) wherein a part of said sample in which said mRNA species of ADC-associated biomarkers was determined in step (a) is subjected to a laser capturing prior to performing step (b);

   (d) preferably said method is an *ex vivo* or *in vitro* method.

2. The method of any one of the preceding claims, wherein said ADC-associated biomarkers from step (a) comprise: *Estrogen receptor (ESR1)*, *Progesterone receptor (PGR)*, *Receptor tyrosine-protein kinase erbB-2 (ERBB2)* and *Proliferation marker protein Ki-67 (MKI67)*, preferably said ADC-associated biomarkers from step (a) used for visualization of cancer in step (a) (e.g., visualization of cancer cells and/or tissue).

3. The method of any one of the preceding claims, wherein the ADC-associated biomarker/s from step (a) are different from the ADC-associated biomarkers from step (b).

4. The method of any one of the preceding claims, wherein said determining comprises comparing levels (e.g., expression levels) of said mRNA species of said ADC-associated biomarker obtained from said subject to corresponding levels (e.g., expression levels) of said mRNA species obtained from a heathy subject/s and/or cohort/s.

5. The method of any one of the preceding claims, wherein the presence and/or upregulation of ADC-associated biomarker/s in step (a) and step (b) is indicative of the efficacy of said ADC, wherein said efficacy preferably comprises treatment responsiveness to said ADC.

6. The method of any one of the preceding claims, wherein said laser capturing is carried out by the means of a laser capture microdissection (LCM), preferably said method is carried out by the means of multiplexed RNA-FISH-guided laser capture microdissection.

7. The method of any one of the preceding claims, wherein said multiplexed RNA-FISH-guided laser capture microdissection followed by RNA sequencing predicts a presence or absence of a therapeutic response to said ADC.

8. The method of any one of the preceding claims, wherein said method further comprises selecting and/or monitoring a therapy and/or therapeutic regimen based on said presence or absence of said therapeutic response to said ADC and/or said method further comprises administering said ADC to said subject.

9. The method of any one of the preceding claims, wherein said method is a method of cancer subtyping, preferably said method is a method of breast cancer subtyping, further preferably said breast cancer subtyping is selecting a breast cancer subtype from the group consisting of:

a) luminal A-like breast cancer (e.g., ER-positive and/or PR-positive and HER2-negative);
b) luminal B-like breast cancer (e.g., ER-positive and/or PR-positive/PR-negative and HER2-positive);
c) HER2-enriched type (e.g., ER-negative, PR-negative or HER2-positive); and
d) triple negative breast cancer (TNBC), (e.g., ER-, PR- and HER2-negative).

10. The method of any one of the preceding claims, wherein said ADC-associated biomarkers are ADC-associated cancer biomarkers, preferably said ADC-associated cancer biomarkers are ADC-associated breast cancer biomarkers.

11. The method of any one of the preceding claims, wherein said ADC-associated biomarkers (e.g., from step (a) and/or step (b)) are one or more of the following:

a) said ADC-associated biomarkers comprising or consisting of: ADC-associated biomarkers related to or associated with: ADC antigen targets, endocytosis, lysosomal function, payload targets, resistance pathways and/or linker;

b) said ADC-associated biomarkers comprising or consisting of: ESR1, PGR and/or MKI67, optionally, additionally ERBB2;

c) said ADC-associated biomarkers comprising or consisting of: estrogen receptors (ERs); progesterone receptors (PRs), HER2 and/or Ki67;

d) said ADC-associated biomarkers comprising or consisting of: ADC-associated biomarkers related to or associated with one or more of the following: estrogen (e.g., ESR1 and/or PGR); Her2 (e.g., ERBB2 and/or NRG1); CDK4/6 (e.g., CCNB1); Immune function (e.g., Chemokine12 and/or CD274); chemotherapy (e.g., SLC29A1, SLC19A1 and/or MAPs); DNA repair; Topoisomerase inhibitor/s (e.g., UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1 and/or PROM1); ABC transport (e.g., ABCB1, ABCC1, ABCC3 and/or MTUS1); endocytosis (e.g., RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2 and/or FLOT1 to FLOT2 ratio); enzymatic cleavage (e.g., CTSB, GUSB and/or GLB1); TME (e.g., VEGFA); lysosome (e.g., Cathepsins, LAMP1, LAMP2, SLC46A3 and/or SLX4); proliferation (e.g., Module11_Prolif); ADC target (e.g., ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75 and/or ROR1); molecular subtyping (e.g., Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA); microtubule genes (e.g., TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and/or TUBA1B);

e) said ADC-associated biomarkers are selected from the group consisting of: ESR1, PGR, ERBB2, NRG1, CCNB1, Chemokine12, CD274, SLC29A1, SLC19A1, MAPs, UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1, PROM1, ABCB1, ABCC1, ABCC3, MTUS1, RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2, FLOT1 to FLOT2 ratio, CTSB, GUSB, GLB1, VEGFA, Cathepsins, LAMP1, LAMP2, SLC46A3, SLX4, Module11_Prolif, ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75, ROR1, Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA), TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and TUBA1B; and/or

f) said ADC-associated biomarkers are one or more (e.g., all) of the following: ESR1, PGR, ERBB2, NRG1, CCNB1, Chemokine12, CD274, SLC29A1, SLC19A1, MAPs, UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1, PROM1, ABCB1, ABCC1, ABCC3, MTUS1, RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2, FLOT1 to FLOT2 ratio, CTSB, GUSB, GLB1, VEGFA,

Cathepsins, LAMP1, LAMP2, SLC46A3, SLX4, Module11_Prolif, ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75, ROR1, Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA), TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and TUBA1B.

12. The method of any one of the preceding claims, wherein said biomarker is: (i) one or more of the biomarkers listed in Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and/or 17 and/or 18-31; and/or (ii) genes/signatures as listed in Tables 4, 5 and/or 6.

13. The method of any one of preceding claims, wherein said ADC is Sacituzumab-Govitecan, Trastuzumab-Emtansine and/or Trastuzumab-Deruxtecan, optionally said subject in step (a) is suffering from a cancer (e.g., breast cancer).

14. A method for stratification of a subject for treatment (e.g., cancer treatment, e.g., breast cancer treatment) with an ADC, said method comprising:

(a) carrying out the method for predicting the efficacy of an antibody-drug-conjugate (ADC) according to any one of the preceding claims, wherein:
(b) the presence of one or more of said ADC-associated biomarkers is indicative that said ADC may be effective; and/or
(c) wherein the absence of one or more of said ADC-associated biomarkers is indicative that said ADC may not be effective;
(d) preferably said method is an *ex vivo* or *in vitro* method.

15. The method of any one of the preceding claims, wherein the ADC-associated biomarker/s from step (a) are same or different from the ADC-associated biomarkers from step (b); preferably wherein the ADC-associated biomarker/s from step (a) are different from the ADC-associated biomarkers from step (b).

16. A method for improving performance of a multi-gene signature test (e.g. PAM50, AIMS), said method comprising: (a) carrying out the method for predicting the efficacy of an antibody-drug-conjugate (ADC) according to any one of the preceding claims, thus subjecting a part of a sample obtained from a subject to a laser capturing prior to carrying out said test, preferably said multi-gene signature test is a cancer (e.g., breast cancer) multi-gene signature test, preferably said method is an *ex vivo* or *in vitro* method.

17. The method for improving performance of a multi-gene signature test of any one of the preceding claims, further comprising calculating an average overall score of said multi-gene signature test.

18. An Antibody-Drug-Conjugate (ADC) for use in the treatment of a subject suffering from cancer (e.g., breast cancer), wherein said subject is stratified in accordance with the method of any one of the preceding claims.

19. A kit for performing the method of any one of the preceding claims.

20. The kit of any one of the preceding claims, comprising glass slides, preferably positively charged glass slides, frame slides, frame slides with PET membrane, or glass slides with membrane.

21. The kit of any one of the preceding claims, comprising formaldehyde and/or paraffin.

22. The kit of any one of the preceding claims, comprising probes for fluorescence in situ hybridization of the following RNA: Estrogen receptor (ESR1), Progesterone receptor (PGR), Receptor tyrosine-protein kinase erbB-2 (ERBB2) and Proliferation marker protein Ki-67 (MKI67).

23. The kit of any one of the preceding claims, comprising probes for FISH of RNA specific for one or more of the following:

a) ADC-associated biomarkers related to or associated with: ADC antigen targets, endocytosis, lysosomal function, payload targets, resistance pathways and/or linker;

b) ESR1, PGR and/or MKI67, optionally additionally ERBB2*;*

c) estrogen receptors (ERs); progesterone receptors (PRs), HER2 and/or Ki67;

d) ADC-associated biomarkers are related to or associated with one or more of the following: estrogen (e.g., ESR1 and/or PGR); Her2 (e.g., ERBB2 and/or NRG1); CDK4/6 (e.g., CCNB1); Immune function (e.g., Chemokine12 and/or CD274); chemotherapy (e.g., SLC29A1, SLC19A1 and/or MAPs); DNA repair; Topoisomerase inhibitor/s (e.g., UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1 and/or PROM1); ABC transport (e.g., ABCB1, ABCC1, ABCC3 and/or MTUS1); endocytosis (e.g., RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2 and/or FLOT1 to FLOT2 ratio); enzymatic cleavage (e.g., CTSB, GUSB and/or GLB1); TME (e.g., VEGFA); lysosome (e.g., Cathepsins,

LAMP1, LAMP2, SLC46A3 and/or SLX4); proliferation (e.g., Module11_Prolif); ADC target (e.g., ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75 and/or ROR1); molecular subtyping (e.g., Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA); microtubule genes (e.g., TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and/or TUBA1B);

e) said ADC-associated biomarkers are selected from the group consisting of: ESR1, PGR, ERBB2, NRG1, CCNB1, Chemokine12, CD274, SLC29A1, SLC19A1, MAPs, UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1, PROM1, ABCB1, ABCC1, ABCC3, MTUS1, RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2, FLOT1 to FLOT2 ratio, CTSB, GUSB, GLB1, VEGFA, Cathepsins, LAMP1, LAMP2, SLC46A3, SLX4, Module11_Prolif, ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75, ROR1, Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA), TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and TUBA1B; and/or

f) said ADC-associated biomarkers are one or more of the following: ESR1, PGR, ERBB2, NRG1, CCNB1, Chemokine12, CD274, SLC29A1, SLC19A1, MAPs, UGT1A1, CYP3A4, CYP3A5, TOP1, RAD51, ERCC1, PROM1, ABCB1, ABCC1, ABCC3, MTUS1, RAB5B, ATG9A, HTT, ATP2B2, HSP90AB1, STUB1, CUL5, FLOT1, FLOT2, FLOT1 to FLOT2 ratio, CTSB, GUSB, GLB1, VEGFA, Cathepsins, LAMP1, LAMP2, SLC46A3, SLX4, Module11_Prolif, ERBB2, NECTIN4, TACSTD2, ERBB3, FOLR1, MSLN, F3, MET, SLC39A6, TPBG, ROR2, CD276, VTCN1, CEACAM5, CCR2, STING1, FOLH1, KAAG1, LY75, ROR1, Genefu_Luminal_index_Correlation_to_centroid_PAM50_LumA), TUBB, TUBE1, TUBB2A, TUBB3, TUBA1A, TUBB4B, TUBB6, TUBD1, TUBA4A, TUBG1, TUBA1C, TUBG2 and TUBA1B.

24. The kit of any one of the preceding claims comprising one or more probes for: (i) one or more of the biomarkers listed in Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and/or 17 and/or 18-31 herein; and/or (ii) genes/signatures as listed in Tables 4, 5 and/or 6 herein.

25. Use of a kit of any one of the preceding claims, for performing the methods of any one of the preceding claims, preferably wherein said use is an ex vivo or in vitro use.

26. Use of laser capturing for improving one or more of the following: (i) prediction of efficacy of an Antibody-Drug-Conjugate (ADC), (ii) stratification of a subject for treatment (e.g., cancer treatment) with an ADC, or (iii) performance of a multi-gene signature test (e.g., in cancer); wherein said use comprises the method according to any one of preceding claims, preferably wherein said use is an ex vivo or in vitro use.

27. The method according to any one of the preceding claims, wherein said method comprising: carrying out a laser capture on said sample, wherein said laser capture is a laser capture microdissection (LCM), preferably said LCM is a multiplexed RNA-FISH-guided laser capture microdissection.

28. A method for predicting survival and/or assessing status, risk and/or prognosis of cancer (e.g., breast cancer), said method comprising:

a) determining by multiplex fluorescence in situ hybridization whether or not mRNA species of cancer-associated biomarkers (e.g., as disclosed herein, e.g., in Example 3 herein, e.g., Table 6) are present in a sample obtained from a subject; and

b) determining by multiplex sequencing whether or not mRNA species of cancer--associated biomarkers (e.g., as disclosed herein, e.g., in Example 3 herein, e.g., Table 6) are present in said sample;

c) wherein a part of said sample in which said mRNA species of cancer-associated biomarkers was determined in step (a) is subjected to a laser capturing prior to performing step (b);

d) wherein the presence of one or more of said cancer-associated biomarkers is indicative that the survival, risk and/or prognosis of cancer is low; and/or

e) wherein the absence of one or more of said cancer-associated biomarkers may be indicative that the survival, risk and/or prognosis of cancer is high;

f) preferably wherein said method is an ex vivo or in vitro method.