



- (51) **International Patent Classification:**
G01N 33/574 (2006.01) *G06F 19/00* (2011.01)
- (21) **International Application Number:**
PCT/GB2015/053571
- (22) **International Filing Date:**
24 November 2015 (24.11.2015)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
1420859.9 24 November 2014 (24.11.2014) GB
- (71) **Applicant: THE INSTITUTE OF CANCER RE-
SEARCH: ROYAL CANCER HOSPITAL** [GB/GB];
123 Old Brompton Road, London SW7 3RP (GB).
- (72) **Inventor: YUAN, Yinyin;** The Institute of Cancer Re-
search: Royal Cancer Hospital, 123 Old Brompton Road,
London, Greater London SW7 3RP (GB).
- (74) **Agents: GREEN, Katherine et al.;** Mewburn Ellis LLP,
City Tower, 40 Basinghall Street, London, Greater London
EC2V 5DE (GB).
- (81) **Designated States** (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,

[Continued on next page]

(54) **Title:** SCORING OF TUMOR INFILTRATION BY LYMPHOCYTES

(57) **Abstract:** A method of providing a prognosis in a cancer patient comprising analysing a tumour image to calculate a metric of immune infiltration for the tumour, and a method of analysing a tumour image.

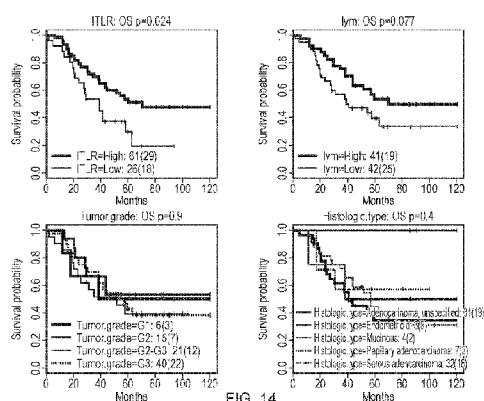


FIG. 14

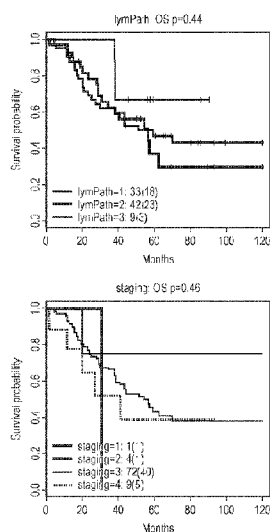


FIG. 14 (continued)



SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

SCORING OF TUMOR INFILTRATION BY LYMPHOCYTES

Field of the Invention

The present invention relates to tumour analysis, and to cancer prognosis. In particular, the present invention relates to methods of analysing tumours for determining a prognosis in cancer.

Background

Cancer is a complex and dynamic disease, and many different ways of analysing and classifying tumours have been developed with the aims of determining the degree of tumour progression or invasiveness and the prognosis for the patient, and informing treatment decisions.

Methods of analysing tumours include the assessment of cell morphology in tumours (typically performed by pathologists), measurement of gene expression in tumours (e.g. by microarray analysis), determination of gene mutation status in tumour cells, and evaluating protein expression within tumours (e.g. by immunohistochemical assessment of tumour sections). These methods of analysing tumours are important not only for predicting clinical outcome, but also for informing decisions on patient therapy.

More recently, it has become apparent that the immunological status of tumours can yield useful prognostic information. Accumulating evidence supports the clinical significance of immune response in many cancer types (Galon et al. 2006, Denkert et al. 2010, Loi et al.). Consistent studies have reported associations between immune activity and disease outcome as well as treatment response (Galon et al. 2006, Denkert et al. 2010, Loi et al. , Liu et al. , Lee et al. , DeNardo et al.).

Furthermore, increasing evidence from clinical trials supports the potential of therapies that target immune activity in certain types of cancer (Robert et al., Stagg et al.). This is perhaps best exemplified in late stage melanoma where recent clinical trials have shown an increased survival advantage in patients receiving the monoclonal antibody ipilimumab, which targets the CTLA4 protein receptor that is expressed on the surface of T cells (Robert et al.). This has led to the development of more standardised methods of characterising tumour immune infiltrate in cancers such as the "immunescore" that aims to quantify the *in situ* immune infiltrate in addition to standardised clinical parameters to aid prognostication and patient selection for immunotherapy in colorectal cancers (Galon et al. 2014).

However, to facilitate the standardisation and reproducibility of scoring immune infiltration, objective approaches are urgently needed (Galon et al. 2014). Furthermore, such approaches need to account for the complexity of immune infiltration into tumours. Abundance, spatial heterogeneity and type of immune cells are the key parameters of

immune infiltration (Galon et al. 2014, Fridman et al.). For example, the spatial locations of immune cells have been shown to be useful in predicting the prognosis of colorectal cancer (Galon et al. 2006). Indeed the pathological "immunescore" is based on the numeration of two lymphocyte populations (CD8+ and CD45RO+ cells), both in the core of the tumour and in the invasive margin that maximises the prognostic power (Galon et al. 2014).

Similarly, large-scale studies of breast cancer have demonstrated that pathological assessment of tumour-infiltration lymphocytes based on Hematoxylin & Eosin (H&E) stained core biopsies is a significant predictor for response to neoadjuvant chemotherapy in 1,058 breast cancer samples (Denkert et al. 2010). Recently, a prospective study demonstrated that in HER2-negative breast cancer stromal lymphocytes can be an independent predictor of response to neoadjuvant chemotherapy (Issa-Nummer et al.). Thus, the spatial organisation of lymphocytic infiltration in the context of nearby cancer cells is an important clinicopathological feature of tumours.

In triple-negative breast cancer (TNBC) an active immune response has been associated with favourable prognosis (Loi et al., Liu et al., Denkert et al.). A large-scale immunohistochemistry study of 3,400 breast cancer samples has showed that TNBC is the only subtype of breast cancer to demonstrate a significant link between CD8-positive immune cells and a good prognosis (Liu et al.). Assessment of lymphocytic infiltration based on whole-tumour H&E sections has been associated with favourable outcome in 256 patients after anthracycline-based chemotherapy (Loi et al.). A recent prospective study showed that the presence of tumour-infiltrating lymphocytes in residual tumours after neoadjuvant chemotherapy is predictive of good prognosis in TNBC (Dieci et al.). Given the current lack of targeted molecular treatment and poor clinical outcome of TNBC, this may suggest new therapeutic opportunities for this aggressive tumour type (Stagg et al.). For instance, accumulating data suggest that anthracyclines mediate their action through activation of CD8+ T-cell responses, hence combination with certain immunotherapies could be especially effective for TNBC (Stagg et al.).

However, despite these advances in understanding of the importance of immune infiltration in cancer, there is a lack of reproducible approaches to objectively assess immune infiltration based on pathological sections.

Summary of the Invention

Lymphocytic infiltration in tumours is often associated with a favourable prognosis and predicts response to chemotherapy in many cancer types. However, it is not well understood because the high levels of spatial and molecular heterogeneity within tumours make it
5 difficult to analyse by traditional pathological assessment.

Identification of cell types by pathologists in the assessment of immune infiltration provides qualitative information on coarse ordinal scales. Such information is poorly suited to analysing large data collections, partly because the high amount human input required
10 renders large scale studies time-consuming and expensive, partly because the subjective nature of the assessment causes an unacceptable degree of variability in the information, and partly because the qualitative data generated do not lend themselves to statistical analysis.

15 The inventor has devised a robust and reproducible method for objectively assessing immune infiltration in tumours. The method is performed on a tumour image in which lymphocytes and cancer cells have been identified.

The method may be performed on images of hematoxylin & eosin (H&E) stained tumour
20 sections. H&E stained sections, and images of H&E stained sections, are often readily available as part of data sets collected for cancer study groups such as the METABRIC group (Curtis, 2012) and the Cancer Genome Atlas (TCGA) group (TCGA, 2012), which makes the methods of the present invention readily adaptable for use in analysing tumours from a variety of cancer types. The method may comprise a step of treating a tumour
25 section with a stain, such as H&E, wherein the presence of subcellular structures such as nuclei creates complexes between the stain and the subcellular structure.

An aspect of the present invention provides a method of measuring immune infiltration in a tumour. In particular, there is provided a method of determining an objective measurement of
30 immune infiltration in a tumour, referred to herein as the ITLR. The ITLR (Intra-Tumour Lymphocyte Ratio) is the ratio of intra-tumour lymphocytes to cancer cells in the tumour expressed as a decimal fraction. For example a ratio of 11 intra-tumour lymphocytes to 1000 cancer cells corresponds to an ITLR of 0.011.

35 Accordingly, an aspect of the invention provides a method of measuring immune infiltration in a tumour, the method comprising:

providing an image of the tumour in which lymphocytes and cancer cells have been identified;

obtaining a lymphocyte-to-cancer measurement for each lymphocyte;

40 classifying a subset of the lymphocytes as intra-tumour lymphocytes according to their lymphocyte-to-cancer ratio;

quantifying the intra-tumour lymphocytes and the cancer cells in the tumour image; calculating the intra-tumour lymphocyte ratio (ITLR) as the ratio of intra-tumour lymphocytes to cancer cells, wherein the ITLR is a measurement of immune infiltration in the tumour.

- 5 A further aspect of the present invention provides a method of determining a cut-off value for ITLR for use in determining a prognosis in cancer, wherein an ITLR below the cut-off value indicates a poor prognosis. The method comprises determining the ITLR for a plurality of tumours, wherein each tumour is from a respective cancer patient in a cohort of cancer patients, and selecting a cut-off value for the ITLR wherein patients with an ITLR lower than
10 the cut-off value have a worse prognosis compared with patients with an ITLR equal to or higher than the cut-off value.

- Accordingly, an aspect of the invention provides a method of determining an ITLR cut-off value for a cancer type or subtype, for use in providing a prognosis in a cancer patient
15 having that cancer type, the method comprising:
measuring immune infiltration in a tumour from each member of a cohort of cancer patients having the cancer type or subtype according to the methods described herein, thereby calculating the ITLR for each tumour;
relating the ITLR for each tumour to the clinical outcome of each cancer patient in the
20 cohort of cancer patients; and
selecting a cut-off value for ITLR, wherein an ITLR equal to or below the cut-off value is associated with a significantly worse clinical outcome in the cohort of cancer patients than an ITLR above the cut-off value

- 25 A further aspect of the present invention provides a method of providing a prognosis in cancer. In particular, there is provided a method of using ITLR as a prognostic biomarker for a cancer patient. The method may comprise measuring the ITLR of a tumour from a cancer patient and using the ITLR to determine a prognosis for the patient. The method may comprise determining the ITLR in a tumour from a cancer patient and using the ITLR to
30 determine a prognosis for the patient, wherein an ITLR below a predetermined cut-off value indicates a poor prognosis.

- Accordingly, an aspect of the invention provides a method of providing a prognosis in a cancer patient, the method comprising:
35 measuring immune infiltration in a tumour from the cancer patient according to the methods described herein, thereby calculating the ITLR for the tumour,
wherein an ITLR below a predetermined ITLR cut-off value indicates a poor prognosis.

- The present invention further provides a method of treating cancer in a patient, the method
40 comprising determining the ITLR in a tumour from the patient or requesting a test providing the results of an analysis to determine the ITLR in a tumour from the patient, and treating the

patient according to a therapeutic regime depending on whether the ITLR is equal to or below, or above, a predetermined-cut-off value.

Summary of the Figures

5 Figure 1. Intra-tumour heterogeneity of cancer cell and lymphocyte distributions.

A. 3D landscapes illustrating the spatial heterogeneity of cancer cells and lymphocytes in an H&E breast whole-tumour section. The height of the hills in the 3D landscape represents the density of cells. **B.** Combined analysis of the spatial distribution of cancer and lymphocytes can lead to quantification of lymphocytic infiltration. Shown are a small H&E image and the
10 corresponding 3D cancer density map, which facilitate the measurement of spatial proximity to cancer for every single lymphocyte in the image.

Figure 2. Quantifying the intra-tumour heterogeneity of lymphocytic infiltration.

A. Schematic depiction of the computational pipeline exemplified with a small region of a
15 breast cancer H&E section: H&E image; classified cells using automated image analysis; a map of cancer density based on image analysis result to quantify cancer-immune spatial relationships. **B.** Discovery of three categories of lymphocytes with unsupervised clustering based on the spatial proximities of lymphocytes to cancer in a subset of TNBC samples. These data were then used to predict the categories of all lymphocytes in all TNBC samples.
20 **C.** Optimal number of cluster K as suggested by BIC over 200 random sampling are 3 in 97% of the time and 5 (3%). BIC curves for the 200 sampling are showed on the left, and boxplot showing means of clusters for $K=3$ solutions in 200 sampling on the right. **D.** Illustration of the distance to the nearest cancer cell d_{min} and the distance to the centroid of convex hull region formed by 10 nearby cancer cells $d_{centroid}$. **E.** Boxplots to show the
25 differences among lymphocyte classes in terms of d_{min} and $d_{centroid}$ (p -values by t-test). **F.** Scatter plot showing d_{min} and $d_{centroid}$ for 1,000 randomly selected lymphocytes, coloured based on the three classes; dashed ellipses showing three clusters fitted to d_{min} and $d_{centroid}$.

30 Figure 3. A representative example illustrating three classes of lymphocytes in cancer density map of a tumour (middle section).

A. Density map of cancer and the spatial distribution of three classes of lymphocytes (spatial points coloured according to the classes). Black contour lines denote cut-off thresholds for the three classes of lymphocytes according to cancer density. **B.** Histogram showing the three types of lymphocytes in this sample. **C.** A higher resolution image of a region in this
35 sample; colour codes follow A.

Figure 4. Association between ITLR and clinical parameters of TNBC.

A. Proportions of three classes of lymphocytes in 181 TNBCs. **B.** Triangle plot to show the lymphocyte composition for each tumour (each black dot represents a tumour; thin lines
40 mark the 50% of corresponding axis). **C.** Boxplot to show correlation between pathological scores and ITLR; p -value from JT-test; n =patient number in each group; whiskers extend to

1.5 interquantile range. **D.** Association between ITLR and tumour size, node status and *TP53* mutations; whiskers extend to 1.5 interquantile range. **E.** Distribution of ITLR in two cohorts with optimal cut-offs marked as dashed red lines. **F.** Kaplan-Meier curves to illustrate the disease-specific survival probabilities of patient groups in two TNBC cohorts stratified by ITLR using the cut-off selected in Cohort 1. Numbers in the legend show the number of patients in each group and numbers in the bracket show the number of disease-specific deaths. **G.** Using Cohort 2 as the discovery cohort and Cohort 1 as the validation cohort yielded similar optimal cut-off.

Figure 5. Comparison of ITLR with other immune signatures.

The optimal cut-off were selected in Cohort 1 and tested in Cohort 2 for **A.** image-based lymphocyte abundance (Lym); **B.** gene expression immune signature by Calabro et al. (18); **C.** Ascierto et al. (19); **D.** *IL8* signature (20); **E.** *CXCL13* expression. **F.** Comparing optimal cut-offs selected in two cohorts. Data were centred at 0 and scaled to have standard deviation 1 and cut-offs were mapped to the centred, scaled data. Signatures close to the diagonal line have similar cut-offs in two cohorts.

Figure 6. ITLR-associated gene modules.

A. Kaplan-Meier curves to illustrate differences in disease-specific survival of patient groups of equal sizes stratified based on the expression of key genes in three modules. **B.** Kaplan-Meier curves to illustrate differences in disease-specific survival of patient groups stratified with *CTLA4* expression by the lower 25, middle 50 and higher 25 percentiles, ITLR, and *CTLA4* and ITLR combined. Survival difference between CTLA low and high stratification within the ITLR high group is given as a p-value.

Figure 7. Kaplan-Meier curves to illustrate the disease-specific survival probabilities of patient groups in in two TNBC cohorts stratified by ATLR (Adjacent) and DTLR (Distal).

The signatures were dichotomised using a cut-off selected over a range of percentiles based on Cohort 1 (the left and middle columns) and tested in Cohort 2 (the right column). Dashed lines in the plots on the left marks the significance threshold of $p=0.05$, and solid vertical lines show the best cut-offs. For the Kaplan-Meier curves, the numbers in the legend show the number of patients in each group and numbers in the bracket show the number of disease-specific deaths.

Figure 8. Kaplan-Meier curves to illustrate the disease-specific survival probabilities of patient groups in in two TNBC cohorts stratified by ATLR (Adjacent) and DTLR (Distal).

The signatures were dichotomised using a cut-off selected over a range of percentiles based on Cohort 2 (the left and right columns) and tested in Cohort 1 (the middle column).

Figure 9. Kaplan-Meier curves to illustrate the disease-specific survival probabilities of patient groups in in two TNBC cohorts stratified by nine immune signatures.

The signatures were dichotomised using a cut-off selected over a range of percentiles based on Cohort 1 (the left and middle columns) and tested in Cohort 2 (the right column). Dashed lines in the plots on the left marks the significance threshold of $p=0.05$, and solid vertical lines show the best cut-offs. For the Kaplan-Meier curves, the numbers in the legend show the number of patients in each group and numbers in the bracket show the number of disease-specific deaths.

Figure 10. Kaplan-Meier curves to illustrate the disease-specific survival probabilities of patient groups in in two TNBC cohorts stratified by nine immune signatures.

The signatures were dichotomised using a cut-off selected over a range of percentiles based on Cohort 2 (the left and right columns) and tested in Cohort 1 (the middle column). Dashed lines in the plots on the left marks the significance threshold of $p=0.05$, and solid vertical lines show the best cut-offs. For the Kaplan-Meier curves, the numbers in the legend show the number of patients in each group and numbers in the bracket show the number of disease-specific deaths.

Figure 11. Scatter plots to show correlation between ITLR and expression of ITLR-associated genes in TNBC.

Figure 12. Compare the prognostic value of top 100 ITLR-associated genes and ITLR by including both in multivariate Cox analysis model, one gene at a time.

Each point denotes analysis for one gene, plotted values are $-\log(\log \text{ rank } p\text{-value})$ for the analysis.

Figure 13. Kaplan-Meier curves illustrating differences in disease-specific survival of TNBC patients stratified with other known parameters including PAM50 (Perou et al., 2000), pathological assessment of lymphocytic infiltration (LI), tumour size, and grade.

Figure 14. Kaplan-Meier curves illustrating differences in 5-year overall survival of ovarian cancer patients of patient groups stratified by ITLR, by "Lym" (Yuan et al., 2012), by lymPath (lymphocyte abundance assessed by pathologist), tumour grade, histologic type, or tumour staging.

Figure 15. A. Histogram showing clustering of cells from TNBC tumours into cluster having relatively high CTLA4 expression and cluster having relatively low CTLA4 expression. B. Kaplan-Meier curve illustrating difference in survival between patients having low CLTA4 expression and high CTLA4 expression.

Detailed Description of the Invention

Certain aspects and embodiments of the invention will now be illustrated by way of example and with reference to the figures described above.

5 The inventor has devised a novel way of statistically modelling the spatial heterogeneity of lymphocytes in tumours, which enables determination of a quantitative measurement of immune infiltration (ITLR). ITLR (intra-tumour lymphocyte ratio) is the ratio of intra-tumour lymphocytes to cancer cells in a tumour. This quantitative measurement of immune infiltration (ITLR) has improved predictive power in cancer prognosis compared with previous
10 indicators of immune infiltration. This measurement of immune infiltration was developed based on a study of tumours from triple negative breast cancer (TNBC) patients from the METABRIC dataset, but is more generally useful in and applicable to other breast cancer sub-types and other cancer types. The generalizable nature of ITLR is demonstrated by the data herein showing that ITLR is also a prognostic indicator in ovarian cancer.

15 Described herein is the first study to statistically identify categories of lymphocytes based on tumour spatial heterogeneity and demonstrate their clinical implications using samples from a large number of patients. This enables a way of modelling spatial heterogeneity in tumours which addresses the need for measuring heterogeneity of lymphocytic infiltration in
20 tumours. The ability to generate reproducible, quantitative scores provides new opportunities for incorporating immune infiltration into staging of cancer (i.e. grading of tumours), as in the use of immunoscore for colorectal cancer (Galon, 2014).

The present invention provides a method of measuring immune infiltration in tumours. In
25 particular, there is provided a method of determining an objective measurement of immune infiltration in a tumour (ITLR), which measurement is the ratio of intra-tumour lymphocytes to cancer cells.

Accordingly, an aspect of the present invention provides a method of measuring immune
30 infiltration in a tumour, the method comprising:

providing an image of the tumour in which lymphocytes and cancer cells have been identified;
obtaining a lymphocyte-to-cancer measurement for each lymphocyte;
classifying a subset of the lymphocytes as intra-tumour lymphocytes according to their
35 lymphocyte-to-cancer ratio;
quantifying the intra-tumour lymphocytes and the cancer cells in the tumour image;
calculating the intra-tumour lymphocyte ratio (ITLR) as the ratio of intra-tumour lymphocytes to cancer cells, wherein the ITLR is a measurement of immune infiltration in the tumour.

40 The methods described herein may be performed using a tumour image in which lymphocytes and cancer cells have been identified. The lymphocytes and cancer cells may have been identified by automated image analysis.

The methods described herein may further comprise a step of identifying cancer cells and lymphocytes in a tumour image by automated image analysis. The methods may comprise steps of generating a tumour image and then identifying lymphocytes and cancer cells in the
5 tumour image by automated image analysis.

The step of identifying cancer cells and lymphocytes in a tumour by automated image analysis may be based on the different nuclear morphologies of cancer cells and lymphocytes. This step may be performed on tumour sections, such as whole-tumour
10 section slides. The tumour section may be H&E stained. The types and/or spatial locations of at least about 10,000 cells may be recorded in this step. The types and/or spatial locations of at least about 20,000, at least about 50,000, at least about 90,000, at least 100,000, at least about 110,000, about 10,000 to 150,000, about 50,000 to 120,000, or about 100,000 to 120,000 cells may be recorded in this step. The types and/or spatial
15 locations of, or about 90,000, about 100,000, or about 110,000 cells may be recorded in this step. The cells may be lymphocytes. This step may use any automated image analysis tool capable of identifying lymphocytes and cancer cells. The automated image analysis tool may be the tool disclosed in Yuan et al, 2012.

20 The image analysis tool disclosed in Yuan et al, 2012, which is hereby incorporated by reference in its entirety, identifies cancer, lymphocytes and stromal cells encompassing fibroblasts and endothelial cells based on their nuclear morphologies in H&E whole-tumour section slides. The main component of this tool is a classifier trained by pathologists over randomly selected tumour regions and validated in 564 breast tumours with 90% accuracy.

25 The image analysis tool described in Yuan et al classifies cells into three categories: cancer, lymphocyte or stromal based on morphological features using a support vector machine.

The image analysis tool described in Yuan et al, 2012 identified cancer cells by their typically large ($>10\ \mu\text{m}$), round nuclei. The stromal class was trained on spindle-shaped stromal cell
30 nuclei (likely to be fibroblasts) and may encompass other stromal cells with similar morphology, such as endothelial cells. The lymphocyte class was trained on immune cells with the distinctive morphology of lymphocytes: small ($<8\ \mu\text{m}$), dark nuclei and not much cytoplasm.

35 The image analysis tool described in Yuan et al was trained using breast tumour images. Automated image analysis tools, such as those described in Yuan et al 2012, can be trained in cancer types other than breast cancer (including those cancer types and subtypes mentioned herein) in order to identify lymphocytes and cancer cells in the tumours of other cancer types.

40 Various automated image analysis tools are known in the art. For example the tools described in Failmezger et al (CRImage) particular Janowczyk et al, and Basavanthally et al,

which are hereby incorporated by reference in their entirety. Any such tool may be suitable for, or adapted for, use in the methods described herein.

As a result of automated image analysis, the types and spatial locations of a large number of cells are recorded in every tumour image. The automated image analysis may enable the mapping of spatial distributions of all, or essentially all, cancer cells and lymphocytes within a tumour image.

Following a step of identifying the cancer cells and lymphocytes using automated image analysis, the spatial relationships of lymphocytes and cancer cells are analysed.

The methods described herein comprise a step of obtaining a lymphocyte-to-cancer measurement for each lymphocyte. This provides a quantitative measurement of each lymphocyte's proximity to cancer cells and spatial location relative to cancer cells.

The step of obtaining a lymphocyte-to-cancer measurement for each lymphocyte may be carried out using the statistical pipeline exemplified in Fig. 1B. First, to globally profile the spatial distribution of the cancer cells, the cancer cell density is quantified, for example using a kernel estimate (Hastie et al, 2001). Alternatively, a mean shift estimate (Cheng, 1995) or scale space (Witkin, 1983) estimate may be used. This builds a 'cancer landscape' where hills indicate tumour regions densely populated with cancer cells. The height of a hill thus correlates with cancer density (tumour density) at a specific location in the tumour (Fig. 1B). Secondly, for every lymphocyte, its spatial proximity to cancer is directly quantified with the cancer density landscape at its specific location to give a "lymphocyte-to-cancer" measurement for each lymphocyte. Thus a quantitative measurement of the spatial proximity to cancer cells is obtained for each lymphocyte (Fig. 1B).

In the studies described herein (see Experimental), cancer cells and lymphocytes were identified, and then their spatial relationships were quantified using a kernel density method. Then, using unsupervised learning, three categories of lymphocytes (intra-tumour, adjacent-tumour and distal tumour) were identified based on their spatial proximities and spatial positioning relative to cancer cells. These lymphocyte categories are consistent with a pathological quantification scheme that considers intratumoral, adjacent stroma and distant stroma compartments (Mahmoud, 2011). Statistically, these clusters are stable, reported as the optimal clustering solution 97% of the time upon repeated sampling.

Accordingly, the methods described herein may comprise a step of obtaining a lymphocyte-to-cancer measurement for each lymphocyte by using a density estimate, such as a kernel estimate, to model the spatial distribution of the cancer cells. The method then comprises a step of determining the proximity of each lymphocyte to cancer by determining the cancer cell density at the location of each lymphocyte, to give a lymphocyte-to-cancer measurement for each lymphocyte. The lymphocytes are then clustered according to their lymphocyte-to-

cancer measurements. An unsupervised learning method, such as Gaussian mixture clustering, may be used to cluster lymphocytes according to their proximity to cancer. The number of clusters may be 2, 3, 4 or more.

5 In the TNBC study described herein (see Experimental), when lymphocytes were clustered according to their lymphocyte-to-cancer measurements the number of clusters was three ($k=3$), corresponding to intra-tumour lymphocytes (ITL), adjacent tumour lymphocytes (ATL) and distal tumour lymphocytes (DTL).

10 In TNBC, lymphocytes having a lymphocyte-to-cancer measurement above the threshold value of 0.10507473 were classified as ITLs, lymphocytes having a lymphocyte-to-cancer measurement below the threshold value of 0.10507473 and above the threshold value of 0.03662728 were classified as ATLs, and lymphocytes having a lymphocyte-to-cancer measurement below the threshold value of 0.03662728 were classified as DTLs. In

15 determining the ITLR, the important distinction is between intra-tumour lymphocytes (ITLs) and non intra-tumour lymphocytes (non-ITLs). Thus in TNBC, lymphocytes having a lymphocyte-to-cancer measurement equal to or above the threshold value of 0.10507473 were classified as ITLs, and the remaining lymphocytes were classified as non-ITLs.

20 In the ovarian cancer study described herein, when lymphocytes were clustered according to their lymphocyte-to-cancer measurements the number of clusters was two ($k=2$), corresponding to intra-tumour lymphocytes and non-intra-tumour lymphocytes.

In ovarian cancer, lymphocytes having a lymphocyte-to-cancer measurement above the
25 threshold value of 0.03114299 were classified as ITLs. Lymphocytes having a lymphocyte-to-cancer measurement below this threshold value were classified as non-ITLs.

As an alternative to using cancer density at a lymphocyte location to give a lymphocyte-to-tumour measurement that is indicative of lymphocyte proximity to cancer (i.e. lymphocyte
30 closeness to cancer), the step of obtaining a lymphocyte-to-cancer measurement for each lymphocyte may be carried out based on a distance measure between a lymphocyte and one or more cancer cells, such as the Euclidean distance. The lymphocytes are then clustered according to their lymphocyte-to-cancer measurements, as described above, for example using an unsupervised learning method, such as Gaussian mixture clustering. In
35 this context, where the lymphocyte-to-cancer measurement is indicative of distance from (rather than proximity to) cancer, a lymphocyte may be classified as an ITL if it has lymphocyte-to-cancer measurement below a threshold value.

The methods described herein may comprise classifying lymphocytes as intra-tumour
40 lymphocytes. That is, the methods may comprise classifying a subset of cells identified as lymphocytes in the tumour image as intra-tumour lymphocytes. Classifying lymphocytes may comprise determining whether the lymphocyte-to-cancer measurement is above a

certain threshold value. The threshold value, for example in TNBC, may be around 0.1, around 0.105 or around 0.10507473. The threshold value, for example in ovarian cancer, may be around 0.03, around 0.0311, or around 0.03114299.

- 5 The methods described herein may comprise determining a threshold value for a lymphocyte-to-cancer measurement, for use in classifying a lymphocyte as an intra-tumour lymphocyte or a non-intra-tumour lymphocyte. For example, where the lymphocyte-to-cancer measurement is indicative of lymphocyte proximity to cancer, the lymphocyte may be classed as an intra-tumour lymphocyte if it has a lymphocyte-to-cancer measurement above the lymphocyte-to-cancer measurement threshold value. Determining a threshold value for a lymphocyte-to-cancer measurement may comprise determining lymphocyte-to-cancer measurements for a population of lymphocytes and clustering the lymphocytes by unsupervised learning, and taking the minimum value of the most cancer proximal cluster (the cluster with the highest measurements) as the threshold value for classifying intra-tumour lymphocytes. A lymphocyte may be classified as an intra-tumour lymphocyte if it has a lymphocyte-to-cancer measurement above (or equal to or above) the threshold value.

- Determining the threshold value may further comprise testing the stability of the clustering by sampling the population of lymphocytes, clustering the sampled population of lymphocytes and determining that the cluster solution ($k=x$ where x is the number of clusters) is stable. The number of clusters is stable where k for the sampled population is the same for 200 repeated samples at least 90%, at least 95% or at least 97% of the time.

- Furthermore, the inventor has shown significant differences between lymphocyte categories both in spatial distance to the nearest cancer cell and spatial positioning of surrounding cancer cells, supporting their biological relevance. For instance, in the presently disclosed study of tumours from TNBC patients from the METABRIC dataset, an intra-tumour lymphocyte is on average $7\mu\text{m}$ away from a cancer cell and $3\mu\text{m}$ from the centroid of convex hull region formed by nearby cancer cells. An adjacent-tumour lymphocyte may be also close to the nearest cancer cells but would be further away from the centroid of convex hull region because it is not surrounded by cancer cells. Thus, the new classification approach disclosed herein is based on spatial measures that account for spatial positioning of cancer cells whilst being computationally efficient enough to analyse whole-tumour sections. Compared to a previously reported measure of lymphocyte abundance as a direct output from image analysis (Yuan, 2012), an advantage of this new approach is that it accounts for the spatial heterogeneity of immune infiltration, which is recognised as an important property of immune infiltration (Galon, 2006) but rarely quantitatively analysed.

- Following the step of classifying lymphocytes as intra-tumour lymphocytes, the ratio of intra-tumour lymphocytes to cancer cells is calculated. This ratio is the ITLR (the intra-tumour lymphocyte ratio), which is an objective and quantitative measurement of immune infiltration in tumours. The ITLR is the ratio of intra-tumour lymphocytes to cancer cells in the tumour

expressed as a decimal fraction. For example, an ITLR of 0.011 represents a 1.1% of intra-tumour lymphocytes to cancer cells i.e. a ratio of 11 intra-tumour lymphocytes to 1000 cancer cells.

- 5 The inventor has shown that ITLR is a robust and powerful prognostic indicator in triple negative breast cancer (TNBC), as discussed below, and also in ovarian cancer. Since immune infiltration is implicated in many cancer types, as discussed in more detail below, including breast cancer, ovarian cancer, colorectal cancer (Galon, 2014), melanoma and non-small cell lung cancer, ITLR may also be used as a prognostic indicator in various
10 cancer types.

- For prognosis in TNBC, the ITLR cut-off of 0.011 was selected based on tumour images from the METABRIC cohort. Patients whose tumours had an ITLR below the cut-off value of 0.011 had a significantly worse clinical outcome in terms of disease-specific survival
15 compared with patients whose tumours had an ITLR above the cut-off value.

- For prognosis in ovarian cancer, the ITLR cut-off of 0.06086 was selected based on tumour images from an unpublished tumour cohort. Patients whose tumours had an ITLR below the cut-off value had a significantly worse clinical outcome in terms of overall survival compared
20 with patients whose tumours had an ITLR above the cut-off value.

- An aspect of the present invention provides a method of determining a cut-off value for ITLR for use in determining a prognosis in cancer, wherein an ITLR below the cut-off value indicates a poor prognosis. The method comprises determining the ITLR for a plurality of
25 tumours, wherein each tumour is from a respective cancer patient in a cohort of cancer patients, and selecting a cut-off value for the ITLR wherein patients with an ITLR equal to or below the cut-off value have a significantly worse prognosis compared with patients with an ITLR above the cut-off value.

- 30 Accordingly, an aspect of the present invention provides a method of determining an ITLR cut-off value for a cancer type or subtype, for use in providing a prognosis in a cancer patient having that cancer type or subtype, the method comprising:

- measuring immune infiltration in a tumour from each member of a cohort of cancer patients having the cancer type or subtype according to the method of claim 1, thereby
35 calculating the ITLR for each tumour;

- relating the ITLR for each tumour to the clinical outcome of each cancer patient in the cohort of cancer patients; and

- selecting a cut-off value for ITLR, wherein an ITLR equal to or below the cut-off value is associated with a significantly different clinical outcome in the cohort of cancer patients
40 than an ITLR above the cut-off value.

An ITLR equal to or below the cut-off value may be associated with a significantly worse clinical outcome than an ITLR above the cut-off value. An ITLR equal to or below the cut-off value may be associated with a significantly better clinical outcome than an ITLR above the cut-off value.

5

The selection of the cut-off value for ITLR serves to dichotomise the continuous range of ITLR values for the tumour images from a patient cohort. The ITLR cut-off value is selected such that there is a significant difference in clinical outcome between patients with an ITLR below the cut-off and patients with an ITLR above the cut-off value. In general, the ITLR cut-off value is selected such that patients having an ITLR below or equal to the cut-off value (i.e. patients having a tumour with an ITLR below or equal to the cut-off value) have a significantly worse prognosis than patients having an ITLR that is above the cut-off value (i.e. patients having a tumour with an ITLR equal to or above the cut-off value).

10

15

In the context of the present invention a significant difference in prognosis refers to a clinical outcome that is significantly different according to the Log rank test. Preferably $p < 0.0500$, $p < 0.0250$, $p < 0.0100$, $p < 0.0090$, $p < 0.0065$, $p < 0.0010$, or $p < 0.0001$ according to the log rank test.

20

Selection of the ITLR cut-off value may comprise identifying an ITLR value wherein about 20% to 80% of the patient cohort has an ITLR below that value. Selection of the ITLR cut-off value may comprise identifying an ITLR value wherein about 20% to 80% of the patient cohort has an ITLR below that value and wherein patients having an ITLR below the cut-off value have a significantly worse prognosis than patients having an ITLR that is above the

25

cut-off value.

The clinical outcome may be disease-specific survival, disease free survival, overall survival, relapse-free survival, progression-free survival, survival rate or survival time. The clinical outcome may be disease-specific survival. Disease-specific survival may be defined with time as maximum 5 years or 10 years from diagnosis and event as death due to cancer (the 5 year disease specific survival and 10 year disease specific survival respectively). Overall survival may be defined with time as maximum 5 years or 10 years from diagnosis and event as death due to any cause. Relapse-free survival may be defined with time as maximum 10 years from diagnosis and event as tumour relapse. A poor prognosis refers to a prediction of a poor clinical outcome, whereas a positive prognosis refers to a prediction of a positive clinical outcome.

30

35

The methods described herein may use a cohort of cancer patients from The Cancer Genome Atlas (TCGA) as the "discovery" cohort. This dataset, with its H&E and matched molecular profiling data will be an extremely useful cohort to validate the utility of ITLR and to select and refine ITLR cut-off values for use in prognostic and/or therapeutic methods. TCGA has chosen cancers for study based on criteria that include poor prognosis and

40

overall public health impact and the availability of human tumour and matched-normal tissue samples that meet TCGA standards for patient consent quality and quantity.

The experiments disclosed herein show the utility of ITLR in TNBC and in ovarian cancer.

- 5 ITLR is a generalizable measure for ITLs and will therefore be useful as a measure of immune infiltration in other cancer types / subtypes, especially given that manual assessment of ITLS has reported value in many cancer types / subtypes.

- 10 As already mentioned above, immune infiltration is implicated in many cancer types, including breast cancer, ovarian cancer, colorectal cancer, melanoma and non-small cell lung cancer, ITLR may also be used as a prognostic indicator in various cancer types.

- Immune infiltration is implicated in many cancers including breast cancer (including breast ductal carcinoma breast and breast lobular carcinoma) (Dieci 2014; Loi S 2013; Kruger JM 2013; Liu S,2012; Ascierto ML 2012, Rody A,2011; Mahmoud SMA, 2011; Denkert C, 2010; Ueno T, 2000) central nervous system cancer (including glioblastoma multiforme and lower grade glioma) (Kmiecik J, 2013; Yang I, 2010; McNamara MG, 2014; Crane CA, 2014; Bambury RM; Alexiou GA, 2013; Vauleon E, 2013) endocrine cancer (including adrenocortical carcinoma, papillary thyroid carcinoma, paraganglioma & pheochromocytoma) (Papewalis C; Huang CT; Mukherji B) gastrointestinal cancer (including Cholangiocarcinoma, Colorectal Adenocarcinoma, Liver Hepatocellular Carcinoma, Pancreatic Ductal Adenocarcinoma, Stomach-Esophageal Cancer)(Kono K, 20116; Wu G; Gao Q; Hiraoka N) gynecologic cancer (including Cervical Cancer (Zhang Y , 2014; Ancuta E, 2009), Ovarian Serous Cystadenocarcinoma (Townsend KN, 2013; Milne K, 2009; Clarke B, 2009), Uterine Carcinosarcoma, Uterine Corpus Endometrial Carcinoma)(Ohno S, 2004) head and neck cancer (including Head and Neck Squamous Cell Carcinoma, Uveal Melanoma) (Spanos WC, 2009; Pretscher D, 2009) hematologic cancer (including Acute Myeloid Leukemia, thymoma, lymphoma) (Yong AS, 2011; Dave SS, 2004;) skin cancer (including Cutaneous Melanoma) (Tjin EP,2014; Erdag G , 2012; Bystryn JC, 1992; Halliday G, 1995) soft tissue cancer (including Sarcoma)(Kim JR, 2016; Sorbye SW 2011; Fiorelli V, 1998) thoracic cancer (including Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, Mesothelioma) (Suzuki K , 2013; Welsh TJ , 2005; Villegas FR , 2002; Hegmans JP, 2006; Dieu-Nosjean MC) urologic cancer (including Chromophobe Renal Cell Carcinoma, Clear Cell Kidney Carcinoma, Papillary Kidney Carcinoma, Prostate Adenocarcinoma, Testicular Germ Cell Cancer, Urothelial Bladder Carcinoma)(Davidsson S , 2013; Thompson RH, 2007; Webster WS, 2006; Gannon PO, 2009; Sjö Dahl G, 2014).

The methods of the present invention may be applied in any of the cancer types or subtypes mentioned above.

ITLR is an objective quantitative indicator of lymphocytic infiltration in tumours. The inventor has shown the importance of using such a quantitative measurement of lymphocytic infiltration in predicting clinical outcome in cancer.

- 5 ITLR is a new spatial and quantitative measure of intra-tumour lymphocytes (ITLs). This measure is a consistent, stable and independent predictor of disease-specific survival across two independent cohorts of 181 TNBC patients in total. This measurement may use a cut-off of 0.011 (1.1% of intra-tumour lymphocytes to cancer cells) that dichotomises the ITLR score. The ~20% of TNBC patients with ITLR scores lower than this cut-off have
- 10 significantly worse disease-specific survival than patients with higher scores, and this association is independent of standard clinical parameters. Taken together, these data support the utility of ITLR as a prognostic biomarker for cancer, including TNBC. Accordingly, disclosed herein is an objective and fully automated scoring system for the standardised assessment of immune infiltration that can be used in the context of clinical
- 15 trials and subsequently aid the treatment decision making process.

Accordingly, an aspect of the present invention may further comprise using ITLR as a prognostic biomarker. The method may comprise measuring the ITLR of a tumour from a cancer patient and using the ITLR to determine a prognosis for the patient. The method

20 may comprise determining the ITLR in a tumour from a cancer patient and using the ITLR to determine a prognosis for the patient, wherein an ITLR below a predetermined cut-off value indicates a poor prognosis. The method may comprise determining the ITLR in a tumour from a cancer patient and using the ITLR to determine a prognosis for the patient, wherein an ITLR above a predetermined cut-off value indicates a poor prognosis.

25 In particular, an aspect of the present invention provides a method of providing a prognosis in a cancer patient, the method comprising:

- measuring immune infiltration in a tumour from the cancer patient according to a method described herein, thereby calculating the ITLR for the tumour,
- 30 wherein an ITLR below a predetermined ITLR cut-off value indicates a poor prognosis.

An aspect of the present invention provides a method of providing a prognosis in a cancer patient, the method comprising:

- providing an image of the tumour in which lymphocytes and cancer cells have been
- 35 identified;
- obtaining a lymphocyte-to-cancer measurement for each lymphocyte;
- classifying a subset of the lymphocytes as intra-tumour lymphocytes according to their lymphocyte-to-cancer ratio;
- quantifying the intra-tumour lymphocytes and the cancer cells in the tumour image;
- 40 calculating the intra-tumour lymphocyte ratio (ITLR) as the ratio of intra-tumour lymphocytes to cancer cells, wherein the ITLR is a measurement of immune infiltration in the cancer patient's tumour

and wherein an ITLR below a predetermined ITLR cut-off value indicates a poor prognosis.

The inventor has shown that ITLR is an independent predictor of clinical outcome in cancer.

5 That is, the ITLR is predictive of clinical outcome without using any other biomarker (such as a gene expression biomarker) or clinical indicator (such as tumour size). For example, the inventor has shown that ITLR is an independent predictor of clinical outcome in triple negative breast cancer (TNBC) and in ovarian cancer. For example in the studies described herein there was no correlation between ITLR and tumour size, node status and *TP53* mutation status (Fig. 4D), and so ITLR is independent of such clinical indicators and biomarkers. Preferably, if the ITLR is below a predetermined cut-off value (or equal to or below a predetermined cut-off value), this indicates a poor prognosis (i.e. a poor clinical outcome). A poor prognosis, or poor clinical outcome, may be poor disease-specific survival.

15 In the present context, an ITLR below a predetermined cut-off value may be referred to as a low ITLR, or ITLR-low. Conversely an ITLR above a predetermined cut-off value may be referred to as a high ITLR, or ITLR-high. A predetermined cut-off value for an ITLR may simply be referred to herein as an ITLR cut-off value.

20 A poor prognosis means that the patient has a worse prognosis than a patient having an ITLR value above the cut-off ITLR value. For example a poor prognosis may mean that the patient is expected to have shorter disease-specific survival time than a patient having an ITLR value above the cut-off ITLR value. A poor prognosis may mean that the patient has a worse prognosis than a patient having an ITLR value above the cut-off ITLR value. The hazard ratio between the patient group having an ITLR below the ITLR cut-off value and the group having an ITLR above the ITLR cut-off value may be from around 0.2 to around 0.4, may be from around 0.25 to around 0.36, may be around 0.25 or may be around 0.36. This means that a patient with ITLR high than the cut-off value is 0.25-0.36 times less likely to die from breast cancer than a patient with ITLR lower than the cut-off value. A poor prognosis may mean that a patient has a survival probability of around 50%, or around 49%, five years from diagnosis, or ten years from diagnosis. A good prognosis may mean that a patient has a survival probability of around 80% five years from diagnosis or ten years from diagnosis.

35 A predetermined ITLR cut-off value may be about 0.011, or about 0.061. The cut-off value may be from about 0.005 to about 0.070, from about 0.010 to about 0.070, from about 0.010 to about 0.012, or from about 0.050 to about 0.070. A predetermined ITLR cut-off value for TNBC may be about 0.011, and for ovarian cancer may be about 0.061.

40 ITLR was tested in two independent cohorts of TNBC and shown to be predictive of disease-specific survival. When TNBC Cohort 1 was used as the discovery cohort an ITLR cut-off value of 0.011 was selected (that is, patients having an ITLR below this value showed

significantly worse disease-specific survival than patients having an ITLR above this value), and in Cohort 2 an ITLR of below 0.011 was associated with significantly worse disease-specific survival than patients with an ITLR above 0.011 (Log-rank test $p=0.0063$, Fig. 4F). Similarly, when TNBC Cohort 2 was used as the discovery cohort an ITLR cut-off of 0.011 was selected, and in Cohort 1 an ITLR of below 0.011 was associated with significantly worse disease-specific survival than patients with an ITLR above 0.011 ($p=0.0037$, Fig. 4F).

The prognostic power of ITLR compares favourably with that of previously published prognostic indicators. ITLR is a more powerful prognostic indicator than the previously published indicator "Lym" (a tumour section image-analysis based indicator of lymphocyte abundance; Yuan et al., 2012) and several published gene signature-based indicators (Calabro et al., Ascierto et al., Rody et al, Ma et al., Gu-Trantien et al).

The same cut-off selection approach used to select the ITLR cut-off was used to test the prognostic power of "Lym" (an image-based measure of lymphocyte abundance in tumour sections) and several gene expression signatures in TNBC and in ovarian cancer. None of these other prognostic indicators consistently correlated with prognosis in both Cohort 1 and Cohort 2. By contrast, ITLR consistently stratified patients into two groups of different clinical outcome. (See Fig. 4, Fig. 5., Fig 14)

Compared to published gene expression signatures, ITLR was also the only signature to show significant correlation with disease-specific survival in multivariate Cox proportional hazards model together with standard clinical parameters of nodal status and tumour size in both cohorts, whichever cohort was used as the discovery cohort (Tables 1 to 3).

Using samples from both TNBC cohorts, ITLR has a log-rank p -value of 2.1×10^{-4} and HR 0.32 (0.17-0.58). To test the robustness of the Cox model in determining the prognostic value of ITLR, bootstrap analysis was used in randomly perturbed data and the univariate and multivariate regression analysis was repeated 1,000 times. In 95.6% and 94.7% of instances, ITLR remained significantly associated with prognosis in univariate and multivariate analysis, respectively. Taken together, these data support the stability and robustness of ITLR as an independent prognostic biomarker in TNBC.

ITLR measures the ratio of intra-tumour lymphocytes to cancer cells, thus is different to the pathological assessment approach described in previous studies (Denkert, 2010; Loi, 2013; Deici, 2014), where the proportion of tumour nests that were infiltrated by lymphocytes were reported. These previous studies agree with the results described herein, because they show that tumour-infiltration lymphocytes are significantly correlated with favorable outcome in TNBC. These previous approaches, like the experiments reported herein, were based on H&E stained pathological samples and therefore support the position that measures of lymphocytic infiltration can be useful tool to aid clinical decisions in TNBC.

Unlike the methods of the invention (which are based on automated image analysis), the previous methods are based on assessment of tumour sections by pathologists (Denkert, 2010; Loi, 2013; Deici, 2014; Salgado, 2014). The previous methods looking at proportions of tumour nests infiltrated by lymphocytes are thus subjective, and therefore subject to bias and variability, and generate results relatively slowly with higher associated costs, the previous methods are thus unsuitable for very large scale analyses.

The approach taken in the present invention, of identifying lymphocyte subtypes by image analysis, contrasts with previous approaches to assessing immune infiltration in tumours.

Previous approaches using image analysis (Yuan, 2012) have only taken account of abundance of lymphocytes in tumours, whereas approaches that attempt to take account of spatial locations of lymphocytes (Denkert, 2010; Loi, 2013; Deici, 2014) have used only manual (pathologist) based processes and have relied on qualitative and subjective assessment of cancer cell constellations and their relationships with lymphocytes (the presence of cancer cell "nest" and the proportion of such nests containing lymphocytes). By contrast the present inventor has taken the approach of using image analysis techniques to identify lymphocyte subtypes within tumours and use the relative abundance of a subtype of lymphocytes (ITLs) to cancer cells as an objective quantitative measure of immune infiltration. The image analysis techniques of the present invention are preferably automated or computer-implemented techniques, thereby facilitating analysis of large numbers (in the order of 100,000 - preferably at least 10,000, at least 50,000, or at least 100,000) of lymphocytes per tumour image and permitting large-scale analyses of cohorts of patients having various types and subtypes of cancer.

Unlike the methods of the invention, which robustly predict clinical outcome, pathological scores of immune infiltration (including pathological assessment of lymphocytic infiltration) were not significantly correlated with prognosis (Figure 13). The pathological scores tested included PAM50 (Perou et al., 2000), pathological assessment of lymphocytic infiltration, tumour size, and grade. Pathological assessment of lymphocytic infiltration for the purposes of this study was scored as absent, mild, or severe: Absent if there were no lymphocytes, mild if there was a light scattering of lymphocytes, and severe if there was a prominent lymphocytic infiltrate.

The prognostic methods of the invention, which are based on an objective indicator of immune cell infiltration obtained by an automated method, have several advantages over previous prognostic methods for use in cancer. As explained above, ITLR has greater predictive power than several previously known cancer biomarkers and prognostic indicators and greater predictive power than pathological scores of immune infiltration. Because ITLR is determined using automated methods it provides an objective measurement of immune cell infiltration in cancer (i.e. not subject to subjective bias or human error, which causes variability in results), it requires no pathologist scoring (and therefore no pathologist training or following of new guidelines) and is relatively low cost and quick to obtain, which makes it

suited to large scale analysis of cancer data. Although detection of gene-expression signature based signatures may be automated, ITLR, because it can conveniently be based on tumour images such as H&E stained sections (copies of which are easily and cheaply shared and stored long-term), is lower cost and more convenient biomarker than gene expression signature-based biomarkers (which require access to preserved biological samples). The image-based ITLR outperforms several gene expression-based signatures using the optimal cut-off selection method. In addition, considering the cost of microarray data acquisition, the ITLR-based approaches described herein open a new avenue for large-scale analysis on readily available pathological samples.

Table 1. Univariate and multivariate Cox regression results for ITLR and other signatures in two TNBC cohorts.

Shaded sections show results from multivariate regression. Uni-: Univariate Cox regression; HR: Hazard Ratio; CI: lower and higher 95% Confidence Interval; Conc: Concordance; 0(0-Inf): where the Cox model failed to converge. *P*-values that pass the significant threshold of 0.05 are shown in bold.

		Cohort 1			Cohort 2		
		HR(CI)	<i>p</i>	Conc	<i>p</i>	HR(CI)	Conc
ITL	Uni-	0.36(0.17-0.77)	0.0063	0.601	0.25(0.09-0.69)	0.0036	0.659
	ITL	0.32(0.15-0.7)	0.0042	0.668	0.15(0.05-0.43)	0.00051	0.76
	Node	0.63(0.29-1.4)	0.26		4.93(1.61-15.08)	0.0052	
	Size	2.62(1.27-5.41)	0.0092		2.07(0.9-4.74)	0.087	
Lym	Uni-	0.47(0.21-1.02)	0.051	0.574	0.41(0.12-1.43)	0.15	0.575
	Lym	0.48(0.22-1.05)	0.066	0.656	0.23(0.05-1.02)	0.053	0.735
	Node	0.69(0.32-1.5)	0.35		4.65(1.46-14.81)	0.0092	
	Size	2.35(1.16-4.77)	0.018		1.66(0.65-4.25)	0.29	
Calabro	Uni-	0.25(0.12-0.52)	5.2x10⁻⁵	0.66	0.5(0.18-1.39)	0.18	0.587
	Calabro	0.27(0.13-0.56)	3.8x10⁻⁴	0.703	0.41(0.14-1.19)	0.1	0.744
	Node	0.75(0.35-1.6)	0.45		4.57(1.45-14.37)	0.0093	
	Size	2.26(1.07-4.76)	0.032		1.91(0.82-4.46)	0.13	
Ascierto	Uni-	0.34(0.15-0.77)	0.0066	0.621	1.23(0.4-3.83)	0.72	0.51

		Cohort 1			Cohort 2		
		HR(CI)	p	Conc	p	HR(CI)	Conc
	Ascierto	0.39(0.17-0.88)	0.024	0.671	1.18(0.37-3.72)	0.78	0.735
	Node	0.85(0.39-1.84)	0.68		3.6(1.21-10.7)	0.021	
	Size	2.06(1.02-4.16)	0.044		2.16(0.86-5.45)	0.1	
IL8	Uni-	3.09(1.46-6.51)	0.0018	0.615	0(0-Inf)	0.0099	0.645
	IL8	2.79(1.32-5.92)	0.0073	0.679	0(0-Inf)	1	0.808
	Node	0.81(0.37-1.75)	0.59		3.14(1.06-9.34)	0.039	
	Size	2.23(1.08-4.63)	0.031		1.75(0.71-4.28)	0.22	
CXCL13	Uni-	0.21(0.1-0.46)	1.5x10⁻⁵	0.69	0.76(0.28-2.1)	0.6	0.545
	CXCL13	0.24(0.11-0.54)	4.5x10⁻⁴	0.721	0.83(0.29-2.37)	0.73	0.739
	Node	0.69(0.32-1.49)	0.35		3.61(1.22-10.71)	0.021	
	Size	1.71(0.83-3.55)	0.15		2.12(0.86-5.22)	0.1	

Table 2. Univariate and multivariate Cox regression results for ITLR and other eight signatures using the optimal cut-offs selected in Cohort 1 and validated in Cohort 2.

5 **Uni-: Univariate Cox regression; Multi-: Multivariate Cox regression; HR: Hazard Ratio; CI: lower and higher 95% Confidence Interval; Conc: Concordance; Inf: Cox model failed to converge.**

	Cohort1			Cohort2		
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-ITL	0.36(0.17-0.77)	0.0063	0.601	0.25(0.09-0.69)	0.0036	0.659
Multi-ITL	0.32(0.15-0.7)	0.0042	0.668	0.15(0.05-0.43)	0.00051	0.76
Multi-node	0.63(0.29-1.4)	0.26		4.93(1.61-15.08)	0.0052	
Multi-size	2.62(1.27-5.41)	0.0092		2.07(0.9-4.74)	0.087	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Lym	0.47(0.21-1.02)	0.051	0.574	0.41(0.12-1.43)	0.15	0.575
Multi-Lym	0.48(0.22-1.05)	0.066	0.656	0.23(0.05-1.02)	0.053	0.735
Multi-node	0.69(0.32-1.5)	0.35		4.65(1.46-14.81)	0.0092	
Multi-size	2.35(1.16-4.77)	0.018		1.66(0.65-4.25)	0.29	

	Cohort1			Cohort2		
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Calabro	0.25(0.12-0.52)	5.20E-05	0.66	0.5(0.18-1.39)	0.18	0.587
Multi-Calabro	0.27(0.13-0.56)	0.00038	0.703	0.41(0.14-1.19)	0.1	0.744
Multi-node	0.75(0.35-1.6)	0.45		4.57(1.45-14.37)	0.0093	
Multi-size	2.26(1.07-4.76)	0.032		1.91(0.82-4.46)	0.13	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-IL8	3.09(1.46-6.51)	0.0018	0.615	0(0-Inf)	0.0099	0.645
Multi-IL8	2.79(1.32-5.92)	0.0073	0.679	0(0-Inf)	1	0.808
Multi-node	0.81(0.37-1.75)	0.59		3.14(1.06-9.34)	0.039	
Multi-size	2.23(1.08-4.63)	0.031		1.75(0.71-4.28)	0.22	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Bcell	0.6(0.25-1.48)	0.26	0.557	0.51(0.12-2.27)	0.37	0.539
Multi-Bcell	0.57(0.23-1.4)	0.22	0.655	0.48(0.11-2.17)	0.34	0.747
Multi-node	0.7(0.32-1.5)	0.35		3.77(1.27-11.2)	0.017	
Multi-size	2.38(1.19-4.76)	0.014		2.07(0.86-5)	0.11	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Bcell.IL8	0.52(0.24-1.11)	0.086	0.581	1.1(0.41-2.95)	0.86	0.482
Multi-Bcell.IL8	0.53(0.25-1.12)	0.097	0.648	1.22(0.42-3.53)	0.71	0.743
Multi-node	0.74(0.34-1.6)	0.44		3.76(1.24-11.41)	0.02	
Multi-size	2.36(1.17-4.76)	0.016		2.21(0.88-5.54)	0.091	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Ascierto	0.34(0.15-0.77)	0.0066	0.621	1.23(0.4-3.83)	0.72	0.51
Multi-Ascierto	0.39(0.17-0.88)	0.024	0.671	1.18(0.37-3.72)	0.78	0.735
Multi-node	0.85(0.39-1.84)	0.68		3.6(1.21-10.7)	0.021	
Multi-size	2.06(1.02-4.16)	0.044		2.16(0.86-5.45)	0.1	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-CXCR3	0.3(0.14-0.64)	9.00E-04	0.618	0.82(0.3-2.25)	0.7	0.535
Multi-CXCR3	0.31(0.15-0.66)	0.0026	0.683	0.79(0.25-2.45)	0.68	0.73
Multi-node	0.86(0.39-1.87)	0.7		3.81(1.24-11.72)	0.02	
Multi-size	2.24(1.13-4.44)	0.02		2.07(0.82-5.18)	0.12	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-CXCL13	0.21(0.1-0.46)	1.50E-05	0.69	0.76(0.28-2.1)	0.6	0.545
Multi-CXCL13	0.24(0.11-0.54)	0.00045	0.721	0.83(0.29-2.37)	0.73	0.739
Multi-node	0.69(0.32-1.49)	0.35		3.61(1.22-10.71)	0.021	
Multi-size	1.71(0.83-3.55)	0.15		2.12(0.86-5.22)	0.1	

Table 3. Univariate and multivariate Cox regression results for ITLR and other eight signatures using the optimal cut-offs selected in Cohort 2 and validated in Cohort 1.

Uni-: Univariate Cox regression; Multi-: Multivariate Cox regression; HR: Hazard

Ratio; CI: lower and higher 95% Confidence Interval; Conc: Concordance.

	Cohort1			Cohort2		
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-ITL	0.45(0.21-0.96)	0.033	0.587	0.26(0.1-0.71)	0.0048	0.656
Multi-ITL	0.38(0.17-0.84)	0.016	0.654	0.16(0.05-0.48)	0.001	0.76
Multi-node	0.62(0.28-1.37)	0.23		4.64(1.52-14.15)	0.007	
Multi-size	2.62(1.27-5.39)	0.0088		2.07(0.89-4.83)	0.091	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Lym	0.91(0.43-1.91)	0.8	0.524	0.35(0.13-0.98)	0.038	0.63
Multi-Lym	0.92(0.43-1.95)	0.82	0.627	0.29(0.1-0.85)	0.024	0.778
Multi-node	0.72(0.33-1.58)	0.41		3.82(1.29-11.38)	0.016	
Multi-size	2.33(1.18-4.64)	0.015		1.85(0.73-4.73)	0.2	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Calabro	0.53(0.24-1.2)	0.12	0.578	0(0-Inf)	0.04	0.608
Multi-Calabro	0.56(0.24-1.28)	0.17	0.667	0(0-Inf)	1	0.799
Multi-node	0.67(0.31-1.46)	0.31		4.12(1.39-12.23)	0.011	
Multi-size	2.17(1.11-4.24)	0.023		1.91(0.8-4.55)	0.14	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-IL8	1.76(0.86-3.6)	0.12	0.575	0.18(0.05-0.63)	0.0026	0.692
Multi-IL8	1.74(0.84-3.59)	0.14	0.65	0.21(0.06-0.77)	0.018	0.795
Multi-node	0.8(0.37-1.73)	0.57		3.76(1.25-11.34)	0.019	
Multi-size	2.29(1.17-4.51)	0.016		1.87(0.68-5.17)	0.23	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Bcell	0.74(0.35-1.59)	0.44	0.541	0.41(0.09-1.82)	0.23	0.557
Multi-Bcell	0.75(0.35-1.6)	0.45	0.63	0.33(0.07-1.5)	0.15	0.763
Multi-node	0.72(0.33-1.55)	0.39		4.24(1.41-12.76)	0.01	
Multi-size	2.33(1.17-4.61)	0.016		1.94(0.81-4.67)	0.14	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Bcell.IL8	0.82(0.35-1.92)	0.65	0.511	0.45(0.16-1.31)	0.13	0.599
Multi-Bcell.IL8	0.75(0.32-1.76)	0.51	0.629	0.59(0.2-1.76)	0.34	0.753
Multi-node	0.73(0.34-1.59)	0.43		3.28(1.08-9.94)	0.036	
Multi-size	2.4(1.2-4.83)	0.014		2.18(0.85-5.6)	0.11	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-Ascierto	0.83(0.39-1.78)	0.64	0.52	2.26(0.82-6.23)	0.11	0.63

	Cohort1			Cohort2		
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Multi-Ascierto	1(0.46-2.16)	0.99	0.619	2.6(0.87-7.82)	0.089	0.773
Multi-node	0.73(0.34-1.59)	0.43		3.22(1.08-9.61)	0.036	
Multi-size	2.33(1.16-4.66)	0.017		2.46(0.91-6.62)	0.075	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-CXCR3	0.56(0.24-1.29)	0.17	0.569	0(0-Inf)	0.029	0.618
Multi-CXCR3	0.62(0.26-1.48)	0.28	0.658	0(0-Inf)	1	0.805
Multi-node	0.68(0.32-1.49)	0.34		4.17(1.41-12.38)	0.01	
Multi-size	2.16(1.09-4.28)	0.028		1.89(0.8-4.46)	0.15	
	HR(CI)	p-value	conc	HR(CI)	p-value	conc
Uni-CXCL13	0.35(0.17-0.75)	0.0043	0.605	2.21(0.71-6.86)	0.16	0.595
Multi-CXCL13	0.38(0.18-0.79)	0.01	0.663	3.38(0.92-12.45)	0.067	0.773
Multi-node	0.69(0.32-1.49)	0.35		3.71(1.24-11.1)	0.019	
Multi-size	2.29(1.11-4.72)	0.026		2.71(0.94-7.8)	0.064	

ITLR as an unbiased assessment of immune infiltration can facilitate the discovery of molecular correlates with this clinically important phenomenon. While the expression of many immune-related genes in tumours was significantly associated with ITLR, it is unclear whether these genes are expressed on cancer cells or lymphocytes. This is because the microarray data were obtained using whole-tumour materials without micro-dissection.

The data herein show that the RNA expression of cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*), a receptor of the immunoglobulin family and the target of ipilimumab, was significantly associated with ITLR as well as longer disease specific survival in TNBC. This is consistent with the recent observation in non-small cell lung cancers that over-expression of *CTLA4* is associated with reduced death rate (Salvi, 2012). *CTLA4* is expressed in tumour cells in different cancer types (Contardi, 2005). In breast cancer it is expressed in both tumour cells and T cells, and an inverse correlation between *CTLA4* expression and clinical outcome (i.e. high *CTLA4* expression associated with poor clinical outcome) has been previously reported in 60 patients with different breast cancer subtypes (Mao, 2010), which is in contrast with the data herein from TNBC (see below), and which thus highlights the novel molecular insights into cancer yielded by ITLR. A recent study showed that *in situ* mRNA expression of another receptor of the immunoglobulin superfamily, *PDL1*, is associated with increased immune infiltration and favourable recurrence free survival across different breast cancer subtypes (Schalper, 2014).

Taken together, the data herein support the potential of *CTLA4*-targeted therapies in TNBC. *CTLA4* is a negative regulator of T cells, and therefore its expression reduces T cell-mediated killing of cancer cells. The data herein show a positive association between

CTLA4 expression and ITLR, consistent with ITL expression of CTLA4. The expression of CTLA4 in ITLs may explain why in many tumours cancer cells were not eliminated even in the presence of high numbers of ITLs. The use of CTLA4 antagonists to inhibit immune tolerance to cancer and to activate ITLs may be an effective treatment strategy for TNBC.

Unsupervised clustering with Gaussian Mixture modelling for CTLA4 expression in all 1,980 METABRIC tumours revealed two clusters, one with high and one with low level of expression of CTLA4 (Fig. 15 A). Using this clustering definition for TNBC tumours we found that TNBC patients with higher level of CTLA4 expression have significantly better disease-specific survival than patients with lower level of CTLA4 expression ($p=0.018$, $HR=0.61$, $CI=0.41-0.92$, Fig. 15 B).

The gene module analysis also revealed several tightly connected, functionally related modules. For example, one module contains *APOBEC3G* (Apolipoprotein B mRNA Editing Enzyme, Catalytic Polypeptide-Like 3G), which is known to play important roles in adaptive and innate immunity and has been investigated extensively in viral infection (Mangeat, 2003) but its role in breast cancer has not been investigated in detail. It is a member of the apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like editing complex family together with *APOBEC3B*, which was found to be a source of mutagenesis in many major cancer types including breast cancer (Kuong, 2013). In the TNBC samples studied herein, *APOBEC3G* expression is significantly correlated with favourable prognosis (log-rank $p=0.02$) but not other *APOBEC* members including *APOBEC3B* ($p=0.29$). *APOBEC3G* is primarily expressed in $CD4^+$ T lymphocytes, macrophages, and dendritic cells (Monajemi, 2012). The present data revealed strong association between *APOBEC3G* and natural killer cell gene *NKG7* and interleukins in this module and support the importance of *APOBEC3G* in TNBC.

The associations between ITLR and immune-relevant genes, pathways and modules support the validity of ITLR as a measure of lymphocytic infiltration and reveal co-regulations of key immune genes.

An aspect of the present invention provides a method of determining a prognosis in a triple negative breast cancer patient, the method comprising, determining the level of expression of *APOBEC3G* in a tumour sample obtained from the patient, wherein increased expression and/or *APOBEC3G* expression indicates a positive prognosis.

An aspect of the present invention provides a method of determining a prognosis in a triple negative breast cancer patient, the method comprising, determining the level of expression of CTLA4 in a tumour sample obtained from the patient, wherein increased CTLA4 expression indicates a positive prognosis. Increased CTLA4 expression may be CTLA4 expression above the middle (50) percentile for TNBC, or may be CTLA4 expression above the (25) percentile for TNBC. Increased CTLA4 expression may be CTLA4 expression that

is high relative to one or more "housekeeping" genes such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

A method of determining a prognosis based on ITLR as described herein may further
5 comprise a step of measuring CTLA4 expression in a tumour obtained from the cancer patient. The cancer patient may be a TNBC patient. The step of measuring CTLA4 expression may involve nucleic acid hybridisation (e.g. microarray-based analysis) or immunohistochemical techniques. In such methods, the combination of an ITLR above a predetermined cut-off value and increased CTLA4 expression indicates a positive prognosis.

10 The predetermined cut-off value for ITLR may be about 0.03 or about about 0.032.

ITLR is an objective quantitative indicator of lymphocytic infiltration in tumours. This quantitative measurement of immune infiltration is useful in guiding treatment decisions in cancer.

15

Accordingly, an aspect of the invention provides a method of using ITLR in predicting whether or not a cancer patient will respond to a therapy.

Such a method may be a method for predicting whether a cancer patient will respond to a
20 therapeutic regime, the method comprising measuring immune infiltration in a tumour from the cancer patient according to a method described herein, wherein an ITLR above a predetermined ITLR cut-off value indicates that the patient is likely to respond to the therapeutic regime.

25 ITLR is useful in informing treatment decisions for cancer patients. Accordingly, an aspect of the present invention provides a method of treating a cancer patient, wherein the ITLR of the tumour has been determined to be either below, or above, a predetermined cut-off value.

The cancer patient may be an individual from whom an image of a tumour has been obtained. The method may comprise determining the ITLR in a tumour from the patient.

30 The method of treatment may comprise administration of a therapeutic regime.

The therapeutic regime may be radiotherapy or chemotherapy or any combination of these. A therapeutic regime comprising chemotherapy may comprise anthracycline-based chemotherapy. A therapeutic regime may comprise administration of a therapeutic agent.

35 Accordingly, an aspect of the present invention provides a therapeutic agent for use in treating cancer in a cancer patient, wherein a prognosis for the cancer patient has been obtained using a method as disclosed herein.

ITLR provides information for predicting a long-term prognosis and for informing patient
40 treatment decisions. Thus if a patient has a low ITLR and is likely to have a poor prognosis, this patient may be treated more intensively (e.g. more rounds of chemotherapy) than a patient having a high ITLR.

Accordingly, an aspect of the present invention provides a method of treating cancer in a cancer patient according to a therapeutic regime, the method comprising analysing a tumour image from the cancer patient according to a method described herein, and treating the cancer patient according to the therapeutic regime depending on whether the ITLR is below or above a predetermined cut-off value.

ITLR combined with CTLA4 expression provides further prognostic information. Relatively high CTLA4 expression may be associated with a high ITLR, and inhibition of CTLA4 may activate T cells to kill cancer cells. Thus, for a patient having an ITLR above a predetermined cut-off value and having increased CTLA4 expression the therapeutic regime may comprise administration of a CTLA4 antagonist. The CTLA4 antagonist may be an antibody, for example ipilimumab.

An aspect of the present invention provides a CTLA4 antagonist for use in a method of treatment of cancer, wherein a tumour from the patient has been determined to have a high ITLR. An aspect of the present invention provides a CTLA4 antagonist for use in a method of treatment of cancer, wherein a tumour from the patient has been determined to have an ITLR above a predetermined ITLR cut-off value. The cancer may be a specific type or subtype of cancer and the predetermined ITLR cut-off value may be the cut-off value determined for a cohort of patients having that cancer type or subtype. The cancer subtype may be breast cancer. The CTLA4 antagonist may be an antibody, which may be an anti-CTLA4 antibody. The anti-CTLA4 antibody may be ipilimumab (also known as MDX-010 and MDX-101). The cancer patient may be a TNBC patient and the therapy may be ipilimumab.

The prognostic and therapeutic methods described herein may further comprise surgically resecting a tumour from a cancer patient, measuring immune infiltration in the tumour according to a method described herein, and determining a prognosis and/or treating the cancer patient according to a therapeutic regime based on the ITLR of the tumour. A surgically resected tumour is a surgically removed tumour. The method of measuring immune infiltration may use a whole tumour section.

An aspect of the present invention provides a method of determining the efficacy of a therapeutic regime. The method may comprise determining the ITLR of a tumour biopsy obtained from a patient before undergoing the therapeutic regime, determining the ITLR of a tumour biopsy obtained from the patient after undergoing the therapeutic agent, and associating an increased ITLR with therapeutic efficacy (i.e. a therapeutic effect).

The methods of analyzing tumours according to the invention may be modified to yield further information on lymphocyte subtypes and their relevance in cancer. Lymphocytes in tumours are known to encompass diverse subclasses including helper T cells, regulatory T

cells, natural killer cells and B cells with sophisticated implications for treatment response (Fridman, 2012; Gu-Trantien, 2013; Andre, 2013). Immunohistochemistry analysis of tumour sections with immune cell markers may be performed, for which automated immunohistochemistry image analysis and statistical modelling methods could be developed to discern interactions between cancer and anti-/pro-tumoural immune response.

In the context of the methods and therapeutic agents described herein, a pathological section may be a tumour section. A tumour section may be a whole-tumour section. A whole-tumour section is typically a section cut from a surgically resected tumour, thus representing the characteristics of the whole tumour. Thus, a whole-tumour section may be a surgically resected section. A pathological section may be a biopsy obtained from a tumour. The pathological section is preferably stained. Staining facilitates morphological analysis of tumour sections by colouring cells, subcellular structures and organelles. Any type of staining may be used, provided that the staining facilitates morphological analysis. The pathological section may be stained with hematoxylin and eosin (H&E). H&E stain is the most commonly used stain in histopathology for medical diagnosis, particularly for the analysis of biopsy sections of suspected cancers by pathologists. Thus H&E stained pathological sections are usually readily available as part of large data sets collated for the study of cancer. The applicability of the present methods to H&E stained pathological sections makes them particularly adaptable for use in analysing data sets from many types and subtypes of cancer to determine the prognostic value of ITLR and to determine cut-off ITLR values for use in the methods described herein.

Reference herein to the ITLR of a tumour also refers to the ITLR of a pathological section, tumour section, or tumour image.

Reference herein to an ITLR value being below a cut-off value may also refer to an ITLR value being equal to or below a cut-off value.

In the present context the term tumour image refers to an image of a tumour from a patient. A tumour image may be an image of a pathological section or tumour section. In the present disclosure a patient may be referred to as having an ITLR (e.g. an ITLR below a predetermined cut-off value), meaning that an image of a tumour from that patient has been determined to have an ITLR. The tumour image may be of a section of a surgically resected tumour, or may be of a biopsy of a tumour.

In the present context the ratio of intra-tumour lymphocytes to cancer cells (ITLR) is the ratio in the pathological section, in a tumour, in a biopsy from the tumour, in a tumour section, or in an image of the tumour, tumour section or pathological section. The term ITLR may also be attributed to a patient. A patient having an ITLR of a particular value refers to a patient from whom a pathological section has an ITLR of a particular value.

The terms "lymphocytic infiltration" and "immune infiltration" are used interchangeably herein.

In the present context "automated" refers to processes that operate independent of external (human) control or input. In the present context an automated process may be a computer-implemented process. The methods of the present invention may be automated methods. The methods of the present invention may be entirely automated methods, that is, they may operate independently of human control or input in their entirety. The methods of the present invention may comprise a step of identifying lymphocyte and cancer cells by automated image analysis. The methods of the present invention may be performed on a tumour image in which lymphocytes and cancer cells have been identified by automated image analysis.

The methods of the present invention are performed on pathological sections, such as tumour sections. The methods of the present invention are therefore *ex vivo* methods, that is, the methods of the present invention are not practiced on the human body.

A cancer patient in the context of the present invention is an individual having cancer or having been diagnosed with cancer. Reference to cancer may be reference to a particular type or subtype of cancer. The cancer patient may have undergone anthracycline-based chemotherapy, immunotherapy, or a combination therapy comprising anthracycline-based chemotherapy and immunotherapy. The cancer patient may have breast cancer, colorectal cancer, melanoma or non-small cell lung cancer. The cancer patient may have the subtype of breast cancer known as triple negative breast cancer. Triple negative breast cancer may be defined as a breast cancer that is negative for estrogen receptors (ER) and HER2. (TNBC is sometimes defined as breast cancer that is negative for estrogen receptors (ER), HER2 and progesterone receptors (PR), but since cancer that are negative for ER are typically also negative for PR, in the present context TNBC is defined as breast cancer that is negative for ER and HER2.

In the context of the present invention reference to the treatment of a cancer patient refers to treatment of cancer in a patient.

The cancer patient may have, or the cancer type or subtype may be selected from, breast cancer (including breast ductal carcinoma breast and breast lobular carcinoma), central nervous system cancer (including glioblastoma multiforme and lower grade glioma), endocrine cancer (including adrenocortical carcinoma, papillary thyroid carcinoma, paraganglioma & pheochromocytoma), gastrointestinal cancer (including Cholangiocarcinoma, Colorectal Adenocarcinoma, Liver Hepatocellular Carcinoma, Pancreatic Ductal Adenocarcinoma, Stomach-Esophageal Cancer), gynecologic cancer (including Cervical Cancer, Ovarian Serous Cystadenocarcinoma, Uterine Carcinosarcoma, Uterine Corpus Endometrial Carcinoma), head and neck cancer (including Head and Neck

Squamous Cell Carcinoma, Uveal Melanoma), hematologic cancer (including Acute Myeloid Leukemia, and Acute Myeloid Leukemia), skin cancer (including Cutaneous Melanoma), soft tissue cancer (including Sarcoma), thoracic cancer (including Lung Adenocarcinoma, Lung Squamous Cell Carcinoma, Mesothelioma) and urologic cancer (including Chromophobe Renal Cell Carcinoma, Clear Cell Kidney Carcinoma, Papillary Kidney Carcinoma, Prostate Adenocarcinoma, Testicular Germ Cell Cancer, Urothelial Bladder Carcinoma). Each of these cancers is the subject of study as part of The Cancer Genome Atlas project.

In the present context the term “immune signatures” is used to encompass all biomarkers related to immune responses and includes the gene expression signatures studied herein as well as other biomarkers including the “Lym” biomarker (Yuan et al. 2012) and ITLR.

Each and every compatible combination of the embodiments described above is explicitly disclosed herein, as if each and every combination was individually and explicitly recited.

Various further aspects and embodiments of the present invention will be apparent to those skilled in the art in view of the present disclosure.

“and/or” where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. For example “A and/or B” is to be taken as specific disclosure of each of (i) A, (ii) B and (iii) A and B, just as if each is set out individually herein.

Unless context dictates otherwise, the descriptions and definitions of the features set out above are not limited to any particular aspect or embodiment of the invention and apply equally to all aspects and embodiments which are described.

Experimental

Methods

Breast cancer studies

Clinical samples

- 5 The complete set of METABRIC (Curtis et al.) samples contains 1,980 primary frozen breast tumours from five contributing hospitals. Among these, 1,026 of the 1,047 tumours from three hospitals have H&E sections without severe artefacts, whilst all the H&E samples from the other two hospitals are highly fragmented due to long-term frozen storage. Therefore we only considered the 1,026 tumours for this study (long-term follow up median 68.3 months).
- 10 On average three tumour sections were obtained at different locations of each primary tumour and placed onto the same slide (Yuan et al., 2012). Tumour materials sandwiched between these sections were sectioned, mixed and used for molecular profiling, thereby maximising the biological relevance of multiple data types being generated. Further details on experimental procedure, staining and molecular profiling protocols can be found in Yuan et al 2012. Gene expression data for the same set of tumours were profiled using the
- 15 Illumina HT-12 platform. *ER* status was determined based on the bimodal distribution of *ESR1* expression microarray data, and *Her2* amplification status based on microarray SNP6 data from the same tumours. In total, there were 181 *ER*-negative, *Her2*-negative samples and these were defined as triple negative/TNBC. Samples from two of the three hospitals
- 20 were merged to form Cohort 1 (89 samples) and samples from the other hospital were merged to form Cohort 2 (92 samples) in order to obtain a similar population size in each cohort. Immune infiltration was scored for 112 of the 181 samples by the pathologists in the METABRIC consortium into three categories: absent, mild and severe. Absent if there were no lymphocytes, mild if there was a light scattering of lymphocytes, and severe if there was a
- 25 prominent lymphocytic infiltrate. The pathological scores of immune infiltration were not significantly correlated with prognosis (Figure 13).

H&E image analysis

- 30 The accuracy of the automated image analysis tool for H&E breast tumour frozen section images had previously been validated based on pathological tumour scores and cell-by-cell evaluation (Yuan et al. , Natrajan et al.). For METABRIC samples, this tool achieved 90% cross-validation accuracy for cell classification and high correlation with pathological scores of cell proportions ($\text{cor}=0.98$) (Yuan et al.). This tool was used to classify all cell nuclei in 181
- 35 TNBC whole-tumour sections, resulting in an average of 81,810 (standard deviation 80,330) cancer cells, 15,500 (25,133) lymphocytes, and 14,090 (14,180) stromal cells for each image. Lymphocytes have a typical morphology of small, round and homogeneously basophilic nuclei, thus can be reliably differentiated from other cell types in cancer. Since this analysis is based on nuclear morphology only in the H&Es, the identified lymphocytes
- 40 are likely to be a mixture of immune cell types including T- and B-lymphocytes.

Modelling the spatial heterogeneity of cancer-immune interaction

Let $x = x_1, x_2, \dots, x_n$ be the spatial locations of n cancer cells and $y = y_1, y_2, \dots, y_m$ be the spatial locations of m immune cells in a tumour image (e.g. an H&E tumour section image).

- 5 Using a quartic kernel function K one can establish a kernel density estimate over the whole tumour image: $f(x) = \frac{\sum_i^n K(x-x_i)}{h}$, where h is the bandwidth parameter for K . h was optimised

using the Minimum Square Error criteria (Berman et al.) in 10 randomly sampled images.

Thus, the spatial proximity to cancer for an immune cell i is $s_i = f(y_i)$. We can then identify

lymphocyte classes based on s , $s = s_1, s_2, \dots, s_m$, using unsupervised Gaussian Mixture

- 10 Clustering (McLachlan, 2000). This method aims to identify multiple components/clusters within the data with probabilities that quantify the uncertainty of observations belonging to the clusters.

$$p(s) = \sum_{k=1}^K w_k G(s|\mu_k, \sigma_k),$$

where K is the number of clusters, μ_k and σ_k are the mean and variance that define the

- 15 probabilistic density function G for the k th component, and w_k is the weight of a component k . These parameters were estimated by Expectation-Maximization (Dempster, 1977).

Selection of models with different numbers of clusters can be done using statistical criteria, one of the most common being the Bayesian Information Criterion (Schwartz, 1978). It can be used in conjunction with mixture model clustering to select the best number of clusters K :

20

$$BIC = -2L(p(s)) + d \log(m)$$

where $L()$ is the maximum log likelihood function and d is the number of free parameters to be estimated. Effectively, the BIC criterion aims to evaluate modelling error as well as model

- 25 complexity. The higher the value of BIC and better the solution is considered to be. To perform clustering, 100,000 immune cells were randomly sampled. Their spatial proximity to cancer data s were used for clustering with a range of different K , $K=1-5$. This was repeated 200 times, 97% of which the solution with three clusters was considered the optimal by BIC. Mean μ_k of the clusters are consistent (median: 0.011, 0.06, 0.13; standard deviation/SD:

- 30 0.002, 0.0047, 0.0045). Subsequently, we classified all lymphocytes in all tumour samples based on these clusters. We used the ratio of the number of intra-tumour lymphocytes and the number of cancer cell as the final measurement of intra-tumour immune infiltration:

$$ITLR = \frac{N_{\text{Intra-Tumour Lymphocyte}}}{N_{\text{cancer}}}$$

35

Image analysis and modelling spatial heterogeneity – in more detailImage data

CRImage processes a H&E slide by first dividing it into 2,000 pixels by 2,000 pixels sub-images and identifying cells in these sub-images. Therefore the cell locations for these sub-images need to be combined. We provide combined cell identifies and spatial locations for all 181 TNBC whole-section H&E sections as R data files in a 'CellPosAndMask' folder. These files are named by their image ID. Each file contain the x, y and class columns storing x y coordinates as well as the class of each cell in the large H&E slide. There is also a 'mask' binary matrix to denote the tissue area. The resolution of this image is 5µm per pixel.

Identify the optimal bandwidth parameter for computing kernel density

By sampling 10 random samples, the Mean Square Error is computed over a range of different bandwidths h for computing cancer density.

```
library(splancs)
MSE <- NULL
set.seed(10)
ffs <- sample(dir('./data/CellPosAndMask/'), 10)
for (ff in ffs){
  res <- try(load(paste('./data/CellPosAndMask/', ff, sep='')))
  CellPos[,1] <- as.character(CellPos[,1])
  CellPos[,2] <- as.numeric(CellPos[,2])
  CellPos[,3] <- as.numeric(CellPos[,3])
  CellPos <- CellPos[rowSums(is.na(CellPos))==0, ]
  CellPos[,3] <- ncol(Mask) - CellPos[,3] + 1
  CellPos[,3][ CellPos[,3] > ncol(Mask)] <- ncol(Mask)
  CellPos <- CellPos[CellPos[,1]!='a',]
  cell.c <- data.frame(x=as.numeric(CellPos[CellPos[,1]!='c',2]),
    y=as.numeric(CellPos[CellPos[,1]!='c',3]))
  cv <- mse2d(as.points(cell.c), poly=cbind(c(0, 0, nrow(Mask), nrow(Mask)), c(0, ncol(Mask),
    ncol(Mask), 0)), nsmse=40, range=10)
  MSE <- rbind(MSE, cv$mse)
}
save(cv, MSE, file='./data/BandwidthSelection.rdata')
```

h=5 was chosen as the optimal bandwidth for lower variability of Mean Square Error.

Generate spatial proximity to cancer for each lymphocyte

Now, spatial scores can be generated given the cell position data using the following getITL function. getITL function uses the cell position files to infer a cancer density map using the bandwidth selected above.

```
getITL <- function(ff, ...){
  require(EBImage)
  require(splancs)
```

```

res <- try(load(paste('./data/CellPosAndMask/', ff, '.rdata', sep='')))
if (class(res)!='try-error'){
  CellPos[,1] <- as.character(CellPos[,1])
  CellPos[,2] <- as.numeric(CellPos[,2])
5   CellPos[,3] <- as.numeric(CellPos[,3])
  CellPos <- CellPos[rowSums(is.na(CellPos))==0, ]
  CellPos[,3] <- ncol(Mask) - CellPos[,3] +1
  CellPos[,3][ CellPos[,3] > ncol(Mask)] <- ncol(Mask)
  cell.c <- data.frame(x=as.numeric(CellPos[CellPos[,1]!='c',2]),
10 y=as.numeric(CellPos[CellPos[,1]!='c',3]))
  res <- kernel2d(as.points(cell.c), poly=cbind(c(0, 0, nrow(Mask), nrow(Mask)), c(0,
  ncol(Mask), ncol(Mask), 0)), h0=h, nx=dim(Mask)[1], ny=dim(Mask)[2])
  cell.l <- data.frame(x=as.numeric(CellPos[CellPos[,1]!='l',2]),
  y=as.numeric(CellPos[CellPos[,1]!='l',3]))
15   z.l <- unlist(sapply(1:length(cell.l$x), function(x) res$z[cell.l$x[x], cell.l$y[x]]))
  }
  z.l
}

```

- 20 Using this function, measurements for each lymphocyte for each tumour can then be generated.

```

itl <- list()
files <- trait$file
25 for (ff in files)
  itl <- c(itl, list(try(getITL(ff, h=5, w=3, cex=.5, ifPlot=F))))
names(itl) <- files
save(itl, file='./data/ITL.rdata')

```

- 30 By default getITL function uses the cut-offs (threshold values) of 0.10507473 and 0.03662728 to determine intra-tumour (ITL), adjacent to tumour (ATL), and distal-to-tumour lymphocytes (DTL). We will now describe how these cut-offs were selected.

Identify sub-populations of lymphocyte by unsupervised learning

- 35 Gaussian mixture clustering and BIC implemented in the R package mclust were used for the discovery of lymphocyte sub-populations. 100,000 lymphocytes were randomly sampled from the itl object and then clustered.

```

library(mclust)
40 load(file='./data/ITL.rdata')
set.seed(11)
x <- sample(as.numeric(unlist(itl)), 100000)
res <- Mclust(x, G=1:5)

```

The sampling process was repeated to generate clusters 200 times, and evaluated output from Mclust and mclustBIC. BIC values for these 200 runs are obtained. The three-cluster solution $k=3$ remains optimal in 97% of the time, and $k=5$ was chosen 3% of the times. The median of cluster means when there are three clusters are 0.0114, 0.0603 and 0.1322 with standard deviation 0.002 and 0.0047 and 0.0045, respectively.

Therefore, the clustering result of lymphocytes from randomly sampled data is stable. Since the clustering is stable, cut-offs were taken at the maximum value of the first and second clusters from one of the sampling runs as our cut-offs for determining lymphocyte classification for the remaining samples.

Generating ITLR, ATLR, and DTLR

Subsequently, the cut-offs can be used to classify every lymphocyte based on their data stored in the R object `itl`. `mat.l` is a matrix with columns of 'Distal', 'Adjacent', 'Intra' denoting the number of lymphocytes in each class for a tumour.

```

15   th=c(0.03662728, 0.10507473)
      mat.l <- NULL
      for (i in 1:length(itl)){
        z.l <- itl[[i]]
20     cl <- rep(1,length(z.l))
        cl[z.l>th[1] & z.l<th[2]] <- 2
        cl[z.l>=th[2]] <- 3
        mat.l <- rbind(mat.l, c(sum(cl==1), sum(cl==2), sum(cl==3)))
      }
25   colnames(mat.l) <- c('Distal', 'Adjacent', 'Intra')
      rownames(mat.l) <- names(itl)

```

The Intra column of `mat.l` is the number of intra-tumour lymphocytes. This divided by the number of cancer cells (`trait$Ntumour`) is the ITLR measurement

30

Measuring cell distances and spatial arrangement

To identify physical properties of ITLs, ATLs and DTLs that differentiate them, in 10,000 lymphocytes randomly sampled from 20 tumours, we identified the 5 nearest cancer cells and the centroid of the convex hull region formed by these cancer cells. For each lymphocyte, the distance from the lymphocyte to the nearest cancer cell was computed (d_{min}), and the distance to the centroid of cancer convex hull was computed ($d_{centroid}$). Centroid of a convex hull region was calculated as the mean positions of the subset of points that define the convex hull. Differences among lymphocyte classes in terms of $d_{centroid}$ and $d_{centroid}$ were tested with student's t-test.

40

Other immune signatures in comparison

Lymphocyte abundance based on image analysis result was calculated as:

$$lym = \frac{N_{lymphocyte}}{N_{cancer}}$$

The gene expression signatures were calculated as described in the referred papers.

ITLR gene modules

- 5 Hierarchical clustering was used to identify highly correlated gene modules by clustering the correlation matrix of all ITL-associated genes into 100 clusters. Modules were selected from these clusters based on average absolute Pearson correlation exceeding 0.75 and cluster size exceeding five.

10 ***Comparing ITLR and ITLR-associated genes***

To test if ITLR has additional value to ITLR-associated genes, we performed multivariate Cox regression analysis with ITLR paired with expression profile of an ITLR gene. This was performed for all of the top 100 ITLR-associated genes ranked by correlation. ITLR was dichotomised using the threshold reported in the paper, and gene expression was
 15 dichotomised into two equal-size group or three groups (25 lower, 50 middle and 25 upper percentiles). Tables with Hazard ratio, log-rank p-value and 95% interval were produced. In both analysis with two and three patient groups according to gene expression data, *p*-values of ITLR were consistently higher than the *p*-values of gene expression profiles, as well as being higher than significance level of 0.05 ($-\log(p)$ 2.99).

20

Other statistical methods

Monotone trend between ITLR and clinical parameters was tested using the Jonckheere-Terpstra trend test (Jonckheere). Survival analysis was performed with breast cancer-specific 10-year survival data. The Kaplan-Meier estimator was used for patient stratification
 25 and log-rank test was used for testing difference among groups. Cox proportional hazards regression model was fitted to the survival data and hazard ratios and 95% confidence intervals were computed to determine the correlation with disease-specific survival, where the log-rank test with $p < 0.05$ was considered significant. Correlation between ITLR and gene expression was computed with Pearson correlation and *q*-values computed using False
 30 Discovery Rate (FDR) correction using 25% of the data for fitting the null model. Cut-offs for dichotomizing immune signatures were optimised stepwise from 20 to 80 percentiles at an interval of 1.5. The cut-offs that displayed the highest prognostic significance with log-rank test were selected. For consistency test in Fig. 5F, each signature was centred at 0 and scaled to standard deviation 1. Optimal cut-offs were also mapped to the new data before
 35 comparison. MSigDB gene set version 4.0 (Subramanian et al.) was used in conjunction with a hypergeometric test for enrichment analysis.

Ovarian cancer studies

- 40 Samples were obtained from a UK-China collaborative study which aims to study the clinical implications of immune infiltration in a set of 91 ovarian cancer patients with metastatic disease. H&E-stained slides for the primary tumours were obtained, scanned, and subjected

to image analysis using CRImage. Cells in these images were classified into cancer, lymphocyte, and stromal cell categories. Once the spatial locations of these cells were obtained from image analysis, kernel density of cancer was computed for each image, and lymphocyte-to-cancer measurements were obtained for each lymphocyte. The
5 measurements were subjected to clustering and two clusters were found, i.e. intra-tumour and non-intra tumour lymphocytes. ITLR as the ratio of intra-tumour lymphocytes to cancer cells was calculated for each patient. The 29% of patients with ITLR lower than a cut-off of 0.06085726 have significantly worst overall survival than patients with ITLR higher than the cut-off (10-year OS log-rank test $p=0.024$, $HR=0.51$, $CI=0.28-0.92$; 5-year OS $p=0.045$,
10 $HR=0.54$, $CI=0.29-0.99$; Figs Ovarian). Overall survival was defined using death as event regardless of the cause, as this information was unavailable.

Tumours were staged according to the 1988 FIGO staging system (Prat 2013). Lymphocytic infiltration was assessed in five high-power fields, each field is scored as absent, mild, or
15 severe: Absent if there were no lymphocytes, mild if there was a light scattering of lymphocytes, and severe if there was a prominent lymphocytic infiltrate. Median of field-based scores was taken as the score for a tumour.

Results

Statistical modelling of the spatial heterogeneity of immune infiltration – determination of ITLR

An automated image analysis tool identified cancer, lymphocytes and stromal cells encompassing fibroblasts and endothelial cells based on their nuclear morphologies in H&E whole-tumour section slides (Yuan et al. 2012). The main component of this tool is a
25 classifier trained by pathologists over randomly selected tumour regions and validated in 564 breast tumours with 90% accuracy (Yuan et al. 2012). As a result of image analysis, the types and spatial locations of on average 110,000 cells were recorded in every breast tumour section. Thus, this fully automated tool enabled the mapping of spatial distributions of all cancer cells and lymphocytes within a tumour section, which can be subsequently
30 visualised as a 3D landscape (Fig. 1A). The spatial relationships of immune and cancer cells are then analysed with a statistical pipeline exemplified in Fig. 1B. First, to globally profile the spatial distribution of the cancer cells, the cancer cell density was quantified using a kernel estimate (Methods). Intuitively, this builds a 'cancer landscape' where hills indicate tumour regions densely populated with cancer cells. The height of a hill thus correlates with cancer
35 density at a specific location in the tumour (Fig. 1B). Secondly, for every lymphocyte, its spatial proximity to cancer can be directly quantified with the cancer density landscape at its specific location. Thus a quantitative measurement of the spatial proximity to tumour cells can be efficiently obtained for every lymphocyte (Fig. 1B).

40 Using this approach, we quantified the spatial proximity to cancer for every lymphocyte in 181 TNBC samples in the METABRIC study (Methods, Fig. 2A). In principle, lymphocytes that differ in their spatial positioning to cancer can be differentiated based on these

quantitative spatial measurements. The inventor investigated whether data-driven clustering methods based on normal distribution can be used to differentiate different classes of lymphocytes, since cell spatial distribution is a naturally emerged pattern. Unsupervised Gaussian Mixture Model clustering (Fraley, 2003) was employed to identify lymphocyte

5 clusters based on their spatial proximity to cancer using a training set of 100,000 randomly sampled lymphocytes (Fig. 2B, Methods). Subsequently, a three-cluster solution that identify three classes of lymphocytes was considered the optimal by the Bayesian Information Criterion (Schwartz, 1978) (Fig. 2B). This three-class solution is consistently the optimal 97% of the time upon 200 repeated sampling, whilst the five-class solution was considered

10 optimal 3% of the time (Methods, Fig. 2C). In addition, the cluster structure of the three-class solutions was stable (median of cluster mean: 0.011, 0.06, 0.13; standard deviation/SD: 0.002, 0.0047, 0.0045; Fig. 2C), indicating that the same clusters were identified in each random sampling. The three classes of lymphocytes were named as Intra-Tumour Lymphocyte (ITL), Adjacent-Tumour Lymphocyte (ATL) and Distal-Tumour Lymphocyte

15 (DTL). Subsequently, a classifier was trained based on the lymphocyte classes to predict the types of lymphocytes in all TNBC samples (Methods).

To understand the differences of the newly proposed lymphocyte classes, additional measures were derived that are based on direct physical distances. First, for each

20 lymphocyte its distance to the nearest cancer cell can be quantified (d_{min} , Methods, Fig. 2D). It was shown that ITLs have a median distance of 7 μ m (interquartile range 5-10) to the nearest cancer cell, whilst it is 10 μ m (7-11) for ATLs, and 20 μ m (14-26) for DTLs (Fig. 2E). The overlap in distance to nearest cancer cell between ITLs and ATLs suggests that this measure is not the fundamental difference between the two classes. Since the kernel density

25 measure based on which the lymphocyte classes were derived is essentially spatial smoothing, the inventor hypothesised that the spatial arrangement of cancer cells surrounding lymphocytes differs between ATLs and ITLs. To measure spatial arrangement, the inventor examined the convex hull region formed by 5 nearest cancer cells, which is the smallest region that covers these cells (Fig. 2D, Methods). If a lymphocyte is surrounded by

30 cancer cells, it should fall into the convex hull region formed by nearby cancer cells and has a small distance to the centroid of this region (Fig. 2D, left). In contrast, if nearby cancer cells are to one side of a lymphocyte, the distance between the lymphocyte and the centroid of the cancer convex hull region is likely to be large (Fig. 2D, right). Thus, the inventor used the distance between a lymphocyte and the centroid of the cancer convex hull region as a

35 quantitative measure of the spatial arrangement of cancer cells surrounding a lymphocyte ($d_{centroid}$). Three lymphocyte classes displayed significant differences in $d_{centroid}$ with median $d_{centroid}$ 3.6 μ m (2.2-5.1), 7.2 μ m (4.5-10.6), 17.7 μ m (11.0-26.6) for ITLs, ATLs, and DTLs, respectively (Fig. 2E). Therefore, d_{min} and $d_{centroid}$ together better define and aid interpretation of the lymphocyte classes (Fig. 2F). Taken together, these data demonstrated

40 that the proposed kernel-based measure of spatial proximity to cancer can effectively account for spatial proximity and surroundings, and also that the three lymphocyte classes differ not only in the distance to the nearest cancer cell but also in the ways nearby cancer

cells are arranged. A representative case showing spatial distribution of lymphocytes in these three classes is illustrated (Fig. 3A-B). For instance, the ITLs can be observed to locate within regions densely populated with cancer cells (Fig. 3C).

In the 181 TNBC samples, there are overall more ATLs than the other two types of lymphocytes (on average 47% ATLs, 32% ITLs and 21% DTLs, Fig. 4A). The changes in abundance of these three classes in 181 samples can be observed in a triangle plot (Fig. 4B). When the proportion of ITLs is low (0-20%), there are in general more DTLs (40-60%) than ATLs (30-50%). As the amount the ITLs increase (20-50%), ATLs also increase (40-60%) while DTLs decrease (10-40%). When there are large amount of ITLs (>50%), there are still certain amount of ATLs (20-40%) with very few DTLs (<10%). To summarise the degree of lymphocytic infiltration for a given tumour, we first calculated the ratio between the number of ITLs and the number of cancer cells (ITLR; see Methods above). In the 181 TNBC samples, a significant association was observed between ITLR and pathological assessment of lymphocytic infiltration of the tumours in categories of absent, mild and severe ($p=2 \times 10^{-33}$, Fig. 4C). In terms of other clinical parameters, there was no correlation between ITLR and tumour size, node status and *TP53* mutation status (Fig. 4D). Tumour grade was not considered because 87% of the TNBC samples are Grade 3 tumours. These data support ITLR's validity as a measurement of lymphocytic infiltration and its potential value in addition to known clinical parameters for TNBC.

ITLR is a statistical measure of lymphocytic infiltration and an independent predictor of disease-specific survival in two TNBC cohorts.

To investigate the clinical significance of the proposed immune measure ITL, the inventor analysed disease-specific survival as a function of ITL. The TNBC samples can be divided into two independent cohorts based on contributing hospitals (Methods, $n=89$ and $n=92$, distribution of ITLR Fig. 3E). To dichotomise the continuous ITLR, the optimal cut-off was selected to have the best prognostic value in Cohort 1 as the discovery cohort (Methods). The best cut-off was selected to be 0.011 and 20% of the patients have ITLR lower than this cut-off. These patients have significantly worse disease-specific survival compared with patients with higher ITLR in Cohort 1 (Log-rank test $p=0.0063$, Hazard ratio $HR=0.36$, 95% confidence interval $CI=0.17-0.77$; Table 1; Fig. 4F). This observation was verified in the validation cohort, Cohort 2 ($p=0.0037$, $HR=0.25$, $CI=0.09-0.69$; Fig. 4F). Significant stratification was observed upon repeated analysis with Cohort 2 as the discovery and Cohort 1 as the validation cohort (Fig. 3G). The same tests were performed for the ratio of ATLs and DTLs to cancer cells (ATLR and DTLR), but neither showed a significant correlation with disease-specific survival (Discovery and Validation cohort: ATLR $p=0.064$ and 0.75 ; DTLR $p=0.43$ and 0.25 ; Fig. 7-8). We subsequently focused on ITLR. ITLR-high TNBC patients have a survival probability of 80% five year from diagnosis versus 49% for ITLR-low patients (Kaplan-Meier survival estimates, two cohorts combined). ITLR was compared with eight other immune signatures. These include the previously published image-based signature, lymphocyte abundance (Lym), defined as the ratio

between the number of lymphocytes and the number of cancer cells (Methods) (Yuan et al. 2012). A major difference between ITLR and Lym is that Lym does not account for different classes of lymphocytes whilst ITLR considers infiltrating lymphocytes. The remainder of signatures are published gene expression-based signatures from Calabro *et al.* (Calabró et al.) that is predictive of ER-negative breast cancer prognosis, a 5-gene signature from Ascierto *et al.* (Ascierto et al. '12) that predicts recurrence-free survival across breast cancer subtypes, and the B-cell, *IL8* and combined signatures to predict prognosis of TNBC (Rody et al.). *CXCR3* and *CXCL13* expression were also included since they have been shown to correlate with breast cancer prognosis (Ma et al. , Gu-Trantien et al.).

The same cut-off selection approach was applied to test the association between these signatures and disease-specific survival (Table 2). The signatures that showed the best prognostic values are shown in Fig. 5A-E (all are provided in Figure 9) and Table 1. None of these signatures correlated with prognosis in both cohorts. This analysis was repeated using Cohort 2 as the discovery cohort for selecting the optimal cut-offs and Cohort 1 for validation (Figure 10, Table 3). In both experiments, only ITLR consistently stratified patients into two groups of different outcome among the nine signatures (Figures 7 and 8). Furthermore, the best ITLR cut-offs selected in two cohorts for all nine signatures were compared (Methods, Figure 5F). ITLR was among the most consistent signatures in terms of optimal cut-offs in two cohorts, supporting the consistency and the potential use of ITLR as an objective measure for identifying patients with low lymphocytic infiltration.

Compared to published immune signatures, ITLR was also the only signature to show significant correlation with disease-specific survival in multivariate Cox proportional hazards model together with standard clinical parameters of nodal status and tumour size in both cohorts, whichever cohort was used as the discovery cohort (Tables 1 to 3). Using samples from both cohorts, ITLR has a log-rank *p*-value of 2.1×10^{-4} and HR 0.32 (0.17-0.58). To test the robustness of the Cox model in determining the prognostic value of ITLR, we used bootstrap analysis in randomly perturbed data and repeated the univariate and multivariate regression analysis 1,000 times. In 95.6% and 94.7% of the time, ITLR remained significantly associated with prognosis in univariate and multivariate analysis, respectively. Taken together, these results show the stability and robustness of ITLR as an independent prognostic biomarker in TNBC.

ITLR heterogeneity is reflected on the transcriptional level by CTLA4 and APOBEC3G expression

To identify molecular associations of immune infiltration and to test the biological relevance of ITLR, the inventor integrated image-based ITLR with microarray gene expression data profiled for the same set of 181 TNBC tumours. The analysis identified 307 genes positively correlated and 105 genes negatively correlated with ITLR (False Discovery Rate multiple testing correction, *q*-value<0.05; Methods). Genes with the most significant correlations with our immune signature ITLR include kinases (*SH3KBP1*, *LCK*, *MAP4K1*) and receptors

(*FCRL3*, *GPR18*, *TNFRSF13B*, *SEMA4D*, *CXCR3*, *IL2RG*), as well as the known immunotherapy target *CTLA4* (Table 4). Thus, significant correlations between ITLR and immune-related genes demonstrate the biological relevance of the ITLR signature.

5

Table 4. Top 20 genes positively correlated with ITLR and top 10 genes negatively correlated with ITLR (underline).

Symbol	Cytoband	Description	cor	q
<i>SH3KBP1</i>	Xp22.12b	SH3-domain kinase binding protein 1	0.4	0.0011
<i>FCRL3</i>	1q23.1d	Fc receptor-like 3	0.4	0.0011
<i>LCK</i>	1p35.1b	lymphocyte-specific protein tyrosine kinase	0.4	0.0011
<i>GPR18</i>	13q32.3a	G protein-coupled receptor 18	0.39	0.0011
<i>TNFRSF13B</i>	17p11.2h	tumour necrosis factor receptor superfamily, member 13B	0.39	0.0011
<i>SEMA4D/CD100</i>	9q22.2a	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4D	0.39	0.0012
<i>MAP4K1</i>	19q13.2a	mitogen-activated protein kinase kinase kinase 1	0.39	0.0012
<i>RLTPR</i>	16q22.1b	RGD motif, leucine rich repeats, tropomodulin domain and proline-rich containing	0.38	0.0012
<i>UBASH3A</i>	21q22.3b	ubiquitin associated and SH3 domain containing A	0.38	0.0012
<i>IKZF3</i>	17q12c	IKAROS family zinc finger 3 (Aiolos)	0.38	0.0012
<i>CYFIP2</i>	5q33.3a-q33.3b	cytoplasmic FMR1 interacting protein 2	0.38	0.0012
<i>CXCR3</i>	Xq13.1d	chemokine (C-X-C motif) receptor 3	0.38	0.0012
<i>CD3E</i>	11q23.3d	CD3e molecule, epsilon (CD3-TCR complex)	0.38	0.0012
<i>IL2RG</i>	Xq13.1c	interleukin 2 receptor, gamma	0.38	0.0012
<i>CXCR5</i>	11q23.3e	chemokine (C-X-C motif) receptor 5	0.38	0.0014
<i>CTSW</i>	11q13.1d	cathepsin W	0.37	0.0018
<i>SH2D1A</i>	Xq25c	SH2 domain containing 1A	0.37	0.0018
<i>SEPT6</i>	Xq24c	septin 6	0.37	0.0018
<i>CTLA4</i>	2q33.2a	cytotoxic T-lymphocyte-associated protein 4	0.37	0.0019
<i>SIRPG</i>	20p13e	signal-regulatory protein gamma	0.37	0.0019
<u><i>C10orf141</i></u>	<u>10q26.2b</u>		<u>-0.4</u>	<u>0.0011</u>
<u><i>CD151</i></u>	<u>11p15.5c</u>	<u>CD151 molecule (Raph blood group)</u>	<u>-0.39</u>	<u>0.0011</u>
<u><i>SPP1</i></u>	<u>4q22.1b</u>	<u>secreted phosphoprotein 1</u>	<u>-0.39</u>	<u>0.0012</u>
<u><i>ANXA2</i></u>	<u>15q22.2a</u>	<u>annexin A2</u>	<u>-0.39</u>	<u>0.0012</u>
<u><i>P4HA2</i></u>	<u>5q31.1b</u>	<u>prolyl 4-hydroxylase, alpha polypeptide II</u>	<u>-0.36</u>	<u>0.0022</u>
<u><i>MUSK</i></u>	<u>9q31.3b</u>	<u>muscle, skeletal, receptor tyrosine kinase</u>	<u>-0.36</u>	<u>0.0023</u>

<u>POFUT2</u>	<u>21q22.3e</u>	<u>protein O-fucosyltransferase 2</u>	<u>-0.36</u>	<u>0.0025</u>
<u>ITGB5</u>	<u>3q21.2a</u>	<u>integrin, beta 5</u>	<u>-0.35</u>	<u>0.004</u>
<u>MXRA7</u>	<u>17q25.1d-</u> <u>q25.2a</u>	<u>matrix-remodelling associated 7</u>	<u>-0.34</u>	<u>0.0046</u>
<u>CALN1</u>	<u>7q11.22c</u>	<u>calneuron 1</u>	<u>-0.34</u>	<u>0.0046</u>

Subsequently, enrichment analysis was performed on the positively and negatively correlated genes respectively against MSigDB gene set categories (Subramanian, 2005) including KEGG pathways (Kanehisa, 2000), canonical pathways curated by domain experts and immunologic signatures (Methods, Figure 11). Genes positively correlated with ITLR are enriched with natural killer cell mediated cytotoxicity, T cell receptor, Antigen processing and presentation KEGG pathways, *CD8* T cell, *CD4* T cell and B cell up-regulated immunogenic signatures, as well as *IL12* and *CD8 TCR* canonical pathways. Conversely, genes negatively correlated with ITLR were enriched with ECM receptor interaction and focal adhesion KEGG pathways, regulatory T cell and *TGF β* related immunologic signatures as well as integrin related pathways. The molecular analysis on the pathway level suggests ITLR is positively associated with anti-tumour immune activities in TNBC.

To further dissect their interconnected relationships and discover *de novo* molecular modules, tightly connected gene modules were identified within ITLR-associated genes (Figure 11; Methods). As such, seven modules of positively correlated genes (P1-7) and two modules of genes negatively correlated with ITLR (N1 and N2) were identified. Known immune-related genes in the modules include *IFNG* (P1), *RLPTR* (P3), *GPR18* (P4), *CXCR3* (P5), *MAP4K1* (P6), *CTLA4* (P7), *ANXA2* (N1) and *FAP* (N2). Notably, two of the modules contain *APOBEC3G* (P2) and *CTLA4* (P7), which may suggest co-regulation among *APOBEC3G*, *NKG7* and interleukins including *IL21R* and *IL18RAP*, as well as high correlations among *CTLA4*, chemoattractant for B lymphocytes *CXCL13* (Denkert) and *TIGIT* T cell immunoreceptor with *Ig* and *ITIM* domains. Furthermore, expression profiles of these genes were significantly associated with disease-specific survival in TNBC, including *APOBEC3G* as well as *GPR18* (P4) and *MAP4K1* (P6) ranked as the top ITLR-associated genes (Figure 6B, Figure 12). *CTLA4* expression was able to stratify patients into groups with significantly different prognosis, and could further stratify the ITLR high group into two subgroups with significantly different outcome ($p=0.046$, Figure 6C, Figure 12). Comparing ITLR with ITLR-associated genes in terms of prognostic value, multivariate analysis showed that ITLR stratification has additional and in many cases superior value to ITLR-associated genes (Figure 13, Methods).

References

A number of publications are cited above in order to more fully describe and disclose the invention and the state of the art to which the invention pertains. Full citations for these
 5 references are provided below. The entirety of each of these references is incorporated herein.

Alexiou GA, Vartholomatos E, Voulgaris S. Prognostic value of neutrophil-to-lymphocyte ratio in patients with glioblastoma. *Journal of neuro-oncology*. 2013;115:521-2.

- 10 Ancuta E, Ancuta C, Zugun-Eloae F, Iordache C, Chirieac R, Carasevici E. Predictive value of cellular immune response in cervical cancer. *Romanian journal of morphology and embryology = Revue roumaine de morphologie et embryologie*. 2009;50:651-5.

- 15 Andre, F., M. V. Dieci, P. Dubsy, C. Sotiriou, G. Curigliano, C. Denkert and S. Loi (2013). "Molecular pathways: involvement of immune pathways in the therapeutic response and outcome in breast cancer." *Clin Cancer Res* 19(1): 28-33.

Ascierto, M. L., M. Kmiecik, M. O. Idowu, R. Manjili, Y. Zhao, M. Grimes, C. Dumur, E. Wang, V. Ramakrishnan, X. Y. Wang, H. D. Bear, F. M. Marincola and M. H. Manjili (2012). "A signature of immune function genes associated with recurrence-free survival in breast cancer patients." *Breast Cancer Res Treat* 131(3): 871-880.

- 20 Bambury RM, Teo MY, Power DG, Yusuf A, Murray S, Battley JE, et al. The association of pre-treatment neutrophil to lymphocyte ratio with overall survival in patients with glioblastoma multiforme. *Journal of neuro-oncology*. 2013;114:149-54

- 25 Basavanthally, A. N., S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot and A. Madabhushi (2010). "Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology." *IEEE Trans Biomed Eng* 57(3): 642-653.

Berman, M. and P. Diggle (1989). "Estimating Weighted Integrals of the 2nd-Order Intensity of a Spatial Point Process." *Journal of the Royal Statistical Society Series B-Methodological* 51(1): 81-92.

Jonckheere, A. R. (1954). "A test of significance for the relation between m rankings and k ranked categories." *British Journal of Statistical Psychology* 7: 93--100.

- 30 Bystryn JC, Oratz R, Roses D, Harris M, Henn M, Lew R. Relationship between immune response to melanoma vaccine immunization and clinical outcome in stage II malignant melanoma. *Cancer*. 1992;69:1157-64.

- 35 Calabró, A., T. Beissbarth, R. Kuner, M. Stojanov, A. Benner, M. Asslaber, F. Ploner, K. Zatloukal, H. Samonigg, A. Poustka and H. Söhlmann (2009). "Effects of infiltrating lymphocytes and estrogen receptor on gene expression and prognosis in breast cancer." *Breast Cancer Res Treat* 116(1): 69-77.

Cheng, Yizong (August 1995). "Mean Shift, Mode Seeking, and Clustering". *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE)* 17 (8): 790-799.

- 40 Clarke B, Tinker AV, Lee CH, Subramanian S, van de Rijn M, Turbin D, et al. Intraepithelial T cells and prognosis in ovarian carcinoma: novel associations with stage, tumor type, and BRCA1 loss. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2009;22:393-402.

- 45 Contardi, E., G. L. Palmisano, P. L. Tazzari, A. M. Martelli, F. Fala, M. Fabbi, T. Kato, E. Lucarelli, D. Donati, L. Polito, A. Bolognesi, F. Ricci, S. Salvi, V. Gargaglione, S. Mantero, M. Alberghini, G. B. Ferrara and M. P. Pistillo (2005). "CTLA-4 is constitutively expressed on tumor cells and can trigger apoptosis upon ligand interaction." *Int J Cancer* 117(4): 538-550.

Curtis, C., S. P. Shah, S. F. Chin, G. Turashvili, O. M. Rueda, M. J. Dunning, D. Speed, A. G. Lynch, S. Samarajiwa, Y. Yuan, S. Graf, G. Ha, G. Haffari, A. Bashashati, R. Russell, S. McKinney, A. Langerod, A. Green, E. Provenzano, G. Wishart, S. Pinder, P. Watson, F. Markowitz, L. Murphy, I.

- Ellis, A. Purushotham, A. L. Borresen-Dale, J. D. Brenton, S. Tavare, C. Caldas and S. Aparicio (2012). "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups." *Nature* 486(7403): 346-352.
- 5 Crane CA, Austgen K, Haberthur K, Hofmann C, Moyes KW, Avanesyan L, et al. Immune evasion mediated by tumor-derived lactate dehydrogenase induction of NKG2D ligands on myeloid cells in glioblastoma patients. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111:12823-8.
- 10 Dave SS, Wright G, Tan B, Rosenwald A, Gascoyne RD, Chan WC, et al. Prediction of Survival in Follicular Lymphoma Based on Molecular Features of Tumor-Infiltrating Immune Cells. *New England Journal of Medicine*. 2004;351:2159-69.
- Davidsson S, Ohlson AL, Andersson SO, Fall K, Meisner A, Fiorentino M, et al. CD4 helper T cells, CD8 cytotoxic T cells, and FOXP3(+) regulatory T cells with respect to lethal prostate cancer. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. 2013;26:448-55.
- 15 DeNardo, D. G., D. J. Brennan, E. Rexhepaj, B. Ruffell, S. L. Shiao, S. F. Madden, W. M. Gallagher, N. Wadhwani, S. D. Keil, S. A. Junaid, H. S. Rugo, E. S. Hwang, K. Jirstrom, B. L. West and L. M. Coussens (2011). "Leukocyte complexity predicts breast cancer survival and functionally regulates response to chemotherapy." *Cancer Discov* 1(1): 54-67.
- 20 Denkert, C., S. Loibl, A. Noske, M. Roller, B. M. Muller, M. Komor, J. Budczies, S. Darb-Esfahani, R. Kronenwett, C. Hanusch, C. von Tonne, W. Weichert, K. Engels, C. Solbach, I. Schrader, M. Dietel and G. von Minckwitz (2010). "Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer." *J Clin Oncol* 28(1): 105-113.
- 25 Dieci, M. V., C. Criscitiello, A. Goubar, G. Viale, P. Conte, V. Guarneri, G. Ficarra, M. C. Mathieu, S. Delaloge, G. Curigliano and F. Andre (2014). "Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: a retrospective multicenter study." *Ann Oncol* 25(3): 611-618.
- Dieu-Nosjean MC, Antoine M Fau - Danel C, Danel C Fau - Heudes D, Heudes D Fau - Wislez M, Wislez M Fau - Poulot V, Poulot V Fau - Rabbe N, et al. Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures.
- 30 Erdag G, Schaefer JT, Smolkin ME, Deacon DH, Shea SM, Dengel LT, et al. Immunotype and immunohistologic characteristics of tumor-infiltrating immune cells are associated with clinical outcome in metastatic melanoma. *Cancer research*. 2012;72:1070-80
- Failmezger H, Yuan Y, Rueda O and Markowitz F (2012). *CRIImage: CRIImage a package to classify cells and calculate tumour cellularity*. R package version 1.14.0.
- 35 Fiorelli V, Gendelman R, Caterina Sirianni M, Chang H-K, Colombini S, Markham PD, et al. γ -Interferon Produced by CD8+ T Cells Infiltrating Kaposi's Sarcoma Induces Spindle Cells With Angiogenic Phenotype and Synergy With Human Immunodeficiency Virus-1 Tat Protein: An Immune Response to Human Herpesvirus-8 Infection? *Blood*. 1998;91:956-67.
- 40 Fraley, C. and A. E. Raftery (2003). "Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST." *Journal of Classification* 20(2): 263-286.
- Fridman, W. H., F. Pages, C. Sautès-Fridman and J. Galon (2012). "The immune contexture in human tumours: impact on clinical outcome." *Nat Rev Cancer* 12(4): 298-306.
- 45 Gannon PO, Poisson AO, Delvoye N, Lapointe R, Mes-Masson AM, Saad F. Characterization of the intra-prostatic immune cell infiltration in androgen-deprived prostate cancer patients. *Journal of immunological methods*. 2009;348:9-17.
- Galon, J., A. Costes, F. Sanchez-Cabo, A. Kirilovsky, B. Mlecnik, C. Lagorce-Pages, M. Tosolini, M. Camus, A. Berger, P. Wind, F. Zinzindohoue, P. Bruneval, P. H. Cugnenc, Z. Trajanoski, W. H. Fridman and F. Pages (2006). "Type, density, and location of immune cells within human colorectal tumors predict clinical outcome." *Science* 313(5795): 1960-1964.
- 50 Galon, J., B. Mlecnik, G. Bindea, H. K. Angell, A. Berger, C. Lagorce, A. Lugli, I. Zlobec, A. Hartmann, C. Bifulco, I. D. Nagtegaal, R. Palmqvist, G. V. Masucci, G. Botti, F. Tatangelo, P. Delrio, M. Maio, L. Laghi, F. Grizzi, M. Asslaber, C. D'Arrigo, F. Vidal-Vanaclocha, E. Zavadova, L. Chouchane, P. S.

- Ohashi, S. Hafezi-Bakhtiari, B. G. Wouters, M. Roehrl, L. Nguyen, Y. Kawakami, S. Hazama, K. Okuno, S. Ogino, P. Gibbs, P. Waring, N. Sato, T. Torigoe, K. Itoh, P. S. Patel, S. N. Shukla, Y. Wang, S. Kopetz, F. A. Sinicrope, V. Scripcariu, P. A. Ascierto, F. M. Marincola, B. A. Fox and F. Pages (2014). "Towards the introduction of the 'Immunoscore' in the classification of malignant tumours." *J Pathol* 232(2): 199-209.
- Gao Q, Qiu SJ Fau - Fan J, Fan J Fau - Zhou J, Zhou J Fau - Wang X-Y, Wang Xy Fau - Xiao Y-S, Xiao Ys Fau - Xu Y, et al. Intratumoral balance of regulatory and cytotoxic T cells is associated with prognosis of hepatocellular carcinoma after resection.
- Gu-Trantien, C., S. Loi, S. Garaud, C. Equeter, M. Libin, A. de Wind, M. Ravoet, H. Le Buanec, C. Sibille, G. Manfouo-Foutsop, I. Veys, B. Haibe-Kains, S. K. Singhal, S. Michiels, F. Rothe, R. Salgado, H. Duvillier, M. Ignatiadis, C. Desmedt, D. Bron, D. Larsimont, M. Piccart, C. Sotiriou and K. Willard-Gallo (2013). "CD4(+) follicular helper T cell infiltration predicts breast cancer survival." *J Clin Invest* 123(7): 2873-2892.
- Halliday G, Patel A, Hunt M, Tefany F, Barnetson RC. Spontaneous regression of human melanoma/nonmelanoma skin cancer: Association with infiltrating CD4+ T cells. *World J Surg.* 1995;19:352-8.
- Hastie R, Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning* (Springer, New York, 2001).
- Hegmans JP, Hemmes A, Hammad H, Boon L, Hoogsteden HC, Lambrecht BN. Mesothelioma environment comprises cytokines and T-regulatory cells that suppress immune responses. *The European respiratory journal.* 2006;27:1086-95.
- Hiraoka N, Onozato K Fau - Kosuge T, Kosuge T Fau - Hirohashi S, Hirohashi S. Prevalence of FOXP3+ regulatory T cells increases during the progression of pancreatic ductal adenocarcinoma and its premalignant lesions.
- Huang CT, Oyang YJ, Huang HC, Juan HF. MicroRNA-mediated networks underlie immune response regulation in papillary thyroid carcinoma.
- Issa-Nummer, Y., S. Darb-Esfahani, S. Loibl, G. Kunz, V. Nekljudova, I. Schrader, B. V. Sinn, H. U. Ulmer, R. Kronenwett, M. Just, T. Kuhn, K. Diebold, M. Untch, F. Holms, J. U. Blohmer, J. O. Habek, M. Dietel, F. Overkamp, P. Krabisch, G. von Minckwitz and C. Denkert (2013). "Prospective validation of immunological infiltrate for prediction of response to neoadjuvant chemotherapy in HER2-negative breast cancer--a substudy of the neoadjuvant GeparQuinto trial." *PLoS One* 8(12): e79775.
- Janowczyk A, S. C, Madabhushi A. Quantifying local heterogeneity via morphologic scale: Distinguishing tumoral from stromal regions. *Journal of pathology informatics.* 2013;4(Suppl).
- Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." *Nucleic Acids Research* 28(1): 27-30.
- Kmiecik J, Poli A, Brons NH, Waha A, Eide GE, Enger PO, et al. Elevated CD3+ and CD8+ tumor-infiltrating immune cells correlate with prolonged survival in glioblastoma patients despite integrated immunosuppressive mechanisms in the tumor microenvironment and at the systemic level. *Journal of neuroimmunology.* 2013;264:71-83.
- Kruger, J. M., C. Wemmert, L. Sternberger, C. Bonnas, G. Dietmann, P. Gancarski and F. Feuerhake (2013). "Combat or surveillance? Evaluation of the heterogeneous inflammatory breast cancer microenvironment." *J Pathol* 229(4): 569-578.
- Kuong, K. J. and L. A. Loeb (2013). "APOBEC3B mutagenesis in cancer." *Nat Genet* 45(9): 964-965.
- Monajemi, M., C. F. Woodworth, J. Benkaroun, M. Grant and M. Larijani (2012). "Emerging complexities of APOBEC3G action on immunity and viral fitness during HIV infection and treatment." *Retrovirology* 9: 35.
- Kono K, Kawaida H, Takahashi A, Sugai H, Mimura K, Miyagawa N, et al. CD4(+)CD25high regulatory T cells increase with tumor stage in patients with gastric and esophageal cancers. *Cancer Immunol Immunother.* 2006;55:1064-71.

- Lee, H. J., J. Y. Seo, J. H. Ahn, S. H. Ahn and G. Gong (2013). "Tumor-associated lymphocytes predict response to neoadjuvant chemotherapy in breast cancer patients." *J Breast Cancer* 16(1): 32-39.
- 5 Liu, S., J. Lachapelle, S. Leung, D. Gao, W. D. Foulkes and T. O. Nielsen (2012). "CD8+ lymphocyte infiltration is an independent favorable prognostic indicator in basal-like breast cancer." *Breast Cancer Res* 14(2): R48.
- 10 Loi, S., N. Sirtaine, F. Piette, R. Salgado, G. Viale, F. Van Eenoo, G. Rouas, P. Francis, J. P. Crown, E. Hitre, E. de Azambuja, E. Quinaux, A. Di Leo, S. Michiels, M. J. Piccart and C. Sotiriou (2013). "Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98." *J Clin Oncol* 31(7): 860-867.
- 15 McNamara MG, Lwin Z, Jiang H, Templeton AJ, Zadeh G, Bernstein M, et al. Factors impacting survival following second surgery in patients with glioblastoma in the temozolomide treatment era, incorporating neutrophil/lymphocyte ratio and time to first progression. *Journal of neuro-oncology*. 2014;117:147-52.
- Ma, X., K. Norsworthy, N. Kundu, W. H. Rodgers, P. A. Gimotty, O. Goloubeva, M. Lipsky, Y. Li, D. Holt and A. Fulton (2009). "CXCR3 expression is associated with poor survival in breast cancer and promotes metastasis in a murine model." *Mol Cancer Ther* 8(3): 490-498.
- 20 Mahmoud, S. M. A., E. C. Paish, D. G. Powe, R. D. Macmillan, M. J. Grainge, A. H. S. Lee, I. O. Ellis and A. R. Green (2011). "Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer." *J Clin Oncol* 29(15): 1949-1955.
- Mangeat, B., P. Turelli, G. Caron, M. Friedli, L. Perrin and D. Trono (2003). "Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts." *Nature* 424(6944): 99-103.
- 25 Mao, H., L. Zhang, Y. Yang, W. Zuo, Y. Bi, W. Gao, B. Deng, J. Sun, Q. Shao and X. Qu (2010). "New insights of CTLA-4 into its biological function in breast cancer." *Curr Cancer Drug Targets* 10(7): 728-736.
- 30 Milne K, Kobel M, Kalloger SE, Barnes RO, Gao D, Gilks CB, et al. Systematic analysis of immune infiltrates in high-grade serous ovarian cancer reveals CD20, FoxP3 and TIA-1 as positive prognostic factors. *PloS one*. 2009;4:e6412.
- Mittendorf et al (2014) "PD-L1 Expression in Triple Negative Breast Cancer" *Cancer Immunol Res* 2; 361.
- Mukherji B Fau - Guha A, Guha A Fau - Loomis R, Loomis R Fau - Ergin MT, Ergin MT. Cell-mediated amplification and down regulation of cytotoxic immune response against autologous human cancer.
- 35 Natrajan, R., F. M. Mardakheh, H. Sailem, C. Bakal and Y. Yuan (In preparation). "The Ecosystem Diversity Landscape of Breast Cancer."
- Ohno S, Ohno Y, Suzuki N, Kamei T, Koike K, Inagawa H, et al. Correlation of histological localization of tumor-associated macrophages with clinicopathological features in endometrial cancer. *Anticancer research*. 2004;24:3335-42
- 40 Papewalis C, Kouatchoua C Fau - Ehlers M, Ehlers M Fau - Jacobs B, Jacobs B Fau - Porwol D, Porwol D Fau - Schinner S, Schinner S Fau - Willenberg HS, et al. Chromogranin A as potential target for immunotherapy of malignant pheochromocytoma.
- Perou C, Sorlie T, Eisen M, van de Rijn M, Jeffrey S, Rees C, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747-52.
- 45 Prat, J., "Staging classification for cancer of the ovary, fallopian tube and peritoneum." *International Journal of Gynecology and Obstetrics*. 2013. (<http://www.ncin.org.uk/view?rid=2515>).
- Pretschner D, Distel LV, Grabenbauer GG, Wittlinger M, Buettner M, Niedobitek G. Distribution of immune cells in head and neck cancer: CD8+ T-cells and CD20+ B-cells in metastatic lymph nodes are associated with favourable outcome in patients with oro- and hypopharyngeal carcinoma. *BMC cancer*. 2009;9:292.
- 50

- Robert, C., L. Thomas, I. Bondarenko, S. O'Day, D. J. M. C. Garbe, C. Lebbe, J. F. Baurain, A. Testori, J. J. Grob, N. Davidson, J. Richards, M. Maio, A. Hauschild, W. H. Miller, Jr., P. Gascon, M. Lotem, K. Harmankaya, R. Ibrahim, S. Francis, T. T. Chen, R. Humphrey, A. Hoos and J. D. Wolchok (2011). "Ipilimumab plus dacarbazine for previously untreated metastatic melanoma." *N Engl J Med* 364(26): 2517-2526.
- 5 Rody, A., U. Holtrich, L. Pusztai, C. Liedtke, R. Gaetje, E. Ruckhaeberle, C. Solbach, L. Hanker, A. Ahr, D. Metzler, K. Engels, T. Karn and M. Kaufmann (2009). "T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers." *Breast Cancer Res* 11(2): R15.
- 10 Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hanker L, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast cancer research : BCR*. 2011;13:R97
- Salgado, R. et al., (2014) "Harmonization of the evaluation of tumour infiltration lymphocytes (TILs) in breast cancer: recommendations by an international TILs-Working Group 2014. *Annals of Oncology*
- 15 Salvi, S., V. Fontana, S. Boccardo, D. F. Merlo, E. Margallo, S. Laurent, A. Morabito, E. Rijavec, M. G. Dal Bello, M. Mora, G. B. Ratto, F. Grossi, M. Truini and M. P. Pistillo (2012). "Evaluation of CTLA-4 expression and relevance as a novel prognostic factor in patients with non-small cell lung cancer." *Cancer Immunol Immunother* 61(9): 1463-1472.
- Schalper, K. A., V. Velcheti, D. Carvajal, H. Wimberly, J. Brown, L. Pusztai and D. L. Rimm (2014). "In Situ Tumor PD-L1 mRNA Expression Is Associated with Increased TILs and Better Outcome in Breast Carcinomas." *Clin Cancer Res*.
- 20 Spanos WC, Nowicki P, Lee DW, Hoover A, Hostager B, Gupta A, et al. Immune response during therapy with cisplatin or radiation for human papillomavirus-related head and neck cancer. *Archives of otolaryngology--head & neck surgery*. 2009;135:1137-46.
- 25 Sjö Dahl G, Lövgren K, Lauss M, Chebil G, Patschan O, Gudjonsson S, et al. Infiltration of CD3+ and CD68+ cells in bladder cancer is subtype specific and affects the outcome of patients with muscle-invasive tumors. *Urologic Oncology: Seminars and Original Investigations*. 2014;32:791-7.
- Sorbye SW, Kilvaer T, Valkov A, Donnem T, Smeland E, Al-Shibli K, et al. Prognostic impact of lymphocytes in soft tissue sarcomas. *PloS one*. 2011;6:e14611.
- 30 Stagg, J. and B. Allard (2013). "Immunotherapeutic approaches in triple-negative breast cancer: latest research and clinical prospects." *Ther Adv Med Oncol* 5(3): 169-181.
- Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." *Proc Natl Acad Sci U S A* 102(43): 15545-15550.
- 35 Suzuki K, Kadota K, Sima CS, Nitadori J, Rusch VW, Travis WD, et al. Clinical impact of immune microenvironment in stage I lung adenocarcinoma: tumor interleukin-12 receptor beta2 (IL-12Rbeta2), IL-7R, and stromal FoxP3/CD3 ratio are independent predictors of recurrence. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2013;31:490-8.
- 40 The Cancer Genome Atlas Network (2012). "Comprehensive molecular portraits of human breast tumours." *Nature* 490(7418): 61-70.
- Tjin EP, Krebbers G, Meijlink KJ, van de Kastele W, Rosenberg EH, Sanders J, et al. Immune-escape markers in relation to clinical outcome of advanced melanoma patients following immunotherapy. *Cancer immunology research*. 2014;2:538-46.
- 45 Thompson RH, Dong H, Lohse CM, Leibovich BC, Blute ML, Cheville JC, et al. PD-1 is expressed by tumor-infiltrating immune cells and is associated with poor outcome for patients with renal cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2007;13:1757-61.
- Townsend KN, Spowart JE, Huwait H, Eshragh S, West NR, Elrick MA, et al. Markers of T cell infiltration and function associate with favorable outcome in vascularized high-grade serous ovarian carcinoma. *PloS one*. 2013;8:e82406.
- 50

- Ueno T, Toi M, Saji H, Muta M, Bando H, Kuroi K, et al. Significance of macrophage chemoattractant protein-1 in macrophage recruitment, angiogenesis, and survival in human breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2000;6:3282-9.
- 5 Vauleon E, Tony A, Hamlat A, Etcheverry A, Chiforeanu DC, Menei P, et al. Immune genes are associated with human glioblastoma pathology and patient survival. *BMC medical genomics*. 2012;5:41.
- Villegas FR, Coca S, Villarrubia VG, Jimenez R, Chillon MJ, Jareno J, et al. Prognostic significance of tumor infiltrating natural killer cells subset CD57 in patients with squamous cell lung cancer. *Lung cancer (Amsterdam, Netherlands)*. 2002;35:23-8.
- 10 Webster WS, Lohse CM, Thompson RH, Dong H, Frigola X, Dicks DL, et al. Mononuclear cell infiltration in clear-cell renal cell carcinoma independently predicts patient survival. *Cancer*. 2006;107:46-53.
- Welsh TJ, Green RH, Richardson D, Waller DA, O'Byrne KJ, Bradding P. Macrophage and mast-cell invasion of tumor cell islets confers a marked survival advantage in non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2005;23:8959-67.
- 15 Witkin, A. P. "Scale-space filtering", *Proc. 8th Int. Joint Conf. Art. Intell., Karlsruhe, Germany*, 1019-1022, 1983.
- Wu G, Han BI Fau - Pei X-t, Pei XT. Immune response of dendritic cells acquiring antigens from apoptotic cholangiocarcinoma cells induced by mitomycin.
- 20 Yang I, Tihan T, Han SJ, Wrensch MR, Wiencke J, Sughrue ME, et al. CD8+ T-cell infiltrate in newly diagnosed glioblastoma is associated with long-term survival. *Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia*. 2010;17:1381-5.
- Yong AS, Stephens N, Weber G, Li Y, Savani BN, Eniafe R, et al. Improved outcome following allogeneic stem cell transplantation in chronic myeloid leukemia is associated with higher expression of BMI-1 and immune responses to BMI-1 protein. *Leukemia*. 2011;25:629-37.
- 25 Yuan, Y., H. Failmezger, O. M. Rueda, H. R. Ali, S. Graf, S.-F. Chin, R. F. Schwarz, C. Curtis, M. J. Dunning, H. Bardwell, N. Johnson, S. Doyle, G. Turashvili, E. Provenzano, S. Aparicio, C. Caldas and F. Markowitz (2012). "Quantitative Image Analysis of Cellular Heterogeneity in Breast Tumors Complements Genomic Profiling." *Science Translational Medicine* 4(157): 157ra143.
- 30 Zhang Y, Wang L, Liu Y, Wang S, Shang P, Gao Y, et al. Preoperative neutrophil-lymphocyte ratio before platelet-lymphocyte ratio predicts clinical outcome in patients with cervical cancer treated with initial radical surgery. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society*. 2014;24:1319-25.
- 35

Claims:

1. A method of measuring immune infiltration in a tumour, the method comprising:
providing an image of the tumour in which lymphocytes and cancer cells have been
5 identified;
obtaining a lymphocyte-to-cancer measurement for each lymphocyte;
classifying a subset of the lymphocytes as intra-tumour lymphocytes according to
their lymphocyte-to-cancer measurement;
quantifying the intra-tumour lymphocytes and the cancer cells in the tumour image;
10 calculating the intra-tumour lymphocyte ratio (ITLR) as the ratio of intra-tumour
lymphocytes to cancer cells, wherein the ITLR is a measurement of immune infiltration in the
tumour.
2. The method according to claim 1, wherein the step of obtaining a lymphocyte-to-
15 cancer measurement for each lymphocyte comprises:
applying a density estimate to obtain a model of the cancer cell density; and
determining the proximity of each lymphocyte to cancer cell density.
3. The method according to claim 1 or claim 2, wherein a lymphocyte is classified as an
20 intra-tumour lymphocyte if its lymphocyte-to-cancer measurement is above a threshold
value.
4. A method of determining an ITLR cut-off value for a cancer type or subtype, for use
in providing a prognosis in a cancer patient having that cancer type or subtype, the method
25 comprising:
measuring immune infiltration in a tumour from each member of a cohort of cancer
patients having the cancer type or subtype, according to the method of any one of claims 1
to 3, thereby calculating the ITLR for each tumour;
relating the ITLR for each tumour to the clinical outcome of each cancer patient in the
30 cohort of cancer patients; and
selecting a cut-off value for ITLR, wherein an ITLR equal to or below the cut-off value
is associated with a significantly worse clinical outcome in the cohort of cancer patients than
an ITLR above the cut-off value.
- 35 5. A method of providing a prognosis in a cancer patient, the method comprising
measuring immune infiltration in a tumour from the cancer patient according to the
method of one of claims 1 to 3, thereby calculating the ITLR for the tumour,
wherein an ITLR below a predetermined ITLR cut-off value indicates a poor
40 prognosis.

6. The method according to any one of the preceding claims, further comprising a step of identifying the lymphocytes and cancer cells in an image of the tumour by automated image analysis to provide an image of the tumour in which lymphocytes and cancer cells have been identified.

5

7. The method according to any one of the preceding claims wherein the cancer patient has breast cancer, ovarian cancer, colorectal cancer, melanoma or non-small cell lung cancer.

10 8. The method of according to claim 7 wherein the cancer patient has breast cancer which is triple negative breast cancer.

9. The method of providing a prognosis in a cancer patient according to any one of claims 5 to 8, wherein the predetermined ITLR cut-off value is from about 0.005 to about
15 0.070;

10. The method of providing a prognosis in a cancer patient according to any one of claims 5 to 8, the method further comprising determining an ITLR cut-off value for the patient's cancer type or subtype by the method according to claim 4, wherein that ITLR
20 cut-off value is the predetermined ITLR cut-off value for providing a prognosis for the cancer patient.

11. A method for predicting whether a cancer patient will respond to a therapeutic regime, the method comprising
25 analysing a tumour image from the cancer patient according to the method of any one of claims 1 to 3, wherein an ITLR above a predetermined ITLR cut-off value indicates that the patient is likely to respond to the therapeutic regime.

12. A method of treating cancer in a cancer patient according to a therapeutic regime,
30 the method comprising
analysing a tumour image from the cancer patient according to the method of any one of claims 1 to 3, and
treating the cancer patient according to the therapeutic regime depending on whether the ITLR is below or above a predetermined cut-off value.

35

13. The method of providing a prognosis in a cancer patient according to any one of claims 5 to 8, the method of predicting whether a cancer patient will respond to a therapeutic regime according to claim 11, or the method of treating cancer according to claim 12,
40 wherein the method further comprises surgically resecting a tumour from the cancer patient and measuring immune filtration in the surgically resected tumour.

14. The method of treating cancer according to claim 12, wherein the cancer patient has triple negative breast cancer, wherein the therapeutic regime comprises administration of a CTLA4 antagonist, and wherein the cancer patient is treated according to the therapeutic regime if the ITLR is above a predetermined cut-off value.

5

15. The method according to any one of the preceding claims, wherein the image of a tumour is an image of a hematoxylin and eosin stained tumour section.

16. A method of determining a prognosis in a triple negative breast cancer patient, the method comprising, determining the level of expression of CTLA4 in a tumour sample obtained from the patient, wherein increased CTLA4 expression indicates a positive prognosis.

17. A method determining a prognosis in a triple negative breast cancer patient, the method comprising, determining the level of expression of APOBEC3G in a tumour sample obtained from the patient, wherein increased expression and/or APOBEC3G expression indicates a positive prognosis.

18. The method of determining a cut-off value for ITLR according to claim 4, wherein the cohort of cancer patients is a subset of patients from
(a) the METABRIC group; or
(b) the Cancer Genome Atlas group.

25

30

35

40

1/28

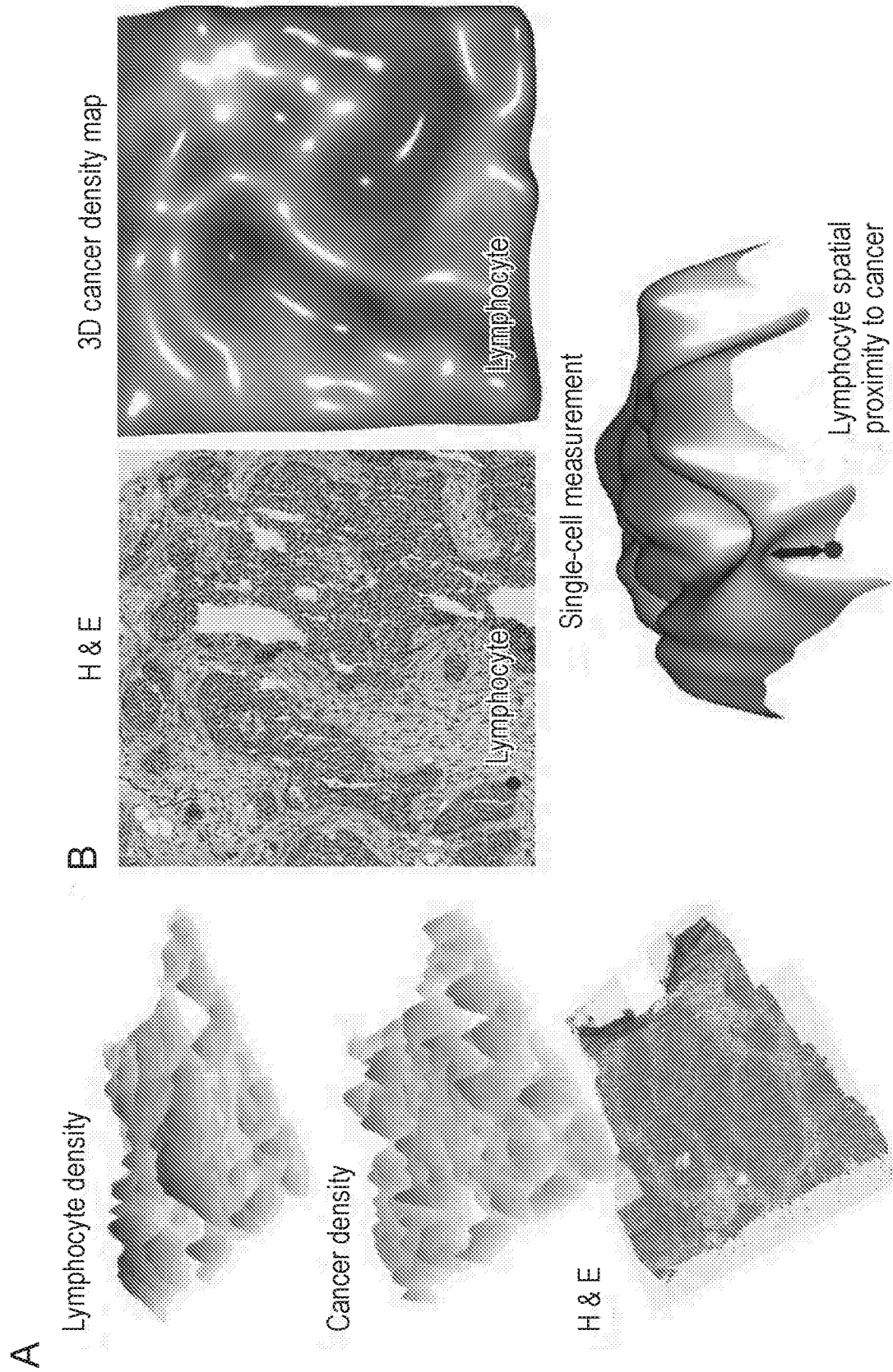
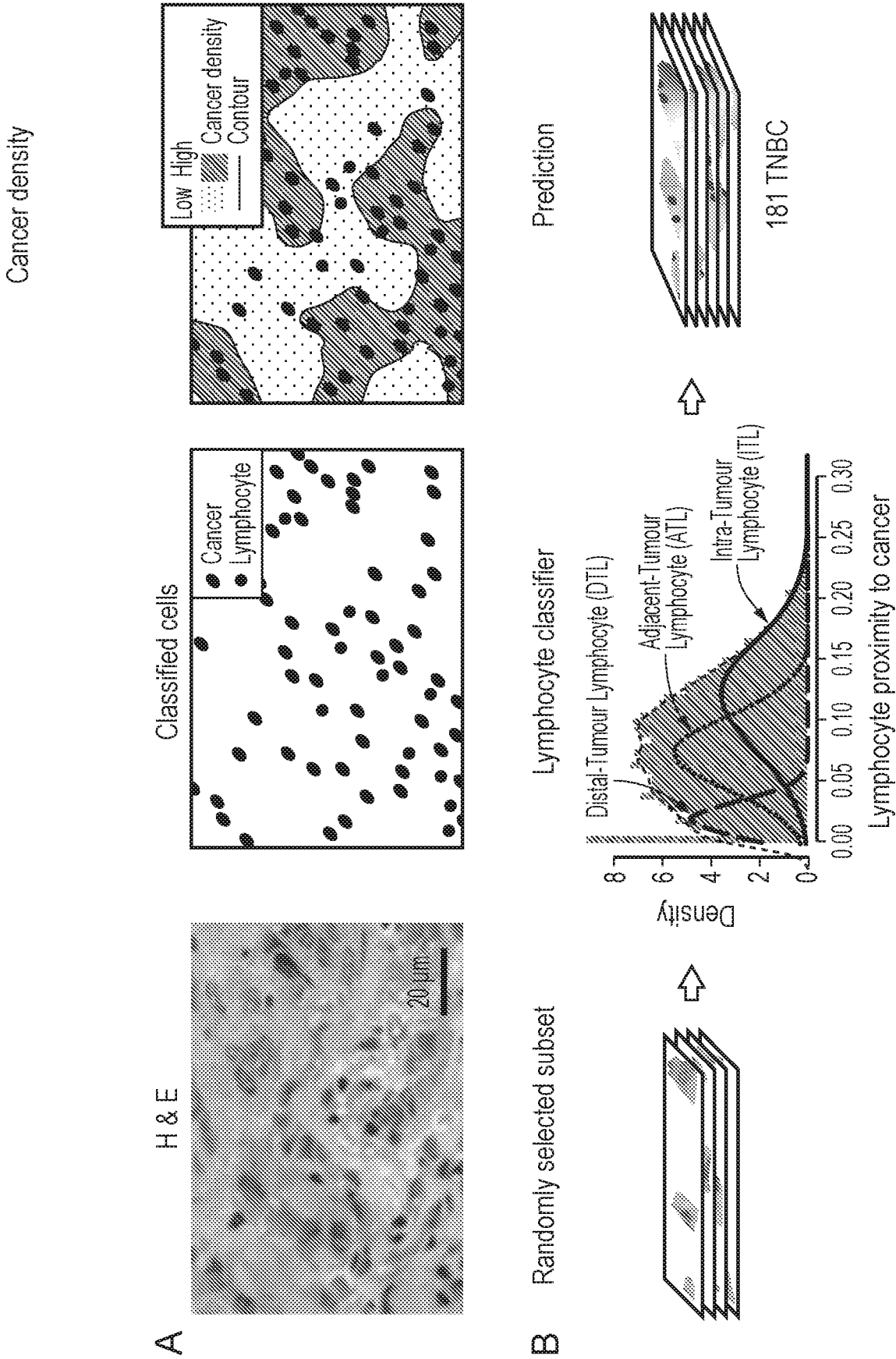


FIG. 1



3/28

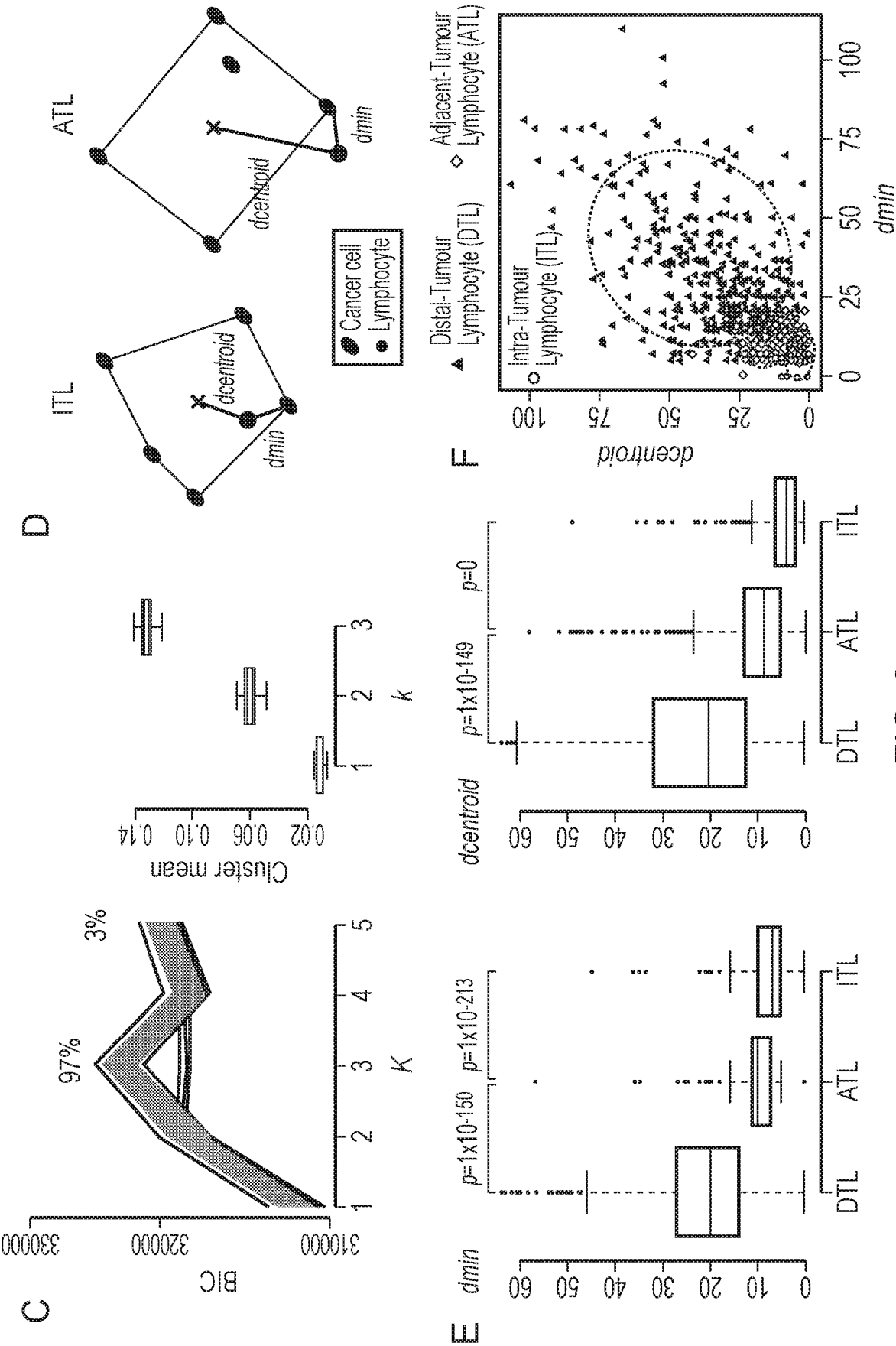
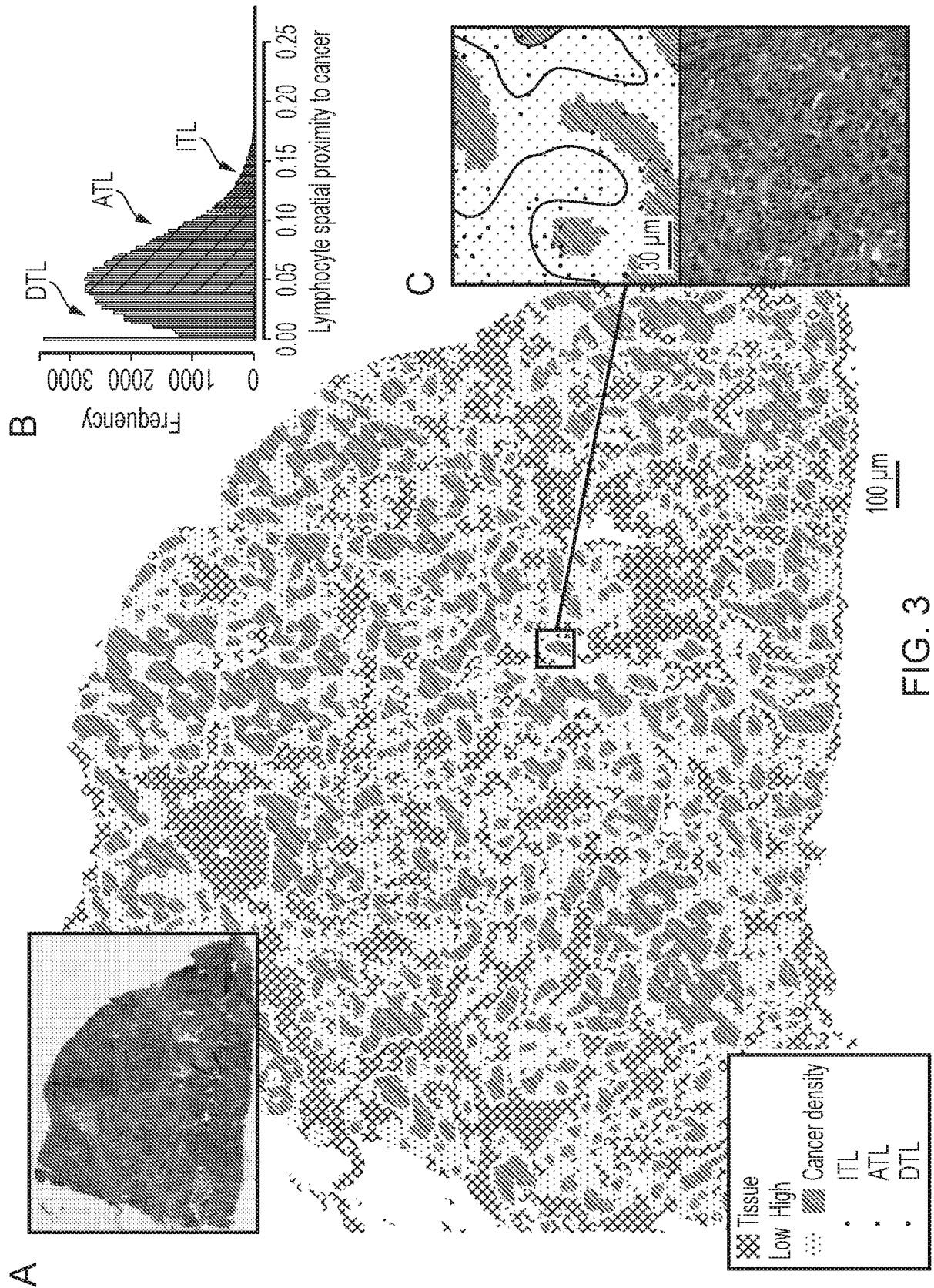


FIG. 2 (Continued)

4/28



5/28

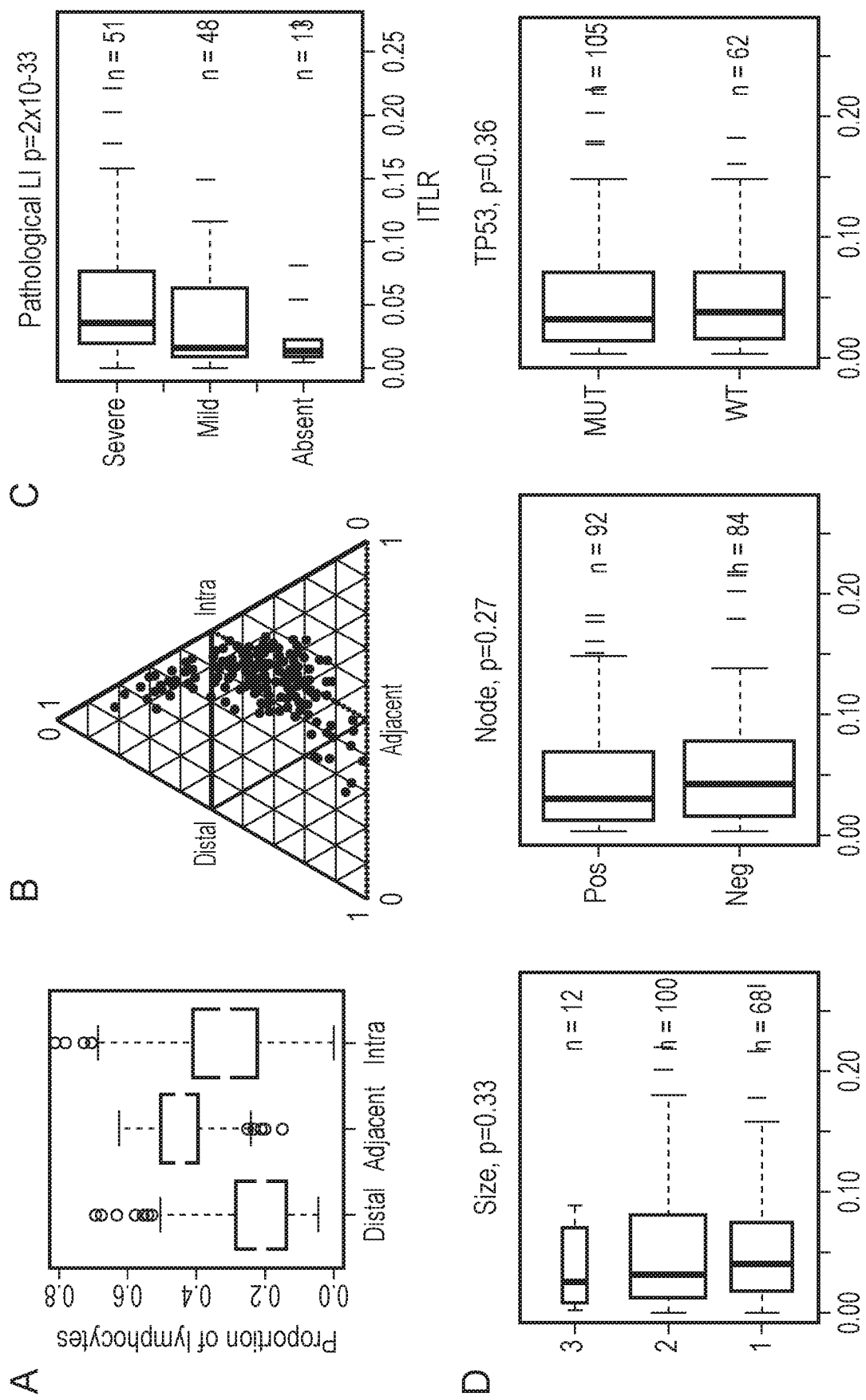


FIG. 4

6/28

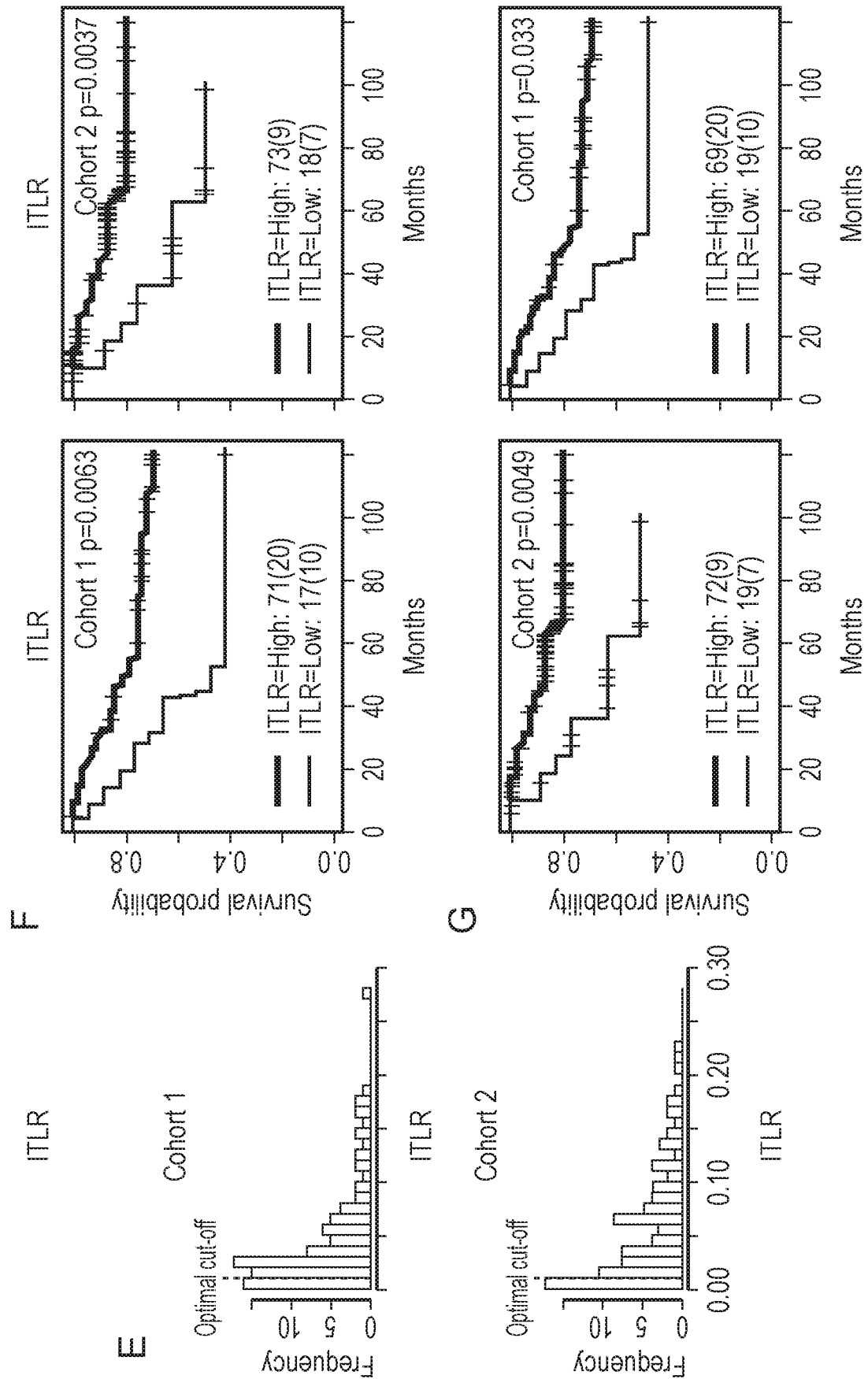


FIG. 4 (Continued)

7/28

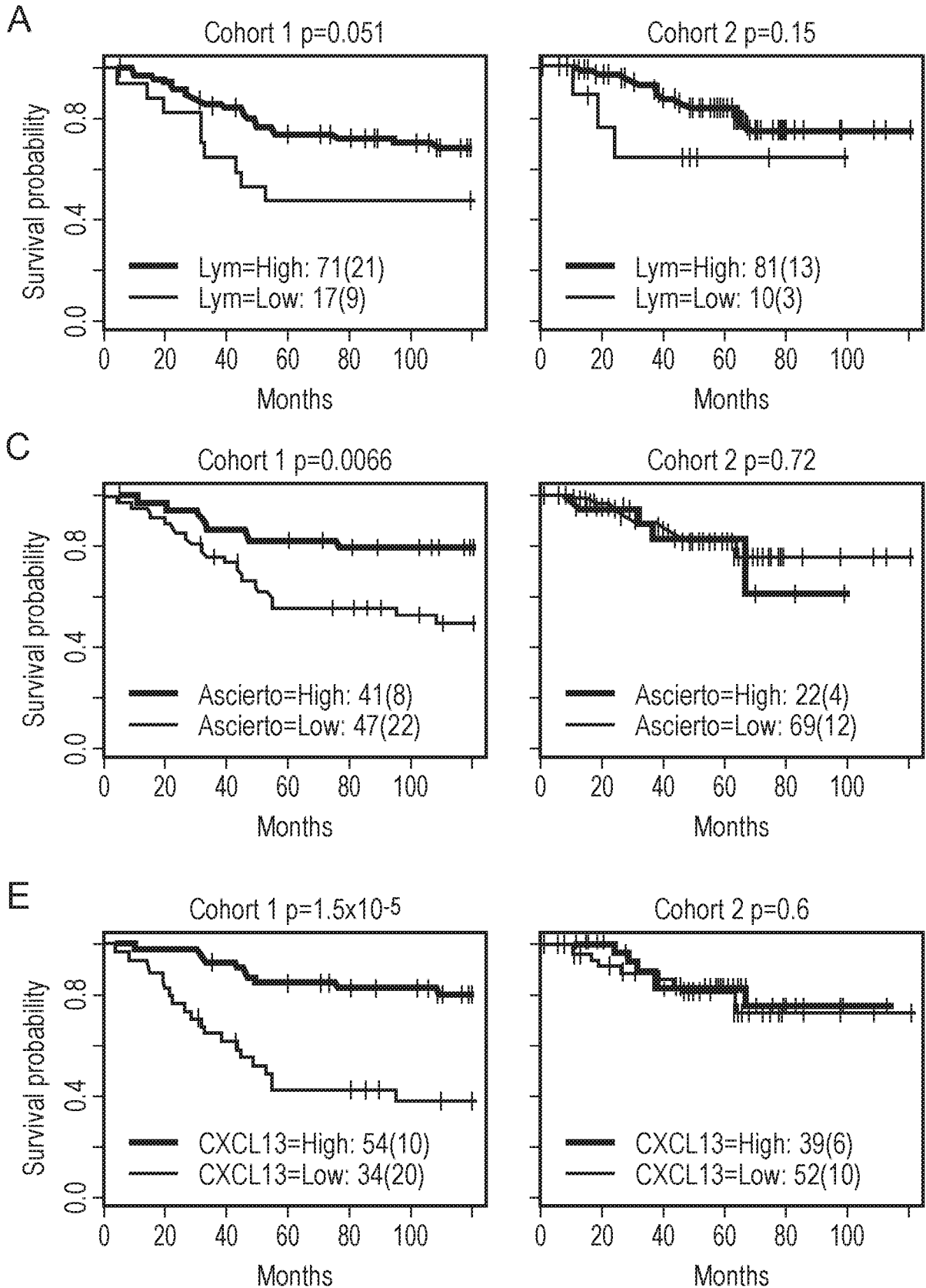


FIG. 5

8/28

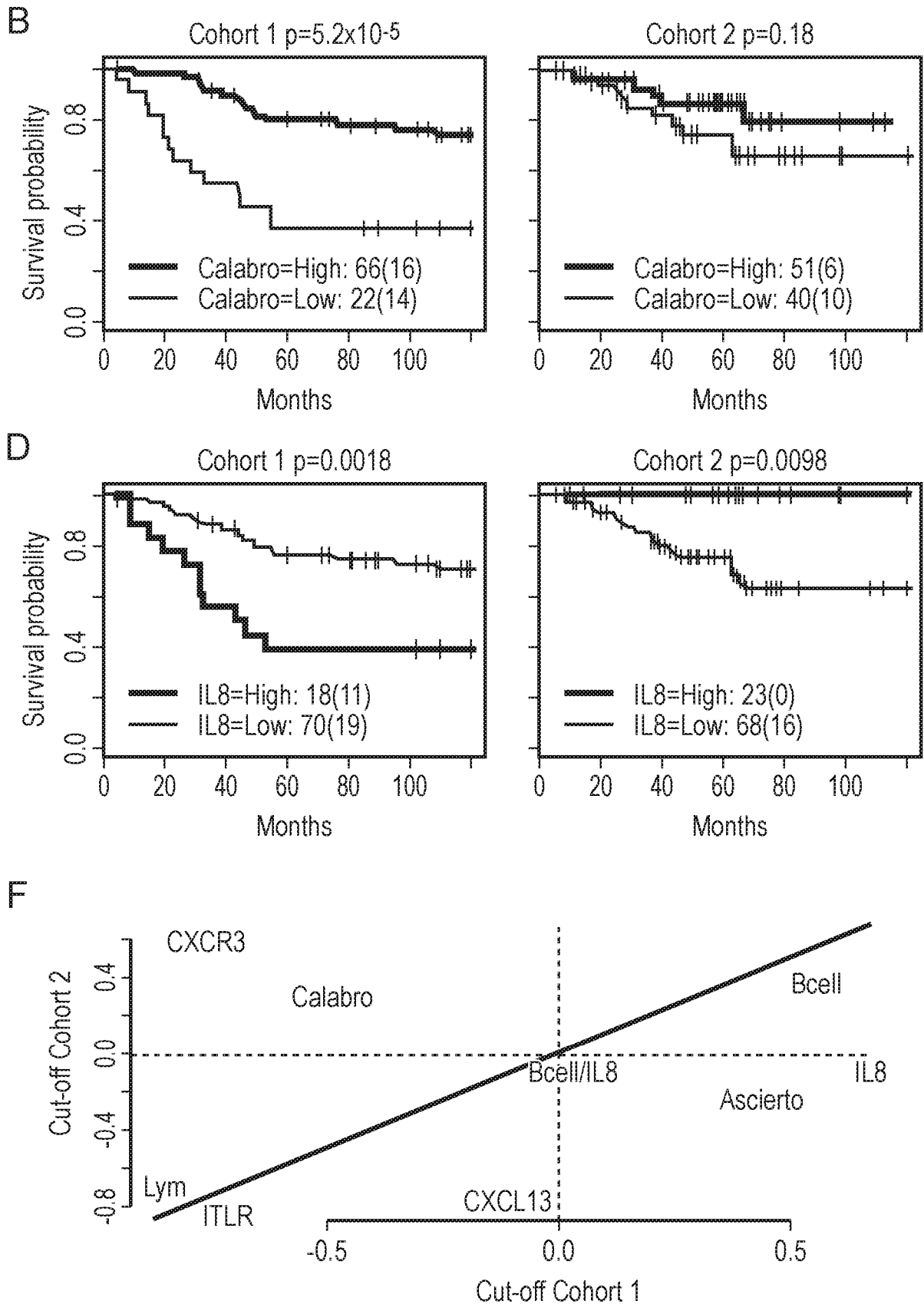
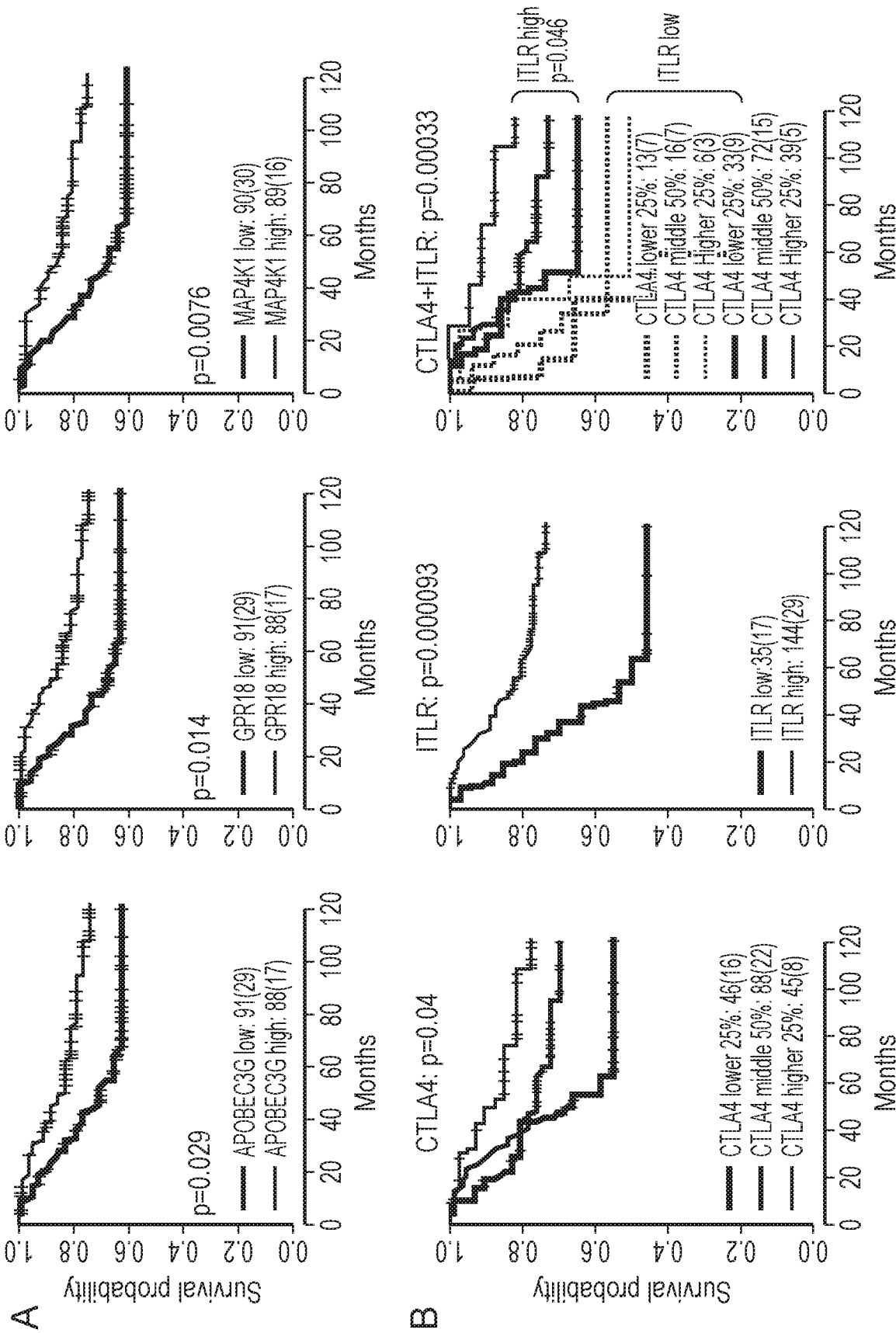


FIG. 5 (continued)

9/28



10/28

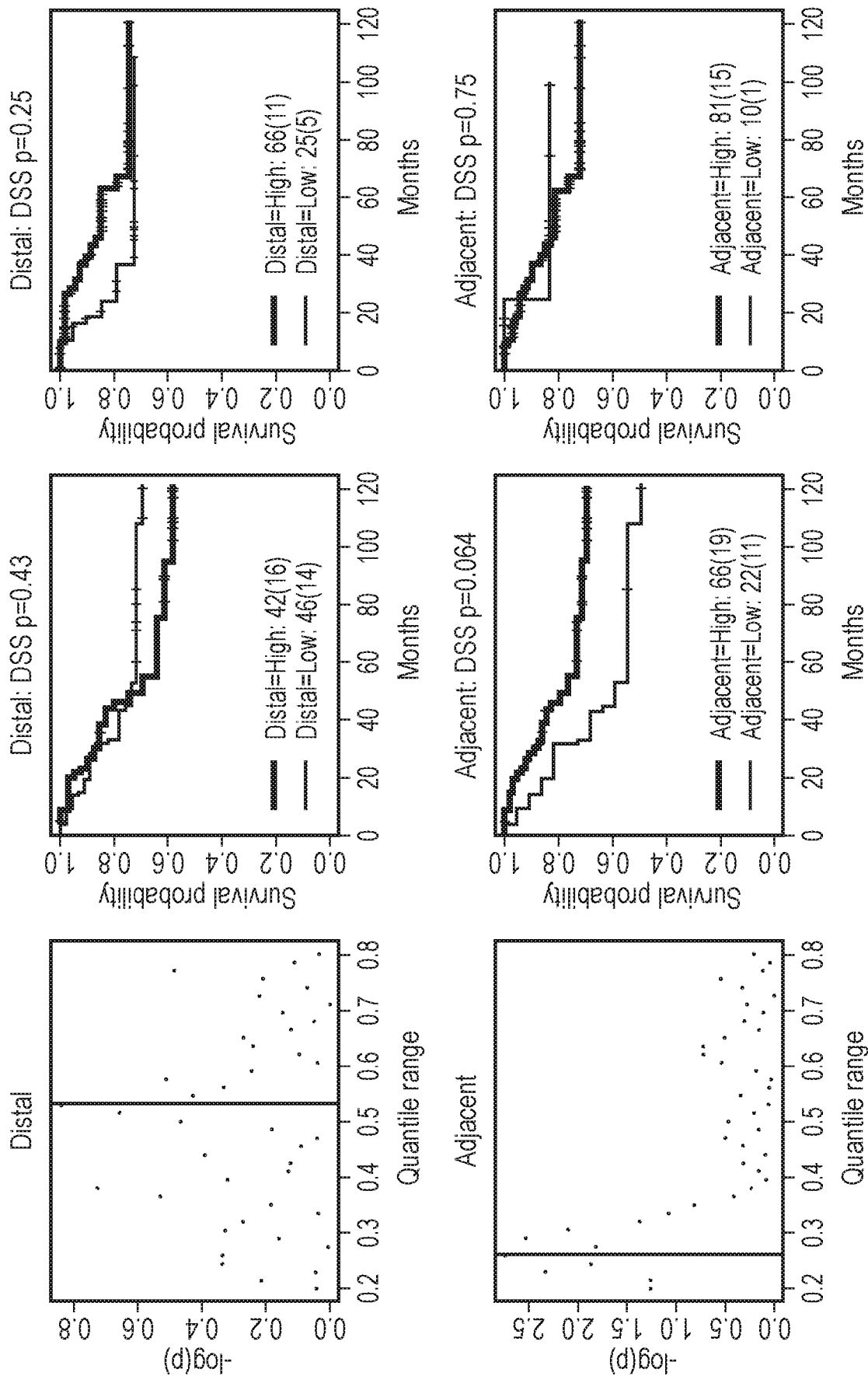


FIG. 7

11/28

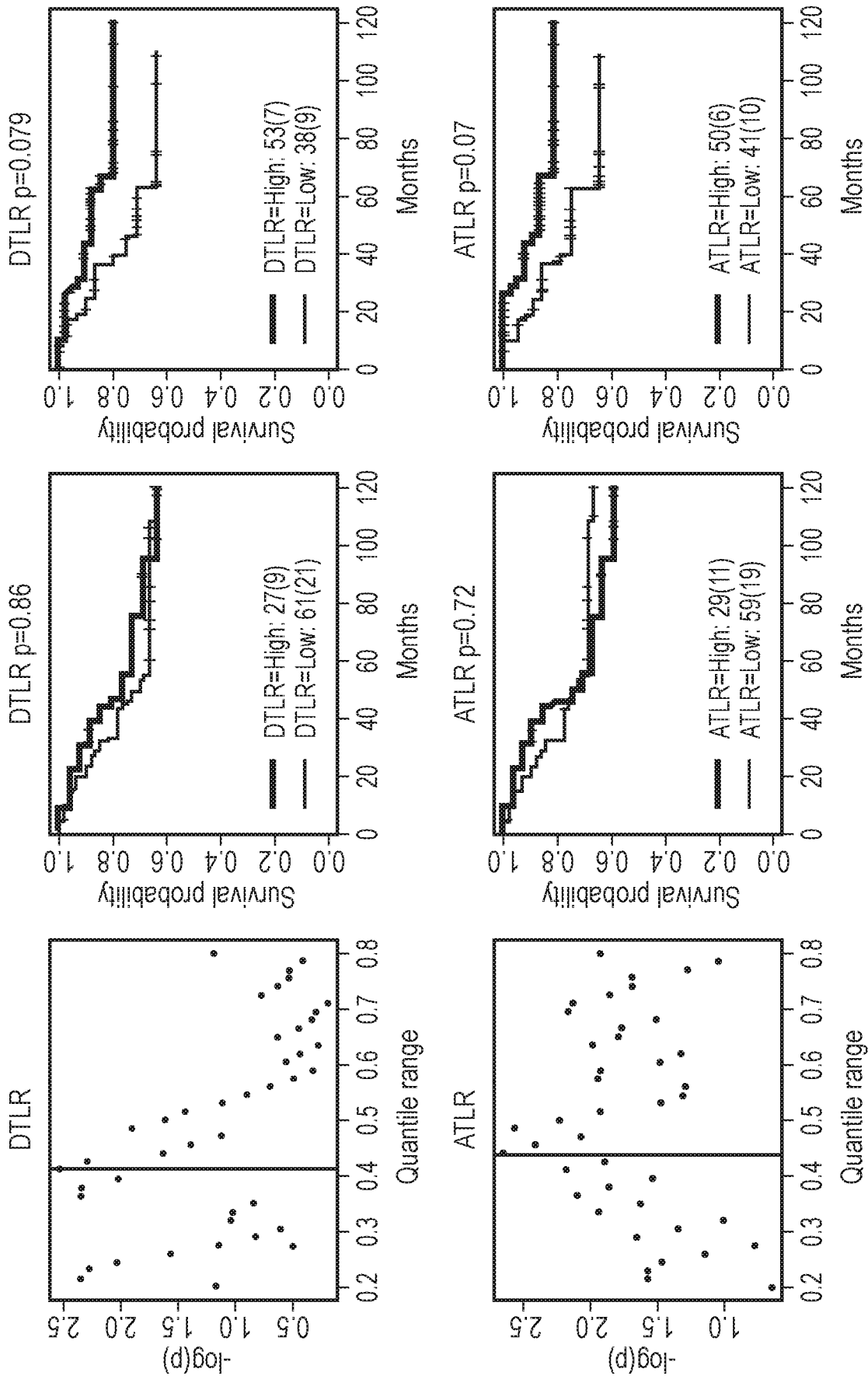
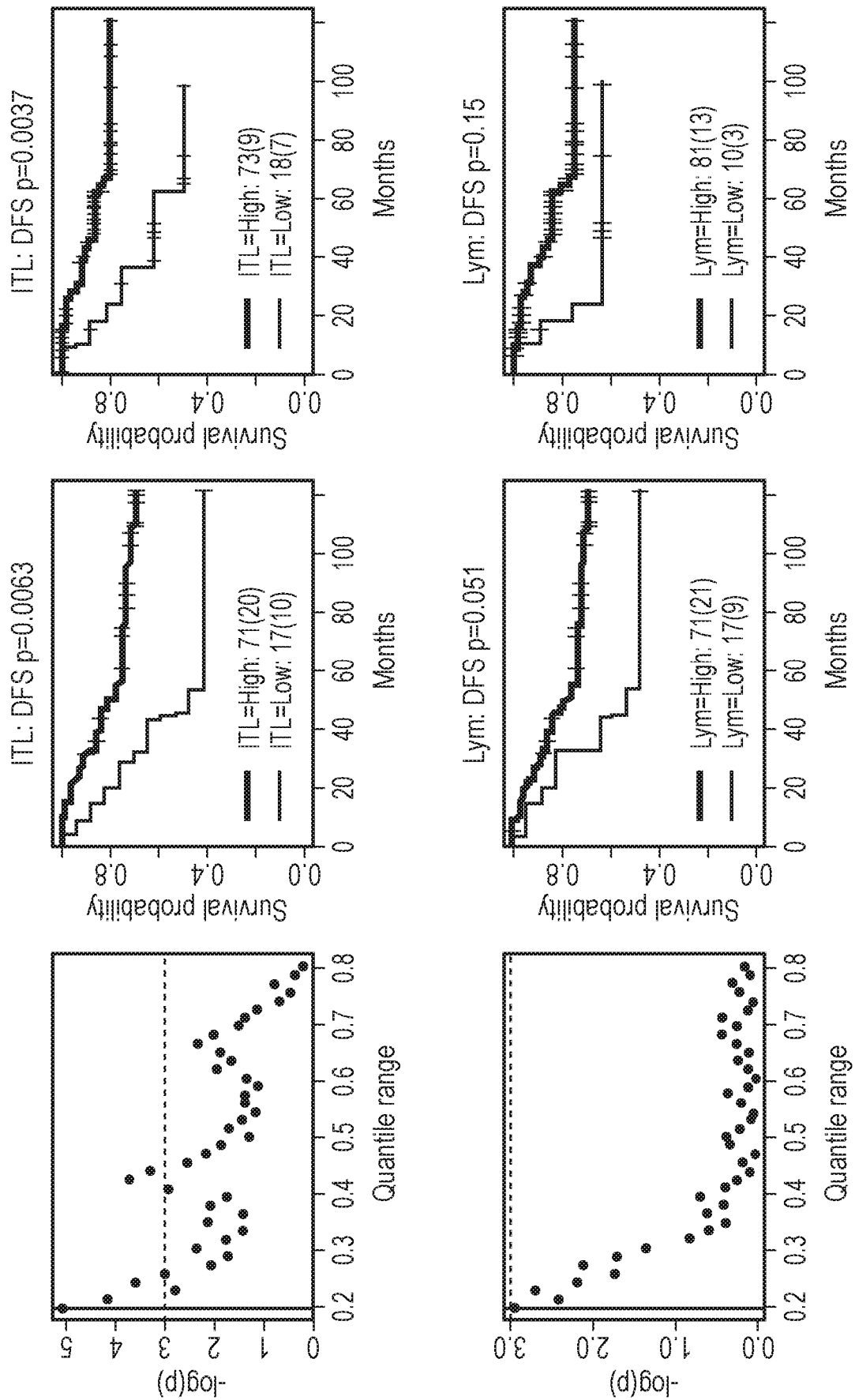


FIG. 8

12/28



௧௮௮௮

13/28

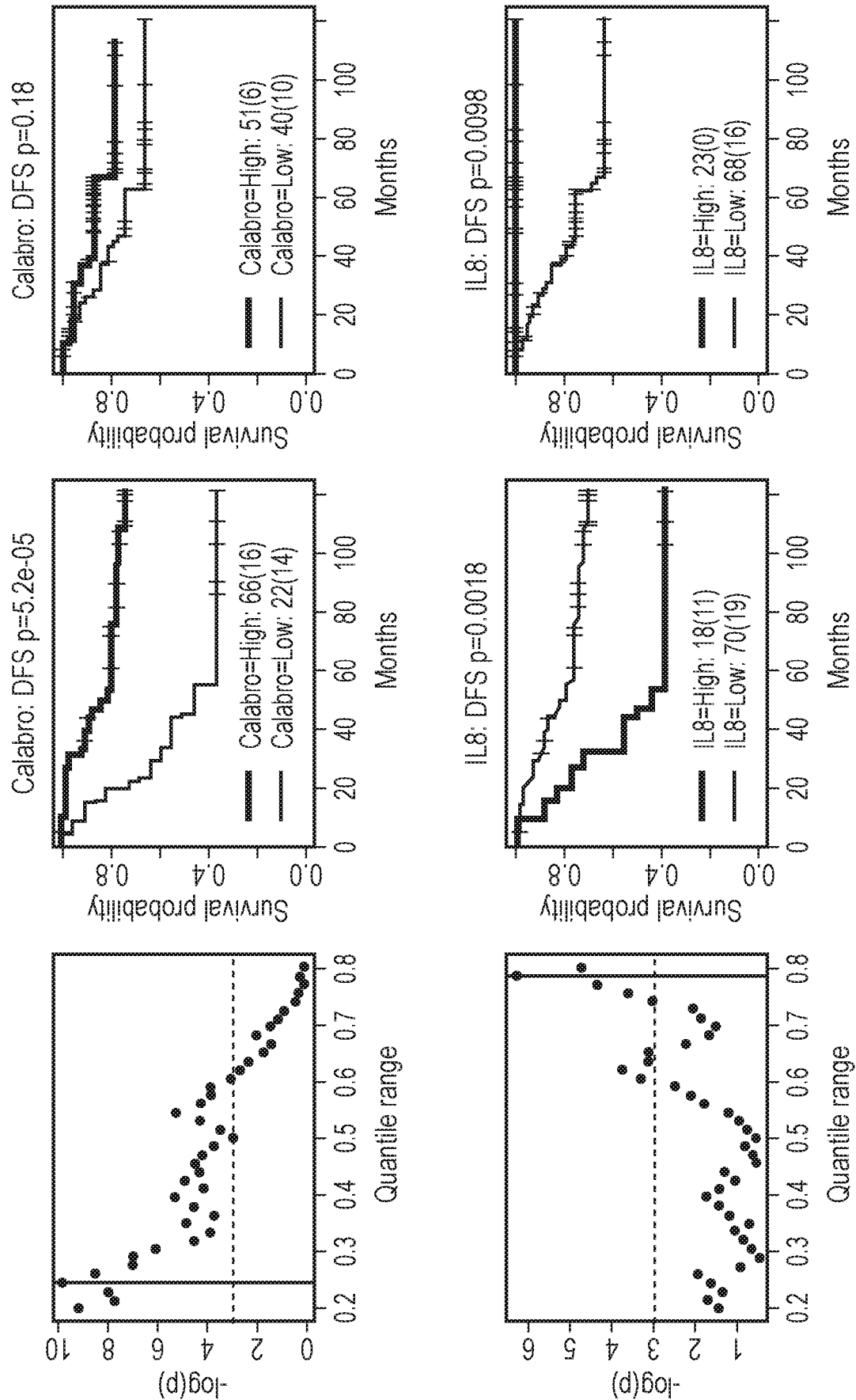


FIG. 9 (continued)

14/28

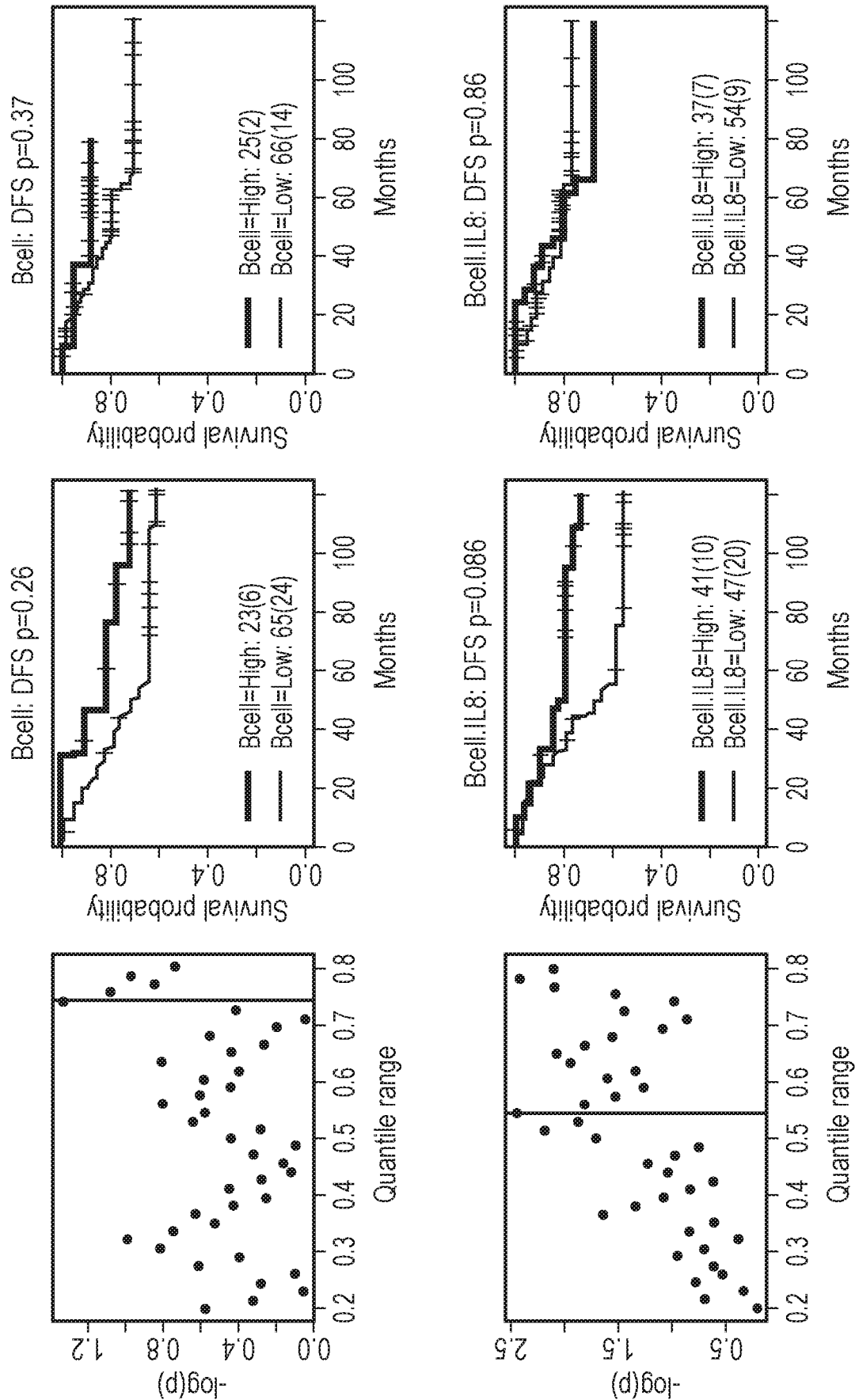


FIG. 9 (continued)

15/28

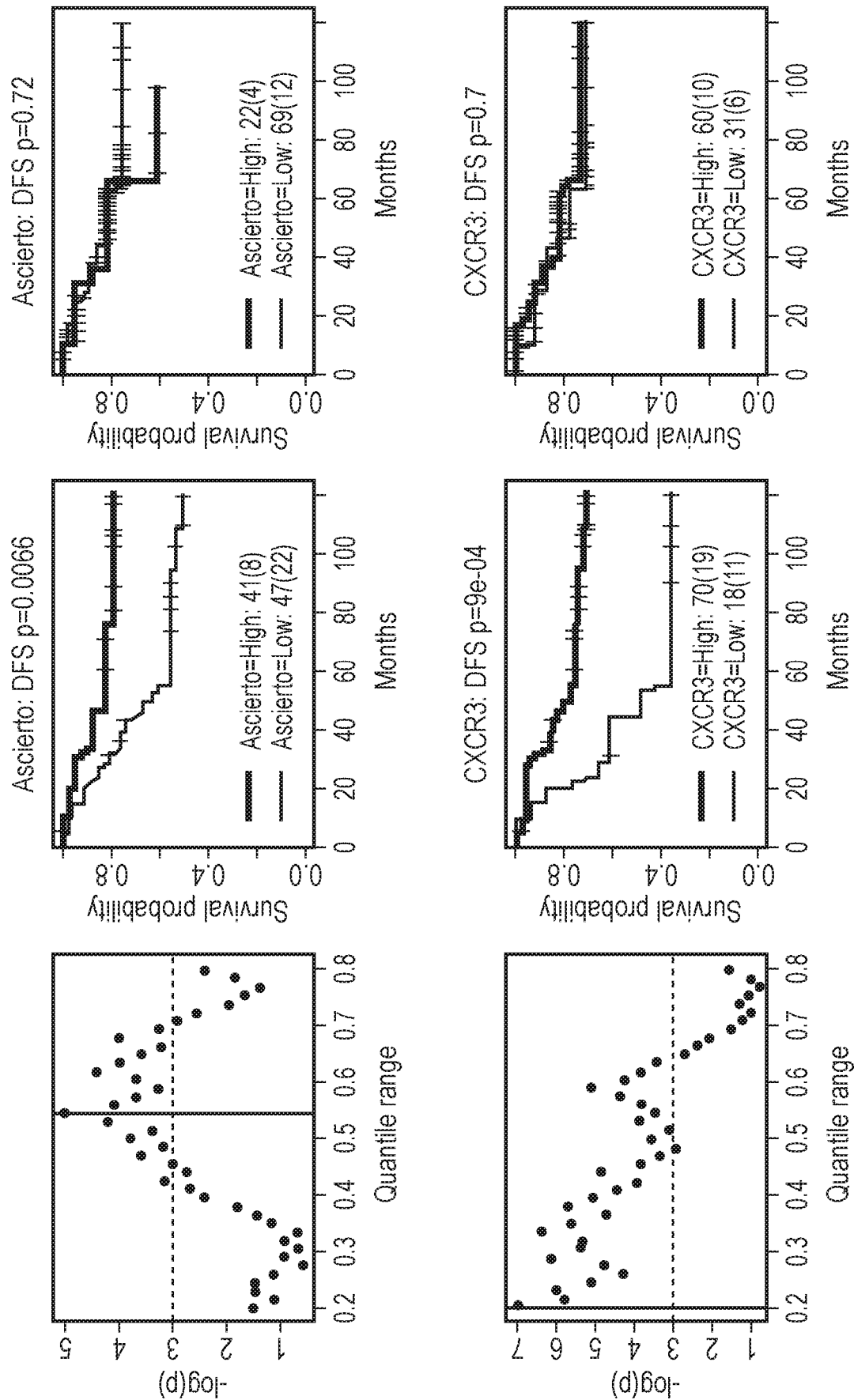


FIG. 9 (continued)

16/28

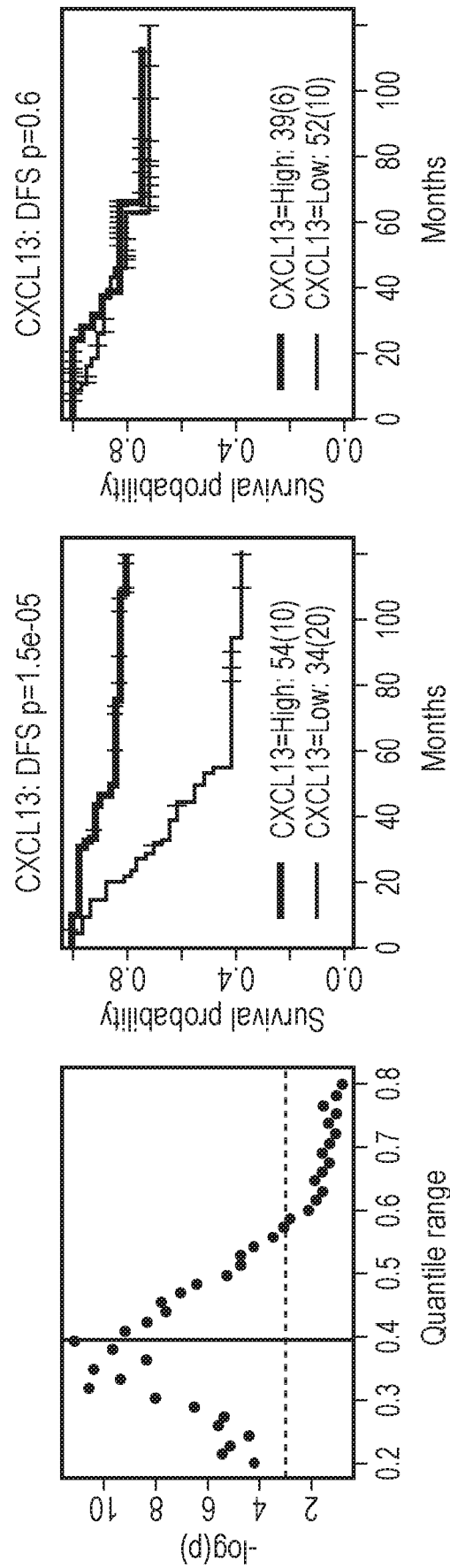


FIG. 9 (continued)

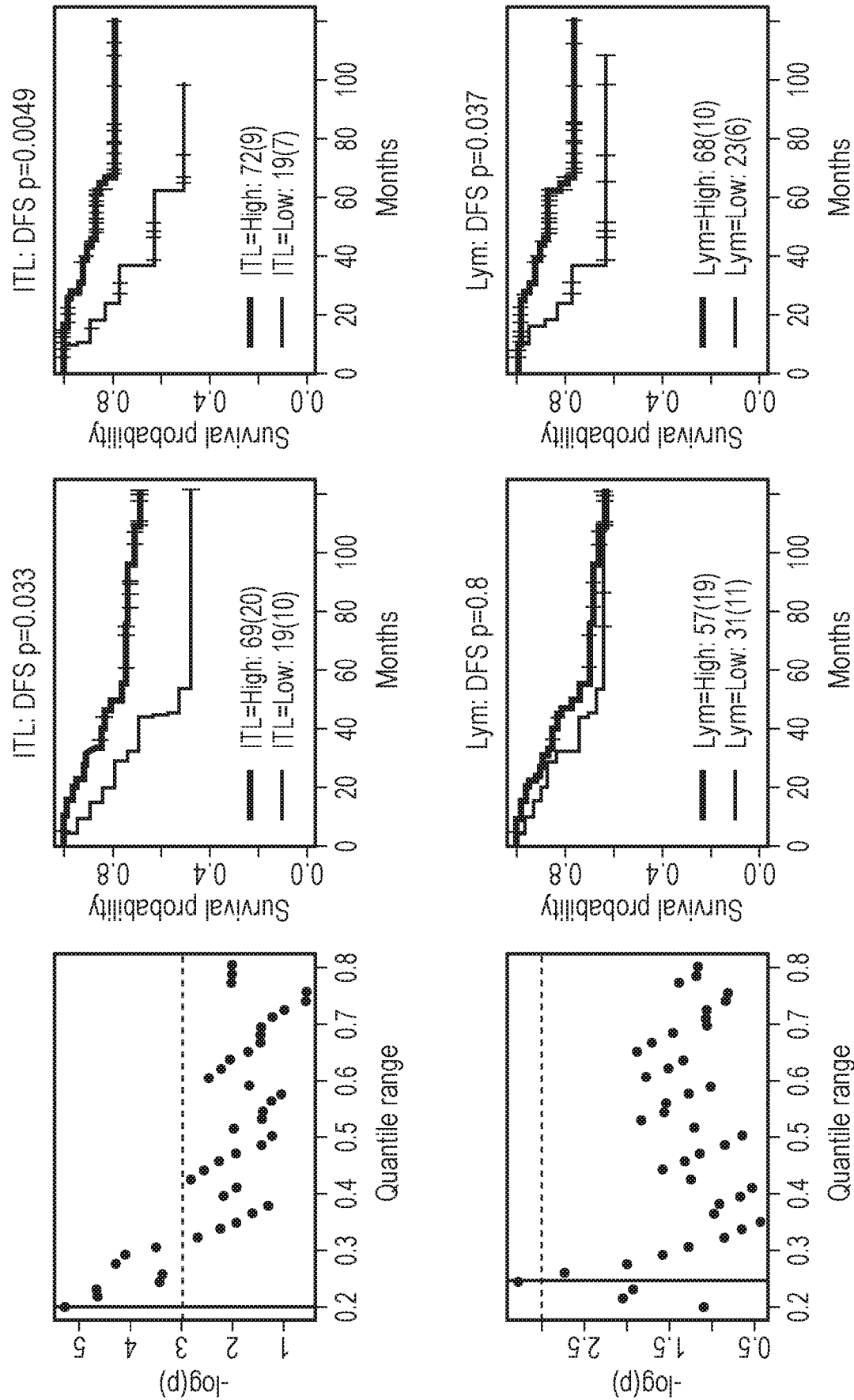


FIG. 10

18/28

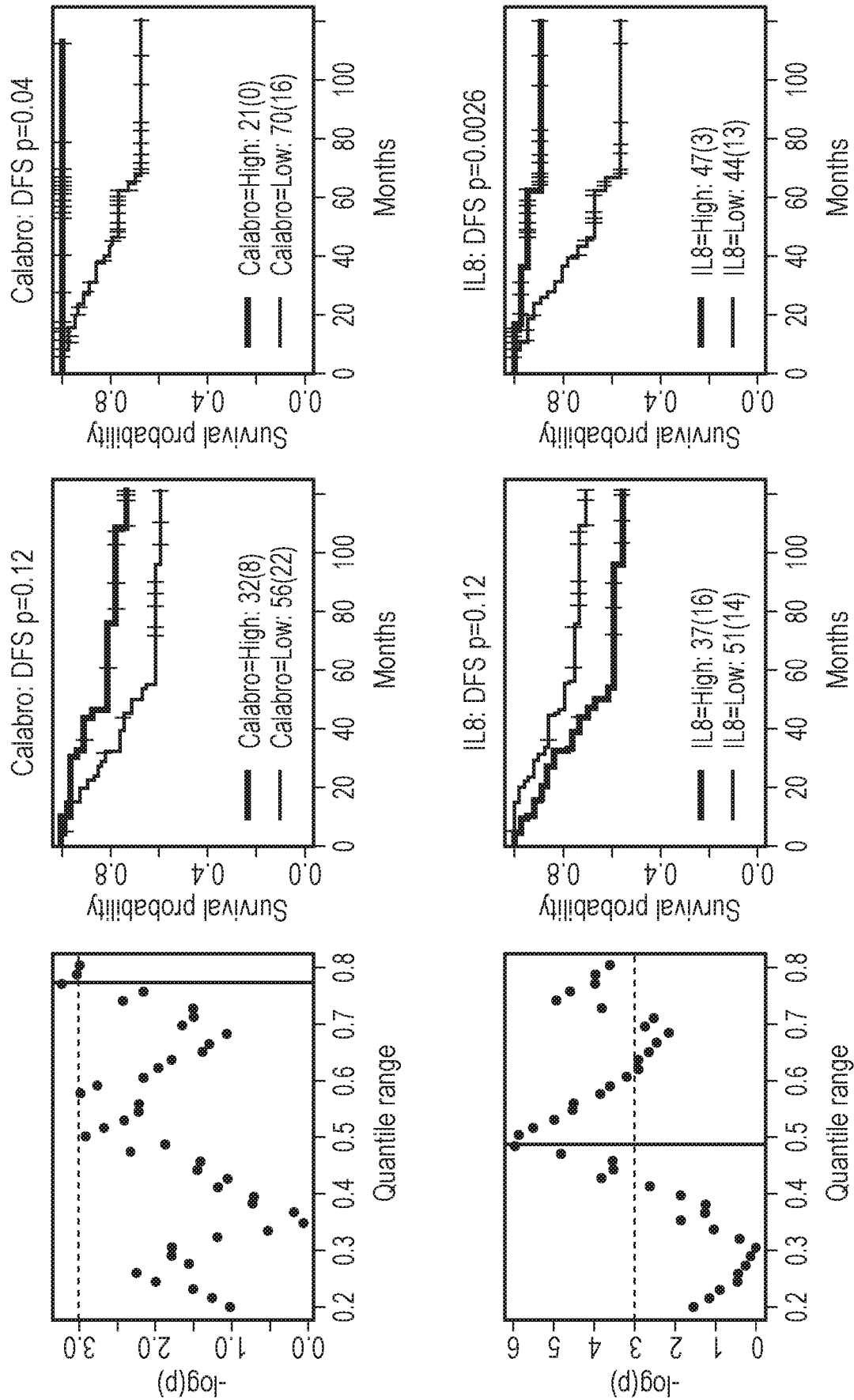


FIG. 10 (continued)

19/28

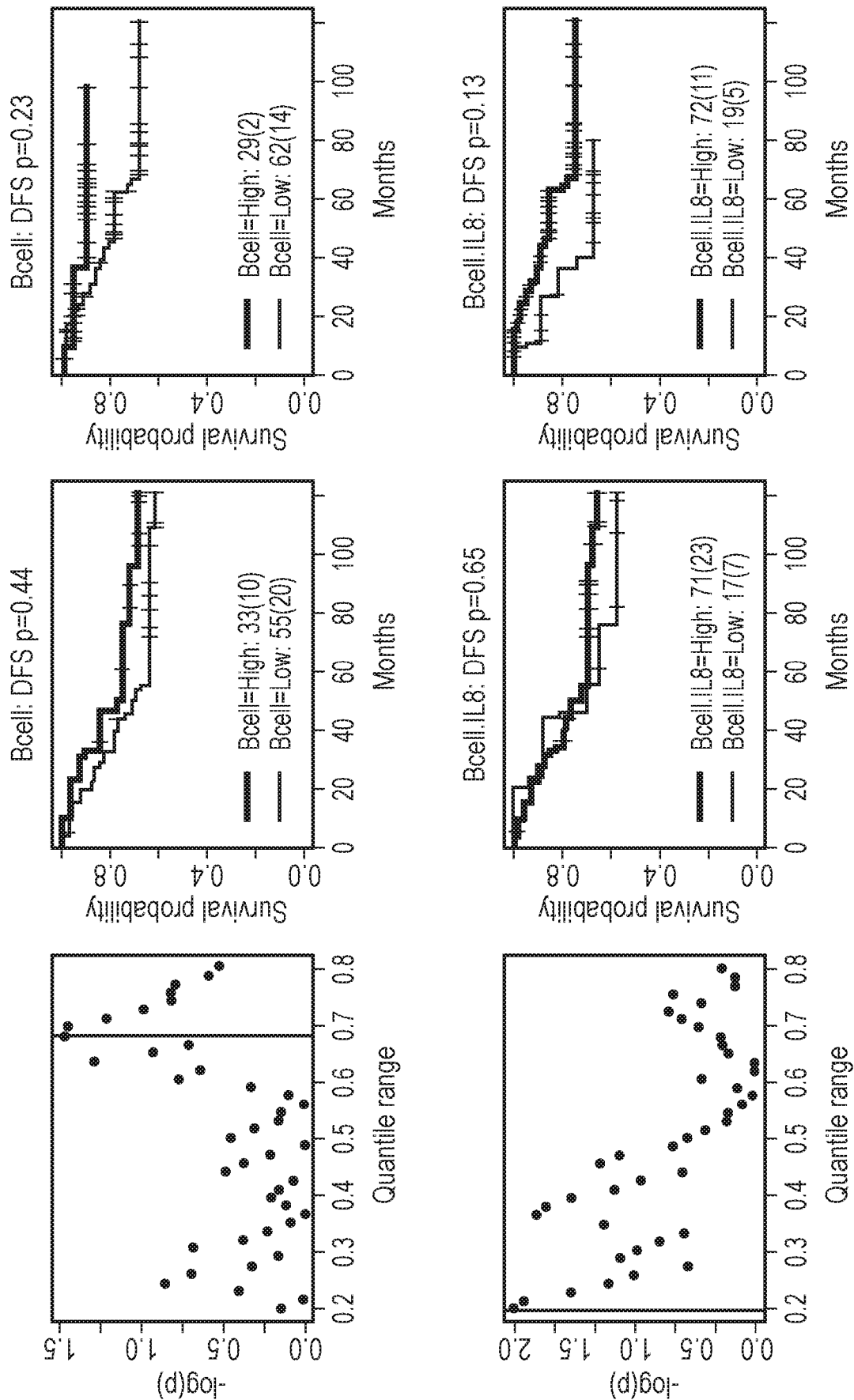


FIG. 10 (continued)

20/28

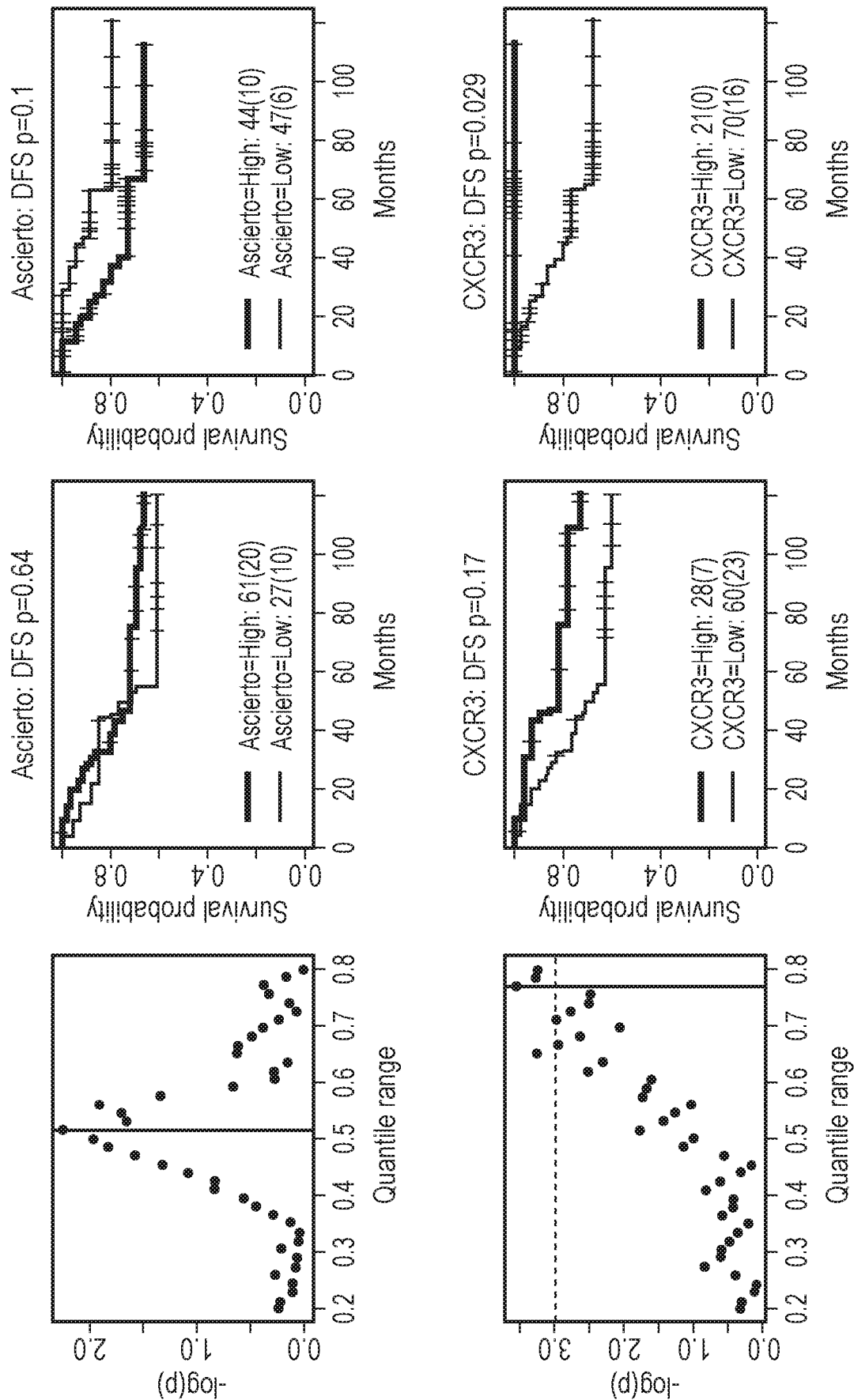


FIG. 10 (continued)

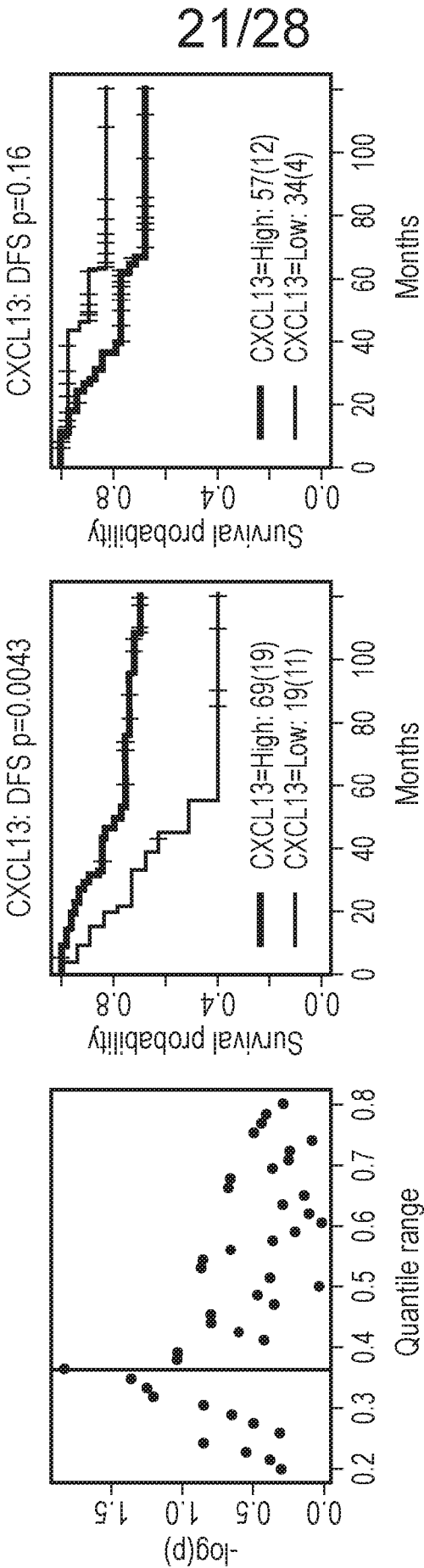


FIG. 10 (continued)

22/28

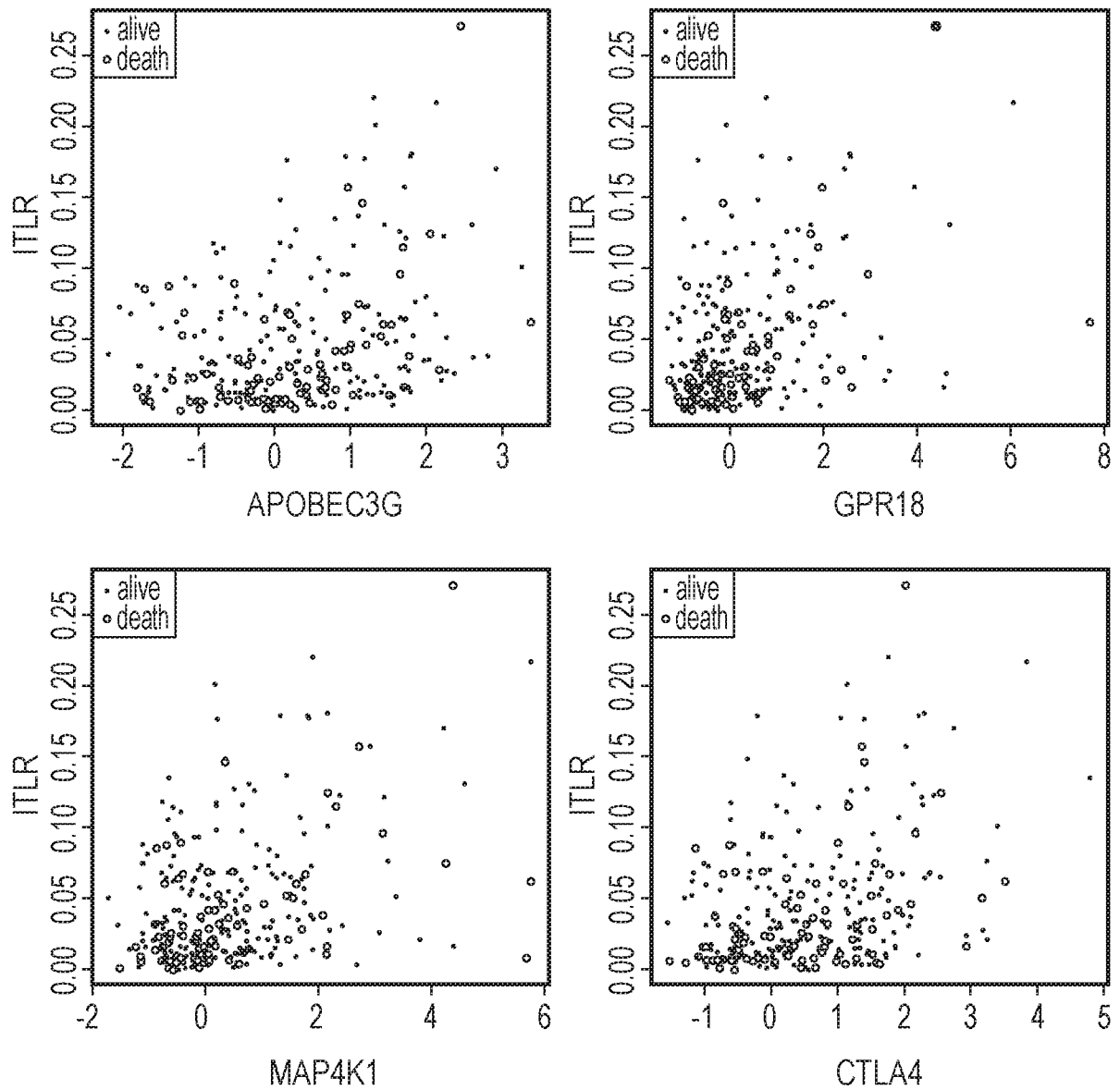


FIG. 11

23/28

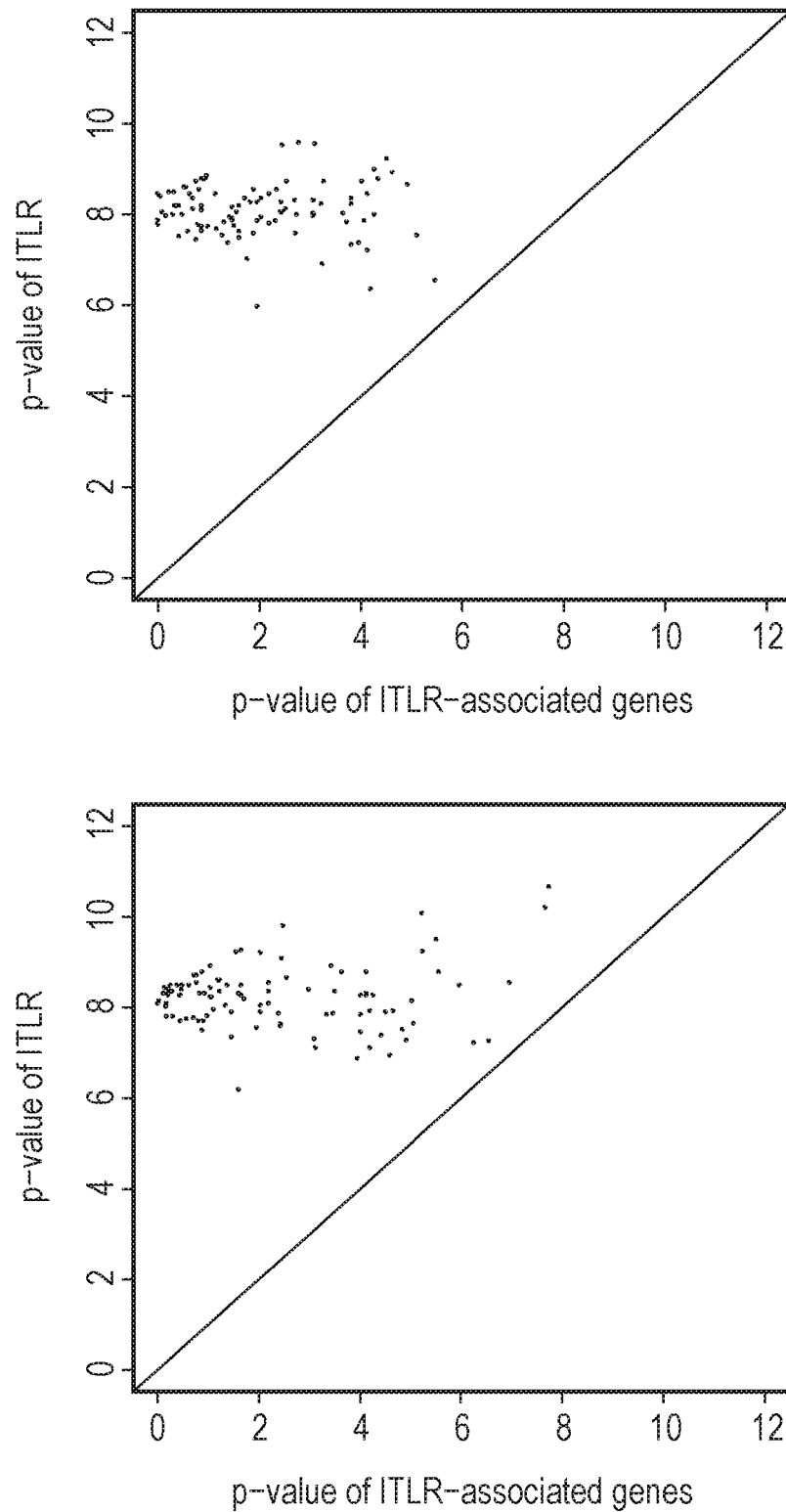


FIG. 12

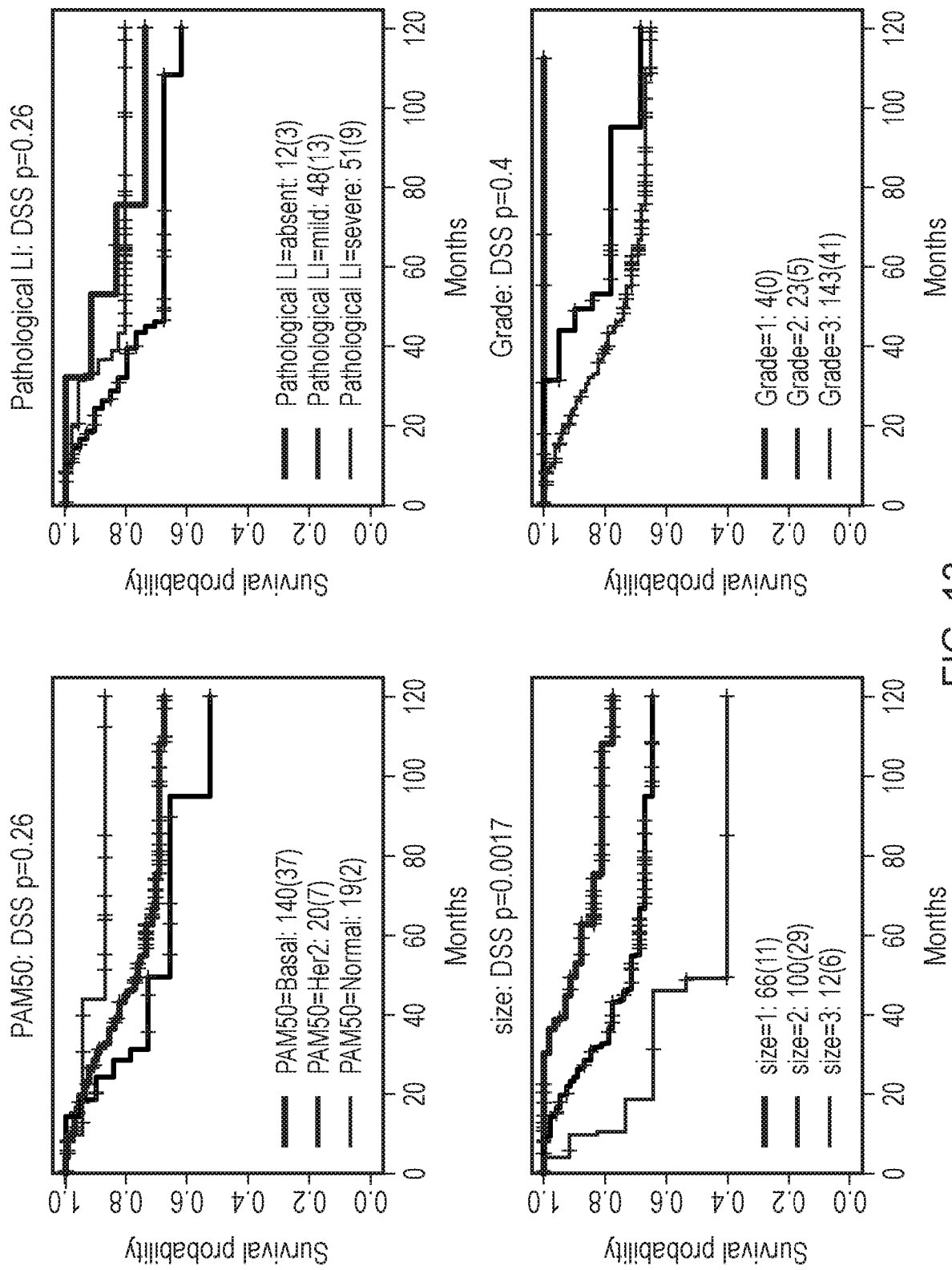


FIG. 13

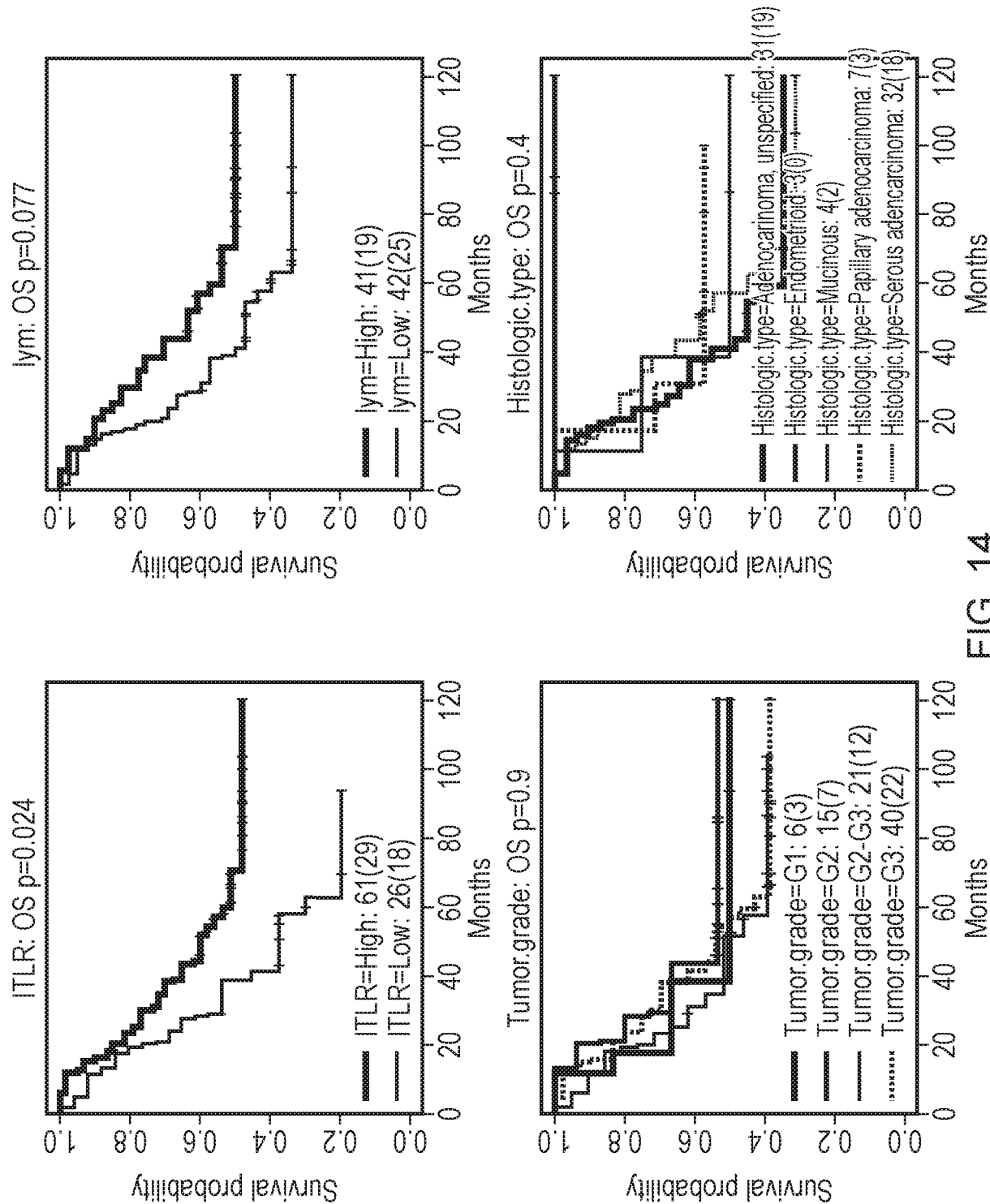


FIG. 14

26/28

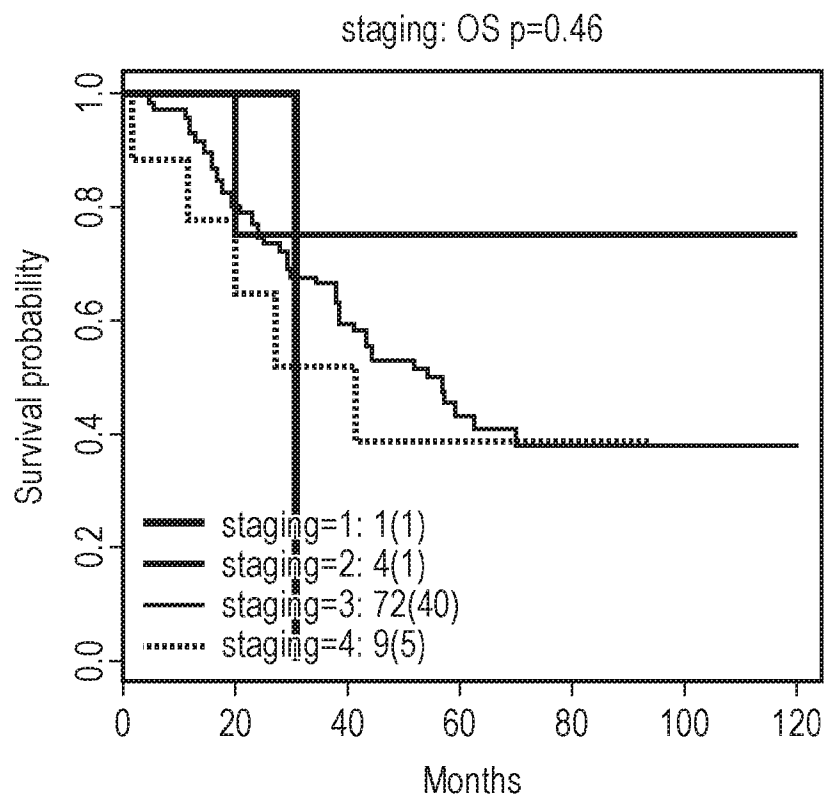
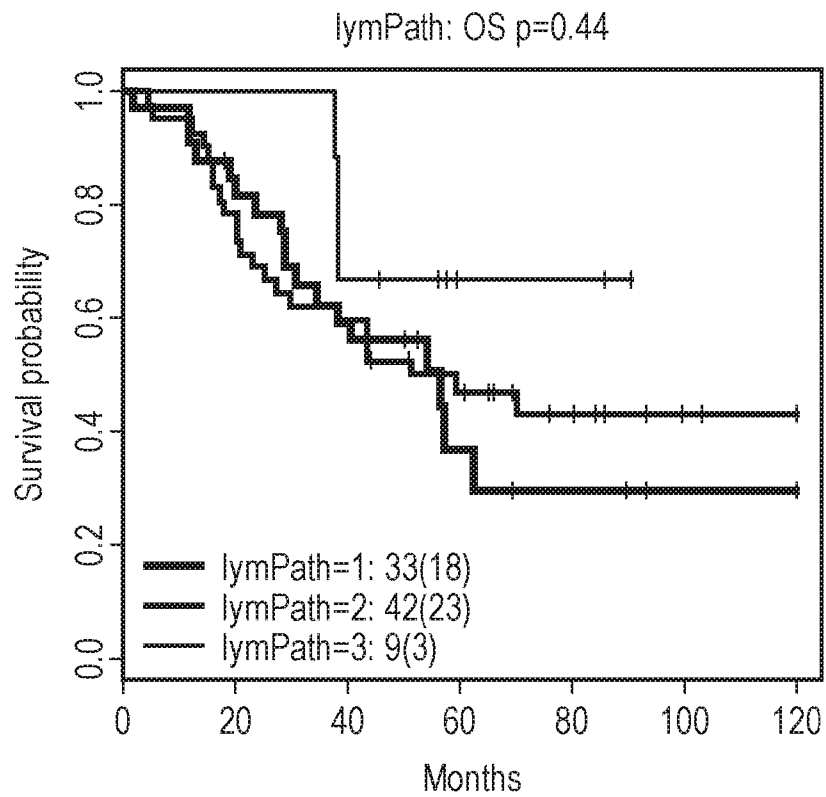


FIG. 14 (continued)

27/28

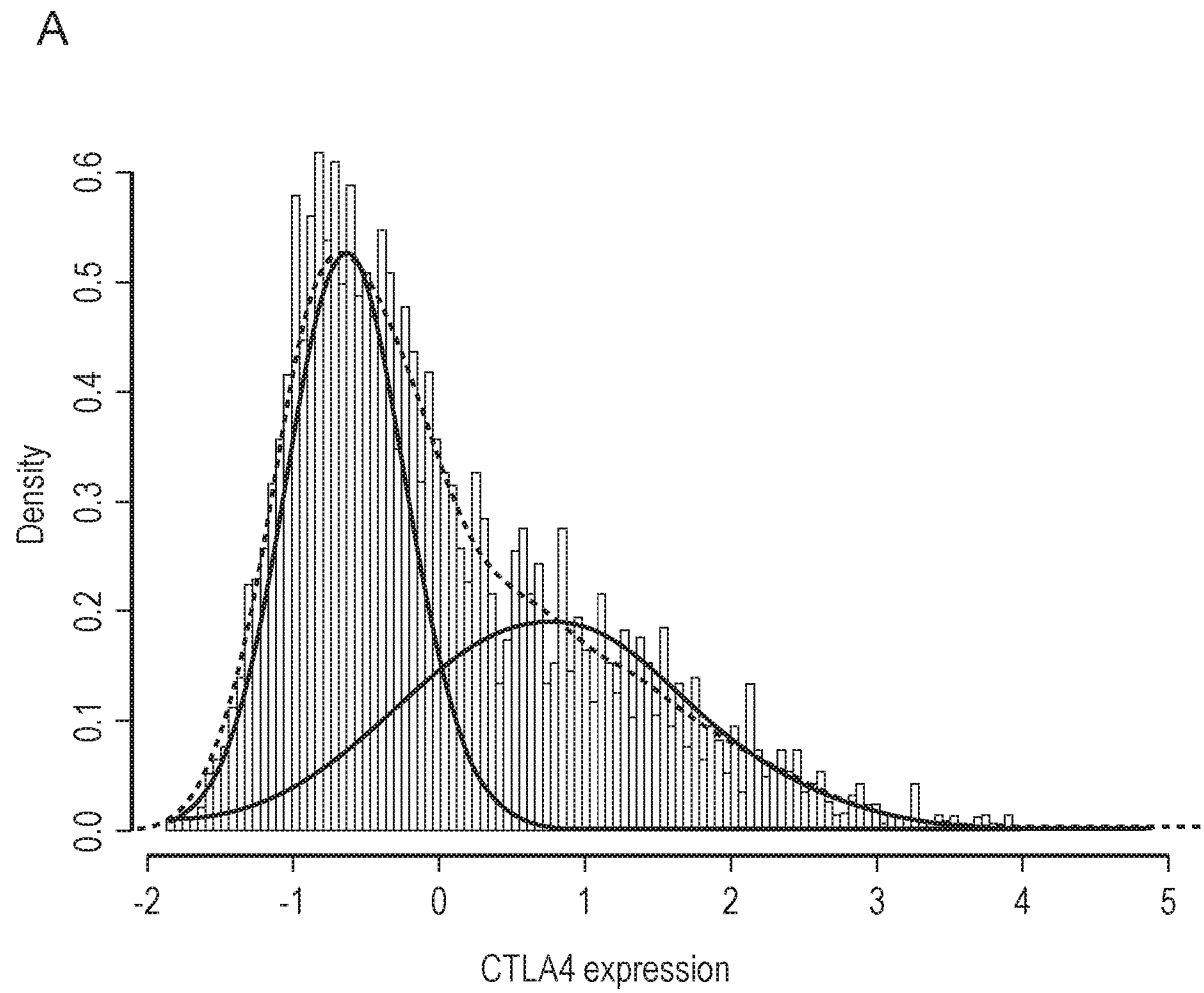


FIG. 15

28/28

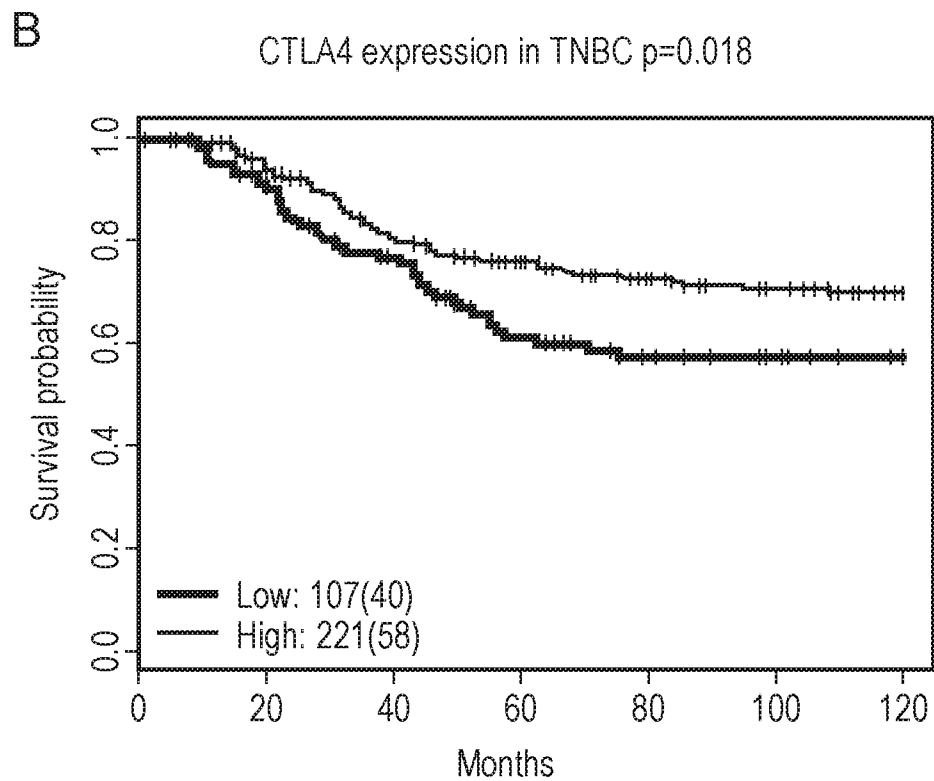


FIG. 15 (Continued)

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2015/053571

A. CLASSIFICATION OF SUBJECT MATTER
INV. G01N33/574 G06F19/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G01N G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JULIANE M KRÜGER ET AL: "Combat or surveillance? Evaluation of the heterogeneous inflammatory breast cancer microenvironment", THE JOURNAL OF PATHOLOGY, vol. 229, no. 4, 15 February 2013 (2013-02-15), pages 569-578, XP055181126, ISSN: 0022-3417, DOI: 10.1002/path.4150 the whole document figure 3 page 571, column 2, paragraph 2 abstract ----- -/--	1-15,18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

1 February 2016

Date of mailing of the international search report

14/04/2016

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Jenkins, Gareth

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2015/053571

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-15, 18

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2015/053571

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JÉRÔME GALON ET AL: "Towards the introduction of the 'Immunoscore' in the classification of malignant tumours", JOURNAL OF PATHOLOGY, vol. 232, no. 2, 10 December 2013 (2013-12-10), pages 199-209, XP055246049, GB ISSN: 0022-3417, DOI: 10.1002/path.4287 the whole document page 204 - page 206 -----	1-15,18
Y	A. M. MULLIGAN ET AL: "Tumoral Lymphocytic Infiltration and Expression of the Chemokine CXCL10 in Breast Cancers from the Ontario Familial Breast Cancer Registry", CLINICAL CANCER RESEARCH, vol. 19, no. 2, 4 December 2012 (2012-12-04), pages 336-346, XP055246060, US ISSN: 1078-0432, DOI: 10.1158/1078-0432.CCR-11-3314 the whole document abstract page 3, last paragraph -----	1-15,18
Y	US 2014/003702 A1 (MCCULLOCH COLIN CRAIG [US] ET AL) 2 January 2014 (2014-01-02) the whole document paragraph [0081] -----	1-15,18
A	WO 2013/154422 A1 (UNIV ERASMUS MEDICAL CT [NL]) 17 October 2013 (2013-10-17) the whole document claim 1 -----	1-15,18
A	S. ADAMS ET AL: "Prognostic Value of Tumor-Infiltrating Lymphocytes in Triple-Negative Breast Cancers From Two Phase III Randomized Adjuvant Breast Cancer Trials: ECOG 2197 and ECOG 1199", JOURNAL OF CLINICAL ONCOLOGY, vol. 32, no. 27, 28 July 2014 (2014-07-28), pages 2959-2966, XP055246066, US ISSN: 0732-183X, DOI: 10.1200/JCO.2013.55.0491 the whole document abstract ----- -/--	1-15,18

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2015/053571

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	ANDREAS HEINDL ET AL: "Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology", LABORATORY INVESTIGATION, vol. 95, no. 4, 19 January 2015 (2015-01-19), pages 377-384, XP055246070, The United States and Canadian Academy of Pathology, Inc. ISSN: 0023-6837, DOI: 10.1038/labinvest.2014.155 the whole document -----	1-15,18
X,P	Y. YUAN: "Modelling the spatial heterogeneity and molecular correlates of lymphocytic infiltration in triple-negative breast cancer", NATURE, vol. 7, no. 7418, 10 December 2014 (2014-12-10), pages 93-20141153, XP055246073, United Kingdom ISSN: 0028-0836, DOI: 10.1038/nature11412 the whole document -----	1-15,18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2015/053571

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014003702	A1	02-01-2014	NONE

WO 2013154422	A1	17-10-2013	CA 2870255 A1 17-10-2013
			CN 104471402 A 25-03-2015
			EP 2836836 A1 18-02-2015
			JP 2015514222 A 18-05-2015
			US 2015079078 A1 19-03-2015
			WO 2013154422 A1 17-10-2013

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-15, 18

Independent method claim 1

2. claim: 16

Independent method claim 16

3. claim: 17

Independent method claim 17
