Next Generation Immunohistochemistry (NGI):

Unlocking the power of immunohistochemistry to improve biomarker analyses in precision oncology

Garazi Serna Alonso
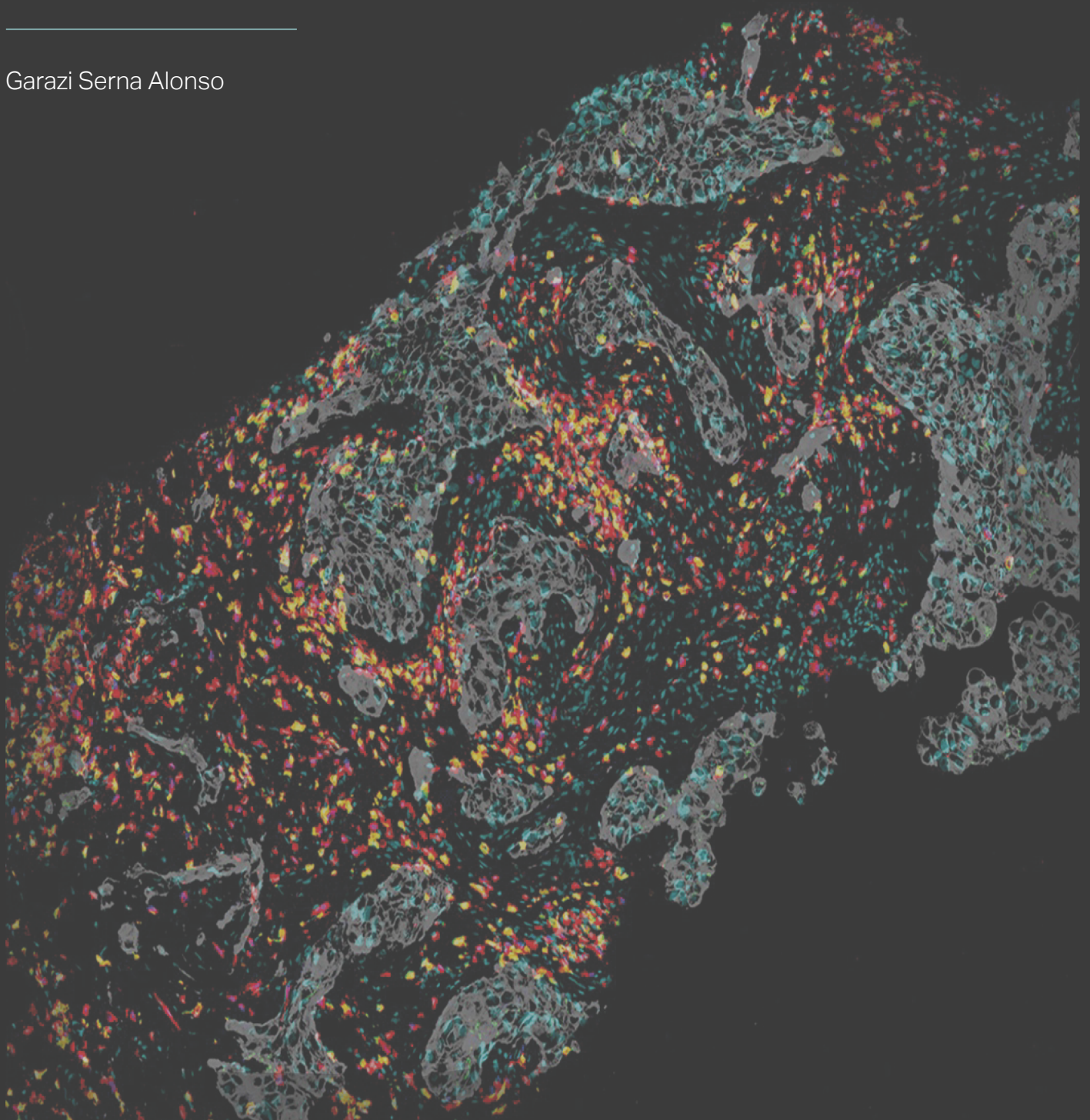
NEXT GENERATION IMMUNOHISTOCHEMISTRY (NGI):

Unlocking the power of immunohistochemistry to improve biomarker analyses in precision oncology

TESIS DOCTORAL

Garazi Serna Alonso

DIRECTORES DE TESIS

Dr. Paolo Nuciforo

Dr. Santiago Ramón y Cajal

TUTOR DE TESIS

Vicente Peg

UAB
**Universitat Autònoma
de Barcelona**

Programa de doctorado de Cirugía y Ciencias Morfológicas
Departamento de Ciencias Morfológicas
Facultad de Medicina
2022

CERTIFICAN:

Que el trabajo de investigación realizado por Garazi Serna Alonso y titulado:

"NEXT GENERATION IMMUNOHISTOCHEMISTRY (NGI): UNLOCKING THE POWER OF IMMUNOHISTOCHEMISTRY TO IMPROVE BIOMARKER ANALYSES IN PRECISION ONCOLOGY"

Se ha realizado bajo nuestra dirección y esté en condiciones para presentar su lectura y defensa delante del tribunal correspondiente para obtener el grado de Doctor.

Y para que conste a los efectos oportunos, firmamos el documento presente en Barcelona, julio de 2022.

PAOLO NUCIFORO   SANTIAGO RAMÓN Y CAJAL

UAB
Universitat Autònoma
de Barcelona

Departamento de Cirugía y Ciencias Morfológicas
Facultad de Medicina
2022

La presente tesis doctoral se ha estructurado siguiendo la Normativa interna de la Universidad Autónoma de Barcelona para la presentación de tesis doctorales como compendio de publicaciones, aprobada por Comisión de Doctorado de la Facultad de Medicina.

Los trabajos que forman parte de la memoria del proyecto de tesis doctoral pertenecen a la misma línea de investigación, enmarcada en la definición y valor del concepto de la tecnología Next Generation Immunohistochemistry (NGI) con objetivo de descubrir el poder de la inmunohistoquímica para mejorar el análisis de biomarcadores en la medicina de precisión. Como resultado de estos trabajos se publicado dos artículos originales en la Literatura internacional y se ha enviado uno pendiente de publicar.

Las referencias bibliográficas son:

Serna, G., Simonetti, S., Fasani, R., Pagliuca, F., Guardia, X., Gallego, P., Jimenez, J., Peg, V., Saura, C., Eppenberger-Castori, S., Ramon Y Cajal, S., Terracciano, L., & Nuciforo, P. (2020). Sequential immunohistochemistry and virtual image reconstruction using a single slide for quantitative KI67 measurement in breast cancer. Breast (Edinburgh, Scotland), 53, 102–110. https://doi.org/10.1016/j.breast.2020.07.002

Griguolo, G., Serna, G., Pascual, T., Fasani, R., Guardia, X., Chic, N., Paré, L., Pernas, S., Muñoz, M., Oliveira, M., Vidal, M., Llombart-Cussac, A., Cortés, J., Galván, P., Bermejo, B., Martínez, N., López, R., Morales, S., Garau, I., Manso, L., … Nuciforo, P. (2021). Immune microenvironment characterisation and dynamics during anti-HER2-based neoadjuvant treatment in HER2-positive breast cancer. NPJ precision oncology, 5(1), 23. https://doi.org/10.1038/s41698-021-00163-6

CONTENTS

LIST OF ABBREVIATIONS

3D: Three-dimensional

AEC: Aminoethyl carbazole

DAB: 3,3'-Diaminobenzidine

ER: Estrogen receptor

FC: Flow cytometry

FFPE: Formalin-fixed paraffin-embedded

HER2:   Human epidermal growth factor receptor 2

IHC: Immunohistochemistry

IMS: Imaging mass spectrometry

IO: Immuno-oncology

MIBI: Multiplexed Ion Beam Imaging

mIF: Multiplex immunofluorescence

mIHC: Multiplex immunohistochemistry

mRNA: Messenger ribonucleic acid

NGI: Next Generation Immunohistochemisty

PR: Progesterone receptor

RNAseq: RNA sequencing
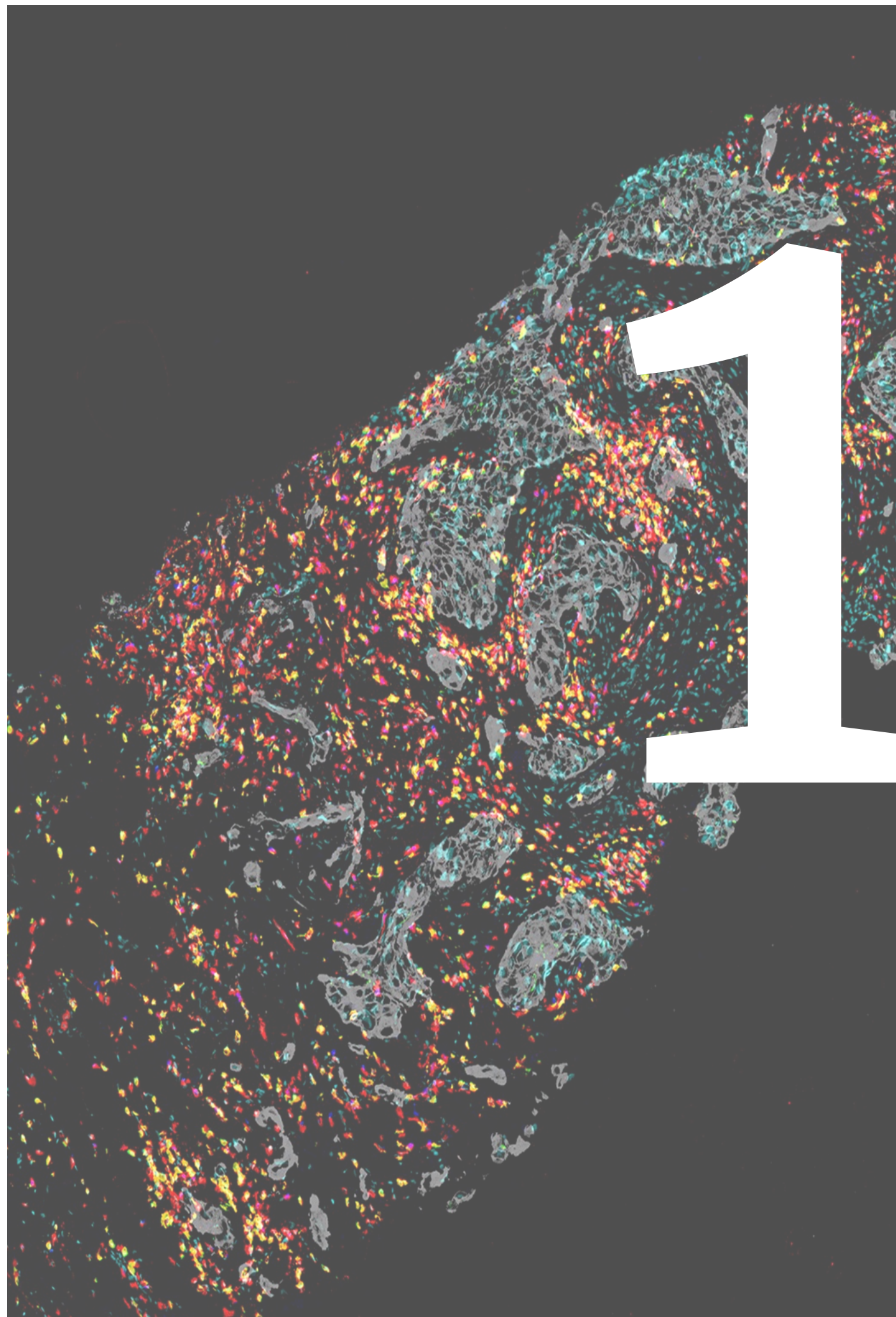
1. INTRODUCTION

Personalized medicine promises diagnosis and treatment of disease at the individual level and relies heavily on clinical specimen and diagnostic assay quality. Today, histopathological and molecular features of the tumor are well integrated to provide the best possible treatment option for the individual patient. Pathologist diagnosis still represents the first step in the patient diagnostic workflow. In routine diagnosis, formalin-fixed paraffin-embedded (FFPE) tissues are used, being the most common form of preserved archived clinical samples (Gerdes et al., 2013).

Modern biopsy techniques often provide only small tumor specimens and serial sample sectioning is usually required to allow the pathologist to determine the morphology and immunophenotype of the tumor which will finally lead to a correct diagnosis. Sample use may be minimal but, in the case of diagnostically challenging cases, additional testing may be required, thus jeopardizing subsequent molecular characterization of the tumor, especially when dealing with small biopsy specimens. The remaining tissue can be used for molecular analysis only after the diagnosis has clearly been established. This could lead to material exhaustion for further analyses and require a re-biopsy of the patient (Pirker & Filipits, 2016).

As more targeted treatment options become available, testing for multiple markers is required and abundant amount of good quality samples need to be acquired. Unfortunately, current

testing techniques do not allow simultaneous detection of gene mutations, gene expression and protein expression on the same starting material and multigene- based test does not provide histological, cellular and subcellular context(Gerdes et al., 2013). Being able to do a complete diagnostic workflow and comprehensively understand the cancer cells, the interactions between the cellular and molecular components of the immune system, as well as the tumor microenvironment as a whole can affect cancer treatment, prognosis, relapse and could help find various mechanisms determining sensitivity and resistance of cancer cells to treatments.

The advent of effective immunotherapy using immune-checkpoint inhibitors has led to a revolution in cancer, with immunotherapeutics now available across multiple tumor types. The characterization of the complexity of tumor microenvironment is becoming an important "biomarker" to combine with tumor genotype to predict response to these drugs. Immune cells consist of different subpopulations with distinct functions (de Obaldia & Bhandoola, 2015) and the nature and structure of the "cancer-immune system" may both promote and suppress tumor growth(Chen & Mellman, 2017).

The limited knowledge of tumor microenvironment composition and spatial organization affect our ability to develop predictive biomarkers of response to immunotherapy, which are urgently needed. Several biomarkers have been proposed to delineate responders from non-responders and to investigate their prognostic role, but most of them have not been incorporated into the clinical practice. This is presumably due to our limited understanding of the complex and heterogeneous spatial organization of the tumor-immune ecosystem and of the cancer mechanisms that evade the immune system (Finotello et al., 2019; Sankar et al., 2022).

The analysis of multiple biomarkers in FFPE samples include serial sectioning and individual staining from the same tissue block by chromogenic immunohistochemistry (IHC) - technique that is used routinely in the majority of the pathology laboratories. Here limitations arise as it requires serial sections that makes co-expression analyses on single cells not possible. Also, the extensive use of the sample - which may become quickly exhausted – adds to the restraints of current approaches.

Multiplex immunofluorescence (mIF) and multiplex immunohistochemistry (mIHC) addresses the aforementioned limitations by allowing expression analysis of multiple biomarkers in a single tissue section. However, the number of biomarkers that can be multiplexed depends on the available fluorophores/chromogens.

By one hand, mIF enables the analysis of more biomarkers than mIHC, more quantitative results and better visualization at an intracellular level. However, autofluorescence makes the evaluation more difficult leaving even image processing algorithms build to remove this less successful. Furthermore, visualizations can't be performed by classical bright-field microscopy, which is the accepted standard in pathology.

By the other hand, mIHC makes the evaluation difficult due to the fact that the different colors are in the same slide and thus not separated in channels, which makes it easier to evaluate. Moreover, the cellular localizations of the various biomarkers must be different in order to distinct them, also leading to the need of complex equipment that allow spectral differentiation if more than one biomarker belongs to the same compartment. While, some translucent chromogens HRP kits (e.g., Ventana medical systems DISCOVERY Green HRP kit) allow a change of color when co-localizing biomarkers, however these are not easy to evaluate and are for pure research purposes thus also not used in diagnostic procedures. Furthermore, the different biomarkers could have different pre-treatments and finding a unique pre-treatment for the technique could make some biomarker not be optimally stained.

Some complex cellular populations (such as specific types of immune cells) are only identifiable by combining several biomarkers. These are routinely studied by flow cytometry (FC) or RNAseq, that look at cells individually but does not provide spatial or morphological context information and require complex computational analyses to obtain and visualize the data(Finotello et al., 2019). Moreover, RNA seq gives the result of the average transcriptome of the cells of the sample, while single-cell RNAseq can give the data from individual cells, though it requires even more complex and expensive equipment.

Metal-based multiplexed imaging systems such as Multiplexed Ion Beam Imaging (MIBI) and Imaging mass spectrometry (IMS)

are effective tools for discovery studies on intact tissue sections providing quantification and imaging of up to 100 biomarkers from tissue sections labelled with metal-tagged antibodies (Giesen et al., 2014; Keren et al., 2018). The issue is that these are very time-consuming techniques requiring high acquisition costs, dedicated infrastructure and expertise, that limit its accessibility outside pure research environment, apart from the fact that they are less sensitive because of the nature of meta-conjugation (Tan et al., 2020).

New technologies of barcoding based multiplexed imaging like CODEX, DSP, Insituplex; fluorescence based multiplexed imaging systems like Vectra; and Chipcy-tometry or 3D imaging technologies are also being developed. These are able to provide comprehensive cellular spatial information but once again, they are expensive and not practical for clinical routine use.

Finally, other multiplexing approaches have tested the possibility of performing iterative cycles of staining and destaining (up to 12) of the same tissue section without losing tissue antigenicity and with absence of cross-reactivity between stainings by the use of an alcohol soluble chromogen(Glass et al., 2009; Remark et al., 2016; Tsujikawa et al., 2017). However, the techniques described are fully manual and time consuming which limit its wider use from outside a research setting.

Building from these observations, we have developed the Next Generation Immunohistochemisty (NGI), an automatized, simple, and flexible IHC based technology for sequential staining using

the combination of Ventana Discovery Ultra (Roche Diagnostics), Nanozoomer slide scanner (Hamamatsu) and Visiopharm image analysis program - equipment that is commonly used in pathology laboratories.

The protocol consists in sequential automatized IHC stainings with an alcohol soluble chromogen (AEC). After each automated IHC staining, the samples are mounted in aqueous medium and digitalized. Subsequently, the section is destained in alcohol and submitted to the following staining cycle (Figure 1). After digitalization, images are aligned in Visiopharm image analysis program for data extraction.

Figure 1. Next Generation Immunohistochemistry (NGI).

A) Protocol of the sequential immunohistochemical staining on a single histological section and image analysis with Discovery Ultra automated slide staining system (Roche Diagnostics), NanoZoomer Digital Pathology Slide Scanner (Hamamatsu) and Visiopharm image analysis program.
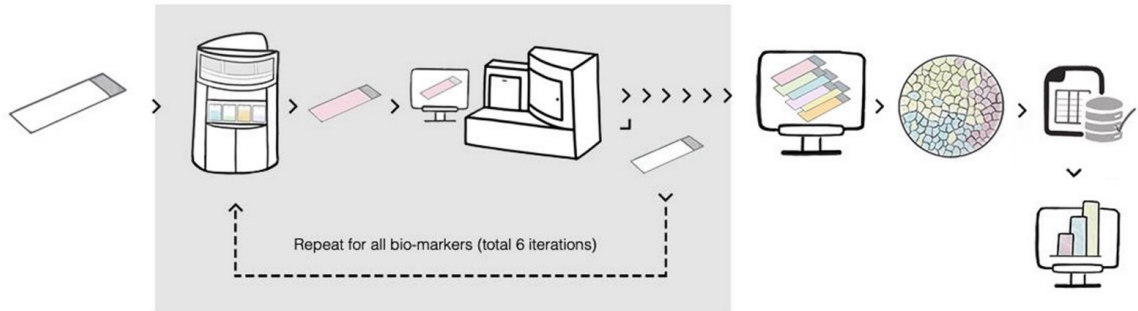
B) Representative example of different sequential stainings (CD3, CD8, ER, HER2, KI67) and color deconvoluted images for virtual color assignment for composite image creation where all stainings and co-expressions are shown with different virtual colors.
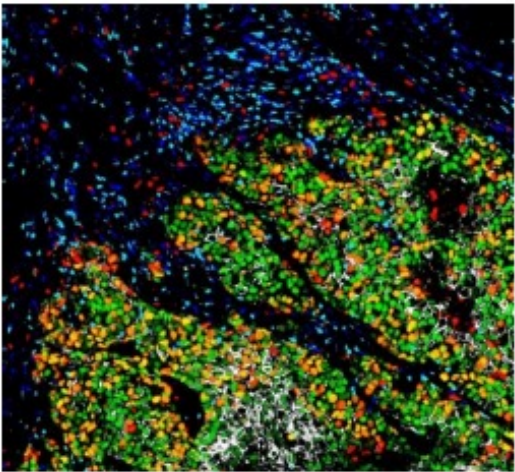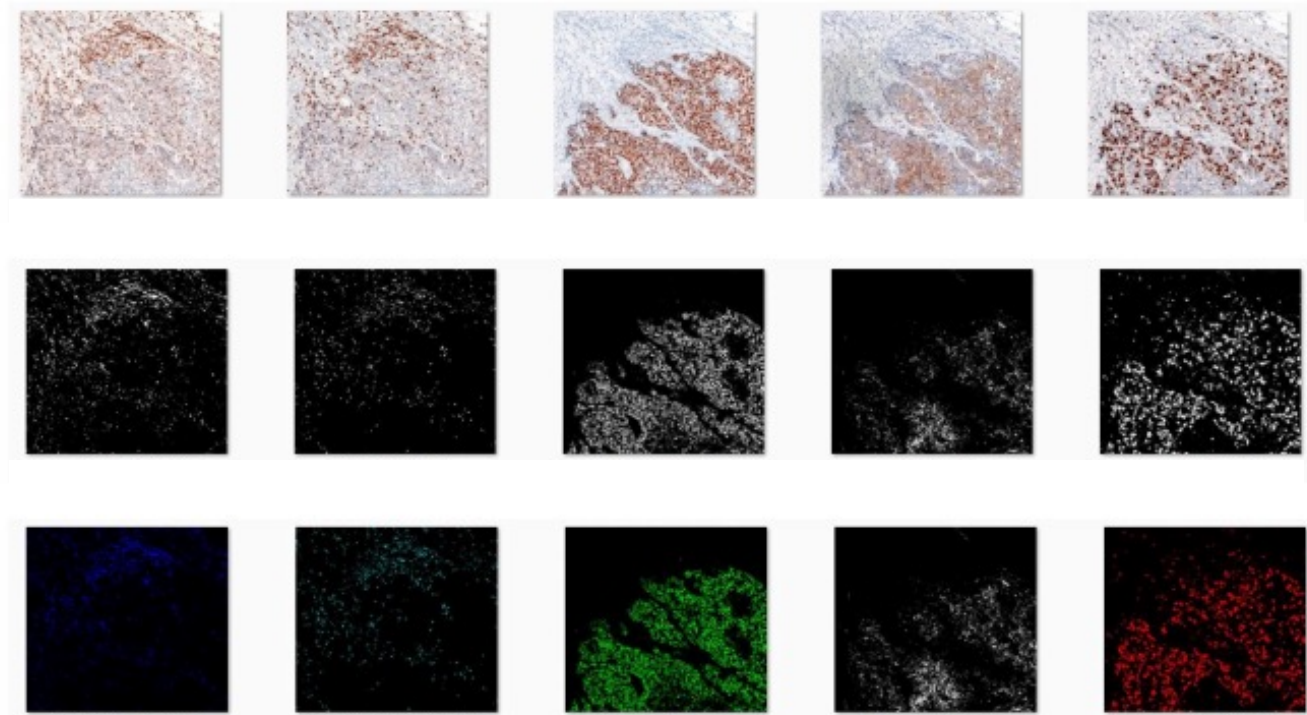
2. HYPOTHESIS

Different multiplexed and non-disruptive imaging technologies are needed for individualized treatment in cancer and other malignancies to extract all the information and biological insights of the disease. The hypothesis is that the automatized technology outlined in this thesis may fill the actual limitations of the current approaches and become one of the multiplexed imaging technologies that could be used in different pathology and research laboratories to improve biomarker analyses in precision oncology.

Image analysis applied to multiplexed imaging is able to extract both qualitative and quantitative individual and combined biomarker data of multiple biomarkers at a cellular level; characterize complex cellular populations and integrate them with spatial and morphological context information; characterize the interaction between tumor and its associated immune microenvironment and explore its applications to immuno-oncology (IO); identify or improve biomarkers to better predict response to chemotherapies and immunotherapies; characterize the impact of intra-tumoral heterogeneity and single cell analysis on response to targeted therapies; and improve the reproducibility and analysis automation of diagnostic, prognostic, and predictive biomarkers.

The results generated by this study can provide a framework for applying this technology to address complex biological questions.

If NGI becomes a useful multiplex imagining technology it may allow for a comprehensive characterization of biological tissue samples at cellular level while maintaining important spatial distribution/interaction between tumor and its microenvironment. These insights would advance our understanding of tumor biology and its complexity and they might help identifying better or new prognostic or predictive biomarkers that predict the response to treatment supporting better patient stratification towards different therapies or their combinations and therefore, increase the number of patients that can benefit from them.

3. OBJECTIVES

The general objective is the validation and application of the high resolution, simple, robust and automated methodology called Next Generation Immunohistochemistry or NGI, that sequentially stains and destains different biomarkers in the same FFPE section allowing together with the image analysis algorithms that are going to be designed the quantification, co-expression and spatial analyses, while preserving the tissue and that would be useful not only in the research field but also in pathology laboratories.

The specific objectives are creating different panels (composed of highly relevant diagnostic and exploratory predictive targets) that are clinically useful and applicable for different tumor types or stages of the disease and even applicable in other malignancies:

- To design and validate a panel that quantifies in a simple, automated and reproducible way the KI67 biomarker in the tumor by using KI67 proliferation biomarker together with the PanCK that allows tumor identification.

- To design and validate a panel that quantifies different t-cell subpopulations. We will focus on the study of the tumoral immune compartment by extensively characterizing immune cell populations on tumor tissue samples. By designing a NGI protocol we aim to characterize the spatial interaction between the tumor and its immune microenvironment and the proliferation of the different populations.

- To design and validate a panel that studies the intratumor intrinsic molecular subtype heterogeneity in breast cancer tissue samples and its impact in patient prognosis and response to targeted therapies.

4. PUBLICATIONS

4.1- Study 1:

"Sequential immunohistochemistry and virtual image reconstruction using a single slide for quantitative KI67 measurement in breast cancer"

In this study we validated the first panel of NGI composed of KI67 with the PanCK and we called it KiQuant, which allows the quantification of the proliferating biomarker ki67 in the tumor cells in an easy and reproducible way that is one of the big problems with ki67 evaluation.

Original article

Sequential immunohistochemistry and virtual image reconstruction using a single slide for quantitative KI67 measurement in breast cancer

Garazi Serna [a], Sara Simonetti [a], Roberta Fasani [a], Francesca Pagliuca [b], Xavier Guardia [a], Paqui Gallego [a], Jose Jimenez [a], Vicente Peg [c], Cristina Saura [d], Serenella Eppenberger-Castori [e], Santiago Ramon y Cajal [c], Luigi Terracciano [e], Paolo Nuciforo [a, *]

[a] Molecular Oncology Group, Vall D'Hebron Institute of Oncology, Barcelona, Spain
[b] University of Naples Federico II, Department of Advanced Biomedical Sciences, Pathology Section, Naples, Italy
[c] Department of Pathology, Vall D'Hebron University Hospital, Barcelona, Spain
[d] Breast Cancer and Melanoma Group, Vall D'Hebron Institute of Oncology, Barcelona, Spain
[e] Institute of Pathology, University Hospital Basel, Basel, Switzerland

ARTICLE INFO

ABSTRACT

Objective: Ki67 is a prognostic and predictive marker in breast cancer (BC). However, manual scoring (MS) by visual assessment suffers from high inter-observer variability which limits its clinical use. Here, we developed a new digital image analysis (DIA) workflow, named KiQuant for automated scoring of Ki67 and investigated its equivalence with standard pathologist's assessment.
Methods: Sequential immunohistochemistry of Ki67 and cytokeratin, for precise tumor cell recognition, were performed in the same section of 5 tissue microarrays containing 329 tumor cores from different breast cancer subtypes. Slides were digitalized and subjected to DIA and MS for Ki67 assessment. The intraclass correlation coefficient (ICC) and Bland-Altman plot were used to evaluate inter-observer reproducibility. The Kaplan-Meier analysis was used to determine the prognostic potential.
Results: KiQuant showed an excellent correlation with MS (ICC:0.905,95%CI:0.878−0.926) with satisfactory inter-run (ICC:0.917,95%CI:0.884−0.942) and inter-antibody reproducibilities (ICC:0.886,95% CI:0.820−0.929). The distance between KiQuant and MS increased with the magnitude of Ki67 measurement and positively correlated with analyzed tumor area and breast cancer subtype. Agreement rates between KiQuant and MS within the clinically relevant 14% and 30% cut-off points ranged from 33% to 44% with modest interobserver reproducibility below the 20% cut-off (0.606, 95%CI:0.467−0.727). High Ki67 by KiQuant correlated with worse outcome in all BC and in the luminal subtype ($P = 0.028$ and $P = 0.043$, respectively). For MS, the association with survival was significant only in 1 out of 3 observers.
Conclusions: KiQuant represents an easy and accurate methodology for Ki67 measurement providing a step toward utilizing Ki67 in the clinical setting.

1. Introduction

Ki67 is a nuclear protein expressed throughout all the phases of the cell cycle from G1 to M-phase [1]. Due to its association with cellular proliferation, Ki67 detection by immunohistochemistry (IHC) has emerged as a useful and inexpensive tool to assess the proliferation index of a tumor. Many studies have shown prognostic and predictive values of Ki67 in a wide range of malignancies [2] [−] [9]. In particular, in breast cancer (BC), Ki67 has been successfully used not only for classification and risk assessment purposes but also to decide therapeutic endpoints in the context of neoadjuvant settings [10] [−] [13].

The promise of Ki67 as a biomarker is affected by technical and scoring reproducibility issues, which make it not ready for clinical use. Despite the efforts of the International Ki67 in Breast Cancer Working Group (IKWG) to standardize the preanalytical, analytical, interpretation, and data analysis steps, variations in protocols and scoring methodologies across laboratories remain large contributors to assay variability [14,15]. Manual counting provides better interobserver reproducibility as compared to visual estimation [16]. However, as scoring the whole section seems impractical, the location and extent of the area that should be scored are controversial and subject to observer's interpretation [17,18]. As a consequence, despite different Ki67 thresholds to define luminal A vs luminal B tumors (14%, 20%, laboratory median values [19] [−] [21]) have been proposed, no absolute standard methodology and cut-off point have been defined so far. In this context, the use of multigene tests [22−24] and digital image analysis (DIA) [25−32] may be valuable, especially across intermediate Ki67 levels where there is high uncertainty.

While computer-assisted methods are expected to provide a more accurate Ki67 assessment, these approaches either rely on significant pathologist's intervention for the area of interest selection or use unique and sophisticated cell segmentation and classification algorithms that require extensive supervised learning.

In this study, we describe a novel methodology for automatic scoring of Ki67 which relies on sequential IHC of Ki67 and cytokeratin using a single slide, followed by virtual image reconstruction for DIA. The use of a cytokeratin mask allows for the precise definition of the region of interest and limits pathologist's intervention. The methodology accuracy was compared with manual scoring (MS) determined by multiple observers to demonstrate equivalence or superiority. Finally, the outcome prediction potential of our method was investigated.

2. Material and methods

2.1. Patients and samples

Clinicopathological features of study cohorts are shown in Table 1. A total of 186 patients from 2 different cohorts was used in this study. Cohort 1 was composed of 99 patients with BC of different subtypes [hormone receptor-positive (HR+), HER2-positive (HER2+), and triple-negative (TN)] retrieved from the Pathology Department of the Vall d'Hebron University Hospital (Barcelona, Spain). No survival data were available for this cohort. Cohort 2 comprised an independent set of 87 BCE patients selected from the archives of the Pathology Department of the University Basel Hospital (Basel, Switzerland), with 58 months median follow up for overall survival (OS).

From the surgical specimens of primary BC of each patient, a representative paraffin-embedded tumor tissue block was selected and five tissue microarrays (TMAs), containing representative BC tissue cores of 1.5 mm, were built. For cohort 1, four TMAs were constructed, containing between 2 and 3 cores of 1.5 mm for each tumor, with a total of 242 cores. Cohort 2 BCE cancer specimens were arrayed in one TMA built using one representative core for each tumor sample (87 cores of 1.5 mm).

The protocol of this study was approved by the Vall d'Hebron University Hospital Ethical Committee (PR(AG)76/2018) and all methods were performed in accordance with relevant guidelines and regulations.

2.2. Immunohistochemistry

The complete sequential IHC and image analysis workflow used in this study (named KiQuant) is illustrated in Fig. 1A. Briefly, after deparaffinization and antigen retrieval (CC1, 64 min at 95 °C), one single slide of each TMA was first stained with a standard anti-Ki67 primary antibody (clone 30−9, prediluted, #790−4286, Ventana, Tucson AZ) for 64 min at 37 °C on an automated staining system (Discovery Ultra, Ventana, Tucson AZ). Antibody binding was amplified with an anti-Rb HRP biotin-free detection system (#760−4315, Ventana, Tucson AZ) ,visualized using 3-amino-9-ethylcarbazole (Mono AEC/Plus, #K050, PALEX), an alcohol soluble substrate that results in red staining, and counterstained with hematoxylin. The slides were mounted and digitalized at 20x using a slide scanner (NanoZoomer 2.0HT, Hamamatsu Photonics, Japan). Following digitalization, the coverslip was removed before destaining using ethanol as previously described (Tsujikawa, Cell Reports 2017; Glass, J Histochem Cytochem 2009). The destained slides were then subjected to an additional antigen retrieval step (CC2, 8 min at 100 °C) to completely strip the first primary antibody (anti-Ki67) before the second staining cycle with another antigen retrieval (CC1, 48 min at 95 °C) was started. Then, the anti-Pan-Keratin primary antibody (clone AE1/AE3/PCK26, prediluted, #760−2135, Ventana, Tucson AZ) was applied for 40 min at 36 °C. Antibody binding was amplified with an anti-MS HRP biotin-free detection system (#760−4313, Ventana, Tucson AZ). For reproducibility analyses, a second anti-Ki67 antibody was also used (clone MIB-1, #M7240, DAKO/Agilent, Santa Clara, CA) for IHC in a non-consecutive slide from a TMA of cohort 1.

Table 1
Clinicopathologic characteristcs of study cohorts.

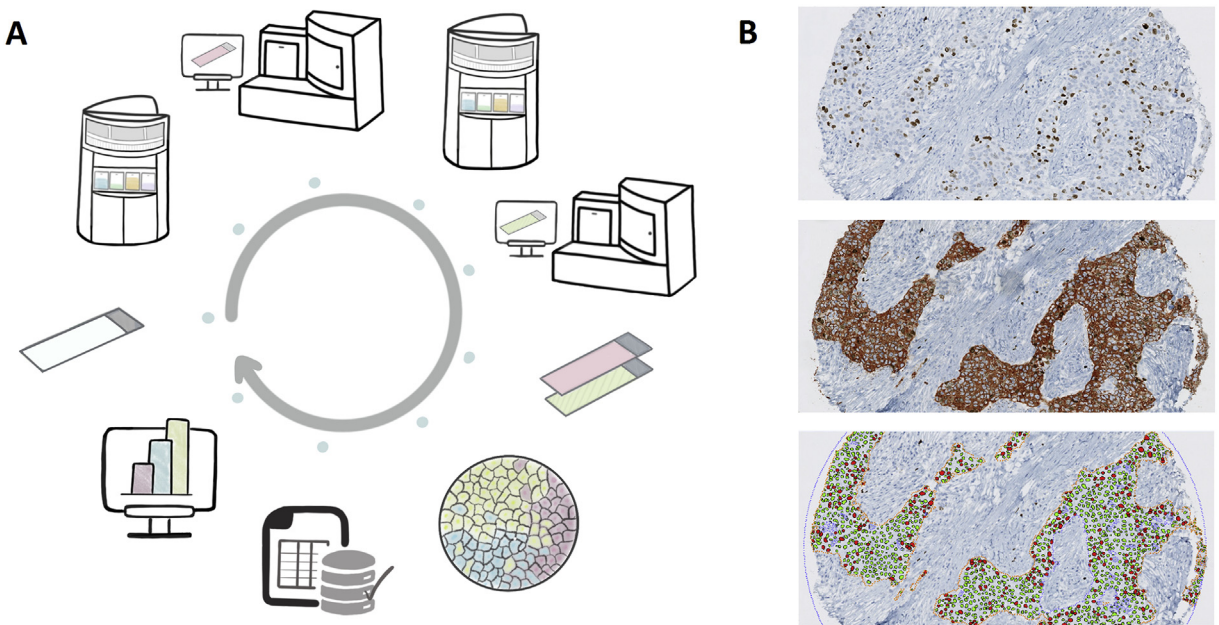|	Cohort 1		Cohort 2	
	N	%	N	%
Patients	99	100	87	100
Diagnosis				
Invasive ductal carcinoma	78	78.8	61	70.1
Invasive lobular carcinoma	12	12.1	14	16.1
Mixed ductal-lobular carcinoma	0	0.0	10	11.5
Medullary carcinoma	1	1.0	1	1.1
Mucinous carcinoma	3	3.0	1	1.1
Metaplastic carcinoma	5	5.1	0	0.0
Grade				
I	1	1.0	8	9.2
II	39	39.4	37	42.5
III	54	54.5	42	48.3
NA	5	5.1	0	0.0
pT				
T1	17	17.2	32	36.8
T2	56	56.6	41	47.1
T3	15	15.2	6	6.9
T4	1	1.0	8	9.2
NA	10	10.1	0	0.0
pN				
Negative	44	44.4	39	44.8
Positive	50	50.5	43	49.4
NA	5	5.1	5	5.7
pM				
0	99	100	83	95.4
1	0	0,0	4	4.6
Subtype				
HER2+	23	23.2	9	10.3
HR+	39	39.4	62	71.3
TNBC	37	37.4	16	18.4

Fig. 1. KiQuant workflow. A) This workflow uses sequential Ki67 and cytokeratin (for precise automatic tumor cells recognition) immunohistochemistry staining on the same tissue section. The steps are: staining of the slide with the first primary antibody anti-Ki67, digitalization of the slide after the first staining, coverslip removal and staining of the slide with the second primary antibody anti-Pan-Keratin, digitalization of the slide after the second staining, image alignment, image analysis, quality check, and data report. B) A representative example of a tumor core sequentially stained with Ki67 (top) and Pan-Keratin (middle). In the virtual digital image (bottom), Pan-Keratin-positive brown areas are used to automatically mark the region of interest (dotted orange line). Green and red cells represent Ki67-negative and -positive nuclei, respectively. Ki67-labelled stromal cells not stained by the Pan-Keratin antibody are excluded from the analysis.

2.3. Digital image analysis (DIA)

To analyze the images, we created an algorithm using the Author® module of VISIOPHARM® (VIS) Image Analysis Software (Visiopharm Integrator System version 2019.January 02, 6005, Visiopharm, Denmark). Both Ki67 and cytokeratin stained digitalized slides were automatically registered, to fuse the information into a single virtual digital image (VDI), using the Tissuealign® module of VIS. Cytokeratin stained images were used to classify tumor and stromal areas within each core. A color deconvolution algorithm enhanced tumoral areas that were extracted using a pixel intensity threshold algorithm. Cytokeratin-based segmentations were transferred to the registered Ki67 image. Cells within the cytokeratin mask were classified into Ki67 positive or negative using a cell classification method based on form and size, and a pixel-color intensity threshold method, which considered only KI67 nuclear staining (algorithms specifications in Supplementary Table 1 and algorithms in Supplementary Material).

Each core on the TMA slides was separately analyzed by locating the tissue, using an automatic thresholding approach, and extracting the area of interest to create separate images. Cytokeratin-positive non cancer areas (such as ductal carcinomas in situ, necrosis or normal ducts) were manually excluded before data extraction. Final results are reported as the cell density, i.e. the total number of cells within the tumor area defined by the mask, number of positive and number of negative cell nuclei, as well as the percentage of positive cells within the corresponding core.

The application was trained by a biotechnologist, expert in image analysis, to identify positive and negative Ki67 cells within the tumor mask in the VDI (Fig. 1B). After setting the optimal conditions, the methodology was evaluated in samples from cohort 1 and validated on an independent group with follow up data. The framework performance was compared to the results obtained by manual scoring (MS) of three expert board-certified pathologists.

MS LI is defined as the ratio between the number of KI67-positive tumor cells and the total number of tumor cells, using either counting or estimation approaches[14] (Supplementary Table 2). Individual images were manually reviewed to exclude non-evaluable cores (n = 24, unpaired cores, cores containing only normal tissue, folded cores or cores without cytokeratin staining).

Method reproducibility was investigated by analyzing KiQuant results obtained by a) staining and analysis of two TMAs (TMA2 and TMA3) during two non-consecutive days (inter-run reproducibility), and b) by staining of two non-consecutive sections of one TMA (TMA2) with two different commonly used anti-Ki67 primary antibodies [inter-antibody reproducibility, clone 30−9 from Ventana (Tucson, AZ, USA) and clone MIB1 from DAKO/Agilent (Santa Clara, CA, USA)].

2.4. Statistical analysis

Agreement between KiQuant and MS was calculated using intraclass correlation coefficient (ICC) and Bland-Altman (BA) plot. We considered ICC values from 0.4 to 0.6 as moderate reliability, from 0.61 to 0.8 as good reliability, and greater than 0.8 as excellent reliability [33]. Spearman rank correlation coefficient and Kruskal-Wallis nonparametric test were used to determine the relationship of the difference between KiQuant and MS scoring with analyzed tumor area and BC subtypes, respectively. Overall survival was modeled using the Kaplan-Meier curves, and the significance of differences between these curves was determined using the log-rank test. Statistical analysis, data preparation, and figures were carried out with R-commander (v.1.9−5) and SPSS software (v.25.0; IBM (Armonk, NY, USA).

3. Results

3.1. Correlation between KiQuant and MS

KiQuant was compared with a reference standard Ki67 LI MS which was determined as the average Ki67 LI between two board-certified pathologists scoring 218 cores of cohort 1 using the counting method [14]. Correlation between KiQuant and MS was excellent (ICC: 0.905, 95% CI: 0.878−0.926). KiQuant returned systematically lower Ki67 LI results compared to MS (mean, 16.6% vs 21.7%, respectively). The correlation was better in HR+ (ICC: 0.934, 95% CI: 0.905−0.955) than in TNBC (ICC: 0.894, 95% CI: 0.835−0.933) and HER2+ (ICC: 0.862, 95% CI: 0.758−0.923) (Fig. 2A). BA plot revealed a significant proportional bias (regression analysis, P < 0.0001) by showing that the distance between KiQuant and MS increased with the magnitude of Ki67 measurement (Fig. 2B). The difference between KiQuant and MS correlated with the analyzed tumor area (Spearman's rho = 0.455, P < 0.0001), BC histology (mean difference, IDC = 5.4; ILC = 1.4; Medullary = 23.85; Metaplastic = 7.7; Mucinous = 0.6; Kruskal-Wallis test P = 0.001, Supplementary Figure S1) and subtype (mean difference, HER2+ = 7.4; TNBC = 6.3; HR+ = 3.1; Kruskal-Wallis test P = 0.001). Thirteen cores (6%) from 9 patients showed a difference between KiQuant and MS Ki67 LI outside the limits of agreement (Supplementary Table 3).

3.2. Concordance of KiQuant and MS across different observers in luminal BC

Three different pathologists scored 100 cores of luminal (HR+) BC and individual observer MS were compared between each other and with KiQuant results (Fig. 3, Supplementary Table 4). Overall, the inter-pathologist concordance was very high (ICC: 0.932, 95% CI: 0.905−0.952). The agreement between KiQuant and each individual observer was excellent for all observers (OBS1 ICC: 0.877, 95% CI: 0.822−0.915; OBS2 ICC: 0.870, 95% CI: 0.810−0.911; OBS3 ICC: 0.842, 95% CI: 0.774−0.891). Then, we determined the agreement between KiQuant and individual MS across different Ki67 LI cut-offs. The highest concordance rates (ranging from 84% to 100%, depending on the observer) were found below the 2.7% and above the 40% cut-offs, whereas the lowest (ranging from 71% to 90%) were observed within 14% and 30% cut-offs (Supplementary table 4). At the clinically relevant cut-off of 20% defined by St Gallen criteria [20], concordance between KiQuant and MS ranged from 71% to 86%, depending on the observer. Inter-observer concordance was lower below the 20% (ICC: 0.606, 95% CI: 0.467−0.727) as compared to equal or above the 20% (ICC: 0.937, 95% CI: 0.893−0.965) cut-point defined by KiQuant.

3.3. KiQuant reproducibility

KiQuant inter-run correlation (n = 120 cores, ICC: 0.917, 95% CI: 0.884−0.942), and inter-antibody reproducibility (n = 64 evaluable cores, ICC: 0.886, 95% CI: 0.820−0.929) were excellent (Supplementary Figure S2).

3.4. Prognostic potential of Ki67 LI as determined by KiQuant and MS

To test the outcome prediction potential of KiQuant, we used an independent cohort of 87 breast invasive carcinomas with available outcome data (Table 1). The analysis was performed using the KiQuant workflow and obtained results were compared with the reference MS LI determined by 3 independent observers. Median Ki67 LI were used as cut-points. Patients with high Ki67 LI by KiQuant had shorted overall survival (all: log-rank, P = 0.028; HR+/ HER2-: log-rank, P = 0.043). For MS LI, the association was statistically significant only in 1 out of 3 observers (Fig. 4, Supplementary
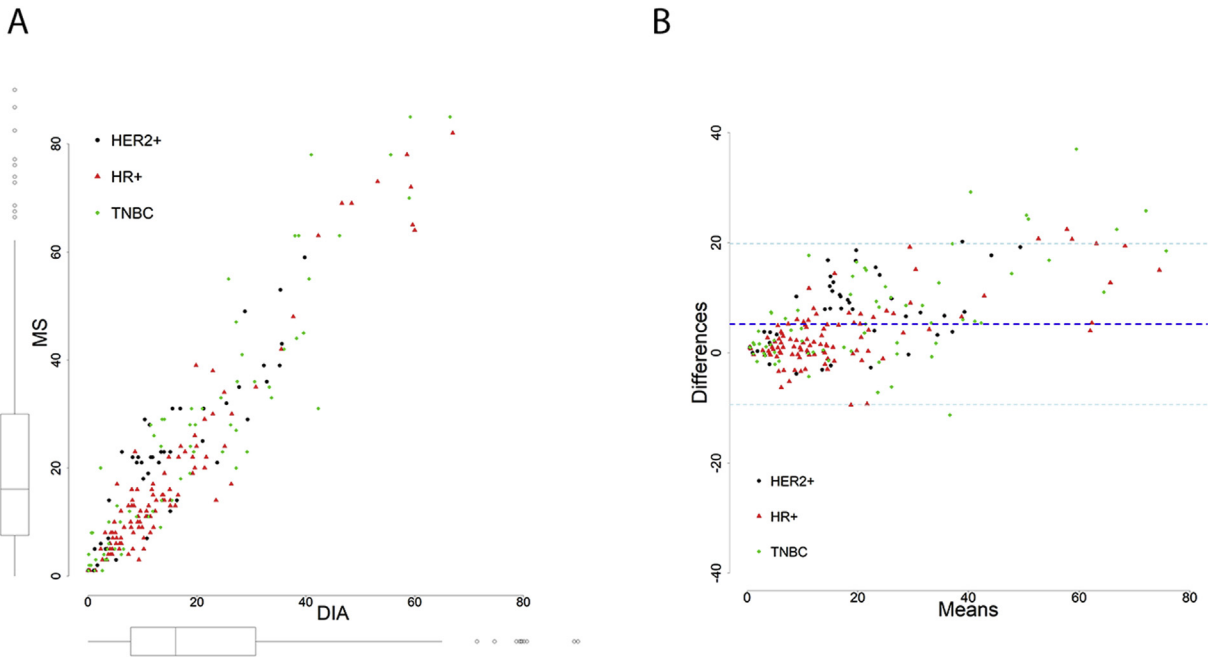
Fig. 2. A) Comparison of manual scoring (MS, y-axis) and KiQuant (DIA, x-axis) in breast cancer. The scattered plots are based on 218 evaluable cores from cohort 1. The breast cancer subtype is indicated in the legend of the top of the plots. Box plots of MS and KiQuant data are shown in the y-axis and x-axis, respectively. B) Bland-Altman plot of agreement between Ki67 labeling index (LI) by MS and KiQuant. In the x-axis, the average Ki67 LI between the two assessment methodologies is shown. The y-axis represents the difference between Ki67 LI scored by manual scoring and digital image analysis. The dotted blue line shows the average difference between the two assessments (4.95). The upper (19.51) and lower (−9.60) limits of agreement are indicated by the dotted light-blue lines.

CORE ID	OBS1	OBS2	OBS3	MS	KiQuant	KI67 range	% agreem KiQuant vs MS	ICC OBS
1	1	1	0	1	0			
2	1	1	0	1	0	0-2.6%	100%	
3	1	1	0	1	1			
4	2	5	2	3	4			
5	5	8	7	7	4			
6	5	8	6	6	4			
7	7	10	6	8	5			
8	15	7	22	15	5			
9	3		3	3	5			
10	7	5	6	6	5			
11	7	5	15	9	5			
12	5	5	5	5	5			
13	5	12	8	8	5			
14	5	7	5	5	6			
15	5	8	4	6	6			
16	7	5	10	7	6			
17	10	5	16	10	6			
18	7	10	6	8	6			
19	10		18	14	6			
20	10	5	20	12	7			
21	7	5	10	7	7	2.7-13%	71%	
22	10	15	15	13	7			
23	7	5	10	7	7			
24	3	5	3,4	4	7			
25	7		10	9	8			
26	12	5	24	14	8			
27	7	10	8	8	9			
28	10	3	12	8	9			
29		20	15	18	9			
30	7	15	20	14	9			
31	10	5	15	10	10			.606 (95% CI: 0.467-0.727)
32	10	15	23	16	10			
33	15		12	14	10			
34	15	5	42	21	11			
35	20	15	25	20	11			
36	15	8	19	14	11			
37	7	5	8	7	11			
38	12	10	11	11	11			
39	12	12	10	11	12			
40	25	10	28	21	12			
41	15	10	12	12	12			
42	15	10	25	17	14			
43	15		27	21	14			
44	20	10	27	19	14			
45	35	15	32	27	14			
46	17	20	15	17	14			
47	20	10	32	21	14			
48	7	5	10	7	14			
49	10	5	10	8	15			
50	15	5	25	15	15			
51	12	7	7	9	15			
52	10	5	15	10	16	14-29%	33%	
53	15	10	25	17	16			
54	30	20	32	27	16			
55	10	10	15	12	16			
56	15	15	25	18	17			
57	20	15	30	22	17			
58	25	10	25	20	17			
59	15	8	15	13	17			
60	12	10	16	13	17			
61	30	20	42	31	18			
62	10	10	12	11	18			
63	15	5	26	15	18			
64	20		50	35	19			
65	15	10	20	15	19			
66	25	15	28	23	20			
67	20	10	35	22	20			
68	35		40	38	21			
69	35	15	50	33	21			
70	35	10	30	20	21			
71	15	8	20	14	22			
72	15	25	18	19	22			
73	25	50	42	39	23			
74	30	20	40	30	23	20%-29%	42%	
75	30	15	33	26	24			
76	20	10	55	28	24			
77	12	15	27	18	24			
78	35	15	30	27	24			
79	30	20	41	30	25			
80	15	15	28	19	25			
81	35	10	41	29	26			
82	30	20	45	32	27			.937 (95% CI: 0.893-0.965)
83	20	20	20	20	28			
84	45	40	54	46	28			
85	35	20	30	28	30			
86	25	20	40	28	30			
87	25	20	30	25	33			
88	50	35	51	45	36			
89	35	15	40	30	38			
90	30	20	40	30	43			
91	25	15	30	23	45			
92	75	60	70	68	51	30%-100%	75%	
93	85	70	97	84	51			
94	90	80	90	87	51			
95	75	60	95	77	52			
96	95	80	90	88	57			
97	95	80	98	91	61			
98	90	85	95	90	71			
99	95	80	95	90	79			
100	95	90	98	94	93			

Fig. 3. Heat map of Ki67 scores. Rows represent cases and columns represent observers. Cases are ordered in ascending order by KiQuant values. Blue color gradients indicate Ki67 score ranges (0–2.6%, 2.7–13%, 14–19%, 20–29%, 30–100%). The percentage of agreement between KiQuant and MS (manual scoring represented by the average value among the three observers) is indicated for each Ki67 score range. Intraclass correlation coefficient (ICC) among the three observers and 95% confidence interval (CI) is shown for each Ki67 score range.
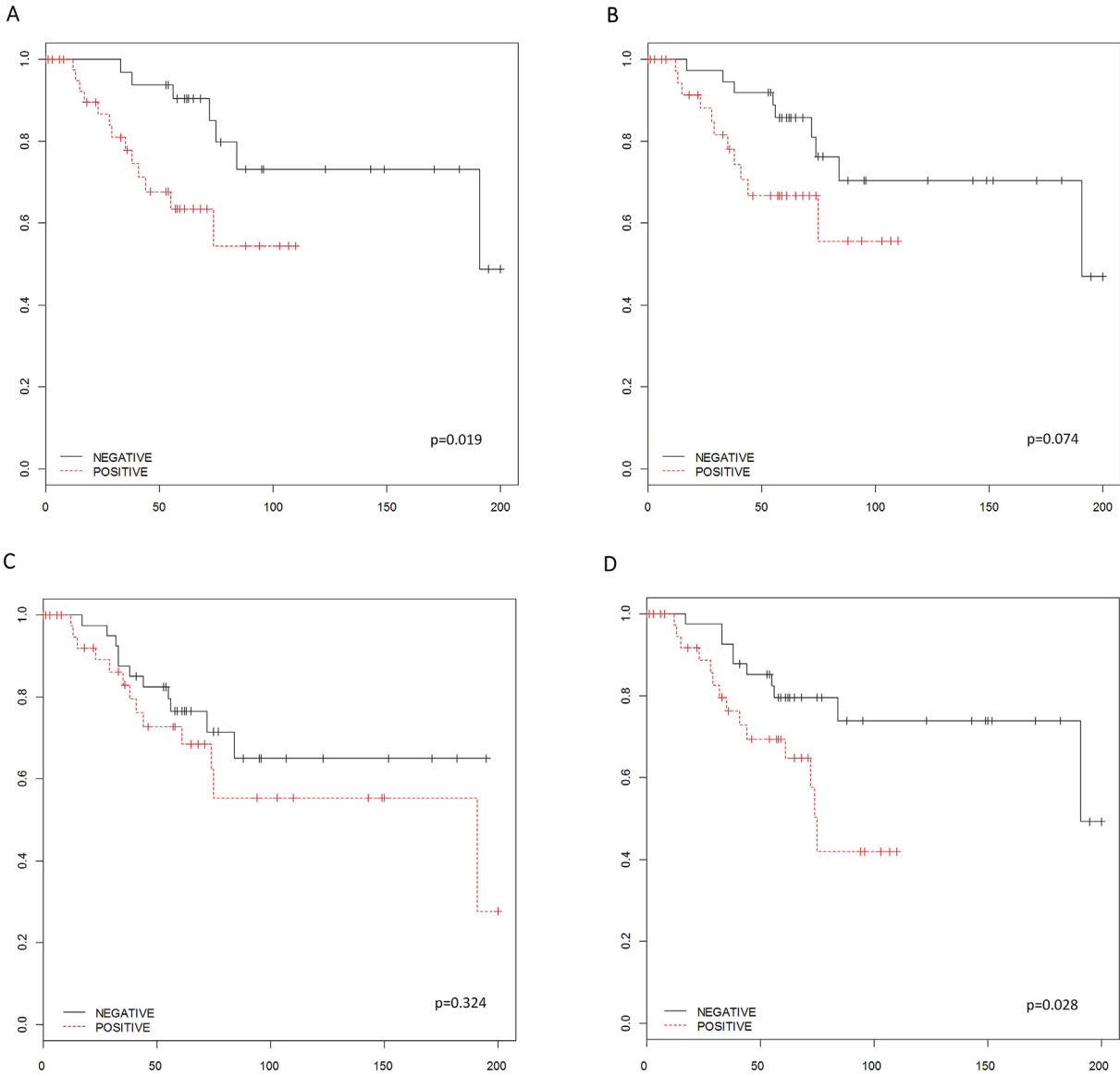

Fig. 4. Kaplan-Meier curves of overall survival according to Ki67 scores determined by three different observers (A, B and C) and KiQuant (D). Negative (black) and positive (red) lines correspond to patients having a Ki67 LI less or above the median Ki67 value, respectively. *P*-values are from the Log-rank test.

Figure S3).

4. Discussion

Ki67 is a useful biomarker for risk stratification, helps to differentiate luminal A- and B-type tumors, and to decide end-of-neoadjuvant-treatment endpoint in clinical trials, thus providing predictive and prognostic information in BC [19,34—38]. However, its implementation in the clinical setting has been hampered by the high technical and interpretation variability, and most significantly, the poor reproducibility across operators and laboratories [15,39]. The European Society for Medical Oncology (ESMO) and the American Society of Clinical Oncology (ASCO) have concluded that Ki67 would be a useful clinical tool if standardized [40,41]. Meanwhile, the use of multigene predictors has given clinicians a more accurate methodology for risk stratification [22—24]. DIA platforms using machine-learning (ML) methods have been proposed as automated systems for Ki67 LI scoring. A recent study comparing different software packages showed an excellent agreement across

the different DIA platforms (ICC: 0.933) which suggests that DIA can be standardized to give highly reproducible, platform-independent Ki67 LI automatic evaluation ([42—44].

In this study, we proposed a fairly easily implementable work-flow named KiQuant for Ki67 LI assessment which showed excellent reproducibility with reference pathologist standard (ICC:0.905) and satisfactory analytical reproducibility (inter-run ICC:0.917; inter-antibody ICC:0.886). KiQuant shows a series of important features that could make standardization of Ki67 interpretation simple and effective. The novelty of our analysis work-flow is the ability to fuse the information contained in two different images derived from sequential IHC on the same tissue slide into one single virtual dual staining (VDS) allowing to superimpose epithelial tumor areas, obtained from cytokeratin, and positive cells information, from Ki67 staining. Previous studies using VDS on serial sections for Ki67 LI determination [45,46] found a high correlation between DIA and MS using both TMA and whole slide analyses. However, a high VDS failure rate (24%) was reported [46] due to the unsuccessful alignment of the Ki67-and cytokeratin-

stained serial sections. By using the same slide, our workflow overcomes many of the factors that can cause VDS misalignments, such as differences between cuts, folding, and twisting. As a matter of fact, in our study, none of the cores was excluded due to alignment issues. Additional advantages of KiQuant over conventional VDS protocols are: a) to preserve material for additional biomarkers staining (useful for example for small biopsy samples with scant material), b) to exploit previously Ki67 stained slides through the application of a mask to the original immunostaining or c) to store a smaller number of slides in the laboratories' archives.

In our analysis, we observed that KiQuant returned lower Ki67 LI counts as compared to pathologists' scoring. This was already described in other studies using DIA [46] and it is very likely related to pathologist's underestimation of the negative over the positive tumor nuclei. We found that this imprecision increased with the magnitude of Ki67 LI. This might also explain the significantly higher difference between DIA and MS in HER2+ and TNBC (high Ki67 LI) compared to HR+ (low Ki67 LI) BC subtypes.

Our analyses were conducted using TMA cores. The KiQuant workflow is not significantly different in whole sections. However, as the size of the scored tumor area affected imprecision, we expect that Ki67 heterogeneity and the choice of the scoring area in the whole section may negatively impact on the correlation between DIA and manual scoring in whole sections [21].

Ki67 cut-offs are needed for proper patient stratification and treatment decision. However, there is no absolute agreement regarding cut-off points. It has been recommended that each pathology department should set its most appropriate cut-off points [14]. A 20% cut-off was recommended for distinguishing between Luminal A-like and Luminal B-like tumor types [19]. A recent meta-analysis concluded that a Ki67 level of over 25% is associated with a worse prognosis [47]. In our cohort, we observed a low concordance among observers below the 20% Ki67 LI (ICC: 0.606) and agreement rates between KiQuant and MS within the clinically relevant 14% and 30% cut-off points ranged from 33% to 44% (Fig. 3). In this scenario, automated approaches like KiQuant might improve Ki67 LI reproducibility, specifically around the grey zone area of Ki67 LI of 10–30% where a high level of inter-observer variability has been documented [15,16,48–51]. Importantly, our study found a significant correlation between high Ki67 by KiQuant and worse survival in both the overall cohort and the luminal BC subtype. This prognostic association could be confirmed only for one out of three pathologists' Ki67 assessment. These data, beside showing KiQuant as a accurate method to stratify BC into good and unfavorable prognostic groups, support the value of such immunohistochemical-based test that, if appropriately performed and standardized, may provide an easy and cheap alternative to more expensive genomic-based prognostic assays.

Our study has some limitations. Analyses were conducted using TMA cores instead of whole sections where the much more complex histology and biological heterogeneity may impact on the correlation between DIA and MS. The segmentation using cytokeratin-based mask may not be effective in rarer breast cancer types, such as metaplastic and medullary carcinomas. The performance of our DIA algorithm was tested only on images acquired using a single platform. Lastly, the analysis workflow is not fully automatic, as it requires the supervision (although limited) of a pathologist.

5. Conclusions

To the best of our knowledge, our study is the first one to investigate reproducibility and prognostic potential between standard and DIA Ki67 LI using sequential IHC on a single slide. Our method is technically feasible and potentially useful for both diagnostic and research use as it relies on a slide scanner and an image analysis software which are today available in many pathology departments. KiQuant may improve the standardization of Ki67 by overcoming some of the factors that determine poor reproducibility of Ki67 assessment, such as the selection of the scoring area and manual counting thus increasing the confidence of oncologists toward the use of Ki67 in the clinical setting for treatment recommendations.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.breast.2020.07.002.

References

[1] Gerdes J, Lemke H, Baisch H, Wacker HH, Schwab U, Stein H. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. J Immunol 1984;133:1710–5.

[2] Luo G, Hu Y, Zhang Z, Wang P, Luo Z, Lin J, et al. Clinicopathologic significance and prognostic value of Ki-67 expression in patients with gastric cancer: a meta-analysis. Oncotarget 2017;8:50273–83. https://doi.org/10.18632/oncotarget.17305.

[3] de Azambuja E, Cardoso F, de Castro G, Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. Br J Canc 2007;96:1504–13. https://doi.org/10.1038/sj.bjc.6603756.

[4] Jones RL, Salter J, A'Hern R, Nerurkar A, Parton M, Reis-Filho JS, et al. The prognostic significance of Ki67 before and after neoadjuvant chemotherapy in breast cancer. Breast Canc Res Treat 2009;116:53–68. https://doi.org/10.1007/s10549-008-0081-7.

[5] Zhao W-Y, Xu J, Wang M, Zhang Z-Z, Tu L, Wang C-J, et al. Prognostic value of Ki67 index in gastrointestinal stromal tumors. Int J Clin Exp Pathol 2014;7:2298–304.

[6] Yamaguchi T, Fujimori T, Tomita S, Ichikawa K, Mitomi H, Ohno K, et al. Clinical validation of the gastrointestinal NET grading system: Ki67 index criteria of the WHO 2010 classification is appropriate to predict metastasis or recurrence. Diagn Pathol 2013;8:752. https://doi.org/10.1186/1746-1596-8-65.

[7] Pollack A, DeSilvio M, Khor L-Y, Li R, Al-Saleem TI, Hammond ME, et al. Ki-67 staining is a strong predictor of distant metastasis and mortality for men with prostate cancer treated with radiotherapy plus androgen deprivation: radiation therapy Oncology group trial 92–02. J Clin Oncol 2004;22:2133–40.

[8] Szentkúti G, Dános K, Brauswetter D, Kiszner G, Krenács T, Csákó L, et al. Correlations between prognosis and regional biomarker profiles in head and neck squamous cell carcinomas. Pathol Oncol Res 2015;21:643–50. https://doi.org/10.1007/s12253-014-9869-4.

[9] Pelosi G, Bresaola E, Bogina G, Pasini F, Rodella S, Castelli P, et al. Endocrine tumors of the pancreas: Ki-67 immunoreactivity on paraffin sections is an independent predictor for malignancy: a comparative study with proliferating-cell nuclear antigen and progesterone receptor protein immunostaining, mitotic index, and other clinicopathologic variables. Hum Pathol 1996;27:1124–34.

[10] Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol 2010;11:174–83. https://doi.org/10.1016/S1470-2045(09)70262-1.

[11] Viale G, Giobbie-Hurder A, Regan MM, Coates AS, Mastropasqua MG, Dell'Orto P, et al. Prognostic and predictive value of centrally reviewed ki-67 labeling index in postmenopausal women with endocrine-responsive breast cancer: results from breast international group trial 1-98 comparing adjuvant tamoxifen with letrozole. J Clin Oncol 2008;26:5569–75. https://doi.org/10.1200/JCO.2008.17.0829.

[12] Kurozumi SI, Yamaguchi Y, Matsumoto H, Kurosumi M, Hayashi S, Fujii T, et al. Utility of Ki67 labeling index, cyclin D1 expression, and ER-activity level in postmenopausal ER-positive and HER2-negative breast cancer with neoadjuvant chemo-endocrine therapy. 2019. https://doi.org/10.1371/journal.pone.0217279.

[13] Ács B, Zámbó V, Vízkeleti L, Szász AM, Madaras L, Szentmártoni G, et al. Ki-67 as a controversial predictive and prognostic marker in breast cancer patients treated with neoadjuvant chemotherapy. Diagn Pathol 2017;12:20. https://doi.org/10.1186/s13000-017-0608-5.

[14] Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in breast cancer working group. JNCI J Natl Cancer Inst 2011;103:1656–64. https://doi.org/10.1093/jnci/djr393.

[15] Polley M-YC, Leung SCY, Gao D, Mastropasqua MG, Zabaglo LA, Bartlett JMS, et al. An international study to increase concordance in Ki67 scoring. Mod Pathol 2015;28:778–86. https://doi.org/10.1038/modpathol.2015.38.

[16] Shui R, Yu B, Bi R, Yang F, Yang W. An interobserver reproducibility analysis of Ki67 visual assessment in breast cancer. PloS One 2015;10:e0125131. https://doi.org/10.1371/journal.pone.0125131.

[17] Christgen M, von Ahsen S, Christgen H, Länger F, Kreipe H. The region-of-interest size impacts on Ki67 quantification by computer-assisted image analysis in breast cancer. Hum Pathol 2015;46:1341–9. https://doi.org/10.1016/j.humpath.2015.05.016.

[18] Leung SCY, Nielsen TO, Zabaglo LA, Arun I, Badve SS, Bane AL, et al. Analytical validation of a standardised scoring protocol for Ki67 immunohistochemistry on breast cancer excision whole sections: an international multicentre collaboration. Histopathology 2019;75:225–35. https://doi.org/10.1111/his.13880.

[19] Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the st gallen international expert consensus on the primary therapy of early breast cancer 2013. Ann Oncol 2013;24:2206–23. https://doi.org/10.1093/annonc/mdt303.

[20] Falck A-K, Fernö M, Bendahl P-O, Rydén L. St Gallen molecular subtypes in primary breast cancer and matched lymph node metastases - aspects on distribution and prognosis for patients with luminal A tumours: results from a prospective randomised trial. BMC Canc 2013;13:558. https://doi.org/10.1186/1471-2407-13-558.

[21] Focke CM, van Diest PJ, Decker T. St Gallen 2015 subtyping of luminal breast cancers: impact of different Ki67-based proliferation assessment methods. Breast Canc Res Treat 2016;159:257–63. https://doi.org/10.1007/s10549-016-3950-5.

[22] Dowsett M, Cuzick J, Wale C, Forbes J, Mallon EA, Salter J, et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. J Clin Oncol 2010;28:1829–34. https://doi.org/10.1200/JCO.2009.24.4798.

[23] Sestak I, Cuzick J, Dowsett M, Lopez-Knowles E, Filipits M, Dubsky P, et al. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian breast and colorectal cancer study group 8 and Arimidex, Tamoxifen Alone or in combination randomized trials using the PAM50 risk of recurrence score. J Clin Oncol 2015;33:916–22. https://doi.org/10.1200/JCO.2014.55.6894.

[24] Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. Lancet Oncol 2010;11:55–65. https://doi.org/10.1016/S1470-2045(09)70314-6.

[25] Ács B, Madaras L, Kovács KA, Micsik T, Tőkés A-M, Győrffy B, et al. Reproducibility and prognostic potential of ki-67 proliferation index when comparing digital-image analysis with standard semi-quantitative evaluation in breast cancer. Pathol Oncol Res 2018;24:115–27. https://doi.org/10.1007/s12253-017-0220-8.

[26] Stålhammar G, Robertson S, Wedlund L, Lippert M, Rantalainen M, Bergh J, et al. Digital image analysis of Ki67 in hot spots is superior to both manual Ki67 and mitotic counts in breast cancer. Histopathology 2018;72:974–89. https://doi.org/10.1111/his.13452.

[27] Stålhammar G, Fuentes Martinez N, Lippert M, Tobin NP, Mølholm I, Kis L, et al. Digital image analysis outperforms manual biomarker assessment in breast cancer. Mod Pathol 2016;29:318–29. https://doi.org/10.1038/modpathol.2016.34.

[28] Klauschen F, Wienert S, Schmitt WD, Loibl S, Gerber B, Blohmer J-U, et al. Standardized Ki67 diagnostics using automated scoring–clinical validation in the GeparTrio breast cancer study. Clin Canc Res 2015;21:3651–7. https://doi.org/10.1158/1078-0432.CCR-14-1283.

[29] Zhong F, Bi R, Yu B, Yang F, Yang W, Shui R. A Comparison of visual assessment and automated digital image analysis of Ki67 labeling index in breast cancer. PloS One 2016;11:e0150505. https://doi.org/10.1371/journal.pone.0150505.

[30] Koopman T, Buikema HJ, Hollema H, de Bock GH, van der Vegt B. Digital image analysis of Ki67 proliferation index in breast cancer using virtual dual staining on whole tissue sections: clinical validation and inter-platform agreement. Breast Canc Res Treat 2018;169:33–42. https://doi.org/10.1007/s10549-018-4669-2.

[31] Røge R, Riber-Hansen R, Nielsen S, Vyberg M. Proliferation assessment in breast carcinomas using digital image analysis based on virtual Ki67/cytokeratin double staining. Breast Canc Res Treat 2016;158:11–9. https://doi.org/10.1007/s10549-016-3852-6.

[32] Joseph J, Roudier MP, Narayanan PL, Augulis R, Ros VR, Pritchard A, et al. Proliferation Tumour Marker Network (PTM-NET) for the identification of tumour region in Ki67 stained breast cancer whole slide images. Sci Rep 2019;9:12845. https://doi.org/10.1038/s41598-019-49139-4.

[33] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

[34] Stuart-Harris R, Caldas C, Pinder SE, Pharoah P. Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. Breast 2008;17:323–34. https://doi.org/10.1016/j.breast.2008.02.002.

[35] Criscitiello C, Disalvatore D, De Laurentiis M, Gelao L, Fumagalli L, Locatelli M, et al. High Ki-67 score is indicative of a greater benefit from adjuvant chemotherapy when added to endocrine therapy in Luminal B HER2 negative and node-positive breast cancer. Breast 2014;23:69–75. https://doi.org/10.1016/j.breast.2013.11.007.

[36] Brown JR, DiGiovanna MP, Killelea B, Lannin DR, Rimm DL. Quantitative assessment Ki-67 score for prediction of response to neoadjuvant chemotherapy in breast cancer. Lab Invest 2014;94:98–106. https://doi.org/10.1038/labinvest.2013.128.

[37] Ács B, Zámbó V, Vízkeleti L, Szász AM, Madaras L, Szentmártoni G, et al. Ki-67 as a controversial predictive and prognostic marker in breast cancer patients treated with neoadjuvant chemotherapy. Diagn Pathol 2017;12:20. https://doi.org/10.1186/s13000-017-0608-5.

[38] Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor. J Clin Oncol 2011;29:4273–8. https://doi.org/10.1200/JCO.2010.31.2835.

[39] Polley M-YC, Leung SCY, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. An international Ki67 reproducibility study. JNCI J Natl Cancer Inst 2013;105:1897–906. https://doi.org/10.1093/jnci/djt306.

[40] Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: recommendations from the international Ki67 in breast cancer working group. JNCI J Natl Cancer Inst 2011;103:1656–64. https://doi.org/10.1093/jnci/djr393.

[41] Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26. https://doi.org/10.1093/annonc/mdv298. v8–30.

[42] Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: open source software for digital pathology image analysis. Sci Rep 2017;7:16878. https://doi.org/10.1038/s41598-017-17204-5.

[43] Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution, vol. 23; 2001.

[44] Acs B, Pelekanou V, Bai Y, Martinez-Morilla S, Toki M, Leung SCY, et al. Ki67 reproducibility using digital image analysis: an inter-platform and inter-operator study. Lab Invest 2019;99:107–17. https://doi.org/10.1038/s41374-018-0123-7.

[45] Røge R, Riber-Hansen R, Nielsen S, Vyberg M. Proliferation assessment in breast carcinomas using digital image analysis based on virtual Ki67/cytokeratin double staining. Breast Canc Res Treat 2016;158:11–9. https://doi.org/10.1007/s10549-016-3852-6.

[46] Koopman T, Buikema HJ, Hollema H, de Bock GH, van der Vegt B. Digital image analysis of Ki67 proliferation index in breast cancer using virtual dual staining on whole tissue sections: clinical validation and inter-platform agreement. Breast Canc Res Treat 2018;169:33–42. https://doi.org/10.1007/s10549-018-4669-2.

[47] Petrelli F, Viale G, Cabiddu M, Barni S. Prognostic value of different cut-off levels of Ki-67 in breast cancer: a systematic review and meta-analysis of 64,196 patients. Breast Canc Res Treat 2015;153:477–91. https://doi.org/10.1007/s10549-015-3559-0.

[48] Varga Z, Diebold J, Dommann-Scherrer C, Frick H, Kaup D, Noske A, et al. How reliable is ki-67 immunohistochemistry in grade 2 breast carcinomas? A qa

study of the Swiss working group of breast- and gynecopathologists. PloS One 2012;7:e37379. https://doi.org/10.1371/journal.pone.0037379.

[49] Cserni G, Vörös A, Liepniece-Karele I, Bianchi S, Vezzosi V, Grabau D, et al. Distribution pattern of the Ki67 labelling index in breast cancer and its implications for choosing cut-off values. Breast 2014;23:259–63. https://doi.org/10.1016/j.breast.2014.02.003.

[50] Laenkholm A-V, Grabau D, Møller Talman M-L, Balslev E, Bak Jylling AM, Tabor TP, et al. An inter-observer Ki67 reproducibility study applying two

different assessment methods: on behalf of the Danish Scientific Committee of Pathology, Danish breast cancer cooperative group (DBCG). Acta Oncol (Madr) 2018;57:83–9. https://doi.org/10.1080/0284186X.2017.1404127.

[51] Hida AI, Oshiro Y, Inoue H, Kawaguchi H, Yamashita N, Moriya T. Visual assessment of Ki67 at a glance is an easy method to exclude many luminal-type breast cancers from counting 1000 cells. Breast Cancer 2015;22: 129–34. https://doi.org/10.1007/s12282-013-0460-8.

4.2- Study 2:

"Immune microenvironment characterisation and dynamics during anti-HER2-based neoadjuvant treatment in HER2-positive breast cancer"

In this study we validated a second panel of NGI composed of different t-cell subpopulations to analyze the immune microenvironment during anti-HER2- based neoadjuvant treatment in HER2-positive breast cancers. We quantified the densities of total t-cells (CD3+), regulatory t-cells (CD3+FOXP3+), helper t-cells (CD3+CD4+), and cytotoxic t-cells (CD3+CD8+) in the tumor sample, obtained the proliferating percentage of each of the subtypes and analyzed their spatial distribution in intratumoral, proximal and distant areas in the tissue.

ARTICLE   OPEN

Immune microenvironment characterisation and dynamics during anti-HER2-based neoadjuvant treatment in HER2-positive breast cancer

G. Griguolo [1,2,3,20], G. Serna [4,20], T. Pascual [3,5,6], R. Fasani[4], X. Guardia[4], N. Chic[3,5], L. Paré[6], S. Pernas[7], M. Muñoz[5], M. Oliveira [8,9], M. Vidal [5], A. Llombart-Cussac[10], J. Cortés[11], P. Galván[3], B. Bermejo[12], N. Martínez[13], R. López[14], S. Morales [15], I. Garau[16], L. Manso [17], J. Alarcón[18], E. Martínez[19], P. Villagrasa[6], A. Prat [3,5,6,21✉] and P. Nuciforo[4,6,21✉]

Despite their recognised role in HER2-positive (HER2+) breast cancer (BC), the composition, localisation and functional orientation of immune cells within tumour microenvironment, as well as its dynamics during anti-HER2 treatment, is largely unknown. We here investigate changes in tumour-immune contexture, as assessed by stromal tumour-infiltrating lymphocytes (sTILs) and by multiplexed spatial cellular phenotyping, during treatment with lapatinib-trastuzumab in HER2+ BC patients (PAMELA trial). Moreover, we evaluate the relationship of tumour-immune contexture with hormone receptor status, intrinsic subtype and immune-related gene expression. sTIL levels increase after 2 weeks of HER2 blockade in HR-negative disease and HER2-enriched subtype. This is linked to a concomitant increase in cell density of all four immune subpopulations (CD3+, CD4+, CD8+, Foxp3+). Moreover, immune contexture analysis showed that immune cells spatially interacting with tumour cells have the strongest association with response to anti-HER2 treatment. Subsequently, sTILs consistently decrease at the surgery in patients achieving pathologic complete response, whereas most residual tumours at surgery remain inflamed, possibly reflecting a progressive loss of function of T cells. Understanding the features of the resulting tumour immunosuppressive microenvironment has crucial implications for the design of new strategies to de-escalate or escalate systemic therapy in early-stage HER2+ BC.

INTRODUCTION

The host immune system has an important role in HER2-positive (HER2+) breast cancer (BC). Prior studies have revealed that ~55% of HER2+ tumours have >10% of stromal tumour-infiltrating lymphocytes (sTILs)[1]. From a clinical point of view, TILs are associated with better survival outcomes in HER2+ early and advanced BC[1–3], higher pathological complete response (pCR) rates after neoadjuvant anti-HER2-based chemotherapy[2,4–7] and higher response to trastuzumab plus pembrolizumab in the advanced setting[8]. Thus, baseline TILs in HER2+ disease determine prognosis and might contribute to the therapeutic effects of anti-HER2-based treatments[9,10].

Despite the recognised role of immune cells in HER2+ BC, the composition, localisation and functional orientation of immune cells within the tumour microenvironment (jointly referred to as immune contexture), as well as the dynamics of TILs during and after anti-HER2 treatment, are largely unknown. Limited and inconsistent evidence is available regarding the prognostic impact of TILs in residual tumours following neoadjuvant anti-HER2-based chemotherapy. High TIL levels in residual disease have been associated both with better outcome[11], worse outcome[12] and no

impact on prognosis[13]. A study by Ladoire S. and colleagues evaluated the prognostic impact of different lymphocytic sub-populations; it reported high CD8 and low Foxp3 cell infiltrates after chemotherapy to be significantly associated with improved long-term outcome[14]. This study thus further highlights the need for a more comprehensive evaluation.

The current treatment standard of early-stage HER2+ BC is anti-HER2-therapy plus chemotherapy. Thus, prior studies have not been able to dissect whether the observed changes in the immune microenvironment are owing to chemotherapy, anti-HER2 therapy, or both. Studies without chemotherapy are the ideal scenario to address the specific role of anti-HER2 therapy[15,16]. The neoadjuvant PAMELA trial (SOLTI-1114)[15] treated 151 patients with HER2+ BC with trastuzumab and lapatinib (and endocrine therapy if the tumour was hormone receptor [HR] positive) for 18 weeks. In this study, sTILs at baseline and at day 15 differed significantly according to PAM50 intrinsic subtype. Moreover, sTILs at baseline were found significantly associated with pCR in the univariable analysis but not in multivariable analysis. At day 15, a significant increase in sTILs was observed in most patients; also, sTILs at day 15 were found independently associated with pCR at multivariate analysis[17].

With this background, several questions can be addressed from the PAMELA trial: (1) which cells compose immune infiltrate in early HER2+ BC and how are they interacting with tumour cells? (2) how does this relate to the probability of achieving a pCR?; (3) which subgroup of patients increases sTILs after 2 weeks of priming with HER2-targeted treatment?; (4) how do these changes relate to the probability of achieving a pCR?; (5) how do immune contexture changes after anti-HER2 priming relate to the probability of achieving a pCR?; (6) how are sTILs expressed at surgery following neoadjuvant treatment? and (7) how is the presence of sTILs associated with immune-related gene expression?

We here investigated changes in the tumour-immune microenvironment following treatment with lapatinib and trastuzumab and the relationship of sTILs with HR status, intrinsic subtype and immune-related gene expression in patients with HER2+ BC from the PAMELA trial. Moreover, we assessed immune contexture at baseline and day 15 in patients with available samples by multiplexed spatial cellular phenotyping (REMARK diagram, Supplementary Fig. 1). The results from this analysis might help design new strategies to de-escalate or escalate systemic therapy in HER2+ early BC.

RESULTS

A multiplexed imaging assay for immune microenvironment characterisation

We developed a multiplexed immunohistochemistry (IHC) work-flow (named next-generation IHC or next-generation impactor

(NGI)) comparable with that of conventional IHC. This workflow is based upon iterative cycles of staining and destaining of the same slide with different primary antibodies, individual slide digitalisation, virtual multiplexed digital image reconstruction and complex image analyses (Fig. 1a). A six-plex panel was specifically designed to interrogate the tumour-immune microenvironment and included a tumour-related protein (cytokeratin), a functional marker for proliferation (Ki67) and four immune-related T-cell lineage markers (CD3, CD4, CD8 and Foxp3). The latter was selected as established markers of T cells with an effector (CD3+ CD8+) and suppressor/regulatory (CD3+CD4+Foxp3+) functions.

The antibody staining of each target was initially optimised using two different tissue controls (a normal tonsil and a BC) and further validated on an independent series of HER2+ BC stained by regular IHC to ensure consistent results among methods (Supplementary Figs. 2 and 3). We evaluated marker specificity by matching the staining obtained with each antibody to the known histologic distribution and to staining with a conventional IHC protocol: Ki67-stained cells in the germinal centre, and CD3-, CD4- and CD8-stained cells predominantly in the mantle zone (Fig. 1b, c). Furthermore, we evaluated the subcellular localisation of the biomarkers and confirmed that staining of the transcription factors Ki67 and Foxp3 were nuclear, whereas CD3, CD4 and CD8 were membrane.

The individual digitalised IHC images were aligned to obtain a virtual multiplexed image (Fig. 1d), which was analysed using
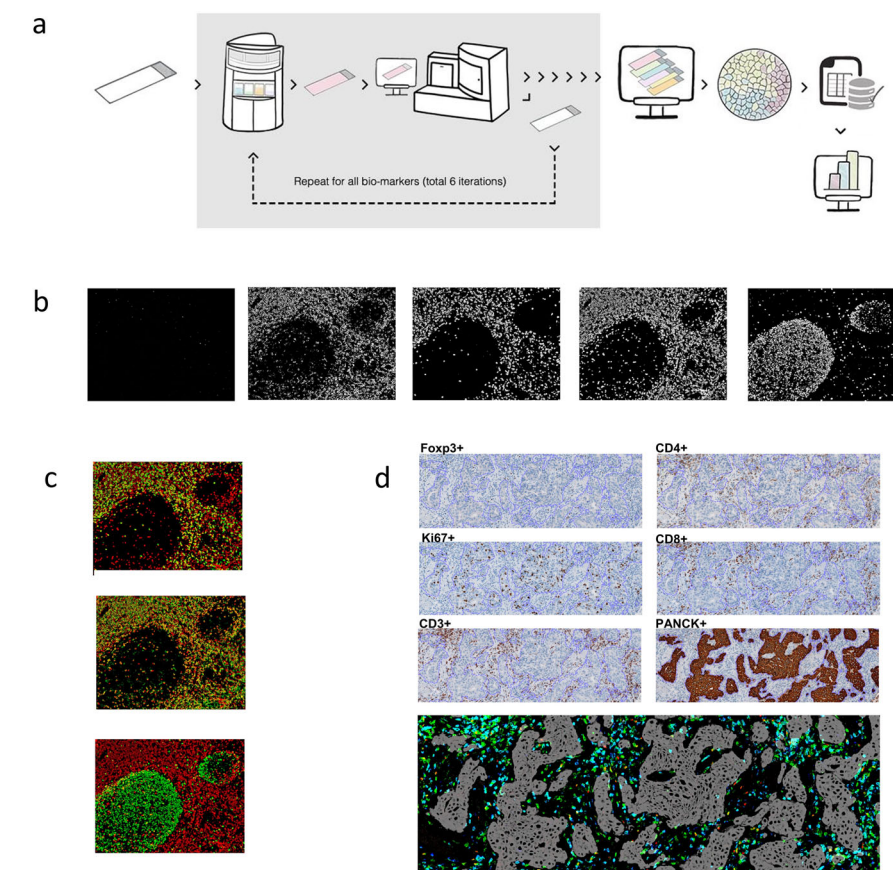
Fig. 1 Multiplexed imaging assay. a Next-generation immunohistochemistry (NGI) workflow. An FFPE tissue section is stained, scanned and destained six times. All the scanned images are aligned, image analysis is done to obtain the data and after doing all the quality check controls, the data are analysed to obtain the final results. **b** Representative colour deconvoluted images of different biomarkers in the tonsil. From left to right Foxp3+, CD3+, CD8+, CD4+ and Ki67+. Images at 6×. **c** Colour overlays of different biomarkers in the tonsil (CD3+ in red, CD8+, CD4+ and Ki67+ in green from top to bottom). Images at 6×. **d** A representative example of all the stainings (Foxp3, CD4, Ki67, CD8, CD3 and cytokeratin) in breast cancer samples and virtual image reconstruction of some of them by assigning virtual colours to the deconvoluted images. Foxp3+ in red, CD8+ in blue, CD3+ in green and cytokeratin in grey. The fine purple line in each image marks the tumour borders. Images at 5×.

[1]Department of Surgery, Oncology and Gastroenterology, University of Padova, Padova, Italy. [2]Division of Oncology 2, Istituto Oncologico Veneto IRCCS, Padova, Italy. [3]Translational Genomics and Targeted Therapeutics in Solid Tumors, August Pi i Sunyer Biomedical Research Institute (IDIBAPS), Barcelona, Spain. [4]Molecular oncology group, Vall d´Hebron Institute of Oncology, Barcelona, Spain. [5]Department of Medical Oncology, Hospital Clínic de Barcelona, Barcelona, Spain. [6]SOLTI Breast Cancer Research Group, Barcelona, Spain. [7]Institut Catala d'Oncologia-H.U.Bellvitge-IDIBELL, Hospitalet, Barcelona, Spain. [8]Medical Oncology Department, Vall d'Hebrón University Hospital, Barcelona, Spain. [9]Breast Cancer and Melanoma Group, Vall d'Hebron Institute of Oncology, Barcelona, Spain. [10]Hospital Universitario Arnau de Vilanova de Valencia, Valencia, Spain. [11]IOB Institute of Oncology, Quironsalud Group, Madrid & Barcelona, Spain. [12]Hospital Clínico Universitario de Valencia/INCLIVA/CIBERONC, Valencia, Spain. [13]Hospital Universitario Ramón y Cajal, Madrid, Spain. [14]Hospital Clínico Universitario de Santiago, IDIS, CIBERONC, Santiago de Compostela, Spain. [15]Hospital Universitario Arnau de Vilanova de Lleida, Lleida, Spain. [16]Hospital Son Llàtzer, Palma de Mallorca, Spain. [17]Hospital Universitario 12 de Octubre, Madrid, Spain. [18]Hospital Universitario Son Espases, Palma de Mallorca, Spain. [19]Consorcio Hospitalario Provincial de Castellón, Castellón de la Plana, Spain. [20]These authors contributed equally: G. Griguolo, G. Serna. [21]These authors contributed equally: A. Prat, P. Nuciforo. ✉email: alprat@clinic.cat; pnuciforo@vhio.net

different image analysis algorithms developed for biomarker densities, function and spatial analyses (Supplementary Fig. 4).

Immune microenvironment contexture

A total of 231 regions of interest (ROI) from 129 unique samples had sufficient material for NGI analysis. All samples were stained with a sequential IHC workflow that included a panel of six antibodies for T cells subtyping (CD3, CD4, CD8 and Foxp3), proliferation (Ki67) and tumour recognition (cytokeratin) plus hematoxylin for counterstaining. Samples with total region of interest (ROI) below 100,000 μm were excluded from the analysis. After filter, a total of 114 samples (65 baseline and 49 day 15) from 75 patients were evaluable. An average (range) ROI of 7,764,692 μm^2 (122,074–37,742,587) was profiled using an image analysis pipeline for data extraction and analysis (see methods).

To identify the immune cell subpopulations within the samples, we clustered the immune cells by canonical markers. We found that our analytical pipeline was able to accurately classify immune cells even when those with opposite identifiers were located in close proximity to each other (Fig. 2a). Over a total of 1,217,249 cells identified, the proportions of CD3$^+$ immune subsets across all patients' samples were 47% CD8$^+$CD4$^-$Foxp3$^-$, 30% CD4$^+$Foxp3$^-$CD8$^-$, 11% CD4$^+$Foxp3$^+$CD8$^-$ and 13% CD8$^-$CD4$^-$Foxp3$^-$ (Fig. 2b).

To chart the immune landscape in HER2+ BC, we then quantified the number of immune cell populations per area across patients. We found large variability in immune cells content, which was confirmed by pathological sTILs scoring on H&E staining (Supplementary Table 1, Fig. 2c–f, Supplementary Figs. 5–6). Densities of all immune cell subtypes were all positively correlated with the number of sTILs (Spearman Rho, CD3$^+$ = 0.63, CD8$^+$ = 0.65, CD4$^+$ = 0.54, Foxp3$^+$ = 0.615, P < .0001, Supplementary Fig. 7) assessed by a board-certified pathologist according to international TILs working group recommendation[18]. Densities of all immune cell subtypes negatively correlated with the percentage of tumour area in the sample (CD3$^+$ = −0.209, P = 0.028; CD8$^+$ = −0.205, P = 0.031; CD4$^+$ = −0.184, P = 0.053, Foxp3$^+$ = −0.203, P = 0.033, Supplementary Fig. 8a).

To evaluate the spatial organisation of the tumour-immune landscape in HER2+ BC, we developed a method for assessing spatial proximity enrichment of each immune cell subtype from the tumour. We quantified the number of positive cells for each marker located within three regions of different distance from the tumour: A, intratumoural; B, proximal stroma within 30 μm and C, distal stroma >30 μm from the tumour, respectively (Fig. 2g).

Higher immune cells densities were found in the proximal peritumoural regions (B) compared with intratumoural (A) and distal peritumoural (C) locations although the difference was statistically significant only between location B and C for all immune subtypes (P values for comparison between immune cell density in area B and C, CD3 = 0.017, CD8 = 0.014, CD4 = 0.045, Foxp3 = 0.023, Fig. 2h, Supplementary Fig. 9).

To determine the level of activation of immune cells within the tumour microenvironment, we developed a method to quantify the proportion of proliferating immune cells in the stromal compartment by virtually multiplexing individual immune cell markers with Ki67 staining obtained from the same slide (Fig. 2i). The proportion of proliferating immune cells over total proliferating cells (Ki67$^+$ tumour and immune cells) was 35%, with tumour cells representing the main proliferative cell subtype within the tissue, as expected (Supplementary Fig. 10). The proportion of proliferating stromal immune cells was positively correlated with the amount of tumour area in the sample (Spearman Rho, CD3$^+$Ki67$^+$ = 0.397, CD8$^+$Ki67$^+$ = 0.331, CD4$^+$Ki67$^+$ = 0.414, Foxp3$^+$Ki67$^+$ = 0.395, P < 0.001, Supplementary Fig. 11).

Spatial analysis revealed a differential distribution according to immune cells proximity to the tumour, with a decreasing proportion of proliferating CD3$^+$ (P < 0.001), CD8$^+$ (P < 0.001),

CD4$^+$ (P < 0.001) and Foxp3$^+$ (P < 0.001) T cells from the intratumoural region (A) to the distal location (C) (Fig. 2j, Supplementary Fig. 12, Supplementary Table 2).

Immune contexture analysis according to HR status

To evaluate if the tumour-immune microenvironment differed according to HR status, we compared the composition, spatial distribution and functional activity of immune subtypes in HR+ and HR− HER2+ BC.

In the whole sample area, densities of all immune cell subtypes were significantly higher in HR− as compared with HR+ tumours [HR− median (interquartile range) CD3$^+$ 1669(1818), CD8$^+$ 689 (955), CD4$^+$ 629(884), Foxp3$^+$ 187(215); HR+ median (interquartile range) CD3$^+$ 828(863), CD8$^+$ 383(520), CD4$^+$ 361(472), Foxp3$^+$ 57 (89); P values, CD3$^+$ = 0.009, CD8$^+$ = 0.009, CD4$^+$ = 0.019 and Foxp3$^+$ < .001, Supplementary Fig. 13a]. As expected, these findings paralleled what observed for sTILs both at baseline and day 15, for which however the difference between HR− and HR+ tumours was only significant after 2 weeks of anti-HER2 treatment (median sTILs at day 15 20% vs 10% in HR− and HR+ tumours, respectively, P < 0.001).

The number of immune cells was higher in HR− compared with HR+ across all peritumoural stroma locations, whereas their intratumoural content did not differ significantly according to HR status (Supplementary Fig. 14, Supplementary Table 3).

We further investigated if immune cells activation was influenced by tumour HR status. The proportion of proliferating immune cells was higher in HR- as compared with HR+ tumours (Wilcoxon, CD3$^+$ P = 0.007; CD8$^+$ P = 0.011; CD4$^+$ P = 0.009) except for Foxp3$^+$ (P = 0.710) (Supplementary Fig. 13b, Supplementary Table 4). These differences were maintained across all peritumoural stroma locations (Supplementary Fig. 15). In HR+, a higher mean proportion of Foxp3$^+$ proliferating cells (13.2%) was observed compared with proliferating CD3$^+$ (5.6%), CD4$^+$ (6.2%) and CD8$^+$ (5.4%) (Kruskal–Wallis test, P < 0.001). The ratio of proliferating Foxp3$^+$/CD8$^+$ was significantly higher in HR+ as compared to HR− tumours (2.29 vs 1.33, P = 0.004).

Immune contexture analysis according to PAM50 molecular subtype

We have previously shown that, within HER2+ early BC, HER2-enriched PAM50 tumours present significantly higher sTILs as compared with other PAM50 subtypes, both at baseline (10% vs 5%; P = 0.006) and at day 15 (20% vs 10%; P < 0.001)[17]. Tumour-immune contexture analysis showed no significant difference in immune cell subsets densities according to PAM50 subtype at baseline (Table 1). The fraction of proliferating (Ki67$^+$) cells for all four immune cell subpopulations (CD3$^+$, CD4$^+$, CD8$^+$, Foxp3$^+$) was numerically higher in basal-like tumours, whereas luminal tumours showed the lowest fraction of proliferating cells (Table 1, Fig. 3a, Supplementary Fig. 16a). However, this difference was not statistically significant.

After 2 weeks of HER2-targeted treatment, tumours classified as HER2-enriched or Basal-like by PAM50 at baseline showed significantly higher density of all 4 immune subsets as compared to Luminal A and B tumours (Table 1, Fig. 3b, Supplementary Fig. 16b). Similar to baseline, the fraction of proliferating (Ki67$^+$) cells for all four immune cell subpopulations (CD3$^+$, CD4$^+$, CD8$^+$, Foxp3$^+$) at day 15 was numerically higher in basal-like tumours, while luminal tumours showed the lowest fraction of proliferating cells, although the difference was not statistically significant (Table 1, Fig. 3c, Supplementary Fig. 16c).

Immune contexture dynamics under anti-HER2 treatment

As previously described[17], overall sTILs levels at day 15 were significantly higher than those in paired baseline samples. When
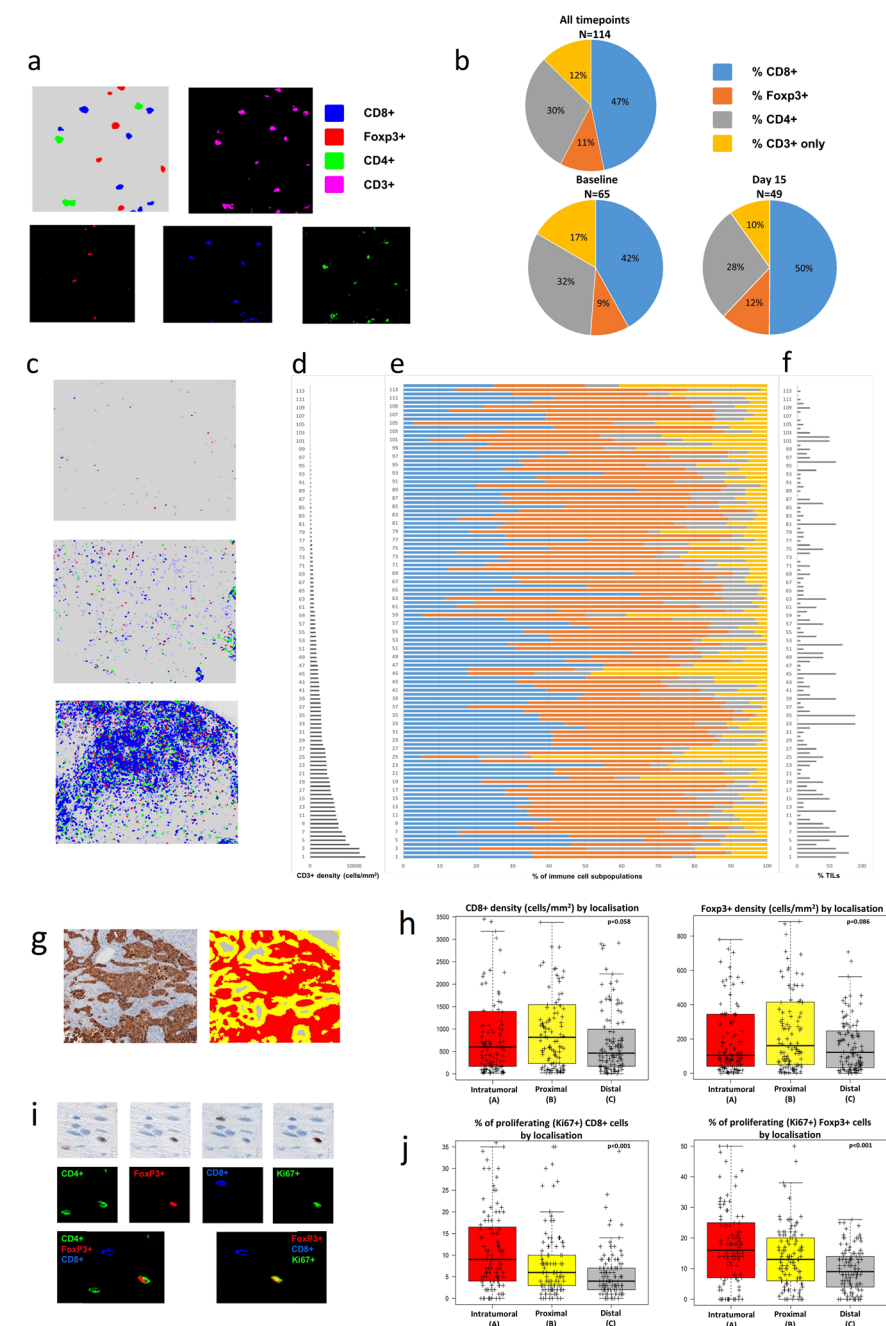
Fig. 2 Multiplexed spatial cellular phenotyping of breast cancer. a Representative example of the analytical pipeline classifying immune cells (top left) and colour deconvoluted images with red, green and blue colours assigned to Foxp3$^+$, CD4$^+$ and CD8$^+$, respectively (bottom) and magenta for CD3$^+$ (top right). Images at 25×. **b** Proportions of CD3$^+$ immune subsets across all patients' samples (up), baseline samples (down to the left) and day 15 samples (down to the right). CD8$^+$, Foxp3$^+$, CD4$^+$ and CD3$^+$-only in blue, orange, gray and yellow. **c** Representative examples of breast cancers with low (up), medium (middle) and high (down) T-cell densities. Images at 5×. **d** CD3$^+$ density results across the entire population of HER2+ breast cancers. Samples are ordered from lowest to highest. **e** Proportions of CD8$^+$ (blue), Foxp3$^+$ (orange), CD4$^+$ (gray) and CD3$^+$ only (yellow) cells for all patients' samples. **f** Stromal tumour-infiltrating lymphocytes (TILs) in breast cancer samples with available NGI data. **g** Representative example of the spatial analysis areas defined by the image analysis algorithm using cytokeratin as tumour mask. Intratumoural (**a**, in red), proximal peritumoural stroma within 30 μm (**b**, in yellow) and distal peritumoural stroma >30 μm from the tumour (**c**, in gray) regions are shown. Images at 5×. **h** Boxplots of immune cells densities (CD8$^+$ and Foxp3$^+$) according to spatial location. Boxplot legend: centre line: median; bounds of box: interquartile range (IQR); whiskers: highest and lowest value excluding outliers (Q3 + 1.5*IQR to Q1 − 1.5*IQR); markers beyond the whiskers: potential outliers. **i** Representative example of CD4$^+$, Foxp3$^+$, CD8$^+$ and Ki67$^+$ sequential immunohistochemistry and co-expression analyses for T cells activity assessment on the same tissue slide (top panel). The colour green, red, blue and green is assigned, respectively, to each individual staining for visualisation purpose, co-expression analyses and virtually multiplexed images (first image composed by CD4, Foxp3 and CD8 and second image composed by Foxp3, CD8 and Ki67; co-expression in yellow). Images at 50×. **j** Boxplots of the proportion of proliferating immune cells (CD8$^+$ and Foxp3$^+$) according to spatial location. Boxplot legend: centre line: median; bounds of box: interquartile range (IQR); whiskers: highest and lowest value excluding outliers (Q3 + 1.5*IQR to Q1 − 1.5*IQR); markers beyond the whiskers: potential outliers.

Table 1. Immune cell density at baseline and after 2 weeks of anti-HER2 treatment according to baseline PAM50 intrinsic subtype.

Immune cell density at baseline					
Immune cell Subpopulation	Immune cell density by intrinsic subtype: median (IQR)				
	Luminal A ($N = 11$)	Luminal B ($N = 6$)	HER2-enriched ($N = 43$)	Basal-like ($N = 5$)	p value
CD3$^+$	781 (200–1035)	1277 (696–1372)	1073 (367–2313)	644 (225–1921)	0.419
CD8$^+$	217 (77–383)	461 (350–481)	475 (125–1018)	166 (64–648)	0.562
CD4$^+$	352 (84–445)	525 (340–773)	507 (151–1096)	403 (144–502)	0.374
Foxp3$^+$	43 (13–93)	28 (28–83)	88 (45–202)	52 (49–176)	0.178
%Ki67$^+$CD3$^+$	6 (4–7)	5 (3–8)	9 (4–14)	8 (5–17)	0.178
%Ki67$^+$CD4$^+$	6 (4–10)	7 (1–8)	9 (4–13)	9 (6–17)	0.184
%Ki67$^+$CD8$^+$	5 (2–9)	3 (3–7)	8 (4–12)	9 (5–19)	0.188
%Ki67$^+$Foxp3$^+$	18 (14–20)	15 (0–15)	15 (9–18)	19 (15–27)	0.064
Immune cell density at day 15					
Immune cell subpopulation	Immune cell density by intrinsic subtype: median (IQR)				
	Luminal A ($N = 7$)	Luminal B ($N = 6$)	HER2-enriched ($N = 33$)	Basal-like ($N = 3$)	p value
CD3$^+$	435 (193–747)	749 (710–896)	1669 (899–2461)	1544 (501–1544)	**0.008**
CD8$^+$	296 (66–367)	442 (405–492)	769 (481–1417)	583 (275–583)	**0.007**
CD4$^+$	204 (102–322)	272 (228–295)	713 (361–1136)	629 (158–629)	**0.011**
Foxp3$^+$	44 (13–48)	37 (22–39)	214 (93–300)	195 (46–195)	**0.004**
%Ki67$^+$CD3$^+$	3 (2–3)	1 (1–1)	4 (2–7)	7 (3–7)	0.051
%Ki67$^+$CD4$^+$	3 (3–3)	1 (1–2)	5 (2–7)	10 (5–10)	0.062
%Ki67$^+$CD8$^+$	2 (1–2)	1 (1–1)	3 (2–8)	6 (2–6)	0.058
%Ki67$^+$Foxp3$^+$	7 (5–13)	2 (1–9)	7 (3–12)	15 (5–15)	0.412
Normal-like $N = 0$; significant p values in bold.					

we looked at differences according to subtype, a statistically significant increase in sTILs was observed in HR-negative ($P < 0.001$) and HER2-enriched subtype ($P = 0.001$), but not in HR-positive and non-HER2-enriched PAM50 subtypes (Fig. 4a, Table 2). Within the HER2-enriched subtype, an increase in sTILs levels was more evident in HR-negative disease (Table 2).

To identify which immune component was responsible for the increase in sTILs levels after priming with 2 weeks of anti-HER2 treatment, we analysed changes in the density of each immune cell subpopulation between these two timepoints by multiplexed spatial cellular phenotyping. We found large differences in both activity and densities of the immune cells when comparing untreated tumours with on-treatment samples. In fact, densities of all immune cells subtypes increased at day 15 [median (interquartile range) CD3$^+$ 1462(1453), CD8$^+$ 688(526), CD4$^+$ 534 (684), Foxp3$^+$ 166(219)] as compared to baseline [median (interquartile range) CD3$^+$ 832(1795), CD8$^+$ 364(841), CD4$^+$ 445 (684), Foxp3$^+$ 83(141)] tumours (P values, significant only for CD8$^+$ = 0.04, Supplementary Table 1, Supplementary Fig. 13c). When individual patient immune cell density data from the 39 patients with paired baseline-day 15 samples were considered, a significant increase in both CD8$^+$ and Foxp3$^+$ cell density was observed at day 15 (Supplementary Fig. 6).

In on-treatment samples (day 15), 50% of all immune cells were CD8$^+$ as compared with 42% of baselines samples (Fig. 2b). CD3$^+$ and CD8$^+$ immune cell densities were inversely correlated with tumour area in on-treatment (day 15) samples (Spearman's rho, CD3$^+$ −0.367 $P = 0.020$, CD8$^+$ −0.368; $P = 0.010$) but not in baseline samples (CD3$^+$ −0.060; $P = 0.620$, CD8$^+$ −0.014; $P = 0.915$; Supplementary Fig. 8b–c).

Spatial analysis revealed that the increase in the number of immune cells at day 15 was significant in the intratumoural and proximal peritumoural regions but not in the distal stromal region (Supplementary Fig. 17). Upon treatment, the number of proliferating immune cells per area uniformly decreased across all location compared to baseline pretreatment samples ($P < 0.001$ for all comparisons, Supplementary Figs. 13d and 18, Supplementary Table 5). Lower percentages of proliferating immune cells at day 15 were not significantly associated with lower tumour cellularity evaluated on the same sample (Spearman Rho, CD3$^+$Ki67$^+$ = 0.089, CD8$^+$Ki67$^+$ = 0.053, CD4$^+$Ki67$^+$ = 0.185, Foxp3$^+$Ki67$^+$ = 0.246, $P > 0.05$ for all comparisons) (Supplementary Fig. 11b, c).

Paired multiplex IHC (mIHC) data from both baseline and day 15 samples were available from 39 patients and were used to assess changes in densities of immune cell subpopulations in day 15 and baseline paired samples according to baseline PAM50 intrinsic subtype and HR status (Supplementary Table 6). Although the decrease in percentages of proliferating immune cells (all four immune subpopulations) at day 15 was consistently observed across all subgroups, a statistically significant increase in immune cells density (all four immune subpopulations) at day 15 was only observed in HR-negative and HER2-enriched subtype (all P values < 0.05 except for CD4+ in HER2-enriched tumours $P = 0.055$), but not in HR-positive and non-HER2-enriched PAM50 subtypes. Similar to prior observations of sTILs levels, the increase was numerically more evident in HR-negative disease than in HER2-enriched tumours.

Tumour-immune contexture analysis and the probability of achieving a pCR

As previously described[17], higher sTILs were significantly associated with pCR and lower residual cancer burden scores, both at baseline and after 2 weeks of anti-HER2 treatment.

In HR-positive disease or non-HER2-enriched subtype, no consistent change in sTILs at day 15 versus baseline was found
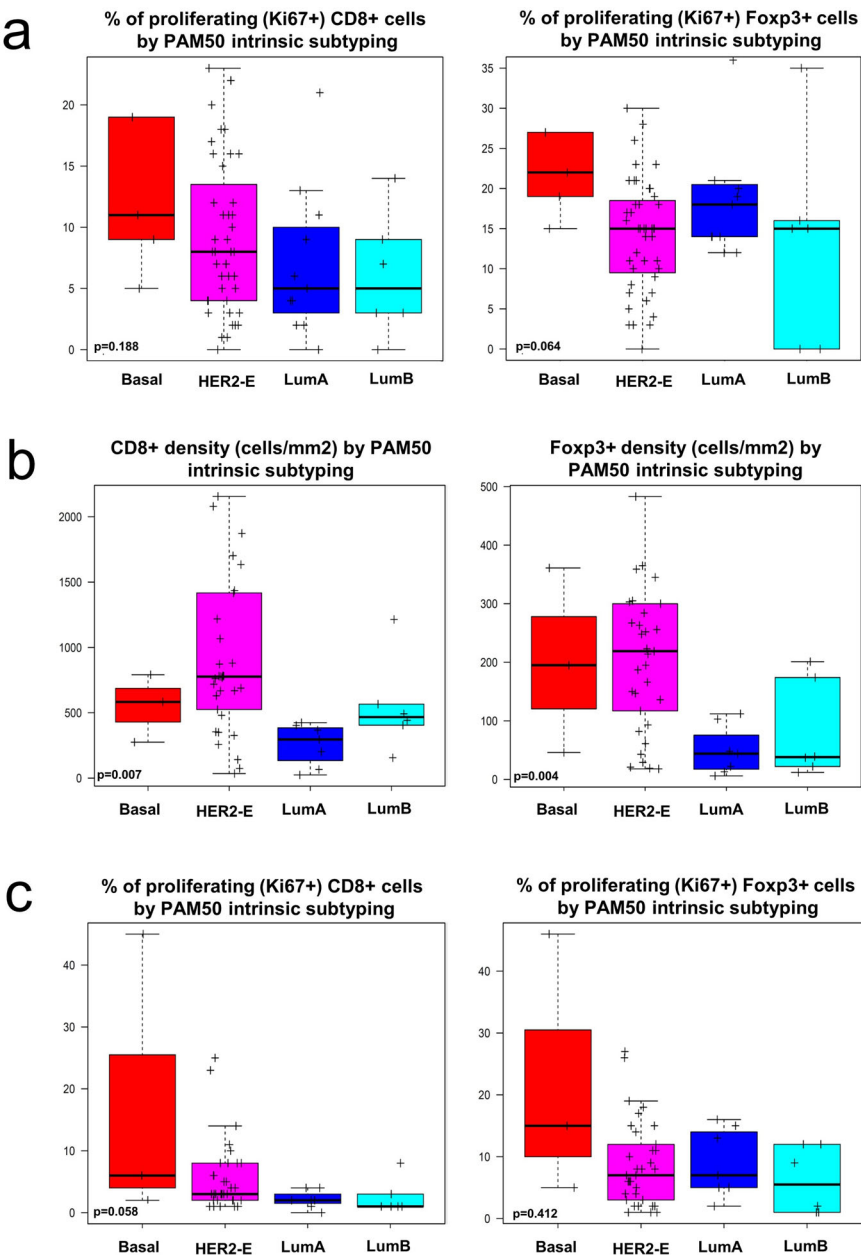
Fig. 3 Immune cell density according to intrinsic subtype. a Boxplots of the proportion of proliferating immune cells (CD8$^+$ and Foxp3$^+$) at baseline according to baseline intrinsic subtyping. **b** Boxplots of immune cells densities (CD8$^+$ and Foxp3$^+$) at day 15 according to baseline intrinsic subtyping. **c** Boxplots of the proportion of proliferating immune cells (CD8$^+$ and Foxp3$^+$) at day 15 according to baseline intrinsic subtyping. Boxplot legend: centre line: median; bounds of box: interquartile range (IQR); whiskers: highest and lowest value excluding outliers (Q3 + 1.5*IQR to Q1 − 1.5*IQR); markers beyond the whiskers: potential outliers.

according to the type of pathological response (Supplementary Table 7). Both in HER2-enriched subtype and HR-negative disease, a consistent increase in sTILs at day 15 versus baseline was found regardless of the type of pathological response (Fig. 4a and Supplementary Table 7).

To identify if the prognostic impact of sTILs might be different according to immune contexture, we then analysed the impact of immune cells composition, activity and spatial interaction with tumour cells on pCR. Both at baseline and at day 15, no significant difference in immune cell subpopulation densities was observed between tumours achieving or not achieving pCR, despite numerically higher densities of all four immune subpopulations were observed in tumours achieving pCR at day 15 (Fig. 5a).

After that, to determine whether T-cell activation status associated with the probability of pCR, we compared the rates of proliferating (Ki67$^+$) immune cells in tumours achieving or not achieving pCR. No statistically significant association was observed, although Odds ratios (OR) for % of proliferating cells and pCR were consistently higher at baseline as compared with day 15 for all immune cell subpopulations except Foxp3$^+$ (Fig. 5b).

Finally, we analysed immune cells subpopulation densities separately according to their spatial distribution. Although only some of the associations reached statistical significance, the association between higher immune cell density and pCR was consistently stronger for more proximal compartments (intratumoural and proximal peritumoural stroma) as compared with the
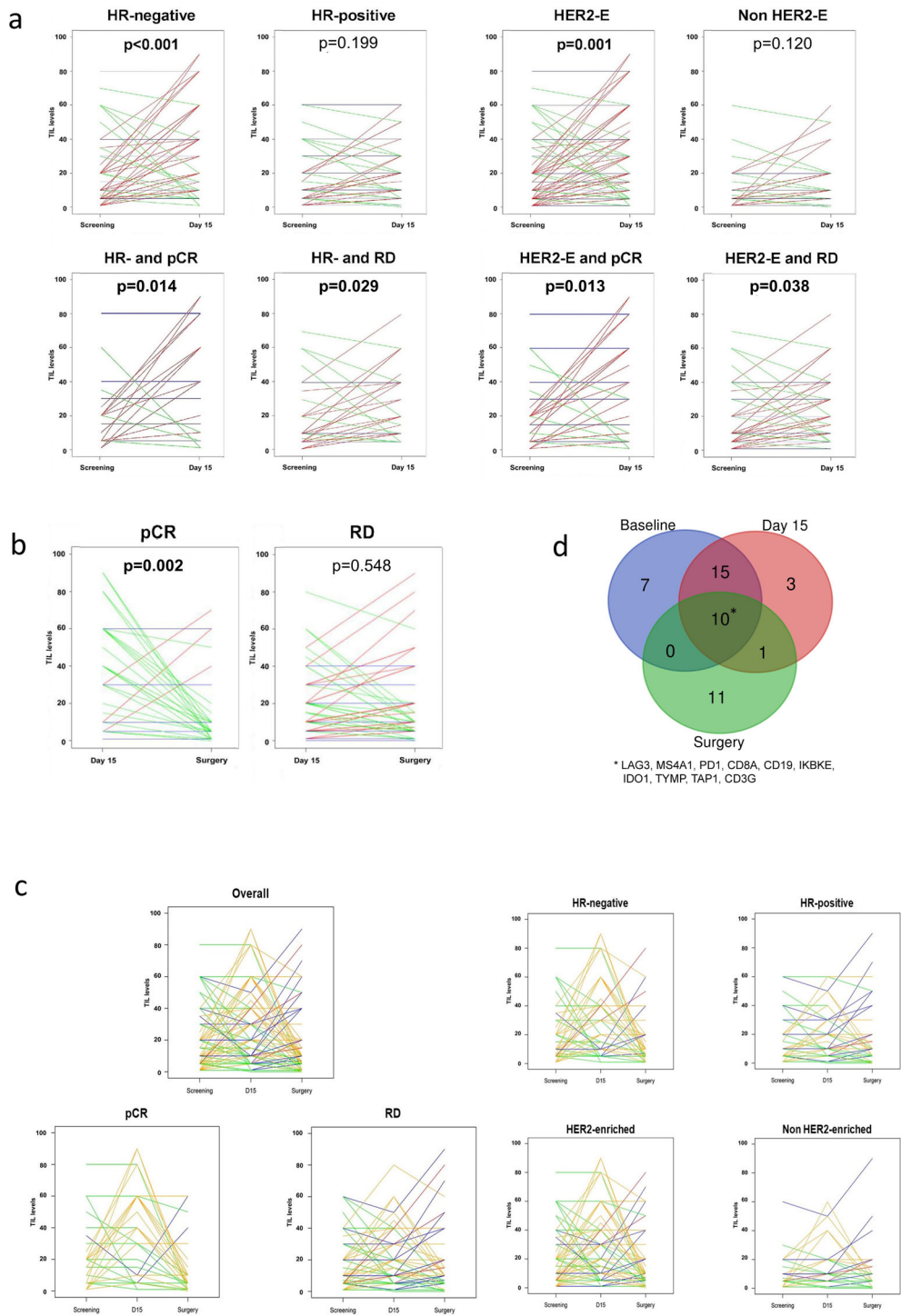
Fig. 4 Changes in sTILs across timepoints and correlation with gene expression. a Changes in TILs between baseline and day 15 according to hormone receptor [HR] status, subtype (HER2-enriched [HER2-E]) and response (pathological complete response [pCR] vs residual disease [RD]). Lines are coloured according to TIL dynamics: increase (red), stable (blue) or decrease (green). **b** Changes in TIL levels between day 15 and surgery according to response: pathological complete response [pCR] vs residual disease [RD]. Lines are coloured according to TIL dynamics: increase (red), stable (blue), or decrease (green). **c** Changes in TIL levels between baseline, day 15 and surgery in the overall study cohort and according to response (pathological complete response [pCR] vs residual disease [RD]), hormone receptor [HR] status and PAM50 subtype. Lines are coloured according to TIL dynamics: increase between baseline and day 15 followed by an increase between day 15 and surgery (red); increase between baseline and day 15 followed by stable or decrease between day 15 and surgery (orange); stable or decrease between baseline and day 15 followed by an increase between day 15 and surgery (blue); stable or decrease between baseline and day 15 followed by stable or decrease between day 15 and surgery (green). **d** Venn diagram representing overlaps in genes upregulated in relation to increase in TIL levels across the three timepoints.

Table 2.	Changes in TILs between day 15 and baseline according to PAM50 intrinsic subtype and hormone receptor (HR) status.			
	N (pairs)	mean difference	95% confidence interval	*p* value
HR-positive	70	+2.1%	−0.7−+4.8	0.199
HR-negative	61	+12.5%	+5.5−+19.6	**<0.001**
HER2-enriched	85	+8.8%	3.5−14.0	**0.001**
HER2-enriched and HR−	53	+12.1%	4.3−20.0	**0.004**
HER2-enriched and HR+	32	+3.2%	−2.0−+8.4	0.235
Non-HER2-enriched	46	+3.5%	−0.3−+7.4	0.120
Non-HER2-enriched and HR−	8	+15.0%	−3.7−+33.7	0.106
Non-HER2-enriched and HR+	38	+1.1%	−1.6−+3.9	0.584
Basal-like	7	+11.6%	−8.6−+31.7	0.201
Normal-like	2	+9.5%	−365.3−+384.3	1.000
Luminal B	15	+3.9%	−1.7−+ 9.6	0.236
Luminal A	22	+0.18%	−2.16−+2.5	0.959
Significant *p* values in bold.				

distal stroma compartment for all four immune cell subpopulations. Moreover, the association between higher immune cell density and pCR was stronger at day 15 as compared with baseline (Fig. 5c).

Changes in sTILs between day 15 and surgery
At surgery, median sTILs levels were 10% (quartile 1–3: 5–20). Paired sTILs data from day 15 and surgery samples were available for 124 patients (82.1%). An increase and a decrease of sTILs between day 15 and surgery were observed in 26.4% (24/91) and 33.0% (30/91) of tumours, respectively, whereas for 40.7% (37/91) of tumours the same percentage of sTILs was reported at day 15 and at surgery. Compared with day 15, a significant decrease in sTILs was observed at surgery in tumours achieving a pCR (mean difference −21.5%, 95% CI −33.3 to −9.7, P = 0.002, Fig. 4b), but not in patients with residual disease at time of surgery (mean difference −0.9%, 95% CI −4.1−+2.4, P = 0.548, Fig. 4b). 89.7% of residual tumours (non-pCR) had sTILs above ≥5%. The distribution of residual tumours (non-pCR) according to sTILs at surgery was 10.3% (sTILs < 5%), 38.1% (sTILs 5–9%), 22.7% (sTILs 10–19%), 16.5% (sTILs 20–39%) and 12.4% (sTILs ≥ 40%). Distribution of tumour samples (non-pCR and pCR) according to sTIL levels at the three timepoints is presented in Supplementary Tables 8–9 and Supplementary Fig. 19. A decrease in sTILs in tumours achieving pCR was observed irrespectively of HR status and intrinsic subtype (Supplementary Table 10). In tumours not achieving a pCR, no significant tendency was observed. Finally, TILs at surgery were not found statistically significantly different according to the type of pathological response (median sTIL levels at surgery 10% (5–20) vs 5% (1–10) in patients with residual disease and achieving pCR, respectively; P = 0.662).

sTILs dynamics across timepoints
Different tumour-infiltrating lymphocytes dynamics observed across the three timepoints (N = 122 patients with sTILs data from all three timepoints) are recapitulated in Supplementary Table 11 and Fig. 4c. The most frequently observed pattern (N = 32, 26%) was an increase in TILs from baseline to day 15 followed by a decrease from day 15 to surgery. This pattern was observed especially in HR-negative tumours, HER2-enriched tumours and tumours achieving pCR (Fig. 4c). Changes in sTILs levels between surgery and baseline paired samples according to the achievement of pCR, HR status and PAM50 subtypes are reported in Supplementary Table 12.

sTILs vs gene expression
To evaluate genes associated with sTILs, we explored data from 413 samples with paired gene expression and sTIL data (from all three timepoints mixed: baseline N = 148; day 15 N = 133; surgery N = 132). A total of 555 BC-related genes were evaluated, including 72 immune-related genes (Supplementary Table 13). Using a quantitative significance of microarrays (SAM) analysis, 36 upregulated genes were found associated with sTIL levels (false discovery rate (FDR) < 1%) (Supplementary Table 14), the top upregulated gene being MS4A1 (CD20). Functional annotation of the 36 genes using DAVID annotation tool[19] revealed that 50% of them were significantly involved in immune response (e.g., CD3G, CD8A, CD4 and LAG3) and regulation of the immune system process (e.g., IDO1, IL6R, STAT1 and PD1), 33% of them were involved in lymphocyte activation (e.g., CD84, CD86, CD3G and CD4) and 28% of them involved in T-cell activation (e.g., CD8A, PD-L1, RELB and CD4).

When a similar analysis was performed within each timepoint separately, similar results were obtained (Fig. 4d; Supplementary Tables 15–17). Among the different genes significantly associated with sTIL levels, 10 (MS4A1, PD1, CD8A, CD19, IKBKE, IDO1, TAP1, TYMP, CD3G and LAG3) were found consistently associated with sTILs across all timepoints. This 10-gene list was found highly enriched (FDR < 1%) for immune genes tracking activated CD8 T cells (e.g., CD8A, CD3G, LAG3, PD1). The correlation coefficients of the expression of these genes with baseline sTILs ranged from 0.52 in gene CD8A to 0.34 in gene TYMP.

DISCUSSION
To our knowledge, our report is the first one to provide new insights into TIL variations and immune contexture during HER2-targeted therapy in the absence of chemotherapy. Moreover, using a novel mIHC technique, immune infiltrate at baseline and after 2 weeks of anti-HER2 treatment was characterised in its immune cell subpopulations and analysed according to proximity to tumour cells and activity (using co-expression of Ki67 marker to identify proliferating immune cells).

First, in early HER2+ treatment-naive BC, tumour-immune contexture analysis showed no significant difference in immune cell subsets densities according to intrinsic subtyping. However, we observed a significantly higher proportion of proliferating immune cells in HR− as compared with HR+ tumours, except for Foxp3+ and consistently observed a numerically higher percentage of proliferating immune cells in basal-like tumours and HER2-enriched tumours and a numerically lower percentage of proliferating
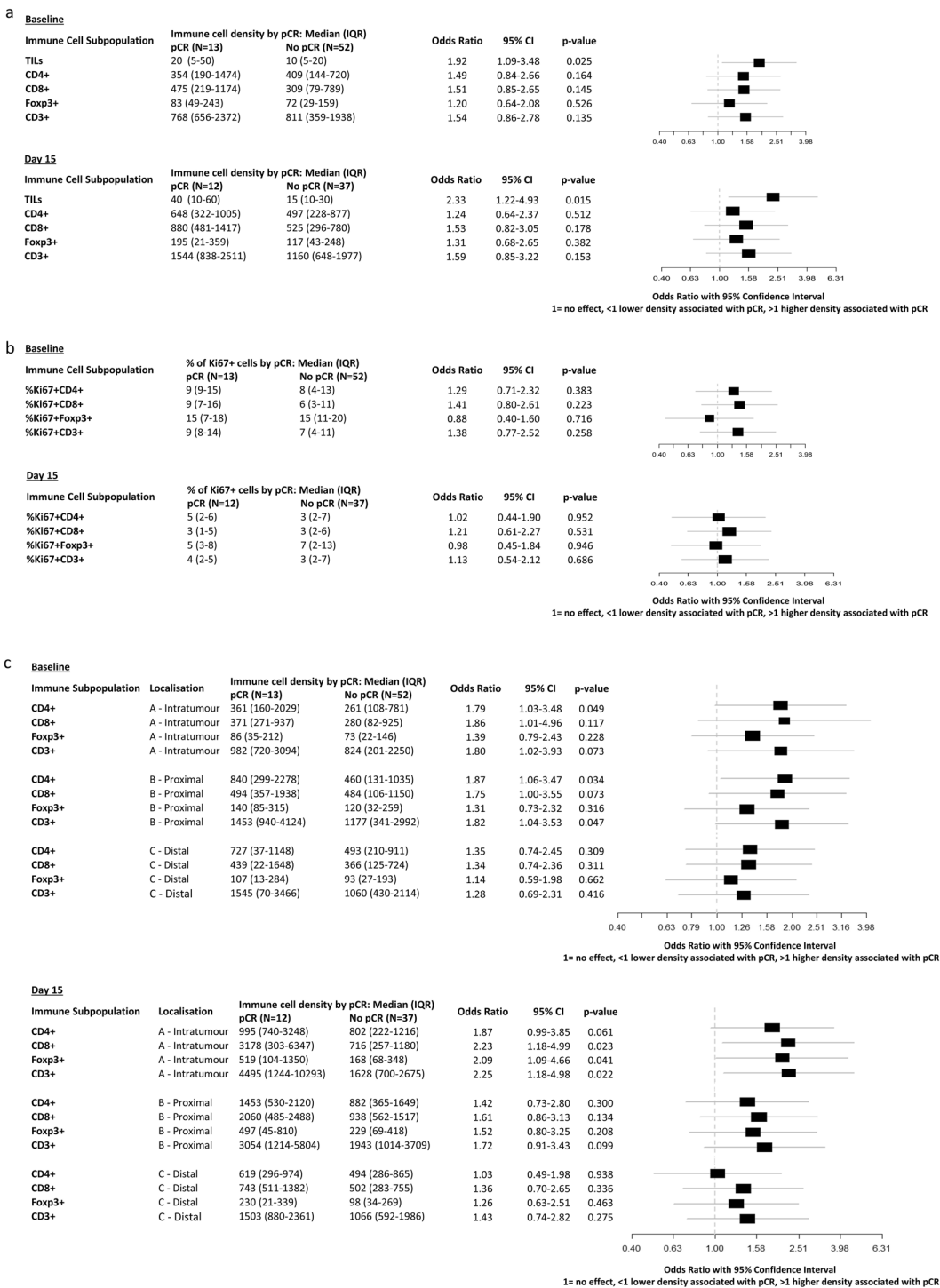
a

Baseline

Immune Cell Subpopulation	Immune cell density by pCR: Median (IQR) pCR (N=13)	No pCR (N=52)	Odds Ratio	95% CI	p-value
TILs	20 (5-50)	10 (5-20)	1.92	1.09-3.48	0.025
CD4+	354 (190-1474)	409 (144-720)	1.49	0.84-2.66	0.164
CD8+	475 (219-1174)	309 (79-789)	1.51	0.85-2.65	0.145
Foxp3+	83 (49-243)	72 (29-159)	1.20	0.64-2.08	0.526
CD3+	768 (656-2372)	811 (359-1938)	1.54	0.86-2.78	0.135

Day 15

Immune Cell Subpopulation	Immune cell density by pCR: Median (IQR) pCR (N=12)	No pCR (N=37)	Odds Ratio	95% CI	p-value
TILs	40 (10-60)	15 (10-30)	2.33	1.22-4.93	0.015
CD4+	648 (322-1005)	497 (228-877)	1.24	0.64-2.37	0.512
CD8+	880 (481-1417)	525 (296-780)	1.53	0.82-3.05	0.178
Foxp3+	195 (21-359)	117 (43-248)	1.31	0.68-2.65	0.382
CD3+	1544 (838-2511)	1160 (648-1977)	1.59	0.85-3.22	0.153

Odds Ratio with 95% Confidence Interval
1= no effect, <1 lower density associated with pCR, >1 higher density associated with pCR

b

Baseline

Immune Cell Subpopulation	% of Ki67+ cells by pCR: Median (IQR) pCR (N=13)	No pCR (N=52)	Odds Ratio	95% CI	p-value
%Ki67+CD4+	9 (9-15)	8 (4-13)	1.29	0.71-2.32	0.383
%Ki67+CD8+	9 (7-16)	6 (3-11)	1.41	0.80-2.61	0.223
%Ki67+Foxp3+	15 (7-18)	15 (11-20)	0.88	0.40-1.60	0.716
%Ki67+CD3+	9 (8-14)	7 (4-11)	1.38	0.77-2.52	0.258

Day 15

Immune Cell Subpopulation	% of Ki67+ cells by pCR: Median (IQR) pCR (N=12)	No pCR (N=37)	Odds Ratio	95% CI	p-value
%Ki67+CD4+	5 (2-6)	3 (2-7)	1.02	0.44-1.90	0.952
%Ki67+CD8+	3 (1-5)	3 (2-6)	1.21	0.61-2.27	0.531
%Ki67+Foxp3+	5 (3-8)	7 (2-13)	0.98	0.45-1.84	0.946
%Ki67+CD3+	4 (2-5)	3 (2-7)	1.13	0.54-2.12	0.686

Odds Ratio with 95% Confidence Interval
1= no effect, <1 lower density associated with pCR, >1 higher density associated with pCR

c

Baseline

Immune Subpopulation	Localisation	Immune cell density by pCR: Median (IQR) pCR (N=13)	No pCR (N=37)	Odds Ratio	95% CI	p-value
CD4+	A - Intratumour	361 (160-2029)	261 (108-781)	1.79	1.03-3.48	0.049
CD8+	A - Intratumour	371 (271-937)	280 (82-925)	1.86	1.01-4.96	0.117
Foxp3+	A - Intratumour	86 (35-212)	73 (22-146)	1.39	0.79-2.43	0.228
CD3+	A - Intratumour	982 (720-3094)	824 (201-2250)	1.80	1.02-3.93	0.073
CD4+	B - Proximal	840 (299-2278)	460 (131-1035)	1.87	1.06-3.47	0.034
CD8+	B - Proximal	494 (357-1938)	484 (106-1150)	1.75	1.00-3.55	0.073
Foxp3+	B - Proximal	140 (85-315)	120 (32-259)	1.31	0.73-2.32	0.316
CD3+	B - Proximal	1453 (940-4124)	1177 (341-2992)	1.82	1.04-3.53	0.047
CD4+	C - Distal	727 (37-1148)	493 (210-911)	1.35	0.74-2.45	0.309
CD8+	C - Distal	439 (22-1648)	366 (125-724)	1.34	0.74-2.36	0.311
Foxp3+	C - Distal	107 (13-284)	93 (27-193)	1.14	0.59-1.98	0.662
CD3+	C - Distal	1545 (70-3466)	1060 (430-2114)	1.28	0.69-2.31	0.416

Odds Ratio with 95% Confidence Interval
1= no effect, <1 lower density associated with pCR, >1 higher density associated with pCR

Day 15

Immune Subpopulation	Localisation	Immune cell density by pCR: Median (IQR) pCR (N=12)	No pCR (N=37)	Odds Ratio	95% CI	p-value
CD4+	A - Intratumour	995 (740-3248)	802 (222-1216)	1.87	0.99-3.85	0.061
CD8+	A - Intratumour	3178 (303-6347)	716 (257-1180)	2.23	1.18-4.99	0.023
Foxp3+	A - Intratumour	519 (104-1350)	168 (68-348)	2.09	1.09-4.66	0.041
CD3+	A - Intratumour	4495 (1244-10293)	1628 (700-2675)	2.25	1.18-4.98	0.022
CD4+	B - Proximal	1453 (530-2120)	882 (365-1649)	1.42	0.73-2.80	0.300
CD8+	B - Proximal	2060 (485-2488)	938 (562-1517)	1.61	0.86-3.13	0.134
Foxp3+	B - Proximal	497 (45-810)	229 (69-418)	1.52	0.80-3.25	0.208
CD3+	B - Proximal	3054 (1214-5804)	1943 (1014-3709)	1.72	0.91-3.43	0.099
CD4+	C - Distal	619 (296-974)	494 (286-865)	1.03	0.49-1.98	0.938
CD8+	C - Distal	743 (511-1382)	502 (283-755)	1.36	0.70-2.65	0.336
Foxp3+	C - Distal	230 (21-339)	98 (34-269)	1.26	0.63-2.51	0.463
CD3+	C - Distal	1503 (880-2361)	1066 (592-1986)	1.43	0.74-2.82	0.275

Odds Ratio with 95% Confidence Interval
1= no effect, <1 lower density associated with pCR, >1 higher density associated with pCR

Fig. 5 Immune cell density and pathological complete response. a Odds ratios (95% confidence interval) for pathologic complete response (pCR) for 10% increases in TIL levels and 1000 cells/mm^2 increases in immune cell density evaluated on baseline and Day 15 (on-treatment) samples. **b** Odds ratios (95% confidence interval) for pathologic complete response (pCR) for increases in % of proliferating immune cells for each immune cell subpopulation evaluated on baseline and day 15 (on-treatment) samples. **c** Odds ratios (95% confidence interval) for pathologic complete response (pCR) for 1000 cells/mm^2 increases in immune cell density according to immune cell localisation evaluated on baseline and day 15 (on-treatment) samples.

immune cells in luminal tumours. However, it should be pointed out that the very limited number of non-HER2-enriched tumours identified in the PAMELA trial, in line with what expected in early HER2+ BC, significantly limits the power of these analyses.

Second, after 2 weeks of dual HER2-targeted therapy, a general increase in sTILs is observed. However, this increase appears to be selectively present in HR-negative and HER2-enriched subtype, regardless of pathological response at surgery, but not in HR-positive and non-HER2-enriched subtypes. Immune contexture analysis highlighted that this increase is not linked to a selective increase of one immune cell subpopulation, but a concomitant increase in cell density of all four immune subpopulations (CD3+, CD4+, CD8+, Foxp3+) after anti-HER2 treatment. However, a significant shift was observed for CD8+ cytotoxic T cells subpopulation, which represented 50% of all tumour-associated immune cells in on-treatment samples.

As previously observed for sTIL levels, increase in specific immune cell subpopulation densities was observed in HER2-enriched subtype and HR-negative tumours, but not in non-HER2-enriched PAM50 subtypes and HR-positive tumours.

Consistently with these trends, immune infiltrate was radically modified after 2 weeks of anti-HER2 treatment. In fact, after priming with anti-HER2 treatment, tumours that were HER2-enriched at baseline showed higher densities of all four immune cell subpopulations, highlighting the differential activation of the immune system towards the disease after priming with anti-HER2 treatment according to tumour biology.

Moreover, the association between pCR and immune infiltrate was stronger at day 15 than at baseline, both for sTILs and specific immune subpopulations, especially when immune cells intratumour/more proximal to the tumour were considered, pointing out the potential biological role of immune activation after anti-HER2 priming in early HER2+ BC.

Subsequently, a general decrease in sTIL levels is observed at surgery. However, this decrease is driven by tumours achieving pCR, whereas no significant trend was seen in patients with residual disease at the time of surgery. Tumours achieving pCR are characterised by an increase in sTIL levels after 2 weeks of anti-HER2 treatment and a decrease in sTIL levels at surgery. This might be linked to downregulation of immune response after clearing tumour cells. However, although the increased infiltration of immune cells observed after HER2 priming is inversely correlated with tumour cellularity at day 15, hinting that these immune cells might have been actively clearing tumour cells during the first 2 weeks of HER2-targeted treatment, their activity (in terms of fraction of proliferating immune cells) was significantly and homogeneously reduced after 2 weeks of HER2-targeted treatment, independently for the amount of residual tumour in the sample and tumour characteristics, thus pointing out that immune exhaustion processes might already be at work at this early timepoint. This observation might by relevant to address the question of which might be the ideal timing of potential combination with immunotherapy in early HER2+ BC.

The association between pCR and decrease in sTIL levels at surgery has also been shown after chemotherapy-containing neoadjuvant treatment for HER2+ BC[12]. However, the same study observed that higher sTIL levels at surgery, in presence of residual disease, were associated with an adverse disease-free survival, suggesting that post-neoadjuvant sTILs might be unable to exert their antitumour function, possibly owing to an immunosuppressive microenvironment or T-cell exhaustion. A limitation of the present study is that immune infiltrate subtyping by mIHC was not available for surgical samples with residual disease and therefore functional assessment of this infiltrate could not be evaluated. Moreover, the true prognostic value of sTIL levels after dual HER2 blockade without chemotherapy remains unknown and long-term follow-up data from the PAMELA trial is not currently available to provide more information on this point.

However, the presence of high sTIL levels in most residual tumours at surgery might imply that these patients might be good candidates for clinical trials evaluating adjuvant immune checkpoint inhibitors. The KATE2 trial, which tested the addition of the anti-PD-L1 antibody atezolizumab to trastuzumab emtansine in metastatic BC HER2+ BC patients previously treated with trastuzumab and taxanes, despite missing its primary endpoint, identified a numerically longer PFS and higher 1-year OS in patients with PD-L1+ and TIL high (≥5%) tumours[20]. These hypothesis-generating data might support the evaluation of PD1/PD-L1 inhibitors to trastuzumab emtansine in the post-

neoadjuvant setting to further improve the prognosis of patients with inflamed residual disease.

Tumour heterogeneity has a predominant role in modulating immune activation in HER2+ BC. Indeed, not only HR-negative and HER2-enriched tumours had higher sTIL levels at baseline, but the impact of tumour biology was observed even more clearly after exposure to HER2-targeted treatment. Non-luminal subtypes showed the highest increases in sTIL levels between baseline and day 15, whereas luminal subtypes showed modest/no increase. Even within HER2-enriched tumours, an increase in sTILs was predominantly seen in HR-negative tumours rather than in HR-positive subgroup. Indeed, HR positivity appeared to be associated with the capacity/incapacity of HER2+ BC to inflame during dual HER2 blockade (without chemotherapy), more than to baseline sTIL levels. These observations were also supported by immune contexture analyses showing a significantly lower proportion of CD8+Ki67+ T cells and higher ratio of proliferating Foxp3+/CD8+ in HR+ as compared with HR- tumours. Whether this is due to specific regulation of the immune system by hormone signalling (or endocrine treatment, as all HR-positive BCs also received hormonotherapy in the PAMELA trial), or if reduced activation of immunity in these tumours is linked to reduced cell death after HER2-targeted treatment and reduced antigen exposure, remains unclear and might hopefully be the subject for further investigation.

In conclusion, in early HER2+ BC, an increase in sTIL levels is observed following 2 weeks of dual HER2 blockade, in HR-negative disease and HER2-enriched subtype. Immune contexture analysis revealed that the strongest impact on pCR was achieved when immune cells spatially interacted with tumour cells. Afterward, sTILs consistently decreased at surgery in patients achieving a pCR, whereas most residual tumours at surgery remained inflamed, possibly reflecting a progressive loss of function of T cells, which is already evident after 2 weeks of treatment. Understanding the features of the resulting tumour immunosuppressive microenvironment has crucial implications for the success of checkpoint blockade and adoptive T-cell transfer therapies. Beyond modulating baseline immune activation, tumour biology also has a role in modulating the dynamic activation of the immune system after exposure to HER2-targeted treatment. This should be taken into account as the role of immunity and immunotherapy is further assessed in HER2+ BC.

METHODS

PAMELA clinical trial

The main results of the neoadjuvant PAMELA phase II trial (NCT01973660) have been previously reported[15]. In this study, 151 early HER2+ BC patients were treated with the combination of lapatinib (1000 mg daily) and trastuzumab (8 mg/kg i.v. loading dose followed by 6 mg/kg) for 18 weeks. Patients with HR-positive disease also received letrozole or tamoxifen according to menopausal status (Supplementary Fig. 1a). In the PAMELA trial, tissue collection was mandatory as it was used for primary endpoint determination. Tumour samples were collected at three timepoints according to the protocol: baseline (within 28 days preceding treatment start), day 15 (a ± 5 days window was admitted, but collection of samples the closest as possible to preplanned timepoint was warmly suggested) and surgery (Supplementary Fig. 1a and sample flow by REMARK diagram in Supplementary Fig. 1b). A minimum of 2 core formalin-fixed paraffin-embedded (FFPE) samples were collected by tru-cut biopsy at each timepoint (except at surgery). In case of multifocality, samples were collected from the same lesion. Samples were primarily used for pre-specified protocol analyses, which included central HER2, ER and PR confirmation by regular IHC, Ki67 by IHC and molecular subtyping by PAM50 gene expression assays. The mIHC analyses performed in the present study were post-hoc and used left-over samples.

The PAMELA trial was conducted under Good Clinical Practice guidelines and the Declaration of Helsinki. The study protocol was approved by independent ethics committees at each centre (trial centres listed at clinicaltrials.gov, NCT01973660). All patients provided written informed consent.

Published in partnership with The Hormel Institute, University of Minnesota

npj Precision Oncology (2021) 23

npj Precision Oncology (2021) 23

Published in partnership with The Hormel Institute, University of Minnesota

44 | Next Generation Immunohistochemistry (NGI)

Garazi Serna Alonso | 45

sTILs evaluation

Stromal TILs at baseline, day 15 and surgery were centrally evaluated on whole sections of tumour tissue stained with H&E blinded from clinical-pathological and outcome data. Percentages (%) of TILs at baseline and day 15 were scored in slides of core biopsies. sTILs were quantified according to the 2014 Guidelines developed by the International TILs Working Group[18,21]. The reproducibility of this method has been described previously[2].

Multiplex IHC (NGI)

Baseline ($N = 65$) and day 15 ($N = 49$) biopsies from 75 patients were analysed using a custom mIHC 6-plex panel, based on iterative cycles of staining and destaining of the same slide with different primary antibodies, individual slide digitalisation, virtual multiplexed digital image reconstruction and complex image analyses.

Before cutting, FFPE blocks were cooled to $-10\,°C$ and 3 µm sections were cut with a microtome. Sections were collected on positively charged Superfrost glass slides and dried overnight at 37 °C. The first IHC staining was performed (information of all the protocols on Supplementary Table 18) in Discovery Ultra Autostainer (Ventana Medical Systems, Tucson AZ). Mono AEC/Plus (#K050; PALEX) was used as the chromogen.

The slides were mounted with aqueous-based mounting medium. The stained slides were digitalised at 20× using the NanoZoomer 2.0HT (Hamamatsu Photonics, Japan). After digitalisation, coverslips were taken and slides were put in increasing alcohol solutions until 100% and then in decreasing alcohols until water before slides were loaded in the Discovery Ultra autostainer for the next immunostaining.

Any residual primary antibody was stripped by an 'extra' antigen retrieval step before the staining process is repeated for the following primary antibody. Heat-induced antigen retrieval was done using ULTRA Cell Conditioning 2 (ULTRA CC2, Ventana Medical Systems, Tucson AZ) for 8 minutes at 100 °C and DISCOVERY Cell Conditioning 1 (DISCOVERY CC1, Ventana Medical Systems, Tucson AZ) for 40 minutes at 95 °C to block the previous antibody and the process was repeated consecutively six times.

To avoid primary antibody cross-reactivity between cycles owing to incomplete stripping, we used several strategies. First, we added an extra antigen retrieval step before the next cycle of staining to prevent any remnant reactivity to primary or secondary antibodies used in the first cycle. Second, the protocol alternated rabbit and mouse primary antibodies to reduce cross-reactivity. Third, the sequence of primary antibodies alternated nuclear (Ki67 and Foxp3), membrane (CD3, CD8, CD4) and cytoplasmic (CK) markers (Supplementary Fig. 20).

The sequential staining procedure was automalised, thus significantly reducing the hands-on time (1-hour per staining cycle) and duration of the entire process (3 days per 6-plex panel run in 30 slides).

The image analysis pipeline was the following one: first, images were uploaded into VISIOPHARM® (VIS) Image Analysis Software (Visiopharm Integrator System version 2019.02.1.6005, Visiopharm, Denmark) for registration. Images were automatically aligned and fused into a single virtual digital image (VDI) using the Tissuealign® module of VIS (order of alignment: Foxp3, CD3, CD8, CD4, Ki67, cytokeratin). After alignment, images were analysed with custom-developed algorithms created using the Author® module of VIS (algorithms in Supplementary Material).

After this, we performed automatic tissue recognition of the aligned slides and the selected areas were reviewed by a pathologist, who manually defined the ROI, which included the tumour and surrounding peritumoural stroma (the tumour bed in case of complete regression) and excluded normal and/or necrotic area. Following this, we ran T-cell density APP was run on the entire slides to obtain global results. The T-cell application (detailed in Supplementary Table 19) uses a cell classification method based on form and size and a pixel-colour intensity threshold method to classify the cells into Foxp3, CD3, CD4, CD8 cells on one hand, and uses the Ki67 staining to inform about the percentage of the cell populations that are proliferating on the other hand. Any brown stained nucleus was considered a positive cell.

We then applied a third APP (location APP), which uses the PANCK staining to divide the ROI created by the pathologist into three different ROIs: the tumour area (A), the stroma within 30 µm from tumour (B) and the stroma >30 µm from tumour (C). For that purpose, we used HDAB-DAB feature, which enhances the brown staining corresponding to the PANCK staining (A). Dilation was used to create B and C ROIs. After applying the location APP to the data set, we used the T-cell APP to obtain densities and proliferation rates of different cell populations across different locations.

Created APPs were trained by a biotechnologist expert in image analysis and the results validated by a board-certified pathologist. Data were finally reported as densities of each category of cells in the tumoural area in general, for each location, and the proliferation rate of each of the cell categories.

Gene expression analysis

Samples from all three timepoints were analysed using the same methodology. First, a section of FFPE breast tissue was examined with H&E staining to confirm the diagnosis and determine the tumour surface area. RNA purification was performed after macrodissection, when needed, to avoid normal breast contamination. RNA was extracted from FFPE material using the High Pure FFPET RNA isolation kit (Roche, Indianapolis, IN, USA) following the manufacturer's protocol. RNA samples were quantified at the NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

A minimum of ~100 ng of total RNA was used to measure the expression of 555 BC-related genes and five housekeeping genes (ACTB, MRPL19, PSMC4, RPLP0 and SF3A1) using the nCounter platform (Nanostring Technologies; Seattle, Washington, USA[22]). Data were log base 2−transformed and normalised using the housekeeping genes. The complete list of genes, which included immune-related genes (e.g., CD8A, CD4, PD1 and PD-L1), can be found in Supplementary Table 13. Intrinsic molecular subtyping at baseline was determined using the previously reported PAM50 subtype predictor[23].

Statistical analysis

Spearman test was used for correlation analysis and Mann–Whitney U test and Wilcoxon test were used for all the density, location and proliferation analyses. To determine differences in the distribution of TIL levels or immune cell density across subgroups Mann–Whitney U and Kruskal–Wallis test were used according to number of subgroups. Significant changes in sTILs or immune cell density between two timepoints were determined using paired Wilcoxon tests. The association of each variable with pCR was determined by univariate logistic regression analysis. As per study protocol, pCR was defined as the absence of residual invasive cancer in the breast following neoadjuvant therapy (ypT0/is). OR with a 95% confidence interval were estimated. All statistical tests were two-sided and considered significant when $p < 0.05$.

When recapitulating TIL dynamics across the three timepoints, any increase/decrease in sTIL levels were taken into account and sTILs were only defined as unchanged if the same % of TILs were present at two subsequent timepoints.

To identify genes whose expression was significantly different according to sTIL levels as a continuous variable, we used a quantitative SAM analysis with an FDR < 1%. Pearson correlations were used to evaluate the association of expression of a single gene with sTILs expression. Biologic analysis of gene lists was performed with DAVID annotation tool (http://david.abcc.ncifcrf.gov/)[19].

All statistical analyses were performed using the R software 3.6.1.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The data generated and analysed during this study are described in the following data record: https://doi.org/10.6084/m9.figshare.13681456[24]. The data files underlying the related study are available from the corresponding authors upon reasonable request. However, several files are not publicly available in order to protect patient privacy. A comprehensive list of data files underlying the related manuscript along with details of their availability is contained in the spreadsheet 'Griguolo_et_al_2021_underlying_datafile_list.xlsx', available as part of the figshare. The custom-developed algorithms (T-cell APP) created using the Author® module of VISIO-PHARM® (VIS) Image Analysis Software (Visiopharm Integrator System version 2019.02.1.6005, Visiopharm, Denmark) are also available as part of the figshare data record.

CODE AVAILABILITY

R codes used for data analysis during this study are available from the corresponding authors on reasonable request.

REFERENCES

1. Denkert, C. et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol.* **19**, 40–50 (2018).
2. Salgado, R. et al. Tumor-infiltrating lymphocytes and associations with pathological complete response and event-free survival in her2-positive early-stage breast cancer treated with lapatinib and trastuzumab. *JAMA Oncol.* **1**, 448 (2015).
3. Luen, S. J. et al. Tumour-infiltrating lymphocytes in advanced HER2-positive breast cancer treated with pertuzumab or placebo in addition to trastuzumab and docetaxel: a retrospective analysis of the CLEOPATRA study. *Lancet Oncol.* **18**, 52–62 (2017).
4. Stanton, S. E., Adams, S. & Disis, M. L. Variation in the incidence and magnitude of tumor-infiltrating lymphocytes in breast cancer subtypes. *JAMA Oncol.* **2**, 1354 (2016).
5. Dieci, M. V. et al. Integrated evaluation of PAM50 subtypes and immune modulation of pCR in HER2-positive breast cancer patients treated with chemotherapy and HER2-targeted agents in the CherLOB trial. *Ann. Oncol. J. Eur. Soc. Med. Oncol.* **27**, 1867–1873 (2016).
6. Ingold, B. & Heppner, B. I. et al. Tumor-infiltrating lymphocytes: a predictive and prognostic biomarker in neoadjuvant-treated HER2-positive breast cancer. *Clin. Cancer Res.* **22**, 5747–5754 (2016).
7. Denkert, C. et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. *J. Clin. Oncol.* **33**, 983–991 (2015).
8. Loi, S. et al. Pembrolizumab plus trastuzumab in trastuzumab-resistant, advanced, HER2-positive breast cancer (PANACEA): a single-arm, multicentre, phase 1b-2 trial. *Lancet Oncol.* **20**, 371–382 (2019).
9. Bianchini, G. & Gianni, L. The immune system and response to HER2-targeted treatment in breast cancer. *Lancet Oncol.* **15**, e58–e68 (2014).
10. Griguolo, G., Pascual, T., Dieci, M. V., Guarneri, V. & Prat, A. Interaction of host immunity with HER2-targeted treatment and tumor heterogeneity in HER2-positive breast cancer. *J. Immunother. Cancer* **7**, 90 (2019).
11. Kurozumi, S. et al. Prognostic utility of tumor-infiltrating lymphocytes in residual tumor after neoadjuvant chemotherapy with trastuzumab for HER2-positive breast cancer. *Sci. Rep.* **9**, 1583 (2019).
12. Hamy, A.-S. et al. Stromal lymphocyte infiltration after neoadjuvant chemotherapy is associated with aggressive residual disease and lower disease-free survival in HER2-positive breast cancer. *Ann. Oncol.* **28**, 2233–2242 (2017).
13. Ochi, T. et al. Predictive and prognostic value of stromal tumour-infiltrating lymphocytes before and after neoadjuvant therapy in triple negative and HER2-positive breast cancer. *Eur. J. Cancer* **118**, 41–48 (2019).
14. Ladoire, S. et al. In situ immune response after neoadjuvant chemotherapy for breast cancer predicts survival. *J. Pathol.* **224**, 389–400 (2011).
15. Llombart-Cussac, A. et al. HER2-enriched subtype as a predictor of pathological complete response following trastuzumab and lapatinib without chemotherapy in early-stage HER2-positive breast cancer (PAMELA): an open-label, single-group, multicentre, phase 2 trial. *Lancet Oncol.* **18**, 545–554 (2017).
16. V. Guarneri, et al. De-escalated therapy for HR+/HER2+ breast cancer patients with Ki67 response after 2-week letrozole: results of the PerELISA neoadjuvant study. *Ann. Oncol.* **30**, 921–926 (2019).
17. Nuciforo, P. et al. A predictive model of pathologic response based on tumor cellularity and tumor-infiltrating lymphocytes (CelTIL) in HER2-positive breast cancer treated with chemo-free dual HER2 blockade. *Ann. Oncol.* **29**, 170–177 (2018).
18. Salgado, R. et al. International TILs Working Group 2014, The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann. Oncol.* **26**, 259–271 (2015).
19. Dennis, G. et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, P3 (2003).
20. Emens, L. A. et al. 305 O Overall survival (OS) in KATE2, a phase II study of programmed death ligand 1 (PD-L1) inhibitor atezolizumab (atezo)+trastuzumab emtansine (T-DM1) vs placebo (pbo)+T-DM1 in previously treated HER2+ advanced breast cancer (BC). *Ann. Oncol.* **30**, v104 (2019).
21. Dieci, M. V. et al. Update on tumor-infiltrating lymphocytes (TILs) in breast cancer, including recommendations to assess TILs in residual disease after neoadjuvant therapy and in carcinoma in situ: a report of the International Immuno-Oncology Biomarker Working Group on Breast Cancer. *Semin. Cancer Biol.* **52**, 16–25 (2018).
22. Geiss, G. K. et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–325 (2008).
23. Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
24. Griguolo, G. et al. Metadata record for the manuscript: Immune microenvironment characterization and dynamics during anti-HER2-based neoadjuvant treatment in HER2-positive breast cancer. figshare https://doi.org/10.6084/m9.figshare.13681456 (2021).

AUTHOR CONTRIBUTIONS

G.G., S.G., A.P. and P.N. contributed conception and design of the study. G.G., G.S., S.G., T.P., R.F., X.G., J.C., S.P., M.M., M.O., M.V., A.L.C., J.C., P.G., B.B., N.M., R.L., I.G., L.M., J.A., E.M., P.V., A.P. and P.N. participated in data collection and/or in biological analysis. G.G., G.S., S.G. and L.P. performed the statistical analysis. G.G. and G.S. wrote the first draft of the manuscript with substantial revision by A.P. and P.N. All authors contributed to manuscript revision, read and approved the submitted version.

COMPETING INTERESTS

A.P. reports consulting fees from Nanostring Technologies, Roche, Pfizer, Oncolytics Biotech, Puma, Lilly, AstraZeneca, Seattle Genetics, Novartis, Amgen, BMS, MSD and Daiichi Sankyo outside the submitted work. P.N. reports consulting fees of advisory boards from Bayer, MSD and Novartis and speaker fees from MSD and Novartis. All other authors declared no competing interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-021-00163-6.

Correspondence and requests for materials should be addressed to A.P. or P.N.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

4.3- Study 3: (See section 8. Annex)

"In situ single-cell analysis of canonical breast cancer biomarkers: phenotypic heterogeneity and implications on response to HER2 targeting agents"

In this study we validated a third panel of NGI to analyze the tumor cell composition and heterogeneity of breast cancer tumors. We did a panel of HER2, ER, PR, KI67 and PanCK that allowed us to classify breast cancer tumor cells into 16 different phenotypes of cells. We used this subclassification to comprehensively analyze the breast cancer tumor heterogeneity in Her2 positive tumors that were treated with an anti-her2 treatment.

Garazi Serna, Eloy García, Roberta Fasani, Xavier Guardia, Tomas Pascual, Laia Paré, Fiorella Ruiz-Pace, Antonio Llombart-Cussac, Javier Cortes, Aleix Prat, Paolo Nuciforo

5. DISCUSSION

Through the different studies presented in this thesis, NGI has shown its value as a multiplexed imaging technology by being able to address different clinically relevant concerns. The different panels that have been used in this thesis show that our NGI approach is a robust, automatized and relatively simple and cost-effective technology, that can be used in research and clinical laboratories that are equipped for digital pathology.

In the first study NGI demonstrated to be an easy and accurate methodology for KI67 quantification, providing a step toward using KI67 in the clinical setting. In the second study, the characterization of the immune microenvironment of Her2-positive breast cancer samples was analyzed during her2-targeted therapy was studied, not only quantifying the different immune populations but analyzing their spatial distribution towards the tumor and their proliferation status. The strongest impact on response was obtained when the immune cells were in contact with the tumor, information that was obtained thanks to the NGI. In the third study, for the first time, tumor cells from Her2-positive breast cancer samples during the treatment were analyzed at a single-cell resolution level providing the tumor cell phenotype composition, heterogeneity and spatial relationships with cytotoxic cells. The identification of tumor phenotypes that are targeted by the treatment, the ones related with response and some resistant ones to the treatment were obtained thanks

to the NGI.

Results successfully show that NGI allows reproducible biomarker quantification together with comprehensive characterization of biological tissue samples at a single cell resolution while maintaining the spatial distribution and the interaction between the tumor and its microenvironment. This is all critical information for understanding tumor biology and complexity, presenting NGI as a promising technology that could be used for comprehensive or reproducible analyses in any pathology or research laboratory.

Hence, NGI presents itself as superior to the traditional imaging techniques used to study protein biomarkers on intact tissue samples. Specifically, NGI is be:

Superior to standard immunohistochemistry, given its ability to use a unique tissue section as shown in the different studies maintaining tumor morphology and overcoming the limitations of serial sectioning and partial reconstruction.

Superior to multiplex immunohistochemistry, where one single slide is stained simultaneously with different chromogens, making interpretation and analysis of the results complex and sometimes not even possible without complex spectral differentiation machines. With NGI, each biomarker can be analyzed individually (as in routine diagnostic practice) while also allowing unlimited multiplexing through virtual image reconstruction and analysis as presented in the different published papers.

Superior to multiplex immunofluorescence, mIF, which is limited by the number of available fluorochromes (traditionally 5, new protocols are able to include up to 7), while in NGI the number of cycles of staining and destaining is virtually unlimited (previous studies have analyzed up to 12 biomarkers in a single slide). Moreover, NGI uses standard protocols routinely used for antibodies already approved for diagnostic use that sometimes do not work for IF, does not need a fluorescence microscope and is readily accepted by pathologists used to work with bright field images.

Superior to Flow Cytometry or RNAseq, because it is able to thoroughly characterize complex cellular populations while maintaining spatial and morphological context information while FC and other bulk technologies disrupt the tissue, that would not have allowed us any spatial analysis we presented

Superior to metal, fluorescence or barcode based multiplex imaging systems because it doesn't need complex equipment and pipelines that they use, and it is a much cost-effective technology, that as shown, did not require any additional equipment in the laboratory.

Superior to other sequential IHC technologies using soluble chromogens because NGI is automatized and clinical routine focused as presented in the different studies, while sequential IHC technologies are manual, time consuming and research focused.

As shown in each paper, the NGI panels that have been developed

are equivalent to the traditional IHC as explained in the different papers. That means that by one hand, antigenicity is not lost during the rounds of stainings, and by the other hand, routine IHCs could be substituted by NGI that would give us more information while still being able to obtain the individual IHC stainings that can be evaluated as they currently are. By using a soluble chromogen, NGI can be easily scored as a traditional single analyte IHC at bright field microscope or computer monitor, and thus a pathologist-friendly approach.

NGI uses 3-4µm of FFPE tissue and is performed on a single slide that allows the preservation of the sample which could revolutionize the standard diagnostic and molecular testing workflow, thus allowing optimization of sample use, modification of current diagrams of prioritization of techniques in samples with limited material and in particular cases, even avoiding the necessity to perform additional biopsies, having an impact on the patient and on the health system.

Slides can be stored and reused for additional diagnostic or research workflow thus representing a tremendous advantage over other techniques. The integration of different modalities associated with single cell data such us epigenomics, proteomics, transcriptomics and spatial information will be critical for cancer understanding. In that line, NGI can be combined with RNAscope technology to analyze not only proteins but mRNA molecules in the same exact cells. Artificial intelligence, that has shown to identify even mutations by the histological features of HE stainings, will be an opportunity to be able to extract the maximum information from NGI registered images.

As a part of the design rationale, we selected methodologies that follow a simple workflow and that all academic pathology laboratories may easily implement and offer to oncologists in the case that an exploratory biomarker panel moves from investigation into clinics. This would allow for a rapid implementation of the assay as a part of the routine molecular diagnostic activity of the laboratory.

These studies are examples of the multiple possible applications that NGI has. In our laboratory, we have already used KiQuant and t-cell panels in more than 15 studies or clinical trials in the laboratory that still haven't been published. We have also developed and used more panels for specific needs for research and clinical trials. The fact that more and more researchers and clinical trials are interested and using NGI is one of the biggest demonstrations of the power of the technique. Further studies using NGI will help address problems like interobserver variability, sample exhaustion, spatial information analyses without the use of expensive and complex machines.

All the studies of the thesis have been done in breast tissues but KiQuant and t-cell panels are nowadays being applied in other tissue types because they are not tissue specific panels. Breast panel is the only one that is tissue specific but it can be applied in other breast cancer subtypes or in samples from patients that have received other treatments to understand more deeply the breast cancer and is heterogeneity.

We developed automatized and flexible IHC protocols for sequential staining using the combination of Ventana Discovery Ultra (Roche Diagnostics), the Nanozoomer slide scanner (Hamamatsu) and Visiopharm image analysis program that were the equipment we had in the laboratory. NGI, then, as it happened in our laboratory, could be easily implemented in any diagnostic pathology laboratory equipped for digital pathology.

The scanning platform wouldn't change the NGI workflow proposed, but if the staining platform is another one, the panels would need to be optimized for the new platform. This is a limitation of NGI, because optimization is the hardest part of NGI technology. Nevertheless, Ventana Discovery Ultra from Roche Diagnostics is one of the most used platforms minimizing the problem.

The change of the used antibodies would also impact the workflow due to the fact that the species of the antibody, the potential change of the pretreatment or primary or secondary conditions, the cross reactivity in successive stainings could affect the technique. The fact that routinely used antibodies have been used for the studies makes it easy to implement in any pathology department.

This last point is linked with the fact that the first step of optimization of the panels is key in NGI as well as in other techniques where different antibodies are applied to the same slide. Lots of efforts has to be put into the panel design so we are sure that each of the staining is equivalent to the optimized and validated golden standard IHC stainings. The fact that we have optimized the panels in a pathology laboratory where we have quality controls for IHC stainings makes NGI robust enough to be sure that AEC stained IHCs are equivalent to the validated DAB ones and this is something we have been sure to prove for the different panels in different studies.

Another key part of the panel design is the fact that no cross-reactivity is happening in the different stainings and that in each of the cycles just the antibody that we want to analyze is the only one that the secondary antibody will recognize, and all the staining comes from the primary antibody of the cycle. For doing so, after each of the stainings, a protocol with no primary antibody is applied to check and confirm that no primary antibody from previous round prevail. This is also something that has been proven in the different studies.

For avoiding cross-reactivity between antibodies, in the creation of the panels we combine anti-mouse and anti-rabbit antibodies apart from a stripping step with the Ventana platform. This stripping step is also being used for mIHC stainings with the Ventana platform. Another point is that the biomarkers sequence in the panel tend to be from the less expressed biomarker to the more expressed one.

All image analysis algorithms in the study are developed with Author module of Visiopharm image analysis program which is one of the most used image analysis programs. A limitation naturally exists in the cost of obtaining the program, if the laboratories

that wants to use it are not already users. Some free programs exist such as QuPath. These could provide similar results though currently the algorithms are not ready to be applied in other image analysis programs.

The scanned images must be carefully saved because the slide is going to be distained and that means we will not have a physically available slide saved. This is a risk if we lose the scanned image that however can be mitigated by also save a space physically.

Already stained slides can be de-stained, mounted and saved and be used in the future for other biomarkers that could be of interest if it's in the context of biomarkers that we know they work in the sequence that we would apply.

The use of the staining platform with the proposed antibodies and the image analysis program makes the applicability of the panels in new laboratories easily applicable. As explained, the laborious parts that require time and expertise are the optimization of the technical panel and image analysis algorithms which are shared thus it would simply be a fact of applying them.

We have for the moment, optimized panels with a maximum of six markers though there could be several more biomarkers in the panel. This was not pursued as our objective was creating panels that were clinically relevant rather than using a slide to stain the maximum possible biomarkers.

The time that the technique requires could be a limitation for the clinical use. A 6-biomarker panel can be done in three days, as

compared to one day if multiple slides are being used, each for one biomarker. If fast results are requested for rapid diagnosis, this could be a limitation and would necessitate an evaluation of any benefits of the added days.

One of the problems of NGI and multiplex imaging technologies in general is the quantity of data that can be generated from them. Storage of big amounts of information is already being a problem in hospitals and research institutes. Then, the analysis of them is the next question that still is being addressed. Even more could be solved if data was shared and was publicly available, this would require complex systems for data protection.

As shown in the different studies and discussed here, NGI has allowed the study of different biomarkers in a single slide and could be applicable in any other pathology laboratory. It has also shown that it solves limitations of the actual technologies and not only that, but has addressed several problems with the different created panels: the reproducibility of KI67 evaluation and the comprehensive characterization of the Her2-positive breast cancer samples during the treatment, analyzing both the immune and tumor compartments and the relationship between them providing information for deeper understanding the tumor biology and complexity that, with further studies, may be used for resistant phenotypes identification, response prediction and better stratification of the patients.of the most used image analysis programs. A limitation naturally exists in the cost of obtaining the program, if the laboratories that wants to use it are not already users. Some free programs exist such as QuPath. These could

provide similar results though currently the algorithms are not ready to be applied in other image analysis programs.

The scanned images must be carefully saved because the slide is going to be distained and that means we will not have a physically available slide saved. This is a risk if we lose the scanned image that however can be mitigated by also save a space physically.

Already stained slides can be de-stained, mounted and saved and be used in the future for other biomarkers that could be of interest if it's in the context of biomarkers that we know they work in the sequence that we would apply.

The use of the staining platform with the proposed antibodies and the image analysis program makes the applicability of the panels in new laboratories easily applicable. As explained, the laborious parts that require time and expertise are the optimization of the technical panel and image analysis algorithms which are shared thus it would simply be a fact of applying them.

We have for the moment, optimized panels with a maximum of six markers though there could be several more biomarkers in the panel. This was not pursued as our objective was creating panels that were clinically relevant rather than using a slide to stain the maximum possible biomarkers.

The time that the technique requires could be a limitation for the clinical use. A 6-biomarker panel can be done in three days, as compared to one day if multiple slides are being used, each for one biomarker. If fast results are requested for rapid diagnosis,

this could be a limitation and would necessitate an evaluation of any benefits of the added days.

One of the problems of NGI and multiplex imaging technologies in general is the quantity of data that can be generated from them. Storage of big amounts of information is already being a problem in hospitals and research institutes. Then, the analysis of them is the next question that still is being addressed. Even more could be solved if data was shared and was publicly available, this would require complex systems for data protection.

As shown in the different studies and discussed here, NGI has allowed the study of different biomarkers in a single slide and could be applicable in any other pathology laboratory. It has also shown that it solves limitations of the actual technologies and not only that, but has addressed several problems with the different created panels: the reproducibility of KI67 evaluation and the comprehensive characterization of the Her2-positive breast cancer samples during the treatment, analyzing both the immune and tumor compartments and the relationship between them providing information for deeper understanding the tumor biology and complexity that, with further studies, may be used for resistant phenotypes identification, response prediction and better stratification of the patients.

6. CONCLUSIONS

1. In precision oncology nowadays, testing for multiple markers is required but current methodologies do not usually provide single-cell data while keeping the spatial information, and if they do, they are complex and expensive technologies.

2. NGI is an automated, simple, and cost-effective IHC-based technology that sequentially stains and destains the sample providing single-cell and spatial information filling the actual limitations of current approaches.

3. This thesis demonstrated that NGI is a multiplexed and non-disruptive imaging technology that can be used in different pathology and research laboratories equipped for digital pathology.

4. NGI allows sample optimization using only 3µm of an FFPE sample per panel, saving material for further analyses.

5. All the designed panels proved no cross-reactivity and all the stainings of the panels were equivalent to pathology gold standard DAB stainings proving no antigenicity loss and applicability in pathology laboratories.

6. NGI allowed the automated quantification of KI67 biomarker in breast cancer tumor cells in an easy and reproducible way.

7. NGI allowed the characterization of the immune microenvironment quantifying the different t-cell populations and providing their proliferation status and spatial interaction with the tumor cells.

8. NGI allowed the characterization of breast cancer tumor cells by analyzing the phenotype composition and heterogeneity and its impact in response to targeted therapies.

9. The results generated in this thesis provide a framework for applying NGI to better understand tumor biology and its complexity.

10. In the future, the same or new NGI panels can be used that may help identify better or new prognostic or predictive biomarkers supporting better patient stratification towards different therapies and increasing the number of patients that can benefit from them.
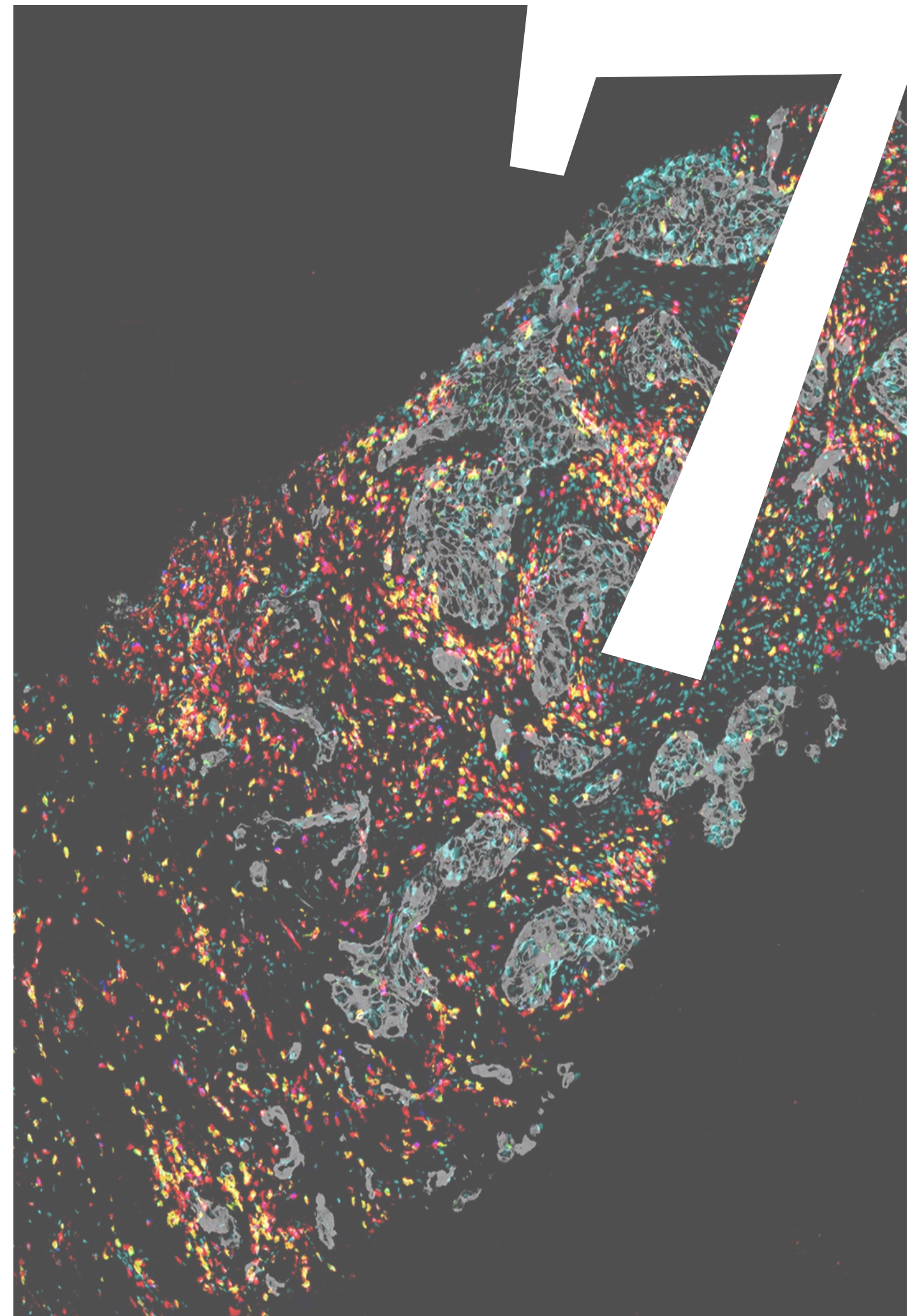
7. BIBLIOGRAPHY

Chen, D. S., & Mellman, I. (2017). Elements of cancer immunity and the cancer-immune set point. Nature, 541(7637), 321–330. https://doi.org/10.1038/NATURE21349

de Obaldia, M. E., & Bhandoola, A. (2015). Transcriptional Regulation of Innate and Adaptive Lymphocyte Lineages. Http://Dx.Doi.Org/10.1146/Annurev-Immunol-032414-112032, 33, 607–642. https://doi.org/10.1146/ANNUREV-IMMUNOL-032414-112032

Finotello, F., Rieder, D., Hackl, H., & Trajanoski, Z. (2019). Next-generation computational tools for interrogating cancer immunity. Nature Reviews Genetics 2019 20:12, 20(12), 724–746. https://doi.org/10.1038/s41576-019-0166-7

Gerdes, M. J., Sevinsky, C. J., Sood, A., Adak, S., Bello, M. O., Bordwell, A., Can, A., Corwin, A., Dinn, S., Filkins, R. J., Hollman, D., Kamath, V., Kaanumalle, S., Kenny, K., Larsen, M., Lazare, M., Li, Q., Lowes, C., McCulloch, C. C., … Ginty, F. (2013). Highly multiplexed single-cell analysis of formalinfixed, paraffin-embedded cancer tissue. Proceedings of the National Academy of Sciences of the United States of America, 110(29), 11982–11987. https://doi.org/10.1073/PNAS.1300136110/SUPPL_FILE/SD04.XLSX

Giesen, C., Wang, H. A. O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P. J., Grolimund, D., Buhmann, J. M., Brandt, S., Varga, Z., Wild, P. J., Günther, D., & Bodenmiller, B. (2014). Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nature Methods 2014 11:4, 11(4), 417–422. https://doi.org/10.1038/nmeth.2869

Glass, G., Papin, J. A., & Mandell, J. W. (2009). Simple: A Sequential Immunoperoxidase Labeling and Erasing Method. Http://Dx.Doi.Org/10.1369/Jhc.2009.953612, 57(10), 899–905. https://doi.org/10.1369/JHC.2009.953612

Keren, L., Bosse, M., Marquez, D., Angoshtari, R., Jain, S., Varma, S., Yang, S. R., Kurian, A., van Valen, D., West, R., Bendall, S. C., & Angelo, M. (2018). A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. Cell, 174(6), 1373-1387.e19. http://www.cell.com/article/S0092867418311000/fulltext

Pirker, R., & Filipits, M. (2016). Personalized treatment of advanced non-small-cell lung cancer in routine clinical practice. Cancer Metastasis Reviews, 35(1), 141–150. https://doi.org/10.1007/S10555-016-9612-6

Remark, R., Merghoub, T., Grabe, N., Litjens, G., Damotte, D., Wolchok, J. D., Merad, M., & Gnjatic, S. (2016). In-depth tissue profiling using multiplexed immunohistochemical consecutive staining on single slide. Science Immunology, 1(1). https://doi.org/10.1126/SCIIMMUNOL.AAF6925/SUPPL_FILE/

AAF6925_SM.PDF

Sankar, K., Ye, J. C., Li, Z., Zheng, L., Song, W., & Hu-Lieskovan, S. (2022). The role of biomarkers in personalized immunotherapy. Biomarker Research, 10(1), 1–13. https://doi.org/10.1186/S40364-022-00378-0/FIGURES/3

Tan, W. C. C., Nerurkar, S. N., Cai, H. Y., Ng, H. H. M., Wu, D., Wee, Y. T. F., Lim, J. C. T., Yeong, J., & Lim, T. K. H. (2020). Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. Cancer Communications (London, England), 40(4), 135–153. https://doi.org/10.1002/CAC2.12023

Tsujikawa, T., Kumar, S., Borkar, R. N., Azimi, V., Thibault, G., Chang, Y. H., Balter, A., Kawashima, R., Choe, G., Sauer, D., el Rassi, E., Clayburgh, D. R., Kulesz-Martin, M. F., Lutz, E. R., Zheng, L., Jaffee, E. M., Leyshock, P., Margolin, A. A., Mori, M., … Coussens, L. M. (2017). Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. Cell Reports, 19(1), 203–217. https://doi.org/10.1016/j.celrep.2017.03.037 AGRADECIMIENTOS

8. ANNEX

Study 3:

"In situ single-cell analysis of canonical breast cancer biomarkers: phenotypic heterogeneity and implications on response to HER2 targeting agents"

In this study we validated a third panel of NGI to analyze the tumor cell composition and heterogeneity of breast cancer tumors. We did a panel of HER2, ER, PR, KI67 and PanCK that allowed us to classify breast cancer tumor cells into 16 different phenotypes of cells. We used this subclassification to comprehensively analyze the breast cancer tumor heterogeneity in Her2 positive tumors that were treated with an anti-her2 treatment.

Garazi Serna, Eloy García, Roberta Fasani, Xavier Guardia, Tomas Pascual, Laia Paré, Fiorella Ruiz-Pace, Antonio Llombart-Cussac, Javier Cortes, Aleix Prat, Paolo Nuciforo

In situ single-cell analysis of canonical breast cancer biomarkers: phenotypic heterogeneity and implications on response to HER2 targeting agents.

Garazi Serna[1], Eloy García[2], Roberta Fasani[1], Xavier Guardia[1], Tomas Pascual[3,4,5], Laia Paré[6], Fiorella Ruiz- Pace[7], Antonio Llombart-Cussac[8], Javier Cortes[9], Aleix Prat[3,4,5], Paolo Nuciforo[1,5]

[1] Molecular oncology group, Vall d´Hebron Institute of Oncology (VHIO), Barcelona, Spain
[2] Department of Mathematics and Informatics, University of Barcelona, Barcelona, Spain
[3] Translational Genomics and Targeted Therapeutics in Solid Tumors, August Pi i Sunyer Biomedical Research Institute (IDIBAPS), Barcelona, Spain
[4] Department of Medical Oncology, Hospital Clínic de Barcelona, Barcelona, Spain
[5] SOLTI Breast Cancer Research Group, Barcelona, Spain
[6] Reveal Genomics, Barcelona, Spain
[7] Oncology Data Science Group, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain
[8] Department of Medical Oncology, Hospital Arnau de Vilanova, Valencia, Spain; FISABIO, Universidad Católica de Valencia, Spain
[9] International Breast Cancer Center (IBCC), Quirónsalud Group, Barcelona, Spain; Medica Scientia Innovation Research (MEDSIR), Barcelona, Spain; Universidad Europea de Madrid, Faculty of Biomedical and Health Sciences, Department of Medicine, Madrid, Spain.

Corresponding author:
Paolo Nuciforo, MD, PhD
Molecular Oncology Group
Vall d'Hebron Institute of Oncology
pnuciforo@vhio.net

ABSTRACT:

Breast cancer is a heterogeneous disease. Tumor cells and the surrounding microenvironment form an ecosystem that determine disease progression and response to therapy. To characterize the breast cancer ecosystem and the changes induced by targeted treatment selective pressure, we analyzed 136 HER2-positive tumor samples for the expression of canonical BC tumor diagnostic proteins at a single cell level without disrupting the spatial context. The combined expression of HER2, ER, PR, and Ki67 in more than a million cells was evaluated using a tumor-centric panel combining the four biomarkers in a single tissue section by sequential immunohistochemistry to derive 16 tumor cell phenotypes. Spatial interactions between individual tumor cells and cytotoxic T cells were studied to determine the immune characteristics of the ecosystem and the impact on response to treatment. HER2-positive tumors displayed individuality in tumor cells and immune cells composition, including intrinsic phenotype dominance which only partially overlapped with molecular intrinsic subtyping determined by PAM50 analysis. This single cell analysis of canonical BC biomarkers deepens our understanding of the complex biology of HER2-positive BC and suggests that individual cell-based patient classification may facilitate identification of optimal responders or resistant individual to HER2-targeted therapies.

INTRODUCTION:

Breast cancer (BC) is a heterogeneous disease accompanied by differences in clinical, molecular, and biological features[1], which creates a challenge for prognosis and treatment[2]. Currently, BC samples are stratified for clinical purposes based on tumor cells' expression of ER, PR, HER2, and the proliferation marker Ki67. These immunohistochemistry (IHC) biomarkers together with clinicopathologic indexes are used to predict disease outcome[3], for treatment decisions, and serve as surrogates for prognostic gene expression profiles (GEP)[4–7] categorizing BC into four basic subtypes which are related – but not equivalent – to GEP-defined intrinsic subtypes[8]. Luminal A and luminal B are roughly equivalent to [ER+|PR+] HER2− and [ER+|PR+] HER2+ tumors, respectively, though a small percentage of [ER+|PR+] HER2− tumors with Ki67 positivity are reported to belong to the luminal B subtype[9]. HER2 enriched tumors refer to [ER−|PR−] HER2+ despite the different methods used on HER2 assessment. The [ER−|PR−] HER2− (also named triple negative tumors) subtype is mainly composed of basal-like tumors, which are highly heterogeneous including at least claudin-low[10], metaplastic breast cancer[11] and interferon-rich tumours[12] in addition to core basal tumors as demonstrated by the accumulated evidence.

Although these stratifications have improved therapy success, patient responses vary within each subtype demanding better characterization of BC ecosystem. Targets of current therapies are heterogeneously expressed within and between patients. This heterogeneity equips cancer cells for proliferation, survival, and invasion and likely underlies differential treatment efficacies. This ecosystem is further shaped by cellular relationships (tumor cell-tumor cell, tumor cell-immune cell, …) and strategies targeting relationships that promote tumor development are promising.

Next generation technologies such as gene expression based molecular profiling and genetic testing are considered as the future of cancer diagnostics. Despite that, the results generated may be significantly affected by the level of intra-tumor heterogeneity in the bulk sample typically analyzed with those methodologies. Given the heterogeneity of cellular phenotypes and relationships, patients classification and treatment should ideally consider the entire tumor ecosystem. Recent single-cell RNA sequencing and mass cytometry studies provided hints into breast cancer complexity and how this may influence prognostic and response to treatment [13–16]. However, no study specifically characterizes the distribution of common breast cancer biomarkers at a single cell level in HER2-positive breast cancer without disrupting the tissue architecture. In the present study, we explored the composition, heterogeneity, and spatial organization of HER2-positive breast cancer at a single-cell level resolution maintaining the spatial information as well as the treatment induced changes following dual HER2 inhibition with lapatinib and trastuzumab. To do so, we took advantage of an innovative technique recently developed in our lab which allows multiplex in situ biomarker analyses in a single FFPE tissue section. Customized algorithms were developed to extract the different phenotypic cell populations within the tissue for subsequent analyses. In addition, we analyzed the spatial distribution and interactions between tumor phenotypes, the interaction with immune cells and their impact in predicting response to treatment.

The results from this analysis might help understand better intra-tumor heterogeneity and improve treatment strategies.

MATERIALS AND METHODS:

1. STUDY POPULATION:

Patients enrolled in the PAMELA phase II trial will be included in the study. Briefly, 151 patients with operable or locally advanced HER2-positive breast cancer were treated with neoadjuvant lapatinib (1000 mg daily) and trastuzumab (8 mg/kg IV loading dose followed by 6 mg/kg) for 18 weeks. Patients with hormone receptor (HR)-positive disease received letrozole or tamoxifen according to menopausal status. Formalin-fixed paraffin-embedded (FFPE) tumor samples at baseline and at D15 of treatment were collected according to protocol. Of the 151 patients enrolled in PAMELA study, 72 had a baseline sample and 64 had an on-treatment (day-15) sample for multiplexed immunohistochemistry analysis. Forty-nine patients had paired baseline and day-15 samples. Summary shown in **Supplementary Figure S1.**

From the all the patients, demographic patient data, tumor histopathological features (histotype, size, pT stage, pN stage), HER2 status (IHC and/or FISH results) and Hormone receptor status, PAM50 intrinsic subtype and treatment response data (pCR) were available. Clinicopathologic data is summarized in **Table 1.**

2. NGI (NEXT GENERATION IHC):

We used an automatized, simple, and flexible IHC protocol developed in our laboratory, named next generation immunohistochemistry (NGI) to study the expression of four canonical breast cancer biomarkers (ER, PR, KI67, HER2) at a single-cell level resolution[17,18]. The NGI protocol consists of iterative cycles of staining/destaining on the same tissue section and uses the combination of Ventana Discovery Ultra (Roche Diagnostics), Nanozoomer slide scanner (Hamamatsu) and Visiopharm image analysis software. Briefly, an alcohol soluble chromogen

(DISCOVERY AEC KIT (#760-258, Roche-Ventana)) was used to allow the destaining of the samples. After each automated IHC, samples were mounted in aqueous medium and digitalized (cycle 1). Subsequently, the section was destained in alcohol and submitted to the following staining cycle as shown in **Figure1A**. The sequence of the stainings were PR, KI67, HER2, ER followed by a PANCK staining for tumor area definition (antibody specification and protocol conditions in **Supplementary Table 1**.

Before image analyses, individual images generated during each NGI staining cycle were aligned into a single virtual image using the Visiopharm® software. Once co-registered, image-analysis algorithms were applied to extract the data. First, a tissue recognition APP was run and then the PANCK staining was used to recognize the tumor for tumor analysis. A pathologist supervised the regions of interest. After that, a breast panel APP was run, which localizes and classifies all the tumor cells in positive or negative for ER, PR, KI67 and Her2 biomarkers. Any brown stain above the background level (average of 210 pixel-intensity) was considered positive for the classification. The number of each of the generated 16 tumor phenotypes (or classes) and the position of them is obtained by the APP. Image analysis algorithms are shared in supplementary material.

For subgroup analyses, the obtained classes were grouped into 4 categories (**Supplementary Table 2**): Her2-enriched (HER2E) for Her2-positive, hormone receptor negative cell phenotypes (classes 6 and 8), luminal A-like (LumA) for hormone receptor positive (ER -positive and/or PR -positive), Her2-negative, and Ki67-negative cells phenotypes (classes 11,12 and 15), luminal B-like (LumB) for hormone receptor positive (ER -positive and/or PR -positive) and HER2 -positive or KI67-positive cells (phenotypes 1-5,7,9,10 and 13), and triple negative (TN) for Her2-negative and hormone receptor negative phenotypes (classes 14 and 16).

For the neighborhood analyses, CD8 staining images generated on a consecutive section from our previous study[18] were aligned with the PR images to extract tumor cell phenotypes and location. CD8 APP was run (image analysis algorithms are shared in **supplementary material**) and the number and location of each of them were obtained for consecutive analyses.

3. STATISTICAL ANALYSES:

R software (v.3.6.1)[19] was used for all statistical analyses. Statistical significance level was set to <0.05. Wilcoxon-Mann-Whitney non-parametric test was used for two group comparisons and Kruskal-Wallis test for three group comparisons. Heterogeneity analyses were done using vegan package. Intratumoral diversity (alpha diversity) was analyzed using richness, Shannon and evenness indexes. Intertumoral diversity (beta diversity) was analyzed using one of the most common metrics, the Bray-Curtis dissimilarity. Significance between groups was tested with Peranova test (adonis). Classes that had a median higher than 1% were used for the heterogeneity analyses. Hierarchical cluster analysis was done using pheatmap package and using Ward's method to group patients with similar compositions. For neighborhood analyses, we connected the cells to each other by means of a Delaunay triangulation algorithm, using the centroid of the segmentation mask. This allowed us to locate the nearest neighbors avoiding those connections that are shielded by nearest cells. In addition, to avoid connections between cells that were too far apart, we established a maximum distance of 20 microns between two neighboring cells. The percentage of connections per sample was obtained and normalized by the total tumor cells for

statistical analyses to ensure that tumor size was not affecting the results. The affinity of the CD8 to the different subtypes was calculated by dividing the percentage of the connections of the cd8 for each subtype by the percentage of that particular subtype in the sample.

RESULTS:

1. Single-cell composition analysis of HER2-positive breast cancer

We analyzed 136 prospectively collected tumor biopsies from the PAMELA trial, including 72 baseline and 64 day-15 samples from HER2-positive breast cancer using our NGI technology (**Supplementary Figure 1A**). Clinicopathological characteristics of the NGI cohort are summarized in **Table1**. In total, 1028974 cells were analyzed (mean=7566, median=4060, IQR=1378-9454). To ensure data quality, we compared NGI results generated with the matched IHC scores available as part of central confirmation analysis of the PAMELA trial. The frequencies of ER+, PR+, HER2+, and Ki67+ cells determined by NGI were comparable with the centrally determined pathological scores (**Supplementary Figure S2**).

To characterize the composition of breast cancer at a single cell level, we used a tumor-centric panel including ER, PR, HER2, and Ki67 to classify individual tumor cells in 16 different phenotypes based on the combined markers expression of the 4 markers. (**Figure 1A, Supplementary Figure S2**). As expected, most cells in HER2-positive breast cancer were expressing HER2 (intrinsic cell phenotypes: 1-8, median=92.3%) followed by ER (intrinsic cell phenotypes 1,3,5,7,9,11,13 and 15, median=57.9%), Ki67 (intrinsic cell phenotypes 1,2,5,6,9,10,13 and 14, median=16.4) and PR (intrinsic cell phenotypes 1-4 and 9-12, median=5.7) (**Supplementary Table S3).**

The analysis of distribution of individual intrinsic cell phenotypes revealed a heterogeneous distribution of tumor cell-intrinsic phenotypes in HER2-positive breast cancers, being the HER2E cell phenotype 8 the predominant cell phenotype (median=11.44, IQR=61.79) followed by TN phenotype 16 (median=3.87, IQR=9.04) and LumB phenotype 7 (median=3.63, IQR=29.01) (**Figure 1B, Supplementary Table S4**).

Analysis of paired samples showed an overall decrease of HER2-positive (median baseline=95.7%, median day-15=87.5%, p=0.022) and Ki67-positive (baseline=20.2%, day-15=7.5%, p<0.0001) cells from baseline to day-15. On the other hand, dual HER2 inhibition did not induce a significant shift in the overall composition of ER-positive and PR-positive cells. The overall decrease of HER2-positive and Ki67-positive cells from baseline to day-15 induced by the treatment was driven, at the intrinsic cell phenotype level, by the significant increase of non-proliferating TN cells (phenotype 16, median baseline=2.1%, median day-15=6.0%, p=0.0016) and decrease of proliferating HER2-positive cells (HER2+Ki67+, median baseline=18.1%, median day-15=5.1%, p<0.0001; phenotype 6, median baseline=2.4%, median day-15=0.9%, p=0.0001; phenotype 5, median baseline=0.3%, median day-15=0.01%, p=0.019) (**Figure 1C**). Interestingly, no treatment effect was observed on proliferation in HER2-negative cells (median baseline=0.4%, median day-15=0.4%, p=0.6492) providing evidence that the treatment is specifically targeting Her2-positive cells (**Figure 1D).**

The cell type frequencies varied among and between tumor molecular intrinsic subtypes determined by GEP, with a higher frequency of HER2-positive cells in HER2E samples (p-value<0.0001), PR-positive and ER-positive cells in luminal samples (p-value <0.0001), and proliferating KI67 positive cells in Basal samples (p-value <0.0001) (**Figure E, Supplementary Table S4).**

At the individual intrinsic cell phenotype level, basal tumors showed a higher frequency of TN (phenotypes 14 and 16) and HER2E (phenotypes 6 and 8) cells compared to other breast cancer subtypes, which was statistically significant for TN phenotypes. Tumors of HER2E subtype were significantly enriched in HER2E cell phenotypes 6 and 8 with a lower frequency of the remaining intrinsic cell phenotypes compared to non-HER2E tumors. A higher frequency of luminal cell phenotypes was observed in luminal samples which were also showing significantly less HER2E cell phenotypes frequency than other tumor subtypes. LumB cell phenotype 5 significantly differentiates luminal A from luminal B tumors (median LumA=0.32, median LumB=6.77, p-value=0.0013). Normal samples were mainly composed of HER2E cells (phenotypes 6 and 8) and TN cells, the latter being significantly more abundant in normal compared to non-normal samples (median phenotype 16, normal=11.67, non-normal=3.22, p-value=0.00026) (**Supplementary Table S5).** When comparing baseline with day-15 samples, we did not find significant differences neither at overall marker levels (**Figure 1F)** nor at individual intrinsic cell phenotype composition (**Figure 1G)** except for molecular intrinsic subtype luminal A where the percentage of ER+/Ki67+ and phenotypes 1 and 9 was significantly lower in on-treatment samples**.**

Lastly, we analyzed whether intrinsic cell phenotypes distribution varied according to the clinicopathological features of the tumor. As expected, clinical HR positive tumors showed a significantly higher proportion of ER-positive and PR-positive tumor cells as compared to HR-negative tumors and a lower proportion of KI67-positive and Her2-positive cells. Luminal intrinsic cell phenotypes were significantly enriched in clinical HR-positive compared to HR-negative tumors except for PR+/ER- phenotypes which were higher (phenotypes 2 and 4) or did not differ significantly (phenotype 10) in HR-negative tumors. HER2E intrinsic cell phenotypes (6 and 8) were also significantly enriched in HR-negative compared to HR-positive tumors (**Figure 1H, Supplementary Table S6**).

Tumor size and nodal stage did not significantly impact on the distribution of, neither the four biomarkers (**Supplementary Table S6**), nor the intrinsic cell phenotype tumor content.

2. HER2-positive breast cancer heterogeneity

Tumor heterogeneity is believed to drive disease progression or resistance to treatment. Intratumoral heterogeneity was determined using alfa-diversity indexes which quantify the number of different intrinsic cell phenotypes co-existing within a sample (richness index), their relative abundance (Shannon index), and how similar the phenotypes are numerically distributed (Pielou's evenness index) within a sample. Intratumoral heterogeneity increased with tumor stage (Baseline eveness index: T1 vs T2, p=0.005; T1 vs T3, p=0.0027) indicating a progressive acquisition of different cell phenotypes with tumor growth. No significant differences were found according to nodal stage. Clinical HR-positive tumors were more heterogeneous than HR-negative tumors. This finding was in line with the higher heterogeneity observed in luminal

tumors compared to other intrinsic molecular subtypes by GEP (**Figure 2A**). Treatment did not induce a significant shift in intratumoral heterogeneity as shown by paired samples analyses (median baseline=0.88, median on-treatment=0.84, p-value=0.52) (**Supplementary FiguresS2A**). Intertumoral heterogeneity was determined using the Bray-Curtis matrix which quantifies the similarity of tumors based on intrinsic cell phenotypes composition and visualized with principal coordinates analysis plot (**Figure 2B**). Analysis of intertumoral heterogeneity failed to show significantly different compositions according to their tumor size or nodal involvement, with tumors segregating together independently of these clinicopathological features. Clinical HR-positive tumors segregated together and separately from HR-negative tumors (p-value <0.0001). This diversity was maintained during treatment in day-15 samples (p-value= 0.00316). Luminal tumors (A and B) clustered together and separately from Basal and HER2-E tumors. The difference was statistically significant at day-15 (p-value= 0.01769) with normal-like tumors clustering together with HER2E and Basal tumors. Intertumoral heterogeneity analyses on paired samples showed the same results (**Supplementary Figure S3**).

Analysis of tumors according to response to anti-HER2 neoadjuvant therapy revealed that tumors from responders shared a similar composition at baseline (p-value=0.008) and separated from tumors from non-responders, suggesting that tumor intrinsic features predictive of pCR may be found before treatment is started. On the other hand, intratumor heterogeneity was not a tumor characteristic predictive of response at baseline but at day-15. Tumors from patients who did not achieve a pCR exhibited a significantly higher intratumoral heterogeneity at day-15 compared to tumors from patients who responded to the treatment (Shannon, p=0.044), indicating that on-treatment survival of different cell phenotypes may predict resistance (**Figure 2A**).

Response analysis by tumor intrinsic cell phenotype composition.

To assess whether tumor intrinsic cell phenotype composition may affect response to neoadjuvant anti-HER2, we calculated whether the frequency of each individual tumor cell phenotype differed between patients achieving or not achieving pathological complete response (pCR).

As our dataset was a subset of the original PAMELA trial, we first determined whether this smaller subset was representative of the entire study in terms of response analyses. At baseline, TILS were significantly higher in responders (median=20) compared to non-responders (median =10, p=0.0017). At day-15, TILs (median responders=50, median non-responders=15, p=0.0036) and the CelTIL (median responders=58.3, median non responders=-1.5, p<0.0001) were significantly higher in responders thus confirming previous finding obtained from analysis of the full dataset.

Comparative analysis of tumor intrinsic cell phenotypes distribution between responders and non-responders did not show significant differences of any individual cell phenotypes between responders and non-responders both at baseline and day-15, except for LumB phenotype 2 at baseline (median responders=0.1%, median non-responders=0.03%, p=0.03), LumB class 11 at baseline (median responders=0.0%, median non-responders=0.04%, p=0.02), and LumB class 1 at day-15 (median responders=0.0% median non-responders=0.08%, p=0.04). However, median values for differentially abundant phenotypes were extremely low, being below 1% in both outcome groups (**Supplementary Figure S4**). When grouping the intrinsic cell phenotypes, we

found a higher proportion of Her2-positive (median responders=95.9%, median non-responders=89.9%, p-value=0.028) and Her2-enriched (phenotypes 6 and 8) cells in responders (median=73.5%) as compared to non-responders (median=10.0%, p-value=0.047). Luminal A phenotypes (11,12 and 15) were enriched in non-responders (median non-responders=1.3%) as compared to responders (median=0.2%p-value=0.003) (**Figure 3A**).

The total number of HER2-positive cells significantly decreased in patients responding to anti-HER2 (median baseline=97.9%, median day-15=87.6%, p=0.009) but not in patients who did not, suggesting a reduction in tumor burden. On the other hand, patients who did not achieve a pCR showed an increase of LumB phenotype 7 with treatment (median baseline=0.8%, median day-15=3.9%, p=0.042) which was not observed in responders, possibly reflecting the expansion of a resistant clone (**Figure 3B**).

Consensus clustering was performed to group samples from HER2+ patients according to the distinct cell phenotypes. Unsupervised hierarchical clustering (**Figure 3C**) using paired samples classified tumors into six groups (C1-C6). Cluster 1 (HER2E) was dominated by tumor cells from phenotypes 6 and 8. Tumors in cluster 1 were clinical HR-negative (86.1%) and molecular intrinsic subtype Her2-E or Normal-like (94.7%). Clusters 2 and 5 (LumA) were enriched in cells from phenotypes 15 and 11, respectively. All tumors were clinical HR-positive. 63.2% were LumA and 21.1% LumB by intrinsic molecular subtyping. Clusters 3 and 4 (LumB) were composed mostly of cells from phenotypes 7 (cluster 3) and 3 (cluster 4). All but two tumor samples in these clusters were clinical HR-positive and exhibited a mixed molecular intrinsic subtyping (32.3% Her2-E, 41.2% Lum-A, 8.8% Lum-B,5.9% Basal and 11.8% Normal). Lastly, cluster 6 (Basal) was enriched in TN phenotypes 14 and 16. Tumors were clinical HR-negative (7/8, 87.5%) and half exhibited a basal intrinsic molecular subtype. Baseline and on-treatment samples from the same patients tended to group together in the same cluster or moved to a similar cluster during treatment (**Figure 3E**) More treatment-induced changes were observed in molecular intrinsic subtyping by PAM50. Similarly, only 20% of LumB tumors remained LumB after treatment. Treatment did not significantly impact on LumA and Basal tumors which tended to maintain the same subtype after treatment (**Figure 3F**).

To determine whether patients responding or not responding to anti-Her2 therapy segregated within a specific cluster, we performed consensus clustering of all (paired and unpaired) baseline and on-treatment samples independently. At baseline, hierarchical clustering classified patients into five groups corresponding to HER2E (C1, phenotypes 6 and 8), LumB (C2 and C3, phenotypes 7 and 3), LumA (C4, phenotype 15) and mixed Luminal/Basal (C5) clusters. Twelve out of 31 (39%), 3 out of 16 (19%), 0 out of 15 (0%) and 4 out of 19 (20%) patients from HER2E, LumB, LumA, and mixed clusters achieved a pCR, respectively. Patients in the HER2E cluster C1 had a significantly higher probability of responding to anti-HER2 therapies as compared to those in other clusters (Fischer exact test, p=0.01) (**Figure 3G**). At day-15, clustering analysis showed six tumor clusters with similar phenotype compositions as the baseline clusters. Six of 19 patients (32%), 2 out of 21 (9.5%), 1 out of 12 (8%), and 2 out of 12 (17%) patients from HER2E, LumB, LumA, and mixed clusters achieved a pathological complete response, respectively. Patients in the HER2E cluster had a non-significant higher probability of responding to anti-HER2 therapies as compared to those in other clusters (Fischer exact test, p=0.06) (**Figure 3H**).

3. Breast cancer tumor and immune cells relationship

Interactions between tumor cells and immune cells within the tumor microenvironment drive disease progression and response to treatment. To study homotypic and heterotypic relationships between tumor and immune cells, we determined the percentage of connections between each individual tumor cell phenotype, the distribution of tumor cell phenotypes according to tumor infiltrating lymphocytes (TILs) and the spatial relationship between cytotoxic (CD8+) immune cells with each tumor cell phenotype.

Homotypic tumor cells relationships were the most common. Most of the connections were found between cells belonging to the same or similar phenotype. Most of the homotypic connections were found between Her2E phenotype 8 cells, followed by connections between LumB phenotype 3 cells, TN phenotype 16 cells, LumA phenotype 11 cells and heterotypic connections between phenotype 8 with phenotypes 6 and 7 (**Figure 4A**).

To determine whether tumor cell phenotypes were differentially enriched in breast cancer according to the level of immune infiltration, we correlated the proportion of TILs with that of each individual cell phenotype in the same sample. Levels of TILs were positively correlated with HER2E phenotypes 6 and 8 both at baseline (0.49 and 0.41, respectively) and at day-15 (0.32 and 0.55, respectively) (**Supplementary Figure S5 A**). When we grouped tumors into inflamed (TILs >50%) and non-inflamed (TILs <50%) based on TILs infiltration, inflamed tumors were again significantly enriched in HER2E cell phenotypes 6 (median high=8.9%, median low=0.7%, p=0.026) and 8 (median high=59.3%, median low=6.3%, p=0.005) and depleted of LUM-B cells (median high=8.9%, median low=35.6%, p=0.048) compared to non-inflamed tumors (**Figure 4B**, **Supplementary Table S7**).

After finding the enriched subtypes in inflamed tumors, we wanted to explore whether TILs levels differed according to tumor intrinsic cell phenotype-defined clusters. We found that patients from clusters 2 and 5 (non-responder clusters) had significantly lower cd8 densities (median cd8 density in cluster 2= 114.5cells/mm2 and cluster 3 202.2 cells/mm2) while cluster 1 patients (responders) had the highest cd8 densities (median 723.5 cells/mm2) (**Figure 4C**).

To determine the spatial relationship between CD8+ immune cells and tumor cell phenotypes, we calculated the proportion of intratumoral immune-tumor cells connections. The most common connections were between CD8 and HER2E tumor cell (phenotypes 8, median=22.75%; phenotype 6, median=2.5%) followed by TN cell phenotype 16 (median=3.93%). Similar results were found when dividing the dataset into baseline and day-15 samples (**Figure 4D, Supplementary Figure S5 B**). Anti-HER2 treatment induced a general increase of number of connections between CD8+ immune cells and tumor cells (median baseline=0.2%, median day15=0.5%, p=0.01) without significant changes in connections of CD8 with any individual tumor cell phenotypes (**Figure 4D, Supplementary Table S8**).

The higher number of connections observed between CD8 immune cells and HER2E and TN tumor cells phenotypes may be attributable to the higher prevalence of these phenotypes within the tumor. To elucidate whether cytotoxic T cells tended to interact preferentially with a specific tumor cell phenotype within a tumor, we determined the affinity of cytotoxic T cells for each

tumor cell phenotype by dividing the percentage of connections of CD8+ cells with each of the phenotype by the percentage of tumor cells that were of the same phenotype.

More affinity of CD8 was found with TN cell phenotype 16 (median=0.58) followed by TN cell phenotype 14 (median=0.50), Her2E phenotype 8 (median=0.47), LumB phenotype 10 (median=0.44) and Her2E phenotype 6 (median=0.43) **(Figure 4E)**. The affinity of CD8 for the Her2E cell phenotype 8 (median baseline=0.3%, median on-treatment=0.6%, p=0.018) and LumB phenotype 7 (median baseline=0.1%, median on-treatment=0.3%, p=0.029) significantly increased with treatment **(Supplementary Table S9)**.

Lastly, we wanted to analyze whether differences in the types of connection and affinity between responders and non-responders could be found. Overall, no significant differences in the number of connections between cd8 and the different cell phenotypes were observed between responders and non-responders. At baseline, no significant differences of CD8 connections or affinity were found between responders and non-responders. On day-15, tumors from patients responding to the treatment showed an overall increase in CD8/tumor cell ratio, connections of CD8 with any tumor cells, with HER2-positive cells and with phenotype 8 cells compared with tumors from non-responders. Similarly, higher CD8 affinity for HER2E phenotypes (phenotype 6 and 8), LumB phenotypes (4, 7, and 13) and LumA phenotype 11 was found in responders compared to non-responders **(Figure 4F, Supplementary Table S10)**. These findings were confirmed in paired sample analysis.

Anti-HER2 treatment induced a significant increase in the number of CD8 connections with any tumor cell (median baseline=0.2%, median on-treatment=1.7%, p=0.007) and, specifically, with TN phenotype 16 (median baseline=0.0%, median on-treatment=0.2%, p=0.02) in responders compared to non-responders **(Figure 4H)**. Similarly, affinity for the Her2E class8 (median baseline=0.3, median on-treatment=1.6, p=0.0059) and class6 cells (median baseline=0.4, median on-treatment=1.5, p=0.022); LumB class7 (median baseline=0.1, median on-treatment=0.7, p=0.04) and class4 cells (median baseline=0.3, median on-treatment=2.7, p=0.013) **(Figure 4I)** increase in the responders but not in non-responders.

Due to the higher number of connections between CD8 and tumor cells observed in responders compared to non-responders, we included this feature to the intrinsic tumor cell phenotypes and performed unsupervised clustering analysis. The addition of immune feature to tumor features did not improve segregation of responder vs non-responders at baseline **(Figure 4J)**. However, at day-15, the number of connections between cytotoxic T cells and tumor cells separated the HER2E cluster in two distinct clusters with different response rate the one with more connections specially enriched in (67% of responders) as expected as compared to the other her2 enriched cluster that became not enriched in responders (15% of responders) **(Figure 4K)**.

DISCUSSION:

In this study, we conducted an *in situ* single-cell analysis of Her2 positive breast cancer using a sequential immunohistochemistry protocol combined with image analysis to provide virtual multiplexed expression of different biomarkers on the same slide. The developed methodology (called NGI or next generation immunohistochemistry) provided information on the expression of 4 canonical breast cancer biomarker at an unprecedented resolution allowing analysis of their

co-expression in the same cell, their distribution, their spatial interaction (tumor cell-tumor cell, tumor cell-immune cells) and their changes during anti-Her2 treatment.

ER, PR, Her2 and Ki67 are standard biomarkers used in clinical practice for diagnosis and prognostication of breast cancer. Their expression is determined by conventional immunohistochemistry and their status defined by consensus cutoffs recommended in clinical guidelines. HER2-positive tumors are defined by >10% tumor cells showing strong (3+) complete membrane positivity or equivocal 2+ weak to moderate complete membrane staining in >10% of the tumor cells with Fish-negative result[20]. Similarly, hormone receptor status is defined by >10% tumor cells showing positivity at any intensity of ER and/or PR. For Ki67, no standard cut-off exists. A 20% threshold is generally used to define a high proliferative tumor and usually associated with poor prognosis[21–23].

Treatment strategies are currently defined according to the individual status of each one of these biomarkers. However, treatment responses are heterogeneous and a better understanding of the breast cancer tumor ecosystem is needed.

Transcriptomics and proteomics approaches have been used to analyze the landscape of breast cancer and dig into the complexity of the tumor. However, these studies are performed on bulk samples, thus lacking contexture analyses[24–26]. Single-cell studies have provided a deeper understanding of the different cell populations distribution, tumor heterogeneity and influence on cancer progression and resistance[27–32], but to the best of our knowledge, no studies have performed single cell level resolution analyses in Her2-positive breast cancers during the anti-her2 treatment.

The approach we described in this study combines a routine, widely used methodology such as IHC with image analyses algorithms to extract complex data from a single slide. Despite immunohistochemistry is traditionally considered a qualitative, single-plex, low-throughput methodology, our NGI protocol allows to virtually multiplex up to 6 (or more) different antibodies on the same slide to simultaneously study the co-expression, distribution, spatial interactions, and function of biomarkers of interest in the target cell (tumor or immune) with a single-cell level resolution and without disrupting the tissue organization by maintaining the spatial information. In this study, we developed a custom-made 5-plex panel which included HER2, ER, PR, Ki67 (classical breast cancer biomarkers) and CK (as tumor mask) to classify more than million cells into 16 different tumor cell intrinsic phenotypes. With an average of 4060 cells/sample analyzed; our approach is at least equivalent (if not superior in term of cells analyzed) to other single-cell breast cancer studies[30–34].

Cell phenotypes prevalence and distribution were investigated in HER2-positive breast cancer from the PAMELA trial and correlated with clinicopathological and molecular features of the tumors. Intratumoral and intertumoral heterogeneity was quantified and impact of anti-HER2 inhibition on the cellular composition of the tumors and response to treatment investigated. Finally, tumor cell phenotype information was integrated with cytotoxic T cells spatial data to study homotypic and heterotypic relationships between tumor and immune cells and how interactions within the tumor microenvironment may predict treatment response or resistance.

Several confirmatory findings validated our experimental approach. First, comparative analysis of biomarkers expression determined by NGI was highly correlated with those obtained by standard immunohistochemistry analysis performed at central laboratory. Second, the overall

reduction of Her2 and ki67 from baseline to day-15 confirmed previous findings reported in the Pamela trial using alternative methodologies as well as those from other neoadjuvant studies in Her2 positive breast cancer. Importantly, higher resolution analysis of tumor individual cell populations showed that Her2-positive breast ecosystem is constituted by multiple tumor cell phenotypes (both her2+ and her2-), which are differently modulated by the treatment. Non proliferating Her2+/ER-/PR- represents the predominant phenotype (class 8) followed by triple negative (16) and luminal B (7) phenotypes. Dual her2 blockade did not affect all phenotypes in the same manner. The significant reduction of proliferative her2 positive phenotypes-only without an effect on her2 negative ones is indicative of a clear pharmacodynamic effect of dual her2 inhibition[24,25,35].

Poor prognosis and therapy resistance are associated with tumor heterogeneity[36,37]. We analyzed intratumor and intertumor heterogeneity according to breast cancer intrinsic cell phenotype composition and their impact on response to treatment. First, we found that intratumoral heterogeneity increases with tumor stage indicating a progressive acquisition of different cell phenotypes with tumor growth[38]. Second, we found that HR-positive tumors (clinical and luminal subtype by PAM50) exhibited higher intratumor and intertumor heterogeneity as compared with HR-negative tumors (clinical HR-negative and HER2E by PAM50). Third, tumor heterogeneity was inversely correlated with the probability of achieving a pCR. These findings may explain why her2-enriched and HR-negative patients, which are homogeneously composed predominantly by tumor intrinsic cell phenotypes 8 and 6, showed the highest rate of response to anti-HER2 treatment[24,25].

On the other hand, clinical HR-positive, luminal tumors by PAM50 were highly heterogenous in composition with significant enrichment of different LumB (3 and 7) and LumA (11, 12 and 15) intrinsic cell phenotypes and depletion of HER2E phenotypes (6 and 8) which, in turn, resulted in reduced sensitivity to HER2 targeted therapy. Interestingly, the association found between higher intratumoral heterogeneity at day-15 and lower response rates indicated that on-treatment survival of different cell phenotypes may predict resistance. Heterogeneity analysis also revealed that tumor intrinsic features predictive of pCR may be found before the treatment is started as shown by the similar ecosystem found at baseline in tumors from responder with beta diversity analysis.

Consensus clustering was performed to group samples from HER2-positive patients according to the distinct cell phenotypes composition and correlated with response. Six different clusters (one HER2E, 2 LumA-like, 2 LumB-like, and one Basal) were identified based on tumor intrinsic cell phenotypes composition which only partially recapitulated molecular intrinsic subtyping by PAM50. Baseline and on-treatment samples from the same patients tended to group together in the same cluster or moved to a similar cluster during treatment. In contrast, PAM50 changes during treatment were notable with 20% of HER2E tumors shifting to LumA molecular subtype. Patients in the HER2E cluster (phenotypes 8 and 6) had a higher probability of responding to anti-HER2 therapies as compared to those in other clusters (Baseline, p=0.01; day-15, p=0.06). Importantly, patients in LumA clusters (enriched in phenotypes 11 and 15) were exquisitely resistant to anti-HER2 inhibition. This data provides more insight into our previous observation that HER2E tumors cells that are sensitive to anti-HER2 therapy but do not die acquire a Luminal A phenotype[35]. Our new results point to tumor intrinsic cell phenotype 15 (ER+/PR-/HER2-/Ki67-) as a resistant clone which pre-exist at low frequencies in HER2-positive breast cancer even before treatment starts, independently of the PAM50 subtype. The observed rapid

acquisition (after 14 days of treatment) of PAM50 LumA molecular subtype may indeed reflect the expansion (as consequence of reduction of HER2E cells due to treatment) of this preexisting resistant clone. On the other hand, PAM50 HER2E tumors that became normal-like, an on-treatment biomarker of tumor responsiveness, are still predominantly composed of HER2E cells (phenotypes 8 and 6) which are diluted by stromal contamination due to decrease in tumor burden as consequence of response.

We also found LumB phenotype 7 (HER2+/ER+/PR-/Ki67-) as a resistant phenotype. This finding reflects the possible effect of the strong interplay between HER2 and ER, which may negatively impact on the response to anti-HER2 [27,30,39,40].

The study of how tumor and immune cells interact within the tumor microenvironment may increase our understanding of treatment response[25,41]. We've previously reported that patients responding to dual anti-HER2 inhibition showed higher TILS compared to non-responders[18]. In the present study, we further dig into tumor-immune cells relationship by spatial analysis. Inflamed tumors were enriched in HER2E cells (phenotypes 6 and 8) as opposed to non-inflamed ones where Luminal phenotypes predominated. In line with these observations, densities of CD8+ immune cells were higher in the HER2E cluster with LumA cluster showing the lowest levels of cytotoxic T cell infiltration. Connectivity analysis showed a higher number of connections between CD8+ immune cells and HER2E and TN tumor cells phenotypes as compared to other phenotypes, more affinity of CD8 for TN cells (phenotypes 14 and 16) and less affinity for hormone receptors positive cells in general. This observation matched with the differences in immunogenicity described according to breast cancer subtypes and partially explains why HER2-positive and TN tumors are more immunogenic[42] as well as why HER2-positive breast cancer enriched in HER2E responded significantly better to anti-HER2 inhibition compared to other phenotypes.

During her2-treatment, cytotoxic T cells increase significantly[18,24,25] and higher densities of intratumoral CD8s associated with pCR[18]. Here, we now showed that this increase reflects a higher number of connections and increased affinity of CD8 with tumor cells and HER2-positive cells in responders which is not found in patients who did not respond to the treatment. Lastly, the addition of the immune microenvironment status to tumor intrinsic cell phenotypes separated the HER2E cluster at day-15 into two subgroups with different response rates based on the connections between CD8 and tumor cells.

One of the biggest limitations of the study is the limited number of cases which did not allowed us to develop a predictive model of response to the treatment. Second, the number of biomarkers analyzed simultaneously on the same section is lower as compared with other single-cell, high throughput methodologies. Third, our study did not consider the different levels of expression of HER2, which is particularly relevant now due to the development of antibody-drug conjugates targeting the HER2-low breast cancer population. Future studies would be needed to address these gaps.

In conclusion, we have analyzed the HER2-positive breast cancer ecosystem at a single-cell resolution using a relatively simple and inexpensive methodology developed in our laboratory. Our findings confirmed previous observation using more expensive and less accessible approaches and expand our knowledge of Her2-positive breast cancer biology and composition

during anti-her2 treatment providing potential clinically important information that could be used to improve targeted therapies.

BIBLIOGRAPHY:

1. Simpson, P. T., Reis-Filho, J. S., Gale, T. & Lakhani, S. R. Molecular evolution of breast cancer. *J Pathol* **205**, 248–254 (2005).
2. Dunnwald, L. K., Rossing, M. A. & Li, C. I. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Research* **9**, R6 (2007).
3. Blows, F. M. *et al.* Subtyping of Breast Cancer by Immunohistochemistry to Investigate a Relationship between Subtype and Short and Long Term Survival: A Collaborative Analysis of Data for 10,159 Cases from 12 Studies. *PLoS Med* **7**, (2010).
4. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**, 10869 (2001).
5. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature 2000 406:6797* **406**, 747–752 (2000).
6. Bernard, P. S. *et al.* Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *Journal of Clinical Oncology* **27**, 1160 (2009).
7. Sørlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100**, 8418 (2003).
8. Dai, X., Chen, A. & Bai, Z. Integrative investigation on breast cancer in ER, PR and HER2-defined subgroups using mRNA and miRNA expression profiling. *Sci Rep* **4**, (2014).
9. Cheang, M. C. U. *et al.* Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer. *JNCI Journal of the National Cancer Institute* **101**, 736 (2009).
10. Prat, A. *et al.* Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* **12**, R68 (2010).
11. Gibson, G. R., Qian, D., Ku, J. K. & Lai, L. L. Metaplastic breast cancer: Clinical features and outcomes. *American Surgeon* **71**, 725–730 (2005).
12. Hu, Z. *et al.* The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7**, 96 (2006).
13. Sinha, V. C. *et al.* Single-cell evaluation reveals shifts in the tumor-immune niches that shape and maintain aggressive lesions in the breast. doi:10.1038/s41467-021-25240-z.
14. Li, X. *et al.* A novel single-cell based method for breast cancer prognosis. *PLoS Comput Biol* **16**, e1008133 (2020).
15. Ali, H. R. *et al.* Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nature Cancer 2020 1:2* **1**, 163–175 (2020).
16. Georgopoulou, D. *et al.* Landscapes of cellular phenotypic diversity in breast cancer xenografts and their impact on drug response. *Nat Commun* **12**, (2021).
17. Serna, G. *et al.* Sequential immunohistochemistry and virtual image reconstruction using a single slide for quantitative KI67 measurement in breast cancer. *Breast* **53**, 102–110 (2020).
18. Griguolo, G. *et al.* Immune microenvironment characterisation and dynamics during anti-HER2-based neoadjuvant treatment in HER2-positive breast cancer. *npj Precision Oncology 2021 5:1* **5**, 1–12 (2021).
19. R Development Core Team (2014) R A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing, Vienna. - References - Scientific Research Publishing. https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=1622068.
20. Ahn, S., Woo, J. W., Lee, K. & Park, S. Y. HER2 status in breast cancer: changes in guidelines and complicating factors for interpretation. *J Pathol Transl Med* **54**, 34 (2020).
21. Goldhirsch, A. *et al.* Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Annals of Oncology* **24**, 2206–2223 (2013).
22. Falck, A.-K., Fernö, M., Bendahl, P.-O. & Rydén, L. St Gallen molecular subtypes in primary breast cancer and matched lymph node metastases - aspects on distribution and prognosis for patients with luminal A tumours: results from a prospective randomised trial. *BMC Cancer* **13**, 558 (2013).
23. Focke, C. M., van Diest, P. J. & Decker, T. St Gallen 2015 subtyping of luminal breast cancers: impact of different Ki67-based proliferation assessment methods. *Breast Cancer Research and Treatment 2016 159:2* **159**, 257–263 (2016).
24. Hurvitz, S. A. *et al.* Pathologic and molecular responses to neoadjuvant trastuzumab and/or lapatinib from a phase II randomized trial in HER2-positive breast cancer (TRIO-US B07). *Nature Communications 2020 11:1* **11**, 1–15 (2020).
25. McNamara, K. L. *et al.* Spatial proteomic characterization of HER2-positive breast tumors through neoadjuvant therapy predicts response. *Nat Cancer* **2**, 400–413 (2021).
26. Brasó-Maristany, F. *et al.* Phenotypic changes of HER2-positive breast cancer during and after dual HER2 blockade. doi:10.1038/s41467-019-14111-3.
27. Ding, S., Chen, X. & Shen, K. Single-cell RNA sequencing in breast cancer: Understanding tumor heterogeneity and paving roads to individualized therapy. *Cancer Commun* **40**, 329–344 (2020).
28. Powell, A. A. *et al.* Single Cell Profiling of Circulating Tumor Cells: Transcriptional Heterogeneity and Diversity from Breast Cancer Cell Lines. doi:10.1371/journal.pone.0033788.
29. Wagner, J. *et al.* A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell* **177**, 1330-1345.e18 (2019).
30. Chung, W. *et al.* Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nature Communications 2017 8:1* **8**, 1–12 (2017).
31. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).
32. Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics 2021 53:9* **53**, 1334–1347 (2021).
33. Wagner, J. *et al.* A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell* **177**, 1330-1345.e18 (2019).

34. Karaayvaz, M. *et al.* Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nature Communications 2018 9:1* **9**, 1–10 (2018).
35. Brasó-Maristany, F. *et al.* Phenotypic changes of HER2-positive breast cancer during and after dual HER2 blockade. *Nat Commun* **11**, (2020).
36. Andor, N. *et al.* Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nature Medicine 2015 22:1* **22**, 105–113 (2015).
37. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* **12**, 323–334 (2012).
38. Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).
39. Bonacho, T., Rodrigues, F. & Liberal, J. Immunohistochemistry for diagnosis and prognosis of breast cancer: a review. *https://doi.org/10.1080/10520295.2019.1651901* **95**, 71–91 (2019).
40. Giuliano, M. *et al.* Upregulation of ER signaling as an adaptive mechanism of cell survival in HER2-positive breast tumors treated with Anti-HER2 therapy. *Clinical Cancer Research* **21**, 3995–4003 (2015).
41. Nederlof, I., Horlings, H. M., Curtis, C. & Kok, M. A High-Dimensional Window into the Micro-Environment of Triple Negative Breast Cancer. *Cancers 2021, Vol. 13, Page 316* **13**, 316 (2021).
42. Disis, M. L. & Stanton, S. E. Triple-negative breast cancer: immune modulation as the new treatment paradigm. *Am Soc Clin Oncol Educ Book* e25–e30 (2015) doi:10.14694/EDBOOK_AM.2015.35.E25.

ACKNOWLEDGEMENTS:

TABLES:

Table1.

	All Patients		Multiplex IHC	
	N	%	N	%
N	151	-	72	-
Age, mean (range)	55 (30-86)		55 (30-82)	
Menopausal status				
Pre-menopausal	61	40.4%	25	34.7%
Post-menopausal	90	59.6%	47	65.3%
Tumor size (mm), median (range)	24 (10-110)		28 (10-110)	
Clinical nodal status				
N0	98	64.9%	46	63.9%
N1	50	33.1%	24	33.3%
N2	3	2.0%	2	2.8%
HR status				
Negative	74	49.0%	33	45.8%
Positive	77	51.0%	39	54.2%
Tumor stage				
T1	60	39.7%	26	36.1%
T2	79	52.3%	40	55.6%
T3	12	8.0%	6	8.3%
Pathologic Complete Response				
Yes	40	26.5%	19	26.4%
No	111	73.5%	53	73.6%

Table 1. Summarized clinicopathologic data of the analyzed samples.

FIGURE LEGENDS:

Figure 1. Her2-positive breast cancer composition.
A) Illustration of NGI methodology used for the study. A single slide is stained for the first biomarker, scanned and destained and the process is repeated five times for each of the biomarkers. All obtained images are aligned and image analysis is used for the extraction of each tumor cell classification and location per sample. Information that is used for composition, clusterization and connectivity analyses. **B)** Box plot of the phenotype composition of her2-positive breast cancer samples, showing classes names, median, IQR for all samples and subtype classification into LumB, LumA, HER2-e and TN. In grey the positivity for each of the analyzed biomarkers (HER2, ER, PR and KI67) for each of the 16 generated classes. **C)** Box plots of phenotype-changes during the treatment in paired samples. **D)** Box plots of grouped Her2+KI67+ and Her2-Ki67+ changes during the treatment. **E)** Box plots of percentages of Her2, ER, PR and KI67 tumoral cells in the PAM50 groups. **F)** Box plots of general-composition of the different PAM50 groups. In pink SCR samples and in blue the day-15 samples. Star for significant differences between scr and day-15 samples medians. **G)** Box plots of phenotype-composition of the different PAM50 groups. In pink SCR samples and in blue the day-15 samples. Star for significant differences between scr and day-15 samples medians. **H)** Box plot of HR status differences in LumB, LumA and TN groups.

Figure 2. Heterogeneity analyses.
A) Box plots showing intratumoral diversity variation for tumor stage, nodal status, hormone receptor status, pam50 classifications and response in all (left), scr (middle) and day-15 (right) samples. Star shows the significant intratumoral heterogeneity between groups.
B) PCoA analysis plots of Bray-Curtis computed distances between all (left), scr (middle) and day-15 (right) samples. Different clinicopathologic features are shown in different colors. Star shows the groups that are significantly different in phenotype composition.

Figure 3. Response analyses. A) Boxplots of significant variables between responders and responders. **B)** Boxplots of treatment induced changes in paired responders and non-responder patients. **C)** Heatmap showing the percentage of each phenotypes of paired scr and day-15 samples. Each cluster is generated with patients with similar phenotype compositions. Heatmap colors represent percentage of each of the phenotypes. Annotation colors are shown next to the heatmap. **D)** Representative images of classified tumors from each cluster. Colors of the cells are shown next to the classes in the heatmap. **E)** Alluvial plot showing the cluster changes of paired patients during the treatment. **F)** Alluvial plot showing the Pam50 group changes during the treatment. **G)** Heatmap of SCR samples. Number of responders and non-responders are shown for the first cluster and for the others. **H)** Heatmap of day-15 samples. Number of responders and non-responders are shown for the first cluster and for the others.

Figure 4. Spatial analyses. A) Dot-plot showing the connections between different tumor phenotypes. The size and color of the dots represent the proportion of connections. **B)** Boxplot of significantly different variables between TIL high and TIL low groups. **C)** Box plot of CD8 densities in the different clusterized groups' samples. **D)** Box plots of the CD8 connections with each of the tumor phenotypes in all samples. Dot plots of the medians in SCR and day-15 are shown in the right. **E)** Box plots of the CD8 affinity with each of the tumor phenotypes in all samples. Dot plots of the medians in SCR and day-15 are shown in the right. **F)** Representative dot plots and virtual multiplexed images of a responder and non-responder patient. The size and color of the dots represent the proportion of connections between CD8 and tumor cells. CD8 cells are shown in red in the image. **G)** Boxplots of treatment induced connection-changes in paired responders and non-responder patients. **H)** Boxplots of treatment induced affinity-changes in paired responders and non-responder patients. **I)** Heatmap of SCR samples with CD8-tumor connection information. Annotation colors are shown next to the heatmap. **J)** Heatmap of day-15 samples with CD8-tumor connection information. Annotation colors are shown next to the heatmap.

Supplementary Figure 1. A) Sample cohort diagram. **B)** Scatter plot and Spearman coefficient of NGI and central lab results for each of the 4 canonical biomarkers. **C)** Box plots of the percentage of positive cells for the different biomarkers with NGI for the central lab categorized results.

Supplementary Figure 2. Heterogeneity analyses. A) Box plots showing non-significant intratumoral diversity variation in paired samples during the treatment. **B)** PCoA analysis plots of Bray-Curtis computed distances in paired samples. Different clinicopathologic features are shown in different colors. Star shows the groups that are significantly different in phenotype composition.

Supplementary Figure 3. Results from response analyses. A) Box plots of the different phenotypes proportions in responder and non-responder patients in SCR and day-15 samples. **B)** Box plots of the different phenotype proportions on responder and non-responder SCR samples. **C)** Box plots of the different phenotype proportions on responder and non-responder day-15 samples.

Supplementary Figure 4. Spatial analyses. A) Scatter plot and Spearman correlation coefficient of TILs and different phenotypes in SCR and day-15 simples. B) Box plots of CD8 connections with the different tumor phenotypes for SCR and day-15 samples. C) Box plots of CD8 affinity to the different tumor phenotypes for SCR and day-15 samples.

Figure 1

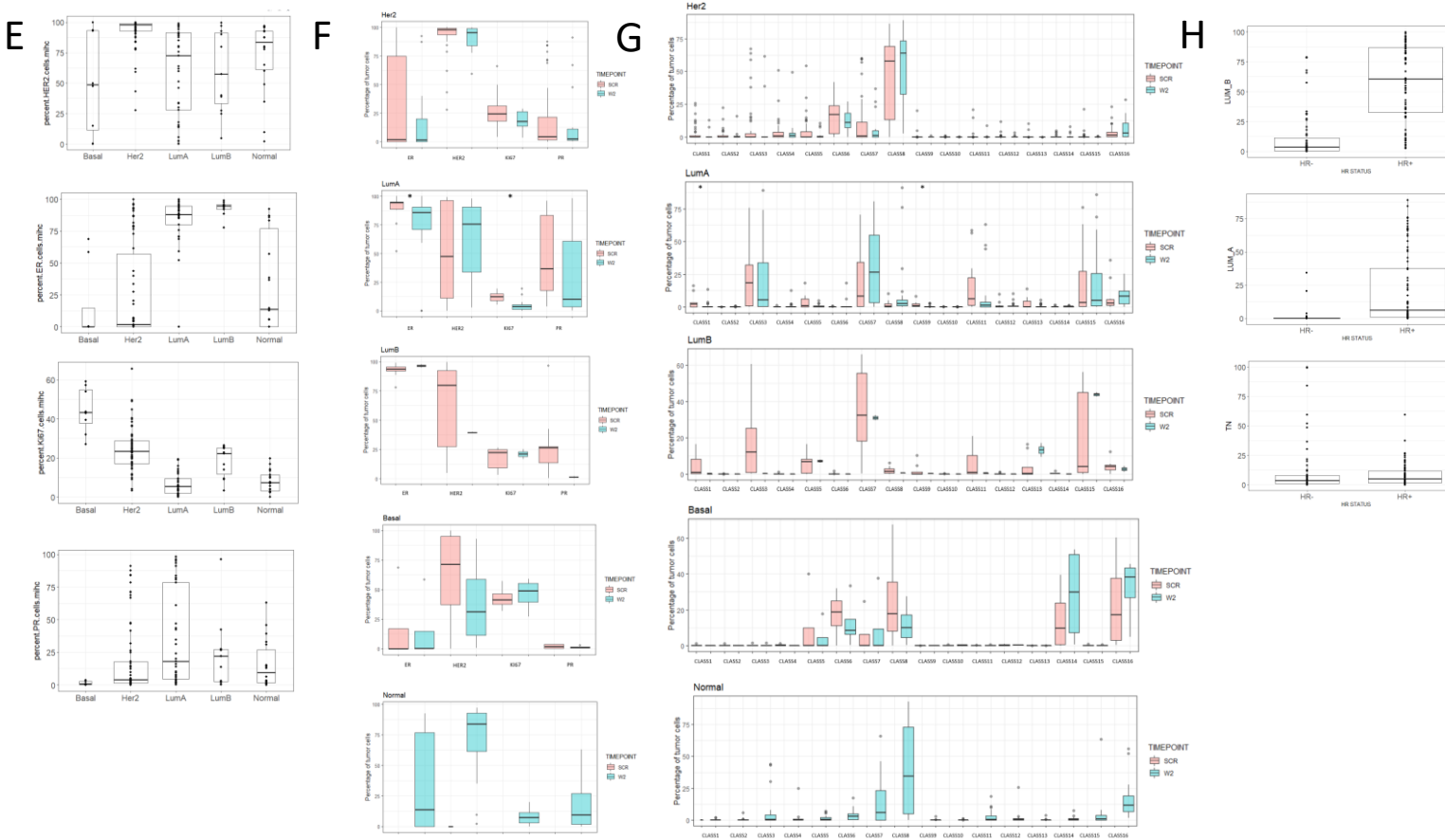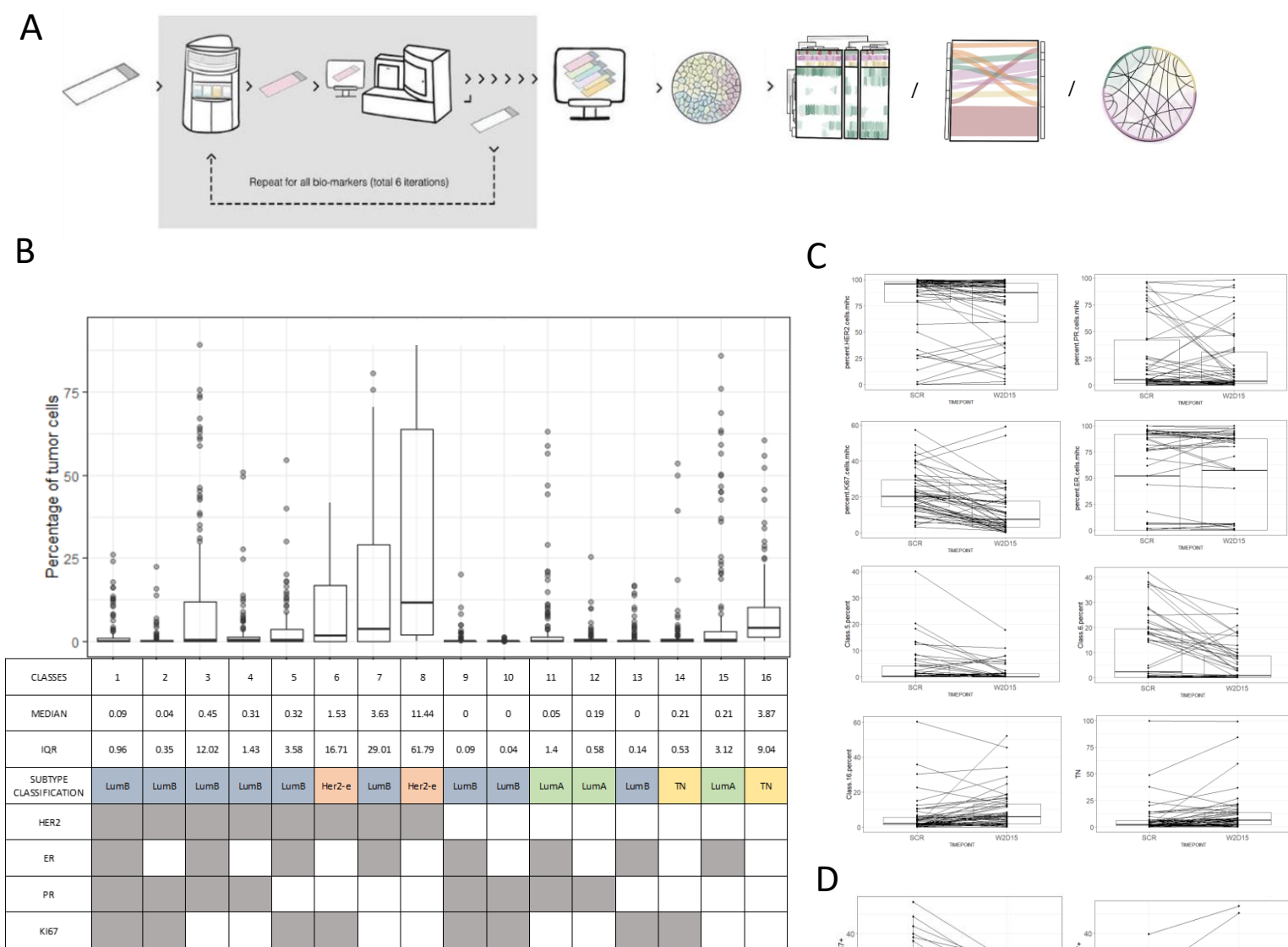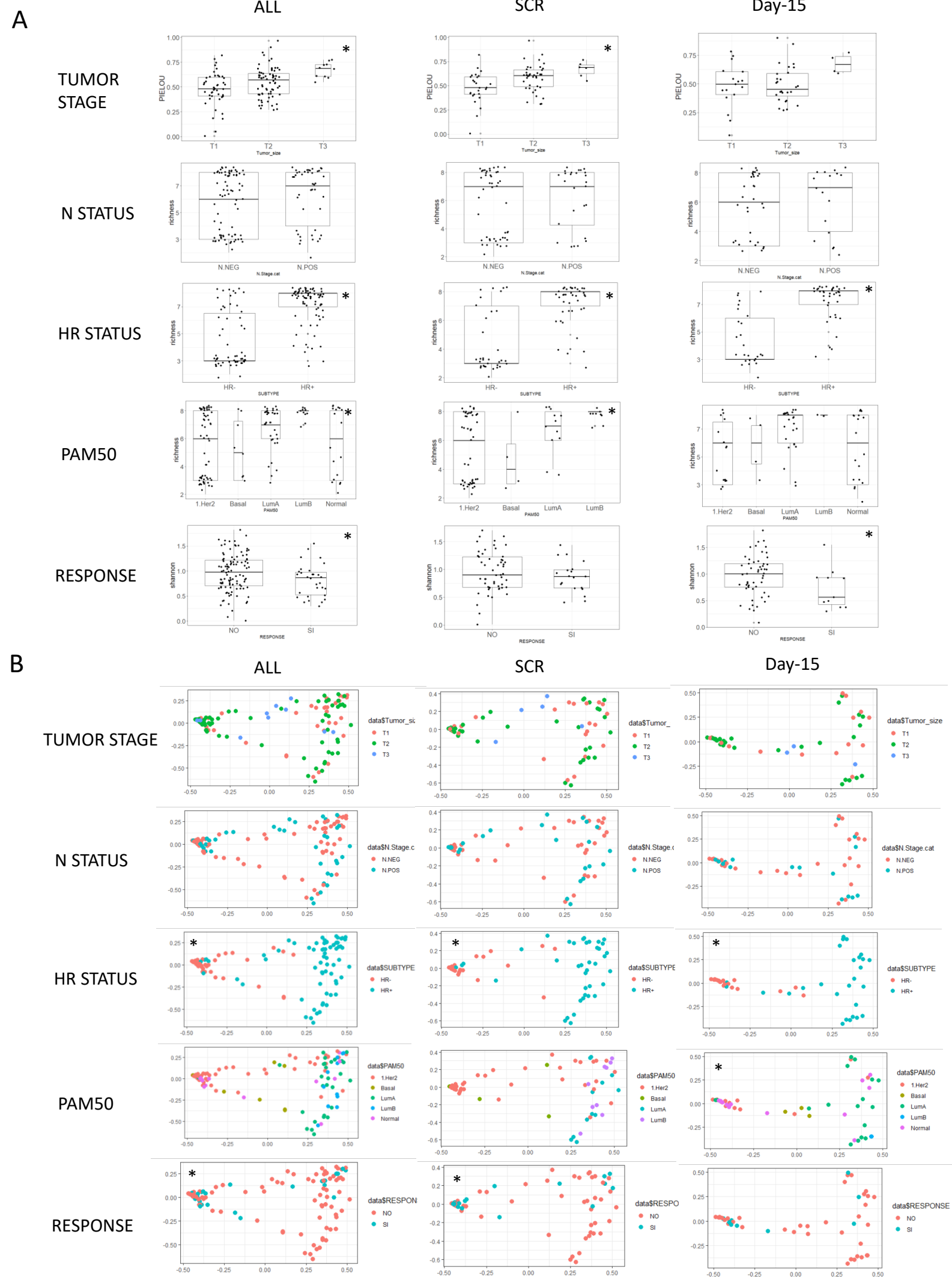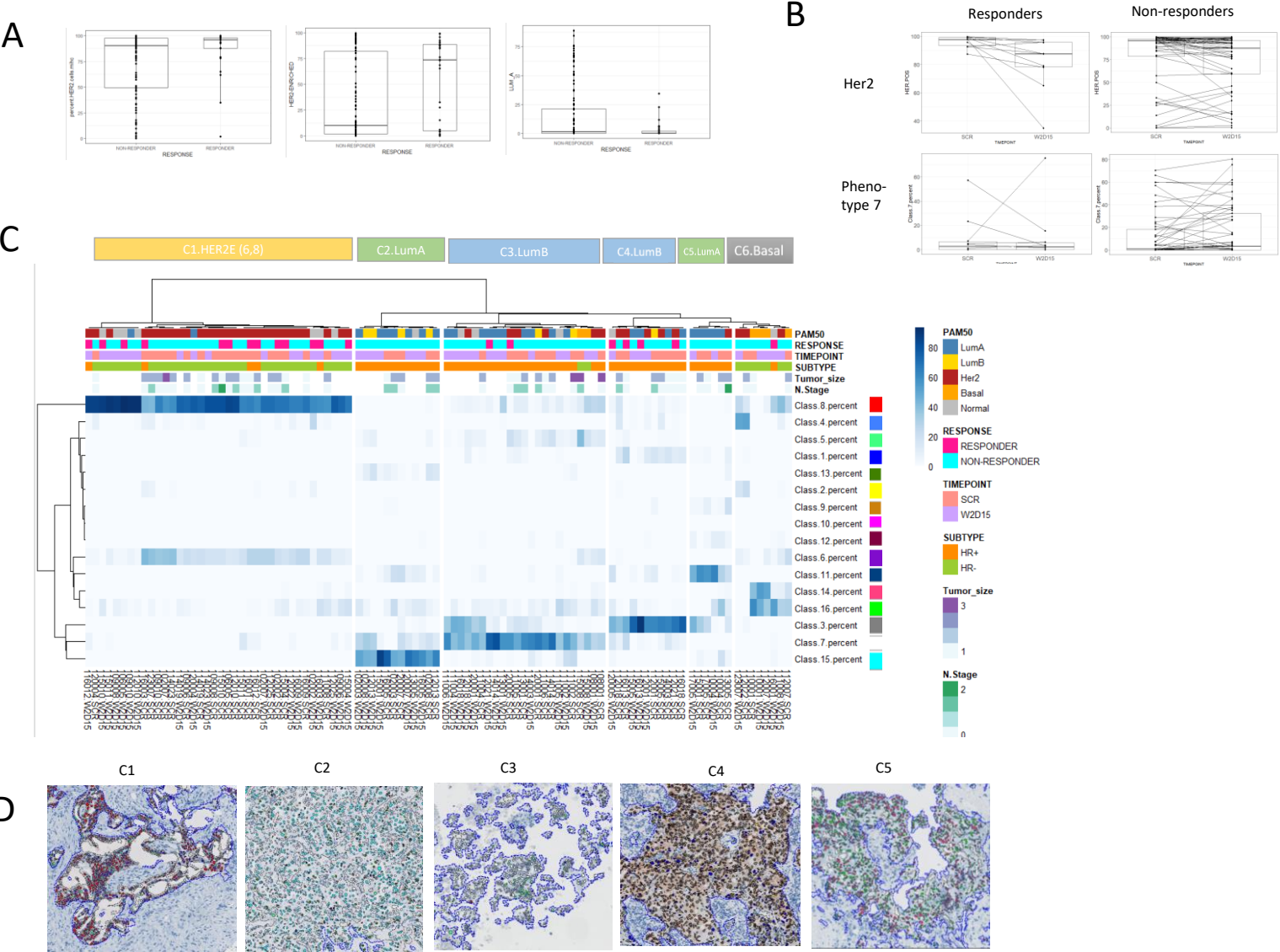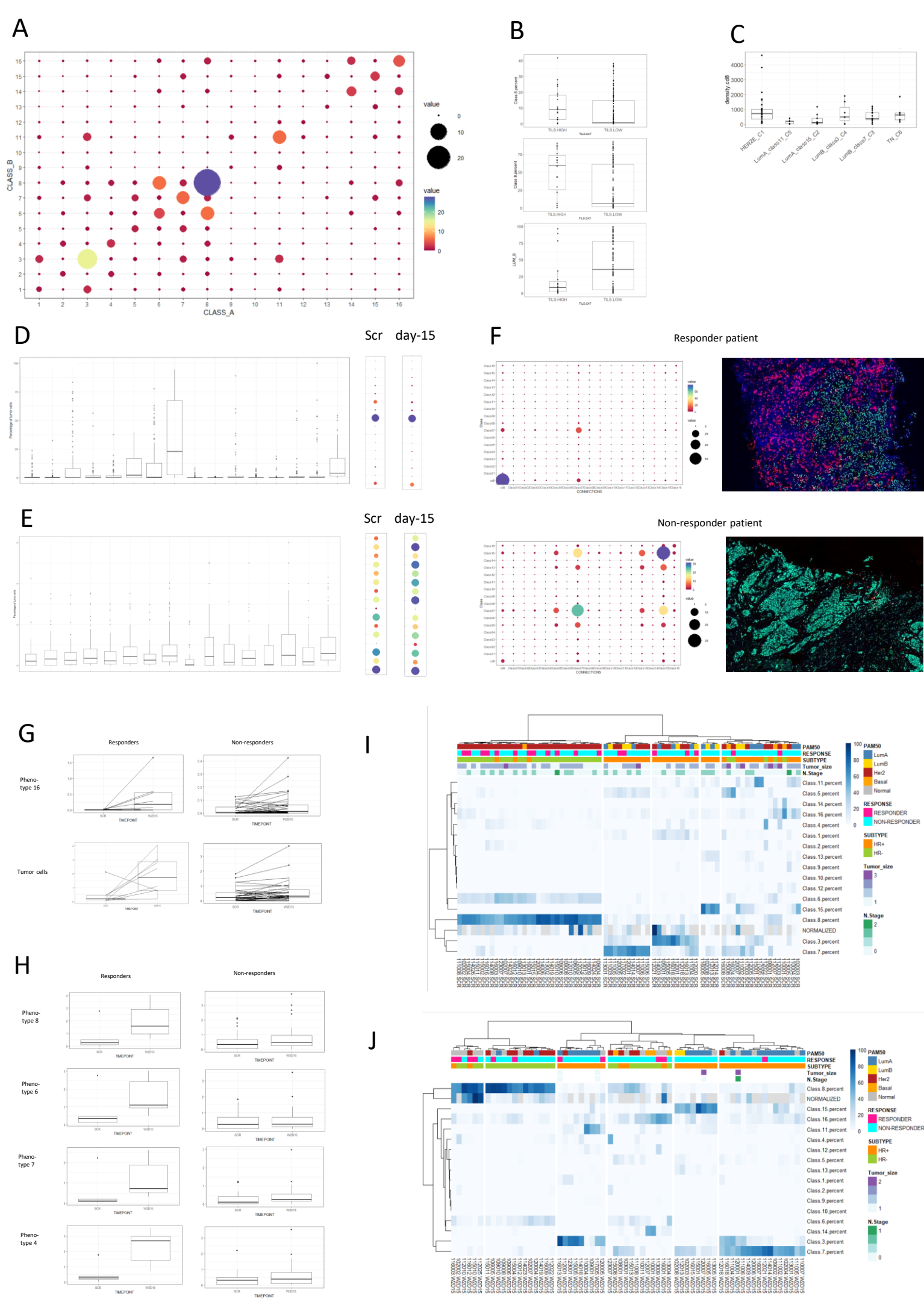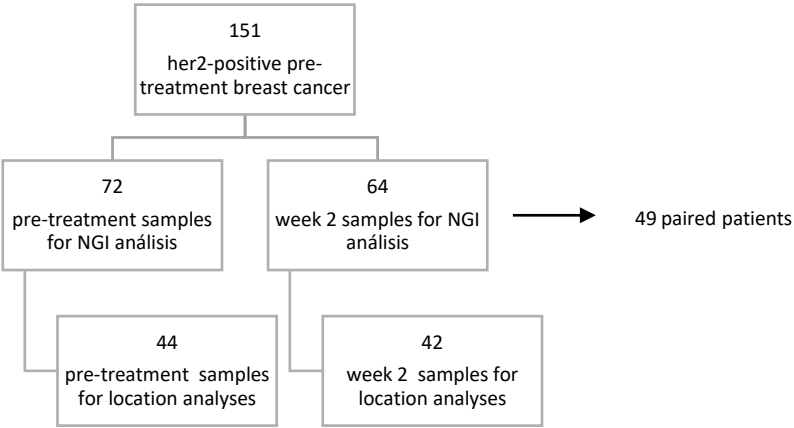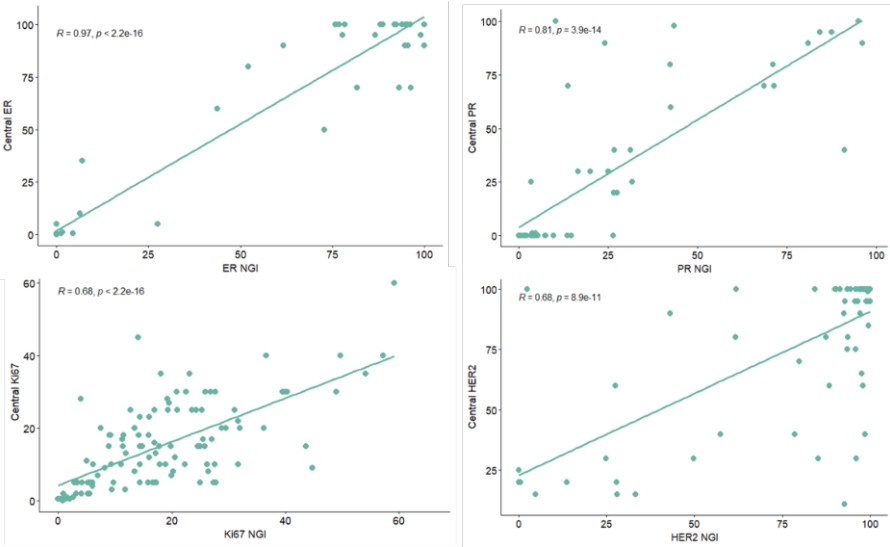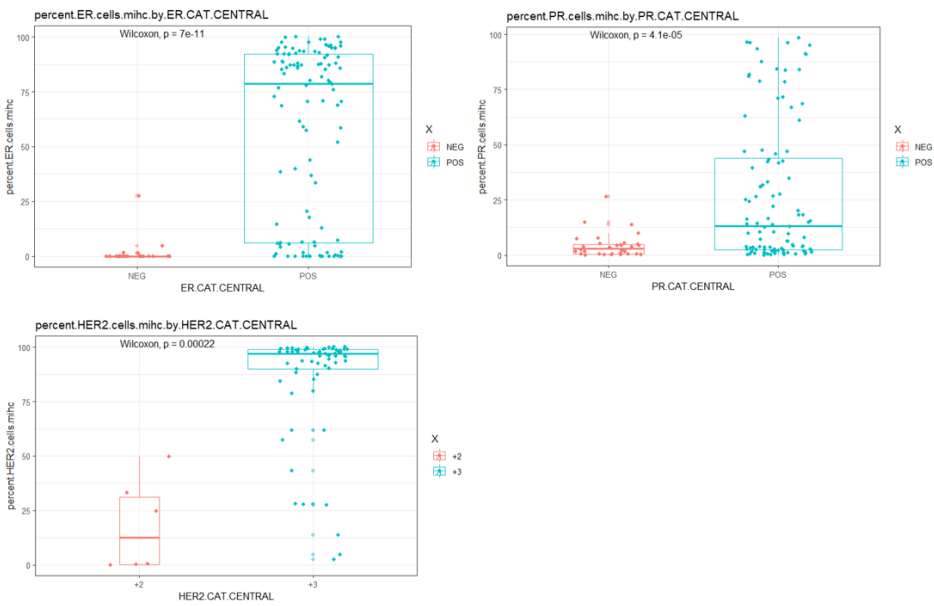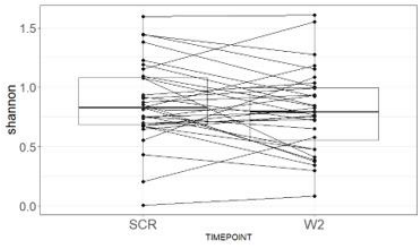Figure 2

Figure 3

Figure 4

Supplementary Figure S1

A

B

C

Supplementary Figure S2

A

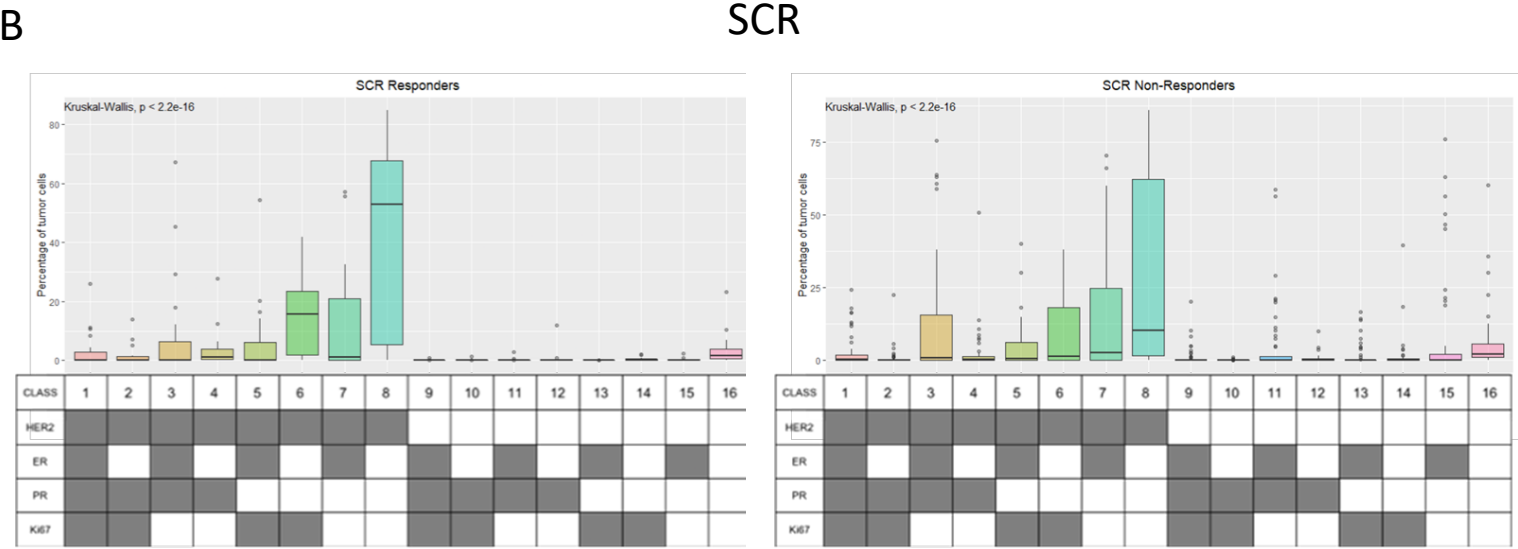
B

PAIRED

Supplementary Figure S3

Supplementary Figure S4

ACKNOWLEDGEMENTS

NEXT GENERATION
IMMUNOHISTOCHEMISTRY (NGI):

Unlocking the power of immunohistochemistry to improve
biomarker analyses in precision oncology

TESIS DOCTORAL
Garazi Serna Alonso

UAB
Universitat Autònoma
de Barcelona

Programa de doctorado de Cirugía y Ciencias Morfológicas
Departamento de Ciencias Morfológicas
Facultad de Medicina
2022