

## ARTICLE OPEN

## Heterocellular gene signatures reveal luminal-A breast cancer heterogeneity and differential therapeutic responses

Pawan Poudel<sup>1</sup>, Gift Nyamundanda<sup>1,2</sup>, Yatish Patil<sup>1,2</sup>, Maggie Chon U Cheang<sup>3</sup> and Anguraj Sadanandam<sup>1,2</sup>

Breast cancer is a highly heterogeneous disease. Although differences between intrinsic breast cancer subtypes have been well studied, heterogeneity within each subtype, especially luminal-A cancers, requires further interrogation to personalize disease management. Here, we applied well-characterized and cancer-associated heterocellular signatures representing stem, mesenchymal, stromal, immune, and epithelial cell types to breast cancer. This analysis stratified the luminal-A breast cancer samples into five subtypes with a majority of them enriched for a subtype (stem-like) that has increased stem and stromal cell gene signatures, representing potential luminal progenitor origin. The enrichment of immune checkpoint genes and other immune cell types in two (including stem-like) of the five heterocellular subtypes of luminal-A tumors suggest their potential response to immunotherapy. These immune-enriched subtypes of luminal-A tumors (containing only estrogen receptor positive samples) showed good or intermediate prognosis along with the two other differentiated subtypes as assessed using recurrence-free and distant metastasis-free patient survival outcomes. On the other hand, a partially differentiated subtype of luminal-A breast cancer with transit-amplifying colon-crypt characteristics showed poor prognosis. Furthermore, published luminal-A subtypes associated with specific somatic copy number alterations and mutations shared similar cellular and mutational characteristics to colorectal cancer subtypes where the heterocellular signatures were derived. These heterocellular subtypes reveal transcriptome and cell-type based heterogeneity of luminal-A and other breast cancer subtypes that may be useful for additional understanding of the cancer type and potential patient stratification and personalized medicine.

*npj Breast Cancer* (2019)5:21 ; <https://doi.org/10.1038/s41523-019-0116-8>

## INTRODUCTION

Breast cancer is the most common female malignancy worldwide. Breast cancers are clinically and molecularly heterogeneous, with five to ten “intrinsic” subtypes now recognized based on gene expression or integrated molecular characteristics, respectively.<sup>1,2</sup> Among the intrinsic gene expression breast cancer subtypes, luminal-A breast cancers represent the majority of estrogen receptor (ER) and progesterone receptor (PR) high/positive tumors.<sup>1,3</sup> Although many luminal-A tumors are highly responsive to endocrine therapies like tamoxifen, a significant proportion possess intrinsic and/or acquired resistance.<sup>4,5</sup> Even this relatively well-characterized breast cancer subtype possesses heterogeneity at the levels of hormone receptor expression,<sup>6,7</sup> treatment response,<sup>5</sup> and genetic variability<sup>2,3</sup> that requires further understanding.

Ciriello et al.<sup>3</sup> defined at least four genetic subtypes of luminal-A tumors involving mutations and somatic copy number alterations (CNAs) potentially associated with tamoxifen resistance. However, genetic changes alone do not explain the entire spectrum of luminal-A heterogeneity. The factors leading to tumor heterogeneity, including in luminal-A tumors, are complex and include interactions between different cell types and the tumor microenvironment along with the genetic changes present within the epithelial compartment.<sup>8</sup> For instance, stroma containing cancer-associated fibroblasts (CAFs) is most associated with basal/claudin-

low breast cancers.<sup>9</sup> However, the exact role of stroma/CAFs in luminal-A breast cancers is unclear.

Moreover, the role of the immune microenvironment in luminal-A tumors requires further exploration. It is particularly important to understand luminal-A heterogeneity and drug resistance at the levels of the immune and stromal microenvironment. Unlike in colorectal and pancreatic cancers,<sup>10–13</sup> no exclusive immune-enriched breast cancer subtype has been reported (to our knowledge). Nevertheless, immune-related genes are often expressed in different subtypes, including the luminal-A subtype (Fig. 1a), with signatures similar to those seen in one of the colorectal cancer (CRC) subtypes—consensus molecular subtypes (CMS)1/inflammatory.<sup>11,13</sup> This prompted us to further interrogate molecular similarities between breast cancer and CRC.

We previously classified CRC into five CRCAssigner subtypes: inflammatory, enterocyte, goblet-like, stem-like, and transit-amplifying (TA).<sup>11,14</sup> Later, we reconciled these five subtypes into four CMS1 to 4 using additional data from independently published CRC subtyping studies.<sup>11,13,15–19</sup> CMS and CRCAssigner subtypes are >90% concordant with certain differences including that the enterocyte and TA subtypes were merged to form the CMS2 subtype.<sup>13,20</sup> Most importantly, the immune-enriched groups (CMS1 and inflammatory) were similar. These CRCAssigner subtypes represent signatures related to stem, mesenchymal, and stromal cells forming the stem-like subtype, immune cells forming

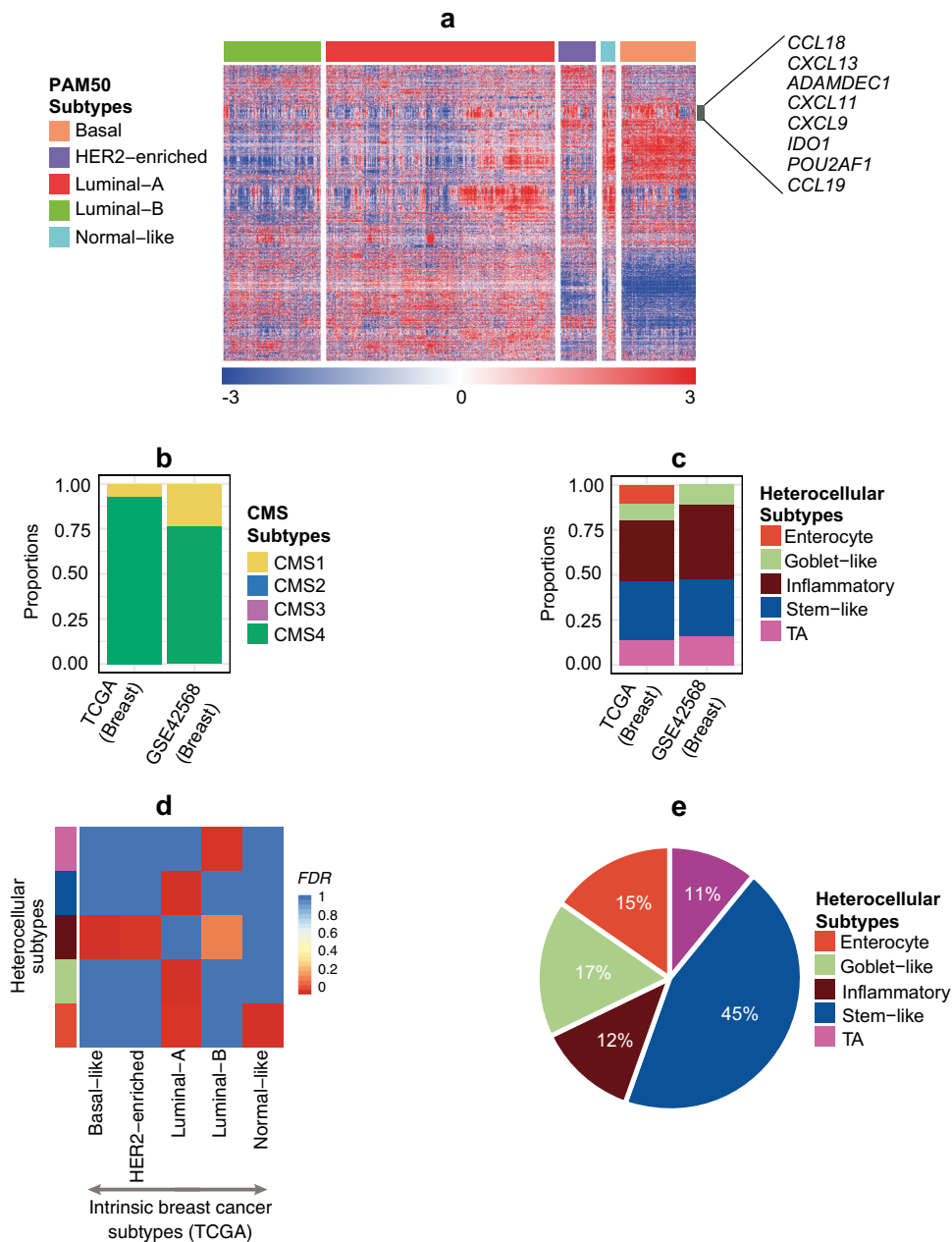
<sup>1</sup>Division of Molecular Pathology, Institute of Cancer Research, London, UK; <sup>2</sup>Centre for Molecular Pathology, Royal Marsden Hospital, London, UK and <sup>3</sup>Division of Clinical Studies, Institute of Cancer Research, London, UK

Correspondence: Anguraj Sadanandam ([anguraj.sadanandam@icr.ac.uk](mailto:anguraj.sadanandam@icr.ac.uk))

These authors contributed equally: Pawan Poudel, Gift Nyamundanda, Yatish Patil

Received: 18 June 2018 Accepted: 25 June 2019

Published online: 02 August 2019



**Fig. 1** Association of breast cancer with heterocellular subtypes. **a** Heatmap showing the expression of the top highly variable genes (standard deviation;  $SD > 2$ ), specifically immune genes, and their association with breast cancer subtype samples ( $n = 817$ ) from TCGA.<sup>23</sup> Highlighted genes represent selected immune specific genes that show high expression in multiple subtypes. **b** Proportion of CMS subtypes in multiple breast cancer data sets—TCGA<sup>23</sup> ( $n = 671$ ) and GSE42568<sup>24</sup> ( $n = 69$ ). **c** Proportion of heterocellular subtypes in multiple breast cancer data sets—TCGA<sup>23</sup> ( $n = 407$ ) and GSE42568<sup>24</sup> ( $n = 63$ ). Although heterocellular signatures were derived from entirely different cancer type (CRC), we observed that about half of the breast cancer samples were classified into all of the five heterocellular subtypes (stringent cutoff was used for mixed/low-confidence sample selection as discussed;<sup>13</sup> Supplementary Table 1c). **d** Heatmap showing sample enrichment analysis using hypergeometric test-based FDR values comparing heterocellular subtypes ( $y$  axis) with intrinsic gene expression subtypes ( $x$  axis) in the TCGA<sup>23</sup> breast cancer data set ( $n = 407$ ; Supplementary Table 1e–g). **e** Pie chart showing proportions of different heterocellular subtypes in luminal-A breast cancer samples (total  $n = 202$ ; enterocyte ( $n = 31$ ), goblet-like ( $n = 34$ ), inflammatory ( $n = 25$ ), stem-like ( $n = 90$ ), TA ( $n = 22$ ); TCGA breast cancer<sup>23</sup>). Only those samples classified into subtypes with high confidence by the CMS and heterocellular classifiers were shown in **b–e**). Summary of low and high confidence samples for both subtype classifications are shown in Supplementary Tables 1a–d and 2a–d and described in Methods section

the inflammatory subtype, a partially differentiated state as the TA subtype, and a differentiated/secretory state as goblet-like and enterocyte subtypes.<sup>11</sup> Therefore, we re-named the CRCAssigner subtypes as “heterocellular” subtypes in this study. Similar to the comparison of breast cancer subtypes to multiple cancer types,<sup>21,22</sup> we sought to use our CRC heterocellular signatures as

surrogates to re-characterize breast cancer subtypes, especially luminal-A breast cancers, and understand their phenotypes according to their differentiated, stem, fibroblast, and immune characteristics. This type of supervised analysis identifies low-frequency or rare intrinsic subtypes that are often difficult to characterize by unsupervised analysis. In addition, interesting sub-

subtypes can be identified that we are reporting in this study for the luminal-A breast cancer subtype with potential personalized treatment associations.

## RESULTS

### Association between breast cancer and heterocellular subtypes

To characterize the breast cancers using heterocellular subtypes, we applied the CMS classifier<sup>13</sup> to two independent breast cancer data sets (The Cancer Genome Atlas; TCGA ( $n = 817$ )<sup>23</sup> and GSE42568 ( $n = 104$ );<sup>24</sup> Fig. 1b and Supplementary Tables 1a, b and 2a, b). Unexpectedly, the CMS classifier was only enriched for the CMS4 (mesenchymal; >75% of high confidence samples; see Methods section) subtype in these data sets, suggesting that this CMS classifier is specific to CRC and may not be applicable to breast cancer. Since our heterocellular (CRCAssigner) signature was derived earlier than and differently to the CMS and describes the phenotypic characteristics of normal colon-crypt cells including immune-enriched inflammatory cells,<sup>11</sup> we applied this signature to the same data sets and observed that all five heterocellular subtypes were present in the TCGA breast cancer data set and four subtypes (except the CRC specialized subtype—enterocyte) in GSE42568 data set (Fig. 1c). There was a similar distribution of the four major subtypes (except enterocyte) in TCGA and GSE42568 data sets, with a variable proportion (between 0% and 10%) of the enterocyte subtype between these data sets. Correspondingly, enterocyte subtype was not present in normal breast tissue (Supplementary Figures 1a–c and Supplementary information). Here, only those samples with statistically high confidence of classification were considered (see Methods section; Fig. 1c; the dominant subtype distribution in mixed/low-confidence samples is shown Supplementary Figures 1d–f and Supplementary Tables 1c, d and 2c, d). This suggests that breast cancer has heterocellular features of different cell types (with variable proportions of enterocyte) that can be characterized with high confidence using our heterocellular signatures and subtypes.

We next sought to understand the relationship between the intrinsic breast cancer subtypes and heterocellular subtypes using hypergeometric sample enrichment analysis of the TCGA data set.<sup>23</sup> The luminal-B intrinsic breast cancer subtype was significantly ( $FDR < 0.05$ ) associated with the TA heterocellular subtype, suggesting that luminal-B cancers might have a transitional phenotype between stem and differentiated cells, like TA in the colon-crypt. Interestingly, the basal-like and human epidermal growth factor receptor 2 (HER2)-enriched intrinsic breast cancer subtypes were significantly associated with the inflammatory heterocellular subtype (Fig. 1d and Supplementary Table 1e–g), suggesting increased immune phenotype in these subtypes. We further validated these results using the GSE42568 data set, with similar results (Supplementary Figure 1g and Supplementary Table 1h–j; the dominant subtype distribution in mixed/low-confidence samples is shown in Supplementary Figure 1h–j). This suggests that breast cancer subtypes are significantly ( $p < 0.05$ ; Chi-squared test) associated with heterocellular signatures and explains additional characteristics of the intrinsic breast cancer subtypes.

### Luminal-A heterogeneity described by heterocellular subtypes

Surprisingly, the heterocellular signatures revealed the most heterogeneity in the relatively well-characterized luminal-A breast cancer subtype (Fig. 1d). This subtype was not only significantly associated with the differentiated goblet-like/enterocyte subtypes but, unexpectedly and interestingly, was also highly enriched for the poorly differentiated stem-like heterocellular subtype: 45% of luminal-A tumors were classified as stem-like tumors followed by 17% goblet-like, 15% enterocyte, 12% inflammatory, and 11% TA subtypes (Fig. 1e;  $n = 202$ ). We further validated our results using an additional data set enriched for ER positive tumors (luminal-A;

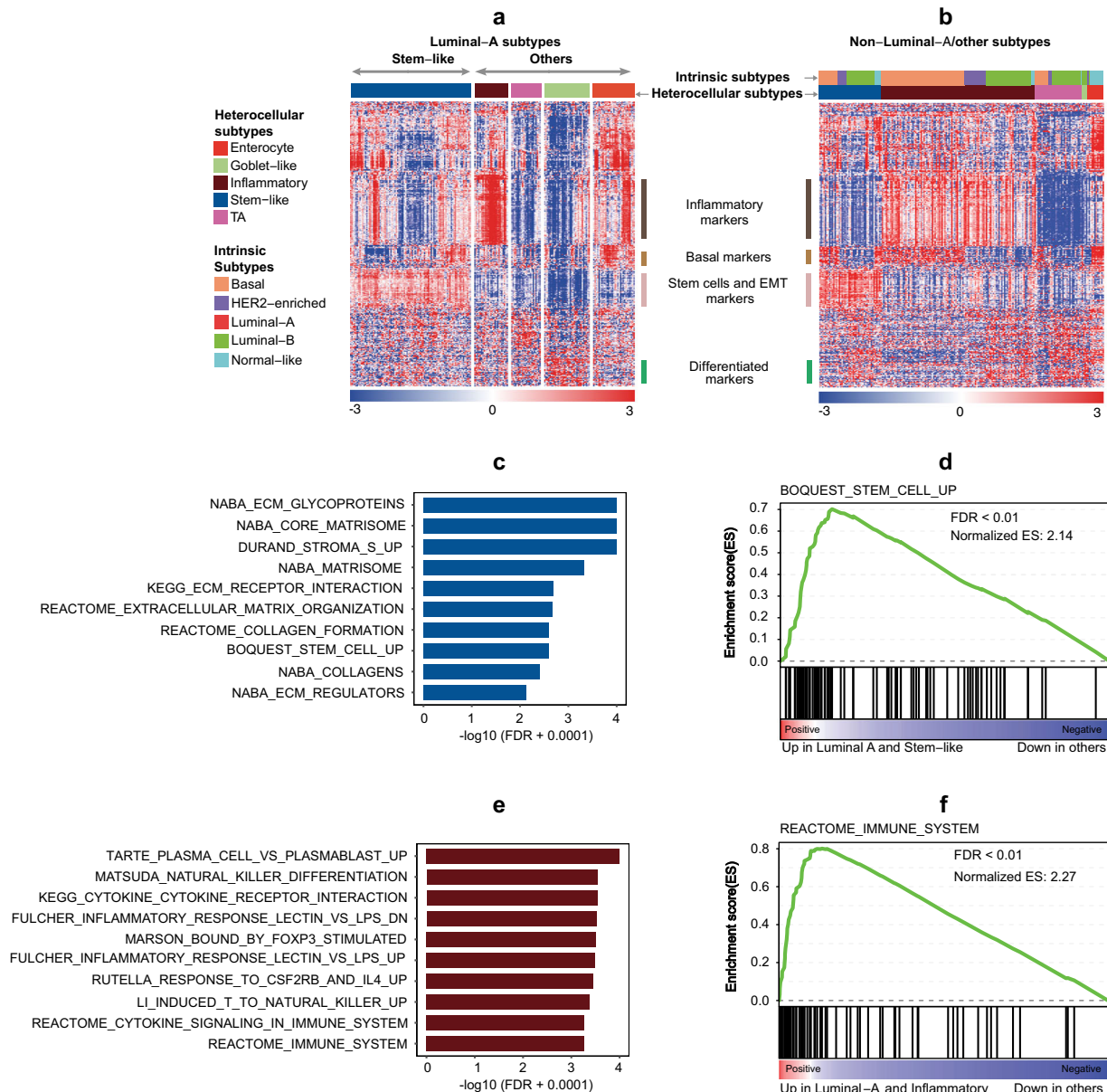
GSE6532<sup>25–27</sup>) observing similar high heterogeneity (Supplementary Figure 1k, l and Supplementary Table 1w–ab: tamoxifen-treated and -untreated samples; >39% stem-like, >24% inflammatory, >16% goblet-like, >8% TA, and >0.8% enterocyte subtype; the distribution of the dominant subtypes in mixed/low confidence and treated samples is shown in Supplementary Figure 1n). The proportions of inflammatory and enterocyte subtypes varied in the validation cohort, with the variable overall enterocyte subtype in luminal-A cancers from different data sets again representing that specialized colonic cells do not exist in breast cancers. Overall, we observed transcriptomic heterogeneity associated with heterocellular signatures in luminal-A breast cancer.

To further characterize these heterocellular subtypes in luminal-A breast cancers, we next performed heatmap analysis of heterocellular gene expression signatures using luminal-A and compared it to non-luminal-A (other subtypes) samples (Fig. 2a, b, Supplementary Figure 1m and Supplementary Table 2e). Here, our goal is to elucidate the heterogeneity in luminal-A using heterocellular subtypes. As expected, the goblet-like subtype contained increased expression of differentiated gene markers compared to the other heterocellular subtypes in luminal-A subtype (Fig. 2a). Although the TA subtype shared some of the differentiated gene markers, they showed increased heterogeneity similar to that of the CRC subtype,<sup>11</sup> with 11% ( $n = 202$ ; Fig. 1e) of the samples representing this subtype in luminal-A subtype.

Nevertheless, there was a consistent enrichment of the stem-like heterocellular subtype in luminal-A breast cancers, suggesting potentially interesting luminal-A characteristics. Of note, the stem-like subtype was enriched for potential luminal progenitor genes,<sup>28</sup> with the presence of stem cell/epithelial-to-mesenchymal transition (EMT), myoepithelial, and basal cancer markers (Fig. 2a). We further confirmed this by geneset enrichment analysis (GSEA), which showed that the stem-like subtype of luminal-A cancers was enriched for stem and stromal fibroblast cells (Fig. 2c, d, and Supplementary Table 1k). Hence, luminal-A tumors represent heterogeneity at the heterocellular level.

### Immune heterogeneity in luminal-A tumors

Although characterizing the immune gene expression heterogeneity in luminal-A tumors, we observed increased expression of immune pathways including chemokine signaling, cytokine–cytokine receptor interaction, immune system, and natural killer cell differentiation in inflammatory luminal-A subtype (Fig. 2e, f, and Supplementary Table 1l). Based on this pathway enrichment analysis, we hypothesized that the inflammatory subtype luminal-A cancers are enriched for the expression of immune checkpoint genes, potentially marking responses to immune checkpoint blockade. As expected, immune checkpoint genes and other immune markers were overrepresented in the inflammatory luminal-A cancers compared to the other subtypes (Fig. 3a). In addition, we observed increased enrichment of certain immune cell types in inflammatory luminal-A subtype (Fig. 3b). In order to predict if these inflammatory luminal-A tumors potentially may respond to anti-immune checkpoint therapy, we used a published ‘expanded immune gene’ signature, which potentially predicts anti-PD1 immune-checkpoint responses in melanoma and other cancers.<sup>29</sup> All 18 expanded immune signature genes were highly expressed in the inflammatory subtype with increased average gene expression for the signature (Fig. 3c, d). Similarly, a proportion of the stem-like subtype showed increased expression of the immune genesets and expanded immune gene signature (Fig. 3d). These results suggest that luminal-A breast cancer subtype is heterogeneous with inflammatory heterocellular subtype showing exclusive immune infiltration.



**Fig. 2** Heterocellular subtype-based heterogeneity in luminal-A breast cancers. **a, b** Heatmap showing the expression of the top highly variable and selected marker genes (371 genes from SD > 1.5 gene list shown in Supplementary Table 2e) between stem-like ( $n = 90$ ) and other subtypes ( $n = 112$ ) within the **a** luminal-A breast cancer subtype ( $n = 202$ ; Supplementary Figure 1m) and **b** subtypes other than luminal-A (non-luminal-A) from TCGA breast cancer data<sup>23</sup> ( $n = 205$ ). **c–f** GSEA analysis showing gene sets enriched in **c, d** stem-like and **e, f** inflammatory heterocellular subtype samples compared with the other subtypes ( $n = 202$ ; stem-like ( $n = 90$ ), inflammatory ( $n = 25$ ), other subtypes ( $n = 87$ )) from TCGA breast cancer.<sup>23</sup> Relevant gene sets that were enriched were shown in **c–f** (See Supplementary Table 1k, l for the top gene sets that were ordered by significance of FDR values). KEGG—Kyoto Encyclopedia of Genes and Genomes; EMT—epithelial-to-mesenchymal transition

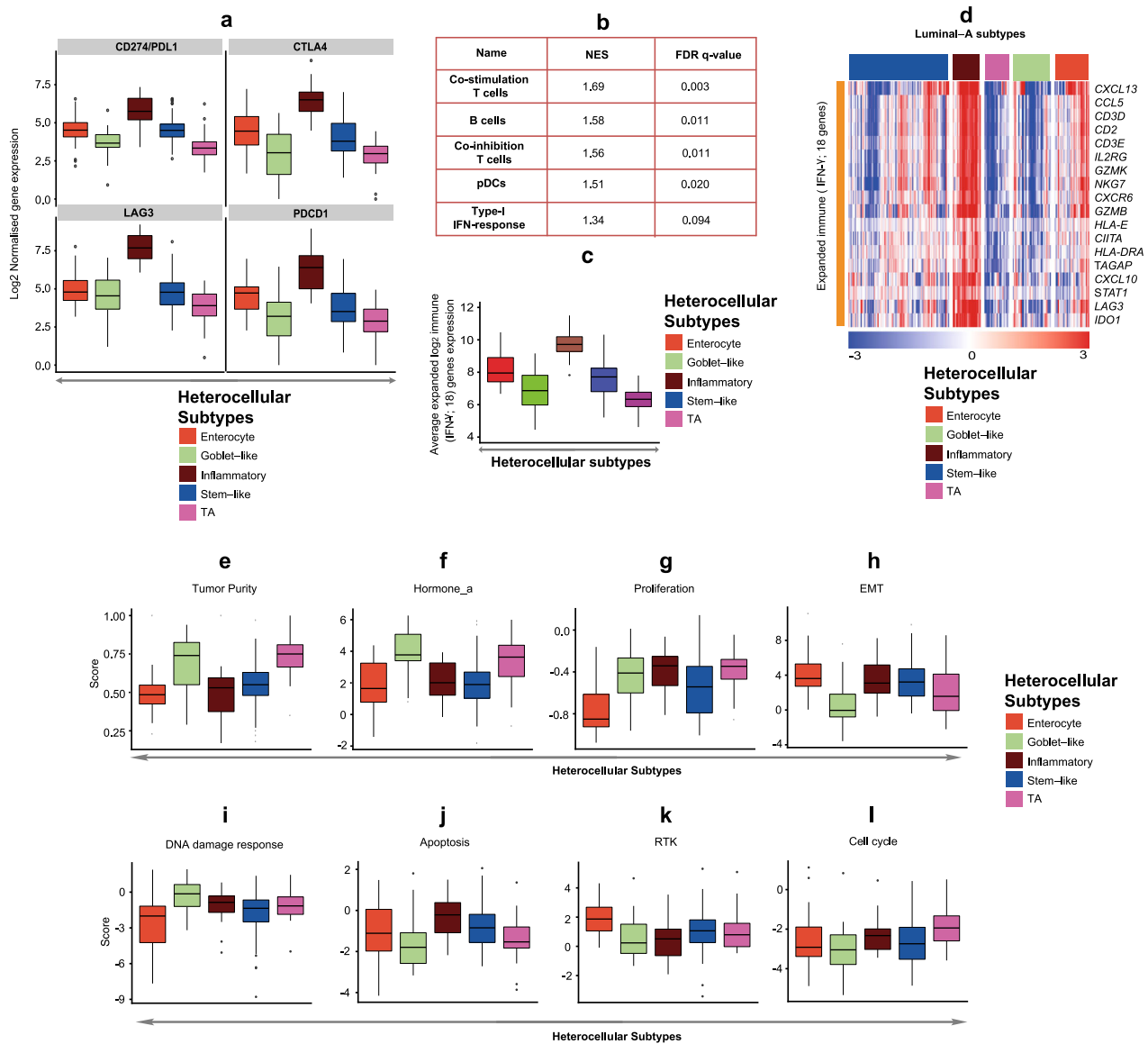
#### Additional characteristics of heterocellular subtypes

Next, we sought to understand if phenotypic changes that were measured as scores by TCGA<sup>23</sup> show difference between our heterocellular subtypes in luminal-A tumors (Fig. 3e–l; scores in Fig. 3f–l are from reverse-phase protein microarray; RPPA as published by TCGA.<sup>23</sup>) Our analysis showed that tumor purity, hormone\_a (represents signatures associated hormone receptors,<sup>30</sup>) proliferation and DNA damage response scores were significantly high in goblet-like and TA subtypes compared with the other subtypes (Fig. 3e–g and i). The inflammatory subtype showed high proliferation score similar to goblet-like and TA subtypes (Fig. 3g). On the other hand, the EMT and apoptosis scores were low in goblet-like and TA, but high in stem-like

subtype (this subtype in CRC is known to have high EMT genes;<sup>11</sup> Fig. 3h and j). We observed increased receptor tyrosine kinase score in enterocyte and stem-like subtypes and significantly increased cell cycle score in TA subtype (Fig. 3k and l). There were other phenotypes from the TCGA that were not significantly associated with the subtypes (Supplementary Table m). These results suggest that these heterocellular subtypes from luminal-A show differences in multiple breast cancer associated phenotypes.

#### Association of heterocellular subtypes with other published luminal-A subtypes

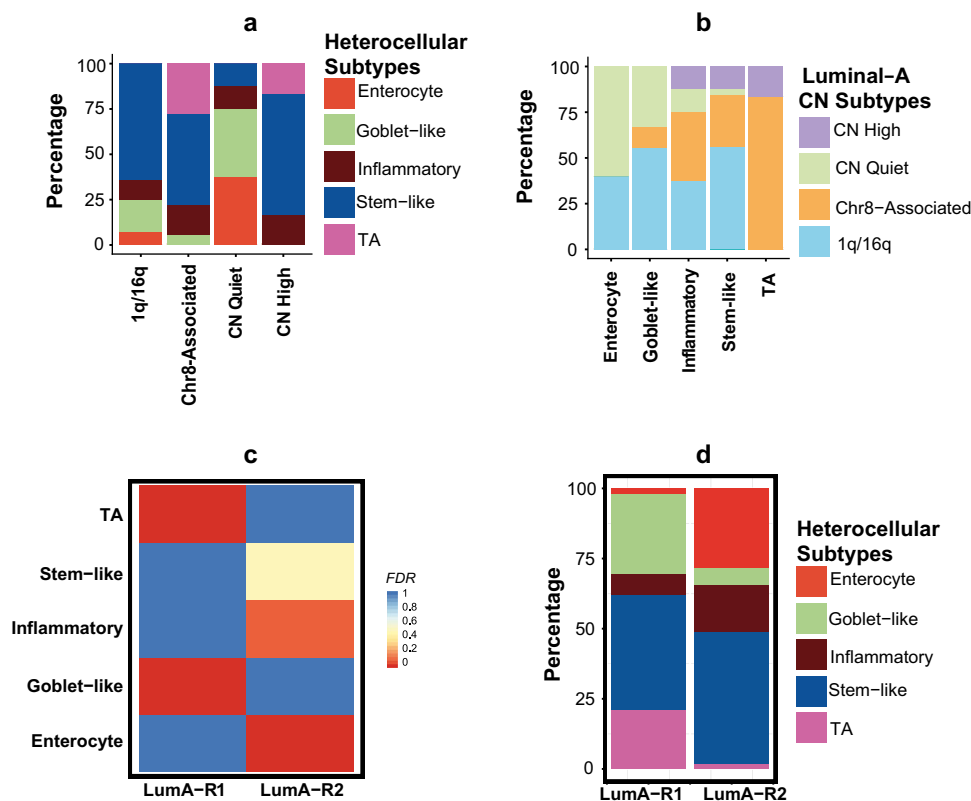
To understand potential mutational and CNA changes in heterocellular subtypes of luminal-A, we next compared our



**Fig. 3** Enrichment of immune checkpoint genes, immune cells, expanded immune (18-gene) signature and other phenotypes in luminal-A heterocellular subtypes. **a** Box plots showing differences in the expression of immune checkpoint genes *CD274* (*PDL1*), *CTLA4*, *LAG3*, and *PDCD1* between heterocellular subtypes ( $n = 202$ ; enterocyte ( $n = 31$ ), goblet-like ( $n = 34$ ), inflammatory ( $n = 25$ ), stem-like ( $n = 90$ ), TA ( $n = 22$ ); TCGA<sup>23</sup> breast cancer). Kruskal–Wallis test was performed to calculate  $p$  and their corresponding FDR values. Those associations with  $FDR < 0.05$  was considered significant. **b** Gene set enrichment analysis (GSEA) showing immune cell types enriched in inflammatory heterocellular subtype samples compared to the other subtypes using the Rooney et al.<sup>60</sup> gene sets ( $n = 202$ ; inflammatory ( $n = 25$ ) and other subtypes ( $n = 177$ ); TCGA breast cancer<sup>23</sup>). Those associations with  $FDR < 0.1$  was considered significant. **c** Boxplot showing differences in sample-wise average gene expression of 18 published expanded immune (18-gene) signature<sup>29</sup> in heterocellular subtypes. Kruskal–Wallis test was performed to calculate  $p$  values.  $p < 0.05$  was considered significant ( $n = 202$ ; enterocyte ( $n = 31$ ), goblet-like ( $n = 34$ ), inflammatory ( $n = 25$ ), stem-like ( $n = 90$ ), TA ( $n = 22$ ); TCGA<sup>23</sup> breast cancer). **d** Heatmap showing the expression of eighteen published expanded immune (18-gene) signature<sup>29</sup> genes between heterocellular subtypes from luminal-A breast cancers ( $n = 202$ ; enterocyte ( $n = 31$ ), goblet-like ( $n = 34$ ), inflammatory ( $n = 25$ ), stem-like ( $n = 90$ ), TA ( $n = 22$ ); TCGA breast cancer<sup>23</sup>). **e–l** Boxplots showing differences in **e** tumor purity, **f** hormone\_a, **g** proliferation, **h** EMT, **i** DNA damage response, **j** apoptosis, **k** RTK, and **l** cell cycle scores from TCGA<sup>23</sup> between heterocellular subtypes. The data from **f–l** were from RPPA data-based scores published by TCGA.<sup>23</sup> Kruskal–Wallis test was performed to calculate  $p$  and their corresponding FDR values. Those associations with  $FDR < 0.05$  was considered significant. pDCs—plasmacytoid dendritic cells; NES—normalized enrichment score; FDR—false discovery rate; EMT—epithelial–mesenchymal transition; RTK—receptor tyrosine kinase

heterocellular luminal-A subtype classification with four Ciriello CNA-based luminal-A subtypes<sup>3</sup> (Fig. 4a, b, and Supplementary Table 1n–p). Regarding the association of the heterocellular subtypes with Ciriello’s subtypes, the well-differentiated goblet-like and enterocyte subtype samples were primarily associated the Ciriello subtypes—1q/16q (characterized by 1q gain and 16q loss chromosomal regions) and CN quite (characterized by quite CNA spectrum). TA heterocellular subtype samples were primarily

associated with Ciriello’s Chr8-associated (characterized by loss of 8p and gain of 8q chromosomal regions) subtype cancers, however, a certain proportion of them also represented CN high (CNH; characterized by multiple focal CNAs) Ciriello subtypes. The stem-like and inflammatory luminal-A subtype samples were heterogeneous and represented all the four Ciriello subtypes, and these subtypes had a scrambled genome such that 12.5% belonged to the Ciriello CNH subtype (Fig. 4b). Though there



**Fig. 4** Association of heterocellular subtypes with published other luminal-A breast cancer subtype classifications. **a**, **b** Bar plots showing the percentage of **a** heterocellular subtypes in Ciriello subgroups of luminal-A subtype<sup>3</sup> (1q/16q ( $n = 28$ ), Chr8-associated ( $n = 18$ ), CN quiet ( $n = 8$ ), CN high ( $n = 6$ )) and **b** vice versa (enterocyte ( $n = 5$ ), goblet-like ( $n = 9$ ), inflammatory ( $n = 8$ ), stem-like ( $n = 32$ ), TA ( $n = 6$ ),  $p < 0.02$ , Chi-square test; Supplementary Table 1n–p). We did not compare our heterocellular subtype to “mixed” subtype of Ciriello, as it is reported to lack any decipherable patterns of chromosomal changes.<sup>3</sup> Data for a–b) are from TCGA<sup>23</sup> breast cancer. **c** Heatmap showing sample enrichment analysis using hypergeometric test-based FDR values comparing heterocellular subtype classification (y axis) with two Netanely et al.<sup>32</sup> luminal-A breast cancer subtypes (x axis). **d** Bar plot showing percentage of different heterocellular subtypes in two Netanely et al., luminal-A breast cancer subtypes (LumA-R1 ( $n = 95$ ), LumA-R2 ( $n = 102$ );  $p < 0.02$ , Chi-square test; Supplementary Table 1q–s). Only those samples classified into subtypes with high confidence from heterocellular subtype classification are shown in **a–d**

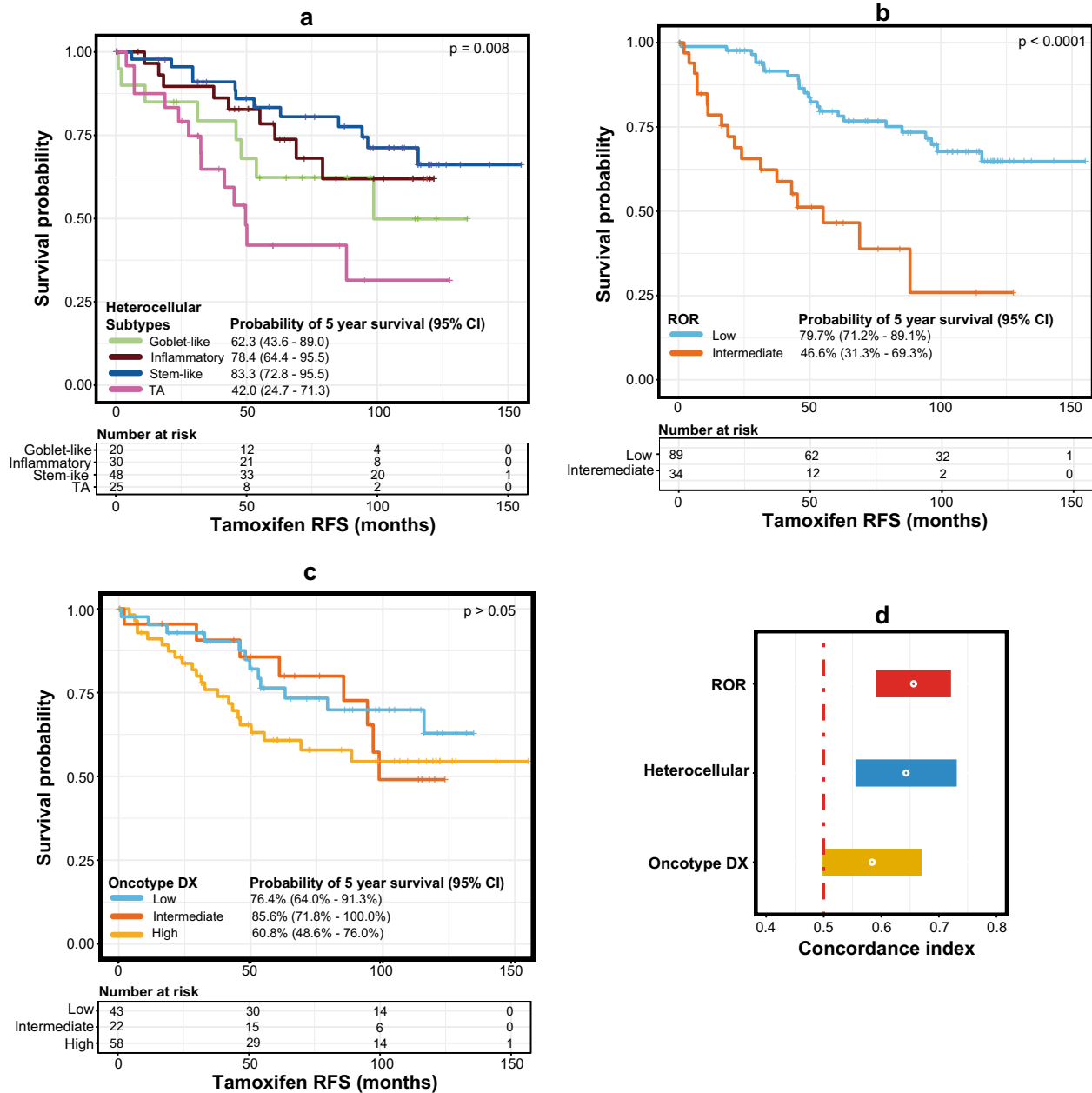
were associations between Ciriello and our heterocellular subtypes, these two represent different classification systems representing genetic and transcriptomic heterogeneity of the luminal-A subtype.

Similarly, we assessed Aure et al.<sup>31</sup> and Netanely et al.<sup>32</sup> luminal-A gene expression subtype classifications. Aure et al.<sup>31</sup> subtype did not show any similarity to our heterocellular subtypes representing that these classification are quite different from each other (Supplementary Figure 2 and Supplementary Table 1t–v). This attributes to the fact that Aure subtypes were not exclusively based on luminal-A cancer samples. They show the enrichment of luminal-A cancer samples in two of their multi-level clusters.<sup>31</sup> On the other hand, our heterocellular subtypes divided two of the Netanely et al.<sup>32</sup> subtypes into sub-subtypes (Fig. 4c and d and Supplementary Table 1q–s). Netanely LumA-R1 was mainly divided into goblet-like and TA, whereas LumA-R2 was divided into inflammatory and enterocyte subtypes. Our stem-like subtype was not significantly associated with any of their two subtypes and substantially present in both the Natanely subtypes. This suggests that our heterocellular subtypes explain additional transcriptomic heterogeneity that these two previous subtype classifications did not reveal.

Heterocellular luminal-A subtypes are associated with tamoxifen treatment-based clinical outcomes

To assess the association of tamoxifen treatment response with heterocellular subtypes, we evaluated the association between

our heterocellular luminal-A subtypes and clinical outcomes in patient samples treated with tamoxifen using GSE6532 data set<sup>25–27</sup> (Fig. 5 and Supplementary Figure 3a; the distribution of the mixed/low confidence subtypes is shown in Supplementary Figure 1n). Heterocellular luminal-A subtypes showed significant ( $p < 0.01$ ) differences in recurrence-free (RFS) and border-line significance ( $p = 0.07$ ) differences in distant metastasis-free survival (DMFS) in patients treated with tamoxifen (Fig. 5a, Supplementary Figure 3b and Supplementary Tables 1w–y and 2f). We considered mixed subtype samples along with high confidence samples only in this case, and for mixed subtype samples only the dominant subtypes were considered. The consideration of mixed subtype was based on our previous report that mixed subtype tumors have a mixture of more than one subtype, and the presence of certain dominant subtype (for example TA) may attribute to prognostic and therapeutic response differences between subtypes/samples in CRC.<sup>33</sup> Unlike in CRC,<sup>11</sup> there was relatively good RFS and DMFS for luminal-A cancer patients with the stem-like subtype, similar to other subtypes including goblet-like and inflammatory subtypes. This may be attributed to the enrichment of expanded immune gene signature<sup>29</sup> in a subset of stem-like subtype samples, similar to the immune-rich inflammatory luminal-A subtype with similar prognosis (Fig. 5a). Conversely, the TA subtype luminal-A tumors showed worse RFS and DMFS with tamoxifen treatment (Fig. 5a). Although there was a significant overall difference between subtypes for RFS/DMFS in tamoxifen-treated patients, there was no significant ( $p \geq 0.5$ ) difference in untreated patient samples



**Fig. 5** Survival differences in heterocellular subtypes from ER-positive tamoxifen-treated samples. **a–c** Kaplan–Meier survival curve showing tamoxifen-treated RFS (GSE6532;<sup>25–27</sup> Supplementary Table 1w–y) between the **a** heterocellular subtypes, **b** groups from Risk of Recurrence (ROR<sup>34</sup>) and **c** risk groups from OncotypeDX<sup>35</sup> from ER-positive breast cancer samples from microarray data. **d** A plot showing concordance index and associated confidence intervals for RFS between heterocellular subtypes and ROR/OncotypeDX groups. Log-rank test was performed for the  $p$  values. RFS–recurrence free survival; CI–confidence interval

(Supplementary Figures 3c, d, and Supplementary Tables 1z–ab and 2g). The lack of prognostic difference in the untreated patients but poor prognosis in treated patients with the TA subtype suggests that TA subtype luminal-A patients may respond less well to tamoxifen.

We next compared these results with RFS from risk of recurrence (ROR<sup>34</sup>) and OncotypeDX.<sup>35</sup> Among the three classifications, there was not much difference in RFS between ROR and our heterocellular subtypes, with similar concordance index (Fig. 5a, b, and d). However, the poor performance of OncotypeDX compared to our heterocellular subtypes could be attributed to the fact that the method was applied to microarray data (Fig. 5a, c, and d), which was not originally intended to be used. Nevertheless, these results warrant further validation using larger

cohorts in the future. Overall, these results confirm the heterogeneity of luminal-A cancers and provide insights into the pathophysiology dictated by different cell types for potential personalized treatment (Fig. 6).

## DISCUSSION

That breast cancers are heterogeneous is well known.<sup>1,2</sup> Clinically, hormone receptor-positive breast cancer patients are treated differently to triple hormone receptor-negative (TNBC) and HER2-positive breast cancer patients.<sup>36</sup> At the molecular level, breast cancer was one of the initial cancer types to be subtyped into intrinsic gene expression subtypes.<sup>1</sup> Similar to clinical breast cancer subtypes, the molecular subtypes have distinct prognostic

Heterogeneity in Luminal-A					
Heterocellular subtypes	Stem-Like	Enterocyte	Goblet-like	Inflammatory	TA
Ciriello CN subtypes	1q/16q Chr8 CN quite CN High	1q/16q CN quiet	1q/16q Chr8 CN quiet	1q/16q Chr8 CN quiet CN High	Chr8 CN High
Molecular features	EMT Stem-cells Stromal cells	Differentiated markers		Basal Chemokines Immune genes IFN- $\gamma$ response	Hetero genous
Tamoxifen-treated RFS	Good	NA	Good/ intermediate	Good	Poor

**Fig. 6** Summary of the luminal-A heterocellular subtypes and their characteristics. EMT—epithelial-to-mesenchymal transition; RFS—recurrence free survival; CN—copy number; chr-8—chromosome 8 associated; TA—transit amplifying; NA—not enough data available to conclude

differences.<sup>1,2</sup> In this study, we further investigated breast cancer heterogeneity, especially in the luminal-A subtype, using heterocellular subtype signatures defined in CRCs. This was done similar to the application of breast cancer subtype signatures to other cancers<sup>21,22,37</sup> and with an intention to identify low frequency and unreported subtypes that are not apparent based on unsupervised approaches.

The basal subtype, which represents the majority of TNBCs, is already known to be highly heterogeneous, with the majority of these patients responding to chemotherapy.<sup>38</sup> However, basal breast tumors often recur with aggressive disease.<sup>39</sup> Similar to other studies,<sup>40,41</sup> our results showed enrichment of immune genes characteristic in the basal/inflammatory breast cancer subtype (Fig. 1d). While no immunotherapy is yet approved, but with immune checkpoint inhibitors being tested clinically in patients with breast cancer,<sup>42</sup> our association of a subset of basal breast cancers with the CRC inflammatory subtype suggests a means to identify patients who might respond to immunotherapy. This potentially aligns with responses to atezolizumab and pembrolizumab immunotherapy in metastatic TNBC patients.<sup>43</sup> Our similar observation of an association between HER2 breast cancers and the inflammatory heterocellular subtype suggests that some HER2-positive patients may similarly be eligible for immunotherapy.

Moreover, we observed an enrichment of inflammatory heterocellular subtype samples in the luminal-A subtype harboring high expression of immune checkpoint genes. Next to the inflammatory subtype, there was a subset of the stem-like subtype with increased expression of immune genes (Fig. 3a and d). Both of these subtypes showed increased expression of expanded immune gene signature,<sup>29</sup> suggesting potential response to immune checkpoint inhibition. Hence, our heterocellular gene signature may be useful for selecting patients within luminal-A breast cancers for immunotherapy, which warrants further exploration in the future. Although there are few indicators of how immunotherapy might work in relatively good prognostic luminal-A subtype cancers,<sup>44</sup> tamoxifen-resistant TA luminal-A tumors do not seem to express many immune genes, suggesting that a combination of tamoxifen plus immunotherapy may not be the treatment of choice for resistant patients. Immune checkpoint inhibitors have now been approved for microsatellite instable CRCs, which are associated with the inflammatory CRC subtype.<sup>45</sup> TA CRC tumors are enriched for microsatellite stable disease,<sup>11</sup> suggesting potential resistance to immunotherapy. However, it may be interesting to find a way to induce this immune dormant TA luminal-A subtype to immune enriched subtype for potential immunotherapy.

Although the epithelial compartment of the breast and colon vary, we observed a significant association between luminal-A tumors and the goblet-like subtype, suggesting an overlap in common gene signatures representing a secretory function.

Specifically, trefoil factor genes were highly expressed in both the luminal-A and goblet-like subtypes.<sup>11,46</sup> Of note, the goblet-like luminal-A subtype enriched for the 1q/16q Ciriello subtype is associated with increased *KRAS* and *PIK3CA* mutations.<sup>3</sup> We have previously shown that the CMS3 (goblet-like) subtype is enriched for *KRAS* mutations.<sup>13</sup> In addition, a subset of TA and stem-like subtype luminal-A cancers was associated with the Ciriello CNH subtype, which is enriched for *TP53* mutations.<sup>3</sup> Enriched *TP53* mutations also exist in TA and stem-like CRCs,<sup>11,13</sup> suggesting that the subtype association between these cancer types is not random and they are associated with similar molecular events both at the transcriptomic and genetic levels. Again, this suggests that different cellular compartments share the same molecular features and perhaps functions. Nevertheless, the lower enrichment of the enterocyte heterocellular subtype in luminal-A cancers suggests the presence of this specialized cell type only in the intestine and not in the breast.

To our surprise, the stem-like subtype of luminal-A breast cancers showed good RFS (Fig. 5a), indicating that the presence of stem cells and fibroblasts (enriched in the stem-like subtype) does not indicate poor survival in differentiated luminal-A breast cancer patients, in contrast to CRC patients.<sup>11,13</sup> On the other hand, the TA luminal-A subtype breast cancer patients showed poor RFS when treated with tamoxifen. However, none of these subtypes showed significantly different prognoses in the untreated patient samples. We recently developed a biomarker assay for CRC subtypes (both CRCAssigner and CMS) that stratify patients into subtypes<sup>20,33</sup> and that potentially may select breast cancer patients for different therapies including immunotherapy. Overall, our current study sheds further light on luminal-A breast cancer heterogeneity that is useful for the personalized diagnosis and treatment of patients with luminal-A and other breast cancer subtypes.

## METHODS

### Gene expression and patient survival data

The raw CEL files containing gene expression data and the corresponding survival data for patient tumors were downloaded from gene expression omnibus (GEO)<sup>47</sup>—GSE42568<sup>24</sup> and GSE6532 (combined Affymetrix Human Genome U133A and U133B Arrays was used).<sup>25–27</sup> Prognostic information for GSE6532 were from the original publications.<sup>25–27</sup> The gene expression profiles for the TCGA breast cancer data (Ciriello et al.<sup>23</sup>) was downloaded from cBioPortal repository<sup>48,49</sup> and other information of the corresponding samples were obtained from the original publication.<sup>23</sup> Those genes with missing values (a value of zero from logarithmically transformed RSEM<sup>50</sup> data) in >30% of the samples were removed, as described.<sup>51</sup> Owing to the retrospective nature of this study using only publicly available data, ethics approval for the study was not required.

### Affymetrix GeneChip® microarray data processing and quality control

The raw gene expression data (CEL files) were processed and normalized using robust multi-array normalization (RMA) from *R*-based Bioconductor<sup>52</sup> package—*affy*.<sup>53</sup> Only the samples having Normalized Unscaled Standard Error (NUSE,<sup>54</sup> from *R*-based bioconductor<sup>52</sup> *affyPLM*<sup>55</sup> package) with a median score of  $1 \pm 0.05$  was considered high-quality arrays and selected for further analysis GSE42568.<sup>24</sup> For GSE6532<sup>25–27</sup> (all samples were considered), data from two different arrays—Affymetrix GeneChip Human Genome U133A and U133B—done for the same set of samples were normalized using RMA<sup>52</sup> and merged by samples. The technical/batch effect in GSE6532<sup>25–27</sup> was corrected using ComBat.<sup>56</sup> Supplementary Figure 3a shows a flow chart of the data processing and analysis for treated samples from GSE6532,<sup>25–27</sup> which also applies for untreated samples.

### CMS and CRCAssigner classifications

For classifying the samples into CMS subtypes, *classifyCMS* function from our published *R* package *CMSClassifier*<sup>13</sup> was used. We applied single sample prediction method from the package, and those samples that were



classified as mixed or undetermined by *CMSClassifier* were considered mixed or low confidence samples, respectively (Supplementary Table 1a, b). For classifying the samples into heterocellular subtypes, the correlation of gene centroids for five subtypes and gene expression data using Pearson method from CRCAssigner subtypes and signatures was applied, as described previously.<sup>11,13</sup> Before Pearson correlation analyses, we used our probe to gene mapping file from our original paper<sup>12</sup> to map to the CRCAssigner PAM centroid genes. After correlation, those samples with maximum correlation coefficient among five of them <0.15 were considered low confidence samples and those with difference in correlation coefficients between first and second subtypes <0.06 were considered mixed samples as described previously.<sup>13</sup> Only those samples qualified otherwise as high confidence samples were mainly considered for further analyses (Supplementary Table 1c, d). Only for GSE6532<sup>25–27</sup> data analysis, high confidence, and mixed samples were considered. In this case of mixed samples, the dominant or the subtype with maximum correlation coefficient was considered for further analysis.

### Breast cancer intrinsic classification

The intrinsic breast cancer classification for GSE42568 data set was performed using an R-based Bioconductor<sup>52</sup> package—*genefu*.<sup>57</sup>

### Reconciliation of subtypes

The association between the heterocellular and published intrinsic subtypes were performed using the hypergeometric test as described by us previously.<sup>14</sup>

### GSEA

This analysis was performed using standalone GSEA package from GenePattern<sup>58</sup> using the c2 geneset from mSigDB<sup>59</sup> and published immune cell specific gene markers from Rooney et al.<sup>60</sup>

### Visualization of gene expression data

For the heatmap, genes were clustered (hierarchical clustering) by Cluster 3.0<sup>61</sup> using the default settings, followed by visualization of the clusters using *GENEE* from GenePattern.<sup>58</sup>

### Association between heterocellular subtypes and breast cancer phenotypes

Breast cancer phenotypes such as proliferation, apoptosis and other features as RPPA scores were from Ciriello et al.<sup>23</sup> Association between these features and heterocellular subtypes were performed using Kruskal–Wallis statistical test and plotted as boxplots.

### Prediction of ROR/Oncotype DX risk groups

Prediction of tumor samples into ROR groups was performed as described.<sup>34</sup> The OncotypeDX Recurrence Score were predicted as described.<sup>35,62,63</sup> For microarray data, most variable probes were selected to represent the 21 OncotypeDX genes.<sup>35</sup> *CD68* gene, which was not annotated in our data set, was replaced with its corresponding probe (*203507\_at*). Five of the 21 OncotypeDX<sup>35</sup> genes were housekeeping genes, whose average expression was subtracted from the other 16 OncotypeDX genes.<sup>65</sup>

### Survival analysis

Kaplan–Meier survival analysis was performed and concordance index was calculated using R package—*survival*<sup>64</sup> and plotted using R package—*survminer*.<sup>65</sup> For statistical test, log-rank test was used.

### Reporting summary

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

### CODE AVAILABILITY

All the analyses in this study were done using R statistical software version 3.5.1. A list of packages used include *CMSClassifier*,<sup>13</sup> for predicting CMS classes and CRCAssigner using our published classifier.<sup>13,20</sup>

### DATA AVAILABILITY

The two Affymetrix gene expression profile data sets, GSE42568 and GSE6532, analyzed during the current study, can be accessed from the NCBI Gene Expression Omnibus (GEO) repository, <https://identifiers.org/geo:GSE42568>, <https://identifiers.org/geo:GSE6532>. Data set GSE42568 supports Fig. 1, Supplementary Figure 1 and Supplementary Tables 1 and 2 of the published article. Data set GSE6532 supports Fig. 5, Supplementary Figures 1 and 3, and Supplementary Tables 1 and 2 in this published article. The TCGA breast cancer data set analyzed during the current study can be accessed from the cBioPortal for Cancer Genomics repository, [https://identifiers.org/cbioportal:brca\\_tcga\\_pub2015](https://identifiers.org/cbioportal:brca_tcga_pub2015), and supports Figs 1, 2, 3 and 4, Supplementary Figures 1 and 2, and Supplementary Tables 1 and 2 in this published article. Additional information for the TCGA breast cancer data set was obtained from the published article <https://doi.org/10.1038/nature11412>. RNAseq data from normal breast cancer samples (cohort: GDC TCGA Breast Cancer (BRCA), gene expression data HTSeq–FPKM–UQ,  $n = 162$  solid tissue normal) analyzed during this study, can be accessed from the University of California Santa Cruz (UCSC) Xena browser, [https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Breast%20Cancer%20\(BRCA\)&removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443](https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Breast%20Cancer%20(BRCA)&removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443), and support Supplementary Figure 1 in this published article. The data analyzed during this study are described in the following data record: <https://doi.org/10.6084/m9.figshare.8256713>.<sup>66</sup>

### ACKNOWLEDGEMENTS

We thank Drs. Sue Eccles, Alan Melcher, Rachael Natrajan, Nagarajan Kannan, Steven Whittaker, Kate Young, Kate Eason, and Anna Wilkins for carefully reading the manuscript. P.P. was supported by Pancreatic Cancer UK Future Research Leaders Fund under the supervision of A.S. We acknowledge NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and the ICR.

### AUTHOR CONTRIBUTIONS

A.S. conceived the idea and designed the experiments. P.P., G.N. and Y.P. performed majority of the experiments and data analyses. M.C.U.C. and A.S. performed certain data analyses. All the authors contributed to writing of the manuscript. P.P., G.N. and Y.P. contributed equally as co-first authors.

### ADDITIONAL INFORMATION

**Supplementary Information** accompanies the paper on the *npj Breast Cancer* website (<https://doi.org/10.1038/s41523-019-0116-8>).

**Competing interests:** A.S. has ownership interest as a patent inventor for a patent entitled “Colorectal cancer classification with differential prognosis and personalized therapeutic responses” (patent number PCT/IB2013/060416). A.S. has research funding from Bristol-Myers Squibb and Merck KgaA. M.C.U.C. has a patent: US Patent No. 9,631,239 with royalties paid. The rest of the authors declare that there are no competing interests.

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

### REFERENCES

- Perou, C. M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
- Ciriello, G. et al. The molecular diversity of Luminal A breast tumors. *Breast cancer Res. Treat.* **141**, 409–420 (2013).
- Higgins, M. J. & Stearns, V. Understanding resistance to tamoxifen in hormone receptor-positive breast cancer. *Clin. Chem.* **55**, 1453–1455 (2009).
- Ring, A. & Dowsett, M. Mechanisms of tamoxifen resistance. *Endocr. Relat. Cancer* **11**, 643–658 (2004).
- Prabhu, J. S. et al. Dissecting the biological heterogeneity within hormone receptor positive her2 negative breast cancer by gene expression markers identifies indolent tumors within late stage disease. *Transl. Oncol.* **10**, 699–706 (2017).
- Collins, L. C., Botero, M. L. & Schnitt, S. J. Bimodal frequency distribution of estrogen receptor immunohistochemical staining results in breast cancer: an analysis of 825 cases. *Am. J. Clin. Pathol.* **123**, 16–20 (2005).
- Rivenbark, A. G., O'Connor, S. M. & Coleman, W. B. Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. *Am. J. Pathol.* **183**, 1113–1124 (2013).

9. Camp, J. T. et al. Interactions with fibroblasts are distinct in Basal-like and luminal breast cancers. *Mol. Cancer Res.* **9**, 3–13 (2011).
10. Bailey, P. et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).
11. Sadanandam, A. et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* **19**, 619–625 (2013).
12. Poudel, P. et al. Revealing unidentified heterogeneity in different epithelial cancers using heterocellular subtype classification. *bioRxiv*, <https://doi.org/10.1101/175505> (2017).
13. Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
14. Sadanandam, A. et al. Reconciliation of classification systems defining molecular subtypes of colorectal cancer: interrelationships and clinical implications. *Cell Cycle* **13**, 353–357 (2014).
15. Roepman, P. et al. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. *Int. J. Cancer* **134**, 552–562 (2014).
16. Schlicker, A. et al. Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. *BMC Med. Genomics* **5**, 66 (2012).
17. Budinska, E. et al. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J. Pathol.* **231**, 63–76 (2013).
18. De Sousa, E. M. F. et al. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat. Med.* **19**, 614–618 (2013).
19. Marisa, L. et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med.* **10**, e1001453 (2013).
20. Ragulan, C. et al. Analytical validation of multiplex biomarker assay to stratify colorectal cancer into molecular subtypes. *Sci. Rep.* **9**, 7665 (2019).
21. Moffitt, R. A. et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* **47**, 1168–1178 (2015).
22. Damrauer, J. S. et al. Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc. Natl. Acad. Sci. USA* **111**, 3110–3115 (2014).
23. Ciriello, G. et al. Comprehensive molecular portraits of invasive lobular breast. *Cancer Cell* **163**, 506–519 (2015).
24. Clarke, C. et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis* **34**, 2300–2308 (2013).
25. Loi, S. et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J. Clin. Oncol.* **25**, 1239–1246 (2007).
26. Loi, S. et al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* **9**, 239 (2008).
27. Loi, S. et al. PIK3CA mutations associated with gene signature of low mTORC1 signaling and better outcomes in estrogen receptor-positive breast cancer. *Proc. Natl. Acad. Sci. USA* **107**, 10208–10213 (2010).
28. Shehata, M. et al. Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res.* **14**, R134 (2012).
29. Ayers, M. et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Investig.* **127**, 2930–2940 (2017).
30. Akbani, R. et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat. Commun.* **5**, 3887 (2014).
31. Aure, M. R. et al. Integrative clustering reveals a novel split in the luminal A subtype of breast cancer with impact on outcome. *Breast Cancer Res.* **19**, 44 (2017).
32. Netanel, D., Avraham, A., Ben-Baruch, A., Evron, E. & Shamir, R. Expression and methylation patterns partition luminal-A breast tumors into distinct prognostic subgroups. *Breast Cancer Res.* **18**, 74 (2016).
33. Fontana, E. et al. Molecular subtype assay to reveal anti-EGFR response sub-clones in colorectal cancer (CRC) (ASCO GI Abstract). *J. Clin. Oncol.* **36**, 658–658 (2018).
34. Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
35. Paik, S. et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
36. Prat, A. et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast* **24**, S26–S35 (2015).
37. Cancer Genome Atlas Research, N. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
38. Masuda, H. et al. Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. *Clin. Cancer Res.* **19**, 5533–5540 (2013).
39. Alluri, P. & Newman, L. A. Basal-like and triple-negative breast cancers: searching for positives among many negatives. *Surg. Oncol. Clin. N. Am.* **23**, 567–577 (2014).
40. Jezequel, P. et al. Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. *Breast Cancer Res.* **17**, 43 (2015).
41. Milioli, H. H., Tishchenko, I., Riveros, C., Berretta, R. & Moscato, P. Basal-like breast cancer: molecular profiles, clinical features and survival outcomes. *BMC Med. Genomics* **10**, 19 (2017).
42. Polk, A., Svane, I. M., Andersson, M. & Nielsen, D. Checkpoint inhibitors in breast cancer – current status. *Cancer Treat. Rev.* **63**, 122–134 (2018).
43. Basile, D. et al. Atezolizumab for the treatment of breast cancer. *Expert Opin. Biol. Ther.* **18**, 595–603 (2018).
44. Miller, L. D. et al. Immunogenic subtypes of breast cancer delineated by gene classifiers of immune responsiveness. *Cancer Immunol. Res.* **4**, 600–610 (2016).
45. Boland, P. M. & Ma, W. W. Immunotherapy for colorectal cancer. *Cancers* **9**, E50 (2017).
46. Lau, W. H. et al. Trefoil factor-3 (TFF3) stimulates de novo angiogenesis in mammary carcinoma both directly and indirectly via IL-8/CXCR2. *PLoS ONE* **10**, e0141947 (2015).
47. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
48. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
49. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
50. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
51. Hoadley, K. A. et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929–944 (2014).
52. Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
53. Kohl, M. & Deigner, H. P. Preprocessing of gene expression data by optimally robust estimators. *BMC Bioinformatics* **11**, 583 (2010).
54. Brettschneider, J., C., F., Bolstad, B. M. & Speed, T. P. Quality assessment for short oligonucleotide arrays. *Technometrics* **50**, 241–264 (2008).
55. Bolstad, B. M. et al. *Quality Assessment of Affymetrix GeneChip Data in Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (2005).
56. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
57. Gendoo, D. M. et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **32**, 1097–1099 (2016).
58. Reich, M. et al. GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).
59. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
60. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
61. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998).
62. Fan, C. et al. Concordance among gene-expression-based predictors for breast cancer. *New Engl. J. Med.* **355**, 560–569 (2006).
63. Fan, C. et al. Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. *BMC Med. Genomics* **4**, 3 (2011).
64. Therneau, T. *A package for survival analysis in S. R package version 2.37-4* (2014).
65. Alboukadel Kassambara, M. K. *Survminer: Drawing Survival Curves using 'ggplot2'* (2018).
66. Poudel, P., Nyamundanda, G., Patil, Y., Cheang, M. C. & Sadanandam, A. Metadata supporting data files of the related article: Heterocellular gene signatures reveal luminal-A breast cancer heterogeneity and differential therapeutic responses. *Figshare*, <https://doi.org/10.6084/m9.figshare.8256713> (2019).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019