BRIEF COMMUNICATION   OPEN

Check for updates

The therapeutic response of ER+/HER2− breast cancers differs according to the molecular Basal or Luminal subtype

François Bertucci [1,2 ✉], Pascal Finetti [1], Anthony Goncalves[1,2] and Daniel Birnbaum [1]

The genomics-based molecular classifications aim at identifying more homogeneous classes than immunohistochemistry, associated with a more uniform clinical outcome. We conducted an in silico analysis on a meta-dataset including gene expression data from 5342 clinically defined ER+/HER2− breast cancers (BC) and DNA copy number/mutational and proteomic data. We show that the Basal (16%) versus Luminal (74%) subtypes as defined using the 80-gene signature differ in terms of response/vulnerability to systemic therapies of BC. The Basal subtype is associated with better chemosensitivity, lesser benefit from adjuvant hormone therapy, and likely better sensitivity to PARP inhibitors, platinum salts and immune therapy, and other targeted therapies under development such as FGFR inhibitors. The Luminal subtype displays potential better sensitivity to CDK4/6 inhibitors and vulnerability to targeted therapies such as PIK3CA, AR and Bcl-2 inhibitors. Expression profiles are very different, showing an intermediate position of the ER+/HER2− Basal subtype between the ER+/HER2− Luminal and ER− Basal subtypes, and let suggest a different cell-of-origin. Our data suggest that the ER+/HER2− Basal and Luminal subtypes should not be assimilated and treated as a homogeneous group.

INTRODUCTION

Breast cancer (BC) is heterogeneous. The treatment and the patients' inclusion in the clinical trials remain based upon clinicopathological features including immunohistochemistry (IHC), insufficient to capture the disease heterogeneity. The genomics-based classifications aim at identifying more homogeneous classes based on the functionality of molecular pathways and associated with a more uniform therapeutic response and outcome. Accumulating evidence suggests that these molecular subtypes provide clinically relevant information beyond clinicopathological classes[1–3]. In a recent study[4], 13.1% of IHC estrogen receptor-positive HER2-negative (ER+/HER2−) BCs were reclassified as molecular Basal subtype by the 80-gene signature (80-GS)[5]. When compared to the ER+/HER2− cases reclassified as Luminal subtype (74.1%), the Basal samples displayed lower *ESR1* mRNA expression and increased relative ERΔ7 dominant-negative variant expression, shorter 3-year distant relapse-free interval (DRFI), and higher pathological complete response rate (pCR) to chemotherapy (CT). But the authors pointed to a few limitations: the limited number of ER+ Basal patients (54 for DRFI, 70 for pCR), the short median 34-months follow-up, and absence of information regarding the sensitivity to hormone therapy (HT). To reinforce these results and extend them to the response and/or potential vulnerability to HT and other systemic therapies of BC, and to assess the degree of difference between these subtypes, we analyzed in silico a meta-dataset including gene expression data from 8982 non-redundant BCs[6], and DNA copy number/mutational and proteomic data from TCGA. Our aim was to compare the Basal versus Luminal samples.

RESULTS

Prognostic analysis according to the molecular subtype

A total of 5836 samples were clinically defined as ER+/HER2−: 4341 (74%) were reclassified as Luminal by the 80-GS, 931 (16%) as Basal, and 564 (10%) as HER2-enriched. Because our aim was to compare the Luminal and Basal samples, the HER2-enriched samples were excluded, leaving 5272 samples for analysis. Regarding the prognostic features, the Basal samples comprised more grade 3 than the Luminal ($p = 7.91E-22$), more pT3 tumors ($p = 3.25E-03$), more *TP53*-mutated tumors ($p = 2.77E-12$), and more "high-risk" tumors according to prognostic gene expression signatures (GES): Mammaprint ($p = 3.94E-56$), Recurrence Score ($p = 2.43E-121$), and EndoPredict ($p = 3.15E-73$; Table 1). ER expression level was lower in the Basal samples than in the Luminal samples, in terms of both mRNA expression ($p = 6.14E-193$) and percentage of positive tumor cells by IHC ($p = 4.33E-02$; $p = 2.33E-19$; Fig. S1).

Regarding DRFI, 2008 patients operated for ER+/HER2− early BC were informative, including 1664 Luminal and 344 Basal. None of them had received any neoadjuvant systemic treatment, whereas 524 (35%) had received adjuvant HT and 342 (21%) adjuvant CT. With a median follow-up of 65 months (range, 1–299), the 5-year DRFI was not different between the two subtypes: 81% (95%CI 77–86) in Basal versus 79% (95%CI 77–82) in Luminal ($p = 0.240$; Fig. 1a). However, the temporal pattern of events differed with 65% of events (41/63) in the Basal within the first 3 years, versus only 44% (159/362) in the Luminal ($p = 2.46E-03$). In univariate analysis, the pathological grade and tumor size, and the use of adjuvant HT were associated with DRFI, whereas grade, axillary lymph node status and use of adjuvant HT were significant in multivariate analysis (Table S1).

---

[1]Laboratoire d'Oncologie Prédictive, Centre de Recherche en Cancérologie de Marseille, Inserm U1068, CNRS UMR7258, Institut Paoli-Calmettes, Aix-Marseille Université, Marseille, France. [2]Département d'Oncologie Médicale, Institut Paoli-Calmettes, Aix-Marseille Université, Marseille, France. ✉email: bertuccif@ipc.unicancer.fr

Table 1. Clinicopathological characteristics of patients and samples according to the molecular subtype.

Characteristics	N	ER+ Luminal	ER+ Basal	p value[a]	N	ER− Basal	p value[b]
Patients' age				0.296			7.89E−26
≤50 years	1335	1115 (33%)	220 (31%)		632	632 (49%)	
>50 years	2795	2296 (67%)	499 (69%)		666	666 (51%)	
Pathological type				0.561			3.48E−16
Ductal	2241	1825 (74%)	416 (76%)		710	710 (85%)	
Lobular	394	324 (13%)	70 (13%)		21	21 (3%)	
Other	380	318 (13%)	62 (11%)		103	103 (12%)	
Pathological grade				7.91E−22			1.86E−251
1	637	562 (18%)	75 (11%)		15	15 (1%)	
2	1941	1660 (54%)	281 (42%)		160	160 (14%)	
3	1137	829 (27%)	308 (46%)		997	997 (85%)	
Pathological tumor size (pT)				3.25E−03			3.38E−10
pT1	1431	1207 (42%)	224 (38%)		288	288 (31%)	
pT2	1681	1398 (49%)	283 (49%)		542	542 (57%)	
pT3	324	248 (9%)	76 (13%)		114	114 (12%)	
Pathological axillary lymph node status (pN)				0.275			0.38
Negative	2168	1791 (58%)	377 (55%)		596	596 (59%)	
Positive	1626	1320 (42%)	306 (45%)		422	422 (41%)	
TP53 mutation status				2.77E−12			6.02E−42
Wild-type	1559	1323 (95%)	236 (84%)		255	255 (71%)	
Mutated	108	63 (5%)	45 (16%)		103	103 (29%)	
Mammaprint relapse risk				3.94E−56			1.99E−231
Low	1879	1757 (40%)	122 (13%)		13	13 (1%)	
High	3393	2584 (60%)	809 (87%)		1647	1647 (99%)	
Recurrent score relapse risk				2.43E−121			<2.00E−255
Low	2110	1968 (45%)	142 (15%)		19	19 (1%)	
Intermediate	1575	1357 (31%)	218 (23%)		1555	86 (5%)	
High	1587	1016 (23%)	571 (61%)		86	1555 (94%)	
EndoPredict relapse risk				3.15E−73			<2.00E−255
Low	2729	2498 (58%)	231 (25%)		19	19 (1%)	
High	2543	1843 (42%)	700 (75%)		1641	1641 (99%)	
Pathological complete response (pCR)				3.72E−08			1.08E−15
No	468	410 (91%)	58 (68%)		271	271 (69%)	
Yes	69	42 (9%)	27 (32%)		123	123 (31%)	
Adjuvant HT				0.369			3.25E−83
No	1375	1134 (47%)	241 (49%)		655	655 (87%)	
Yes	1542	1292 (53%)	250 (51%)		101	101 (13%)	
Adjuvant CT				0.129			1.96E−45
No	3097	2598 (87%)	499 (84%)		756	756 (67%)	
Yes	496	402 (13%)	94 (16%)		367	367 (33%)	
5-year DRFI, % (95% CI)	2008	79% (77−82)	81% (77−86)	0.24	630	62% (58−67)	1.11E−15
DRFI event, yes	2008	362 (22%)	63 (18%)	0.168	630	201 (32%)	1.60E−07

HT hormone therapy, *CT* chemotherapy.
[a] p value for the comparison ER+ Basal versus ER+ Luminal.
[b] p value for the comparison between ER+ Basal, ER+ Luminal, and ER− Basal.

Therapeutic response/vulnerability according to the molecular subtype

Eighty-five Basal BCs and 452 Luminal BCs had received anthracycline-based neoadjuvant CT followed by surgery. We confirmed the higher chemosensitivity of Basal subtype with 32% (27/85) pCR rate versus 9% (42/452) in the Luminal subtype (q = 1.63E−06, Fig. 1c). Such difference was also predicted by the predictive DLDA30 GES[7], which identified 30% (276/931) of Basal samples as "pCR-predicted" versus <1% (19/4341) of Luminal (q = 1.63E−78). Among the 1509 Luminal or Basal ER+/HER2− patients informative for DRFI and adjuvant HT, 524 and 985 had and had not received adjuvant HT, respectively. The Luminal subtype benefited from adjuvant HT with 90% (95%CI 87–94) 5-year DRFI with HT versus 82% (95%CI 79–85) without (p = 1.10E−02), whereas the Basal subtype did not (82% [95%CI 73–92] 5-year DRFI with versus 79% [95%CI 72–86] without; p = 0.422; Fig. 1b).
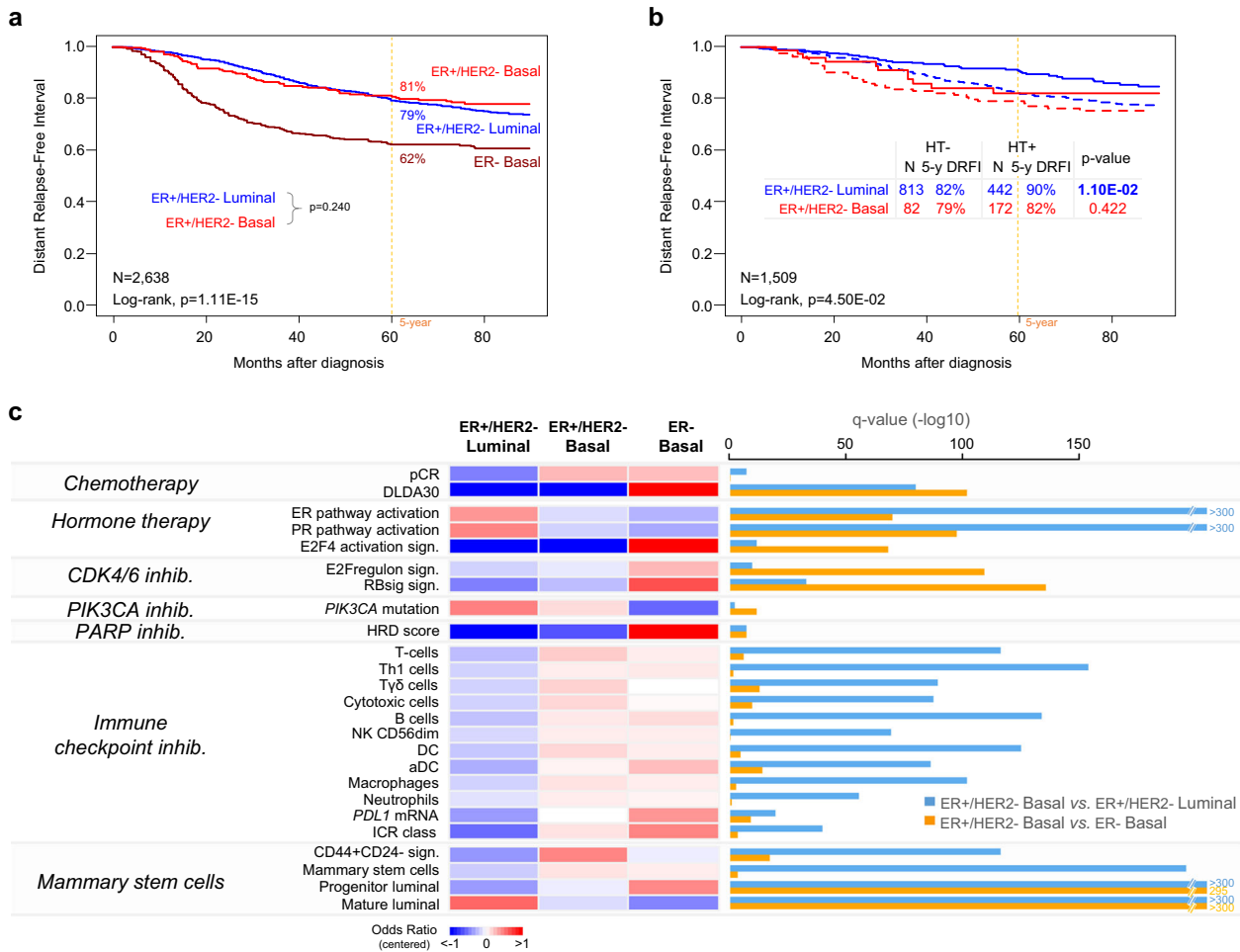
Fig. 1 Comparison of the ER+/HER2− Basal subtype, ER+/HER2− Luminal subtype, and ER− Basal subtype breast cancers. a Kaplan −Meier postoperative DRFI curves in early BCs according to the ER IHC status and the 80-GS molecular type. **b** Similar to (**a**), but in ER+/HER2− early BCs patients only, untreated (dashed curves) and treated (solid curves) with adjuvant HT. **c** Heatmap of the odds ratios (ORs) of regression analysis between the three tumor subtypes (ER+/HER2− Luminal subtype, ER+/HER2− Basal subtype, and ER− Basal subtype used as reference for comparison) for different variables related to the percent of pCR after CT or the probability of therapeutic response of BC to CT, HT, CDK4/6 inhibitors, PIK3CA inhibitor, PARP inhibitors, and immune checkpoint inhibitors. Variables associated to mammary stem cells are also shown. For each variable, the ORs are mean-centered and color-coded according to the color scale shown below. On the right, the bar plots represent the log10-transformed q values of regression analysis for the comparison of each variable between ER+/HER2− Basal subtype versus ER+/HER2− Luminal subtype (blue bar), and between ER+/HER2− Basal subtype versus ER− Basal subtype (orange bar). The longer is the bar, the lower is the q value.

This was confirmed in multivariate analyses (Table S2). In the Luminal subtype, the use of adjuvant HT was an independent favorable prognostic variable ($p = 3.13E−02$), whereas positive pN and higher grade were unfavorable variables. By contrast, in the Basal subtype, the use of adjuvant HT was not associated with DRFI ($p = 0.952$). Such lesser hormone sensitivity of the Basal subtype was also suggested (Fig. 1c) by the higher percentage of "high-risk" cases according to the E2F4-activation signature associated with resistance to HT[8] (47% of Basal versus 34% of Luminal; $q = 8.93E−11$).

We then compared the potential vulnerabilities of ER+/HER2− Basal versus Luminal tumors to targeted therapies. We found higher RBsig score[9] and E2F regulon score[10] in the Basal subtype ($q = 2.44E−31$ and $q = 1.02E−08$ respectively), evocative of RB1-pathway disruption and associated with resistance to CDK4/6 inhibitors, suggesting a likely lower sensitivity to these drugs. By contrast, the percentage of patients with signature evocative of homologous recombination deficiency[11] was higher in the Basal (21%: 25/109) versus Luminal samples (6%: 29/520; $q = 5.18E−06$), suggesting potential higher sensitivity to PARP inhibitors. The percentage of patients with PIK3CA hotspot mutation was higher

in Luminal (37%: 571/1528) versus Basal (30%: 88/290), but the difference lost significance after correction for multiple testing ($q = 0.100$). Regarding the other actionable genetic alterations (AGAs) of BC with clinical evidence level equal to 1−2[12,13], FGFR2 and KRAS amplifications tended to more frequent and PIK3CA amplification more frequent in the Basal subtype (Table S3). Finally, the immune expression profiles were also very different. The Basal subtype displayed enrichment for the Bindea's expression modules[14] representing cell populations associated with adaptive and innate immunity ($q < 1.00E−50$), higher PDL1 mRNA expression ($q = 4.19E−23$), and more frequent ICR4 class (24% versus 7%; $q = 5.75E−39$)[6], suggesting potential for better response to immune checkpoint inhibitors.

Molecular profiles of the ER+/HER2− Basal versus Luminal subtypes

We then compared the whole-exome mutational, whole-genome transcriptional, and proteomic (RPPA) profiles of the two subtypes in the TCGA set. Only one gene, TP53, was differentially mutated, in 39% of Basal versus 12% of Luminal samples ($q = 2.17E−05$:

Table S4). We identified 906 genes as differentially expressed between the two ER+/HER2− subtypes (Table S5). The robustness of this gene list was confirmed by its capacity to differentiate the two subtypes in the independent validation set (Fig. S2). Ontology analysis (Table S6) showed that the genes upregulated in Basal samples were mainly involved in immune response. Comparison of RPPA profiles identified 89 proteins/phosphoproteins as differentially expressed between the two ER+/HER2− subtypes (Table S7). ER-alpha, PR and GATA3 were as expected among the top proteins overexpressed in the Luminal subtype, as well as androgen receptor (AR)[15] and Bcl-2[16], two therapeutic targets under development in ER+/HER2 BC, whereas PARP1 was among the top proteins overexpressed in the Basal samples.

To further compare the extent of differences between the ER+/HER2− Basal versus Luminal samples, we included the ER− Basal subtype. There was a gradient between the three molecular subtypes for nearly all comparisons (Table 1, Fig. 1c), the ER+/HER2− Basal subtype being intermediate between the two other ones, but much more different (as indicated by the q values in Fig. 1c) in terms of response or vulnerability to systemic therapies from the ER+HER2− Luminal subtype than from the ER− Basal subtype. Principal component analysis (PCA) of whole-genome transcriptional profiles (Fig. S3), and the classification of samples based on the 906 gene list (Fig. S2) confirmed that the ER+/HER2− Basal subtype was intermediate between the two other subtypes. Differences were also present with respect to the mammary stem cells signatures[17,18], the ER+/HER2− Basal subtype being more related to the mammary stem cells, the ER+/HER2− Luminal to the mature luminal cells, and the ER− Basal to the progenitor luminal cells.

DISCUSSION

Our in silico analysis shows that the ER+/HER2− Basal subtype is very different from the ER+/HER2− Luminal subtype—and sometimes closer to the ER− Basal subtype—in terms of response and/or potential vulnerability to systemic therapies of BC. These results obtained on a large series reinforce the potential clinical value of the molecular subtypes within ER+/HER2− BCs, already suggested in smaller series[4,19] regarding the prognosis after HT and the sensitivity to chemotherapy and CDK4/6 inhibitors. They also suggest differential therapeutic vulnerability regarding PARP inhibitors and platinum salts, PIK3CA inhibitor, immune therapy and other targeted therapies under development.

The Basal samples were more frequently associated with poor-prognosis features than the Luminal samples. However, the 5-year DRFI was not different, in agreement with the MINDACT trial[20], but in contrast with the Groenendijk's study likely because of a longer follow-up. Other studies have reported decreased survival outcomes in the Basal subtype when compared with the Luminal one in early BC treated with adjuvant tamoxifen[21] and in advanced BC treated with letrozole[22]. In our study, the high percentage of ER+/HER2− patients untreated with adjuvant HT (65%) allowed to compare the benefit of adjuvant HT in the two subtypes, showing higher benefit in the Luminal subtype, even if the number of Basal patients was relatively small (N = 254) and precluded any definite conclusion. This high percentage, notably in the Luminal subtype (65% versus 32% in the Basal subtype), and this observation likely explain the absence of difference observed for DRFI between the two subtypes. Higher sensitivity to HT of the Luminal subtype might be at least in part related to higher ER expression level (mRNA and protein) and a more functional ER pathway (Gatza's activation signatures). The Basal-type was more sensitive to neoadjuvant CT as previously reported in smaller series of ER+/HER2− patients[23,24], likely in part because of higher pathological grade and cell proliferation rate (prognostic GES). Signatures predictive of response to CDK4/6 inhibitors, recently approved in BC[25], suggested higher sensitivity of the Luminal

subtype. In the PALOMA-2 study, the non-Luminal subtypes (20% of the entire population) had very small absolute benefits, if any, from palbociclib, whereas the Luminal subtype benefited substantially from palbociclib plus letrozole versus letrozole[26]. Our data provide new insights regarding the potential vulnerability to other drugs recently approved or in development in BC. For example, no data are available in the literature regarding the sensitivity to PARP inhibitors of the Luminal versus Basal subtypes within ER+/HER2− cases. Our results (HRD score) suggest that the drugs being evaluated in Basal/TN breast cancers (PARP inhibitors and platinum salts) deserve to be tested in the Basal subtype. No signature predictive for sensitivity exists for the alpelisib inhibitor[27], but we found more frequent PIK3CA mutations, a clinically validated selection criterion for alpelisib, in the Luminal subtype, even if the difference was not significant after correction for multiple testing. A few level 2 AGAs, FGFR2 and KRAS amplifications and PIK3CA amplification were more frequent in Basal samples. Finally, the Basal subtype displayed immune profile of "hotter" tumors than the Luminal subtype, suggesting potentially higher sensitivity to immune checkpoint inhibitors. Importantly, all these differences were not dependent on the molecular classification used because we observed similar results when using the PAM50 signature (Table S2, Fig. S4).

To our knowledge, we also report the first comparative analysis of large-scale molecular profiles of ER+/HER2− Basal versus Luminal subtypes. Only one gene (TP53) was differentially mutated, whereas important differences existed at the mRNA and protein levels. Five percent of genes tested were differentially expressed, including many immune genes upregulated in Basal samples. Thirty-nine percent of 226 proteins/phosphoproteins tested were differentially expressed, including therapeutic targets of drugs under development in ER+/HER2− BC such as PARP1, AR and Bcl-2. Inclusion of the ER− Basal subtype showed a gradient between the three subtypes from the ER+/HER2− Luminal subtype to the ER− Basal subtype in terms of clinicopathological correlations and transcriptional profiles. However, the ER+/HER2− Basal subtype was closer to the ER− Basal subtype than to the ER+/HER2− Luminal subtype in terms of response/probability of response to systemic therapies. The extent of differences between subtypes was also suggested by the mammary stem cell signatures, which could suggest a different cell-of-origin. Strikingly, the ER+/HER2− Basal samples were closer to the mammary stem cells than were the ER− Basal samples, warranting further investigations.

In conclusion, our results reinforce the potential clinical value of the different molecular classifications in ER+/HER2− BCs: the Basal and Luminal subtypes are so different with respect to therapeutic response/vulnerability, metastatic risk and cell-of-origin that they cannot continue to be assimilated and treated as a unique homogeneous ER+/HER2− group. Validation in prospective clinical trials is warranted, and caution is required in the interpretation of ongoing trials and the design of future trials.

METHODS

Breast cancer samples, gene profiling and data analysis

We analyzed our BC gene expression database[6] pooled from 36 public datasets (Table S8), comprising 8982 invasive primary BCs. The details of Institutional Review Board and Ethical Committee approval and patients' consent for all 36 studies are present in their corresponding publications listed in Table S8. The preanalytic data processing was done as described[6]. We also collected DNA and proteomic data from TCGA (WES data, array-CGH, HRD score, RPPA)[11] and the PIK3CA and TP53 mutational statutes of TCGA and Metabric[28].

ER and HER2 statutes were preferentially determined by study annotation when in agreement with the recent ASCO guidelines[29,30]. Where unavailable, the normalized gene expression data were used to infer the receptor status. Indeed, several points made impossible to apply the same definition of IHC positivity to all samples: the definition of

ER-positivity by IHC was not available across all studies, and when available, the positivity cut-off was not similar between all studies (1% or 10%); the percentage of tumor cells stained was not available across most of studies; and the ER status was missing for a few samples. The cut-off value for ER status in the most recent ASCO guidelines is 1%[30]. Thus, we have kept the IHC status annotated in the original study when it was based upon a cut-off value of 1%. When it was annotated positive with a 10% cut-off, the sample remained of course positive with the 1% cut-off. When we could not redefine the ER status according to the ASCO guidelines (e.g. in case of a negative status when a 10% cut-off on IHC was used in the original study, or in case of unavailability), we inferred the ER status from transcriptional data of ESR1, by using a two-component Gaussian finite Mixture Model using maximum likelihood estimation on a per-study basis as previously described[31]. This was made possible thanks to the bimodal distribution of mRNA expression level. The mRNA threshold was verified on the 5548 samples of our database ER-annotated according to the 1% ASCO guideline IHC cut-off: the concordance rate was 93%, suggesting good performance. Out of the 8982 samples of our database, 6563 were defined as ER+ (5342 according to IHC and 1221 according to inferred stratus).

The same process was applied for defining the HER2 status, the cut-off of which was based on the recent ASCO guidelines[29]. In accordance with this guideline, the samples were considered HER2+ when they reached an IHC score of 3+ or 2+ with FISH amplification. A HER2 IHC score of +1 or 0 was considered negative. When the information was not available, and thanks to the bimodal distribution of mRNA expression level, the HER2 status was inferred from the mRNA expression level as described for ER[31]. The mRNA threshold was verified within the 6563 ER+ samples on the 3050 HER2-annotated according to the ASCO guidelines: the concordance rate was 91%, suggesting good performance. Out of the 6563 ER+ samples of our database, 727 were defined as HER2+ (421 according to IHC and 306 according to inferred stratus), leaving 5836 ER+/HER2− samples.

We applied to each dataset separately several GES: 80-GS[5] and PAM50[32], and the surrogate signatures of three commercial prognostic signatures (Recurrence Score[33], Mammaprint[34], EndoPredict[35]), Gatza's ER and PR pathways activation signatures[36], E2F4-activation signature associated with resistance to HT[8], two scores associated with resistance to CDK4/6 inhibitors on preclinical models for RBsig[9] and on clinical samples of PALOMA-3 trial for E2F regulon[10], several Bindea's expression modules[14] representing different immune cell populations, the immunologic constant of rejection (ICR) signature[6], and two signatures related to mammary epithelial cell hierarchy and stem cells[17,18] were informative. For all above-quoted GES, all 5272 samples (931 Basal and 4341 Luminal) were informative. The PDL1/CD274 mRNA expression level was assessed as previously described[17,18]. The present in silico study was approved by our Institutional Review Board (No. 19-006; March 15, 2019).

Supervised analyses compared the profiles of ER+/HER2− Basal versus Luminal TCGA samples at several levels: WES mutational and RPPA proteomic using logistic regression with significance thresholds of $p \leq 0.05$ and $q \leq 0.10$, and transcriptional using moderated t test with significance thresholds of fold-change and $|FC| > 2$, $p \leq 0.05$ and $q \leq 0.10$.

Statistical analysis

The data generated and analyzed during this study are described in[37] and correspond to 36 publicly available datasets that are listed in Supplementary Table 8 (all datasets) and in data refs [7,11,28,34,38–105]. The prognostic clinicopathological variables tested included the patients' age (≤50 years versus >50), and pathological type (ductal versus lobular versus other), grade (1 versus 2 versus 3), tumor size (pT1 versus pT2 versus pT3), and axillary lymph node status (negative versus positive), and use of adjuvant HT and CT (yes versus no). Distant relapse-free interval was calculated from the date of diagnosis until the date of distant relapse or death from breast cancer. Event-free patients lost to follow-up or dead from unspecified cause or from cause unrelated to breast cancer were censored at time of last contact. The follow-up was calculated from the date of diagnosis until the date of last news for event-free patients using the reverse Kaplan−Meier method. Survivals were calculated using the Kaplan−Meier method and compared with the log-rank test. Uni- and multivariate analyses for DRFI were done using Cox regression analysis (Wald test). The pCR after neoadjuvant chemotherapy was defined as absence of invasive cancer in both breast and axillary lymph nodes (ypT0/Tis-ypN0). Correlations between molecular subtypes and other variables were analyzed using Fisher's exact test and t test (Table 1) and regression analysis with binomial or Gaussian family (Figs. 1 and S4) for discrete or continuous variables respectively. All statistical tests were two-sided at the 5% level of significance. In the case of multiple testing, the p values were replaced by the corrected q values. Analyses were done in the R software (version 3.5.2) using glm function and the survival package (version 2.44).

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 analyzed during this study are described in the following data record: https://doi.org/10.6084/m9.figshare.11558676. The dataset Breast_cancer_classifications.csv supporting Fig. 1, Table 1, and Supplementary Tables 1−3, and containing the clinicopathological and molecular data of tumors analyzed, is publicly available in the figshare repository as part of the data record mentioned above. (GEO, Array Express and EGA datasets).

CODE AVAILABILITY

All codes and scripts used for applying the gene classifiers are available at https://github.com/predictiveoncologyCRCM/NPJbc_2019_ERposBC.

REFERENCES

1. Bertucci, F. et al. How basal are triple-negative breast cancers? *Int. J. Cancer* **123**, 236–240 (2008).
2. Cheang, M. C. et al. Responsiveness of intrinsic subtypes to adjuvant anthracycline substitution in the NCIC.CTG MA.5 randomized trial. *Clin. Cancer Res.* **18**, 2402–2412 (2012).
3. Prat, A. et al. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. *J. Clin. Oncol.* **31**, 203–209 (2013).
4. Groenendijk, F. H. et al. Estrogen receptor variants in ER-positive basal-type breast cancers responding to therapy like ER-negative breast cancers. *NPJ Breast Cancer* **5**, 15 (2019).
5. Krijgsman, O. et al. A diagnostic gene profile for molecular subtyping of breast cancer associated with treatment response. *Breast Cancer Res. Treat.* **133**, 37–47 (2012).
6. Bertucci, F. et al. The immunologic constant of rejection classification refines the prognostic value of conventional prognostic signatures in breast cancer. *Br. J. Cancer* **119**, 1383–1391 (2018).
7. Hess, K. R. et al. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J. Clin. Oncol.* **24**, 4236–4244 (2006).
8. Guerrero-Zotano, A. L. et al. ER(+) breast cancers resistant to prolonged neoadjuvant letrozole exhibit an E2F4 transcriptional program sensitive to CDK4/6 inhibitors. *Clin. Cancer Res.* **24**, 2517–2529 (2018).
9. Malorni, L. et al. A gene expression signature of retinoblastoma loss-of-function is a predictive biomarker of resistance to palbociclib in breast cancer cell lines and is prognostic in patients with ER positive early breast cancer. *Oncotarget* **7**, 68012–68022 (2016).
10. Turner, N. C. et al. Cyclin E1 expression and palbociclib efficacy in previously treated hormone receptor-positive metastatic breast cancer. *J. Clin. Oncol.* **37**, 1169–1178 (2019).
11. The Cancer Genome Atlas. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
12. Arnedos, M. et al. Precision medicine for metastatic breast cancer-limitations and solutions. *Nat. Rev. Clin. Oncol.* **12**, 693–704 (2015).
13. Condorelli, R. et al. Genomic alterations in breast cancer: level of evidence for actionability according to ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). *Ann. Oncol.* **30**, 365–373 (2019).
14. Bindea, G. et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795 (2013).
15. Kono, M. et al. Androgen receptor function and androgen receptor-targeted therapies in breast cancer: a review. *JAMA Oncol.* **3**, 1266–1273 (2017).
16. Drago, J. Z., Chandarlapaty, S. & Jhaveri, K. Targeting apoptosis: a new paradigm for the treatment of estrogen receptor-positive breast cancer. *Cancer Discov.* **9**, 323–325 (2019).

17. Creighton, C. J. et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl. Acad. Sci. USA* **106**, 13820–13825 (2009).

18. Lim, E. et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913 (2009).

19. Cejalvo, J. M. et al. Clinical implications of the non-luminal intrinsic subtypes in hormone receptor-positive breast cancer. *Cancer Treat. Rev.* **67**, 63–70 (2018).

20. Viale, G. et al. Immunohistochemical versus molecular (BluePrint and Mamma-Print) subtyping of breast carcinoma. Outcome results from the EORTC 10041/BIG 3-04 MINDACT trial. *Breast Cancer Res. Treat.* **167**, 123–131 (2018).

21. Prat, A. et al. Concordance among gene expression-based predictors for ER-positive breast cancer treated with adjuvant tamoxifen. *Ann. Oncol.* **23**, 2866–2873 (2012).

22. Prat, A. et al. Prognostic value of intrinsic subtypes in hormone receptor-positive metastatic breast cancer treated with letrozole with or without lapatinib. *JAMA Oncol.* **2**, 1287–1294 (2016).

23. Prat, A. et al. Response and survival of breast cancer intrinsic subtypes following multi-agent neoadjuvant chemotherapy. *BMC Med.* **13**, 303 (2015).

24. Prat, A. et al. Prediction of response to neoadjuvant chemotherapy using core needle biopsy samples with the prosigna assay. *Clin. Cancer Res.* **22**, 560–566 (2016).

25. Turner, N. C. et al. Palbociclib in hormone-receptor-positive advanced breast cancer. *N. Engl. J. Med.* **373**, 209–219 (2015).

26. Finn, R. S. et al. Comprehensive gene expression biomarker analysis of CDK 4/6 and endocrine pathways from the PALOMA-2 study *Cancer Res.* **78**, Abstract nr P2-09-10 (2018).

27. Andre, F. et al. Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced Breast Cancer. *N. Engl. J. Med.* **380**, 1929–1940 (2019).

28. Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).

29. Wolff, A. C. et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J. Clin. Oncol.* **31**, 3997–4013 (2013).

30. Hammond, M. E. et al. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J. Clin. Oncol.* **28**, 2784–2795 (2010).

31. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J. Clin. Invest.* **121**, 2750–2767 (2011).

32. Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).

33. 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).

34. van de Vijver, M. J. et al. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).

35. Filipits, M. et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin. Cancer Res.* **17**, 6012–6020 (2011).

36. Gatza, M. L. et al. A pathway-based classification of human breast cancer. *Proc. Natl. Acad. Sci. USA* **107**, 6994–6999 (2010).

37. Bertucci, F., Finetti, P., Goncalves, A. & Birnbaum, D. Metadata and data files supporting the published article: the therapeutic response of ER+/HER2- breast cancers differs according to the molecular Basal or Luminal subtype. Figshare https://doi.org/10.6084/m9.figshare.11558676 (2020).

38. Curley, E. Expression Project for Oncology (expO). NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE2109 (2005).

39. Iggo, R. EORTC 10994 clinical trial. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE1561 (2005).

40. Minn, A. J. & Massague, J. Subpopulations of MDA-MB-231 and primary breast cancers. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE2603 (2005).

41. Jatkoe, T. Breast cancer relapse free survival. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE2034 (2005).

42. Ivshina, A. V. et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE4922 (2006).

43. Pawitan, Y. et al. Gene expression of breast cancer tissue in a large population-based cohort of Swedish patients. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE1456 (2006).

44. Sotiriou, C. et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE2990 (2006).

45. Farmer, P., Bonnefoi, H. & Iggo, R. Prediction of breast cancer pathological complete response to anthracycline/taxane chemotherapy. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE6861 (2007).

46. Farmer, P., Bonnefoi, H. & Iggo, R. A stroma-related gene signature predicts sensitivity to epirubicin-containing neoadjuvant chemotherapy in breast cancer. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE4779 (2008).

47. Desmedt, C. et al. Strong time dependence of the 76-gene prognostic signature. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE7390 (2007).

48. Seitz, S. et al. Comparison of gene expression data from human and mouse breast cancers: identification of conserved breast tumor genes. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE6596 (2007).

49. Marty, B. et al. Frequent PTEN genomic alterations and activated PI3K pathway in basal-like breast cancer. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE13787 (2008).

50. Spellman, P. E-TABM-158—Transcription profiling of human breast cancer samples. Array Express https://identifiers.org/arrayexpress:E-TABM-158 (2008).

51. Schmidt, M. et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE11121 (2008).

52. Yu, K. & Ganesan, K. A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE5364 (2008).

53. Smid, M. Expression data from primary breast tumors. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE12276 (2009).

54. Januario, T. & Lackner, M. R. Gene expression profiling of 30 human breast cancers. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE12763 (2009).

55. Zhang, Y., Wang, Y. & Fokens, J. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE12093 (2008).

56. Miller, W. R. et al. Letrozole (Femara) early response to treatment. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE5462 (2007).

57. Datto, M. & Barry, W. Intratumor heterogeneity and precision of microarray-based predictors of breast cancer biology and clinical outcome. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE23593 (2010).

58. Korde, L. A. & Lusa, L. Expression data from human breast cancers pre and post chemothrapy. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE18728 (2009).

59. Prat, A. & Perou, C. M. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE18229 (2010).

60. Richardson, A. L. & Eklund, A. C. Tumor expression data from neoadjuvant trial of cisplatin monotherapy in triple negative breast cancer patients. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE18864 (2009).

61. Pusztai, L., Tabchy, A. & Qi, Y. Expression data from breast cancer FNA biopsies from patients. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE20271 (2010).

62. Jönsson, G. et al. Genomic subtypes of breast cancer identified by array comparative genomic hybridization. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE22133 (2010).

63. Chen, D., Fulp, W. J. & Chinnaibally, R. V. Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE10780 (2009).

64. Shi, L. MAQC-II Project: human breast cancer (BR) data set. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE20194 (2010).

65. Pusztai, L., Iwamoto, T. & Qi, Y. Expression data from breast cancer FNA biopsies from patients. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE22093 (2010).

66. Pusztai, L., Ueno, N. T., Cristofanilli, M., Iwamoto, T. & Qi, Y. Expression data from Fine Needle Aspiration (FNA) biopsies from breast cancer patients. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE22597 (2011).

67. Desmedt, C. et al. Multifactorial approach to predicting resistance to anthracyclines. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE16446 (2010).

68. Petel, F. E-MTAB-365—Transcription profiling by array of breast cancer samples to define breast cancer subsets. Array Express https://identifiers.org/arrayexpress:E-MTAB-365 (2011).

69. Hatzis, C. Genomic predictor of response and survival following neoadjuvant taxane-anthracycline chemotherapy in breast cancer. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE25066 (2011).

70. Sabatier, R. et al. Down-regulation of ECRG4, a candidate tumor suppressor gene in human breast cancer. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE31448 (2011).

71. Hoog, J. W. BRC77 agilent gene expression arrays. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE29442 (2012).

72. Hoog, J. Paradigm test set expression array. NCBI Gene Expression Omnibus https://identifiers.org/geo:GSE35186 (2012).

73. University of Cambridge & Cancer Research UK Cambridge Research Institute. European Genome-phenome Archive https://identifiers.org/ega.study:EGAS00000000083 (2012).

74. van 't Veer, L. J. et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–536 (2002).

75. Curley, E. Expression Project for Oncology (expO). Unpublished http://www.intgen.org/research-services/biobanking-experience/expo/ (2005).

76. Farmer, H. et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature 434, 917–921 (2005).

77. Minn, A. J. et al. Genes that mediate breast cancer metastasis to lung. Nature 436, 518–524 (2005).

78. Wang, Y. et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365, 671–679 (2005).

79. Ivshina, A. V. et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res. 66, 10292–10301 (2006).

80. Sotiriou, C. et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J. Natl. Cancer Inst. 98, 262–272 (2006).

81. Bonnefoi, H. et al. Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial. Lancet Oncol. 8, 1071–1078 (2007).

82. Desmedt, C. et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin. Cancer Res. 13, 3207–3214 (2007).

83. Klein, A. et al. Comparison of gene expression data from human and mouse breast cancers: identification of a conserved breast tumor gene set. Int. J. Cancer 121, 683–688 (2007).

84. Marty, B. et al. Frequent PTEN genomic alterations and activated phosphatidylinositol 3-kinase pathway in basal-like breast cancer cells. Breast Cancer Res. 10, R101 (2008).

85. Chin, K. et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10, 529–541 (2006).

86. Schmidt, M. et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res. 68, 5405–5413 (2008).

87. Yu, K. et al. A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers. PLoS Genet. 4, e1000129 (2008).

88. Bos, P. D. et al. Genes that mediate breast cancer metastasis to the brain. Nature 459, 1005–1009 (2009).

89. Hoeflich, K. P. et al. In vivo antitumor activity of MEK and phosphatidylinositol 3-kinase inhibitors in basal-like breast cancer models. Clin. Cancer Res. 15, 4649–4664 (2009).

90. Zhang, Y. et al. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. Breast Cancer Res. Treat. 116, 303–309 (2009).

91. Miller, W. R. & Larionov, A. Changes in expression of oestrogen regulated and proliferation genes with neoadjuvant treatment highlight heterogeneity of clinical resistance to the aromatase inhibitor, letrozole. Breast Cancer Res. 12, R52 (2010).

92. Barry, W. T. et al. Intratumor heterogeneity and precision of microarray-based predictors of breast cancer biology and clinical outcome. J. Clin. Oncol. 28, 2198–2206 (2010).

93. Korde, L. A. et al. Gene expression pathway analysis to predict response to neoadjuvant docetaxel and capecitabine for breast cancer. Breast Cancer Res. Treat. 119, 685–699 (2010).

94. Prat, A. et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 12, R68 (2010).

95. Silver, D. P. et al. Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. J. Clin. Oncol. 28, 1145–1153 (2010).

96. Tabchy, A. et al. Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. Clin. Cancer Res. 16, 5351–5361 (2010).

97. Jonsson, G. et al. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. Breast Cancer Res. 12, R42 (2010).

98. Chen, D. T. et al. Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. Breast Cancer Res. Treat. 119, 335–346 (2010).

99. Popovici, V. et al. Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. Breast Cancer Res. 12, R5 (2010).

100. Iwamoto, T. et al. Gene pathways associated with prognosis and chemotherapy sensitivity in molecular subtypes of breast cancer. J. Natl. Cancer Inst. 103, 264–272 (2011).

101. Desmedt, C. et al. Multifactorial approach to predicting resistance to anthracyclines. J. Clin. Oncol. 29, 1578–1586 (2011).

102. Guedj, M. et al. A refined molecular taxonomy of breast cancer. Oncogene 31, 1196–1206 (2012).

103. Hatzis, C. et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. JAMA 305, 1873–1881 (2011).

104. Sabatier, R. et al. Down-regulation of ECRG4, a candidate tumor suppressor gene, in human breast cancer. PLoS ONE 6, e27656 (2011).

105. Ellis, M. J. et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 486, 353–360 (2012).

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41523-020-0151-5.

Correspondence and requests for materials should be addressed to F.B.

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.