

1 **Targetable *ERBB2* mutation status is an independent marker of adverse prognosis in**
2 **estrogen receptor positive, *ERBB2* non-amplified primary lobular breast carcinoma: a**
3 **retrospective *in silico* analysis of public datasets**

4 Sasagu Kurozumi^{1,2}, Mansour Alsaleem^{1,4}, Cíntia J. Monteiro^{1,3}, Kartikeya Bhardwaj^{1,3}, Stacey
5 E. P. Joosten⁵, Takaaki Fujii², Ken Shirabe², Andrew R. Green¹, Ian O. Ellis¹, Emad A. Rakha¹,
6 Nigel P. Mongan⁴, David M. Heery³, Wilbert Zwart⁵, Steffi Oesterreich⁶ and Simon J.
7 Johnston^{1,3}

8
9 ¹Nottingham Breast Cancer Research Centre, School of Medicine, University of Nottingham,
10 UK; ²Department of General Surgical Science, Gunma University Graduate School of Medicine,
11 Gunma, Japan; ³Gene Regulation and RNA Biology Laboratory, School of Pharmacy, University
12 of Nottingham, UK; ⁴School of Veterinary Medicine and Science, University of Nottingham,
13 UK; ⁵Division of Oncogenomics, OncoCode Institute, Netherlands Cancer Institute, Amsterdam,
14 Netherlands; ⁶Womens Cancer Research Center, UPMC Hillman Cancer Center and Magee-
15 Women Research Institute, Pittsburgh, PA, USA.

16
17 **Contact Information**

18 Dr Simon J Johnston PhD MBBS
19 Nottingham Breast Cancer Research Centre, University of Nottingham
20 Nottingham Biodiscovery Institute, University Park Nottingham
21 Tel: +44 (0) 115 82 31858. Email: simon.johnston@nottingham.ac.uk

22
23 **Key Words**

24 **lobular breast cancer, *ERBB2*, HER2, mutation, prognosis, therapeutic biomarker, adjuvant.**

25 **Abstract**

26 **Background**

27 Invasive lobular carcinoma (ILC) accounts for 10-15% of primary breast cancers and is typically
28 estrogen receptor alpha positive (ER+) and *ERBB2* non-amplified. Somatic mutations in
29 *ERBB2/3* are emerging as a tractable mechanism underlying enhanced human epidermal
30 growth factor 2 (HER2) activity. We tested the hypothesis that therapeutically targetable
31 *ERBB2/3* mutations in primary ILC of the breast associate with poor survival outcome in large
32 public datasets.

33 **Methods**

34 We performed *in silico* comparison of *ERBB2* non-amplified cases of ER+ stage I-III primary ILC
35 ($N=279$) and invasive ductal carcinoma (IDC, $N=1,301$) using METABRIC, TCGA and MSK-
36 IMPACT information. Activating mutations amenable to HER2-directed therapy with neratinib
37 were identified using existing functional data from *in vitro* cell line and xenograft experiments.
38 Multivariate analysis of 10-year overall survival (OS) with tumor size, grade and lymph node
39 status was performed using a Cox regression model. Differential gene expression analyses by
40 *ERBB2* mutation and amplification status was performed using weighted average differences
41 and an *in-silico* model of response to neratinib derived from breast cancer cell lines.

42 **Results**

43 ILC tumors comprised 17.7% of all cases in the dataset but accounted for 47.1% of *ERBB2*-
44 mutated cases. Mutations in *ERBB2* were enriched in ILC *versus* IDC cases (5.7%, $N=16$ vs.
45 1.4%, $N=18$, $p<0.0001$) and clustered in the tyrosine kinase domain of HER2. *ERBB3* mutations
46 were not enriched in ILC (1.1%, $N=3$ vs. 1.8%, $N=23$; $p=0.604$). Median OS for patients with
47 *ERBB2*-mutant ILC tumors was 66 months *versus* 211 months for *ERBB2* wild-type ($p=0.0001$),
48 and 159 vs. 166 months ($p=0.733$) for IDC tumors. Targetable *ERBB2* mutational status was

49 an independent prognostic marker of 10-year OS – but only in ILC (hazard ratio, HR=3.7, 95%
50 CI 1.2–11.0; $p=0.021$). Findings were validated using a novel *ERBB2* mutation gene enrichment
51 score (HR for 10-year OS in ILC=2.3, 95% CI 1.04–5.05; $p=0.040$).

52 **Conclusions**

53 Targetable *ERBB2* mutations are enriched in primary ILC and their detection represents an
54 actionable strategy with the potential to improve patient outcomes. Biomarker-led clinical
55 trials of adjuvant HER-targeted therapy are warranted for patients with *ERBB2*-mutated
56 primary ILC.

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71 **Background**

72 Invasive lobular carcinoma of the breast (ILC) accounts for 10-15% of all breast cancer with
73 an estimated 250,000 cases per year worldwide [1-4]. Nearly all cases of ILC derive from
74 luminal cells that express estrogen receptor alpha but lack E-Cadherin (*CDH1*) expression or
75 *ERBB2* amplification [5].

76

77 Clinical evidence suggests that despite favorable prognostic indicators, e.g. ER+ and/or
78 progesterone receptor positive, low Ki67 proliferation index and HER2- status, patients with
79 ILC have similar or worse long-term outcomes compared to those with invasive ductal
80 carcinoma (IDC, otherwise known as invasive carcinoma of no special type) [5, 6]. Adjuvant
81 treatment of ER+ ILC with letrozole, an aromatase inhibitor, may be superior to tamoxifen,
82 and ILC cells demonstrate resistance to tamoxifen *in vitro* [7, 8]. However, patients with ILC
83 are treated according to identical protocols as those with IDC [9, 10].

84

85 HER2-targeted therapy is indicated for patients whose tumors are HER2+ by
86 immunohistochemistry (IHC) or, if IHC is equivocal, where *ERBB2* is amplified as detected by
87 *in situ* hybridization (ISH) [11, 12]. Recent evidence indicates that in *ERBB2* non-amplified
88 breast cancer, somatic mutation of *ERBB2* (*ERBB2mut*) and/or *ERBB3* (*ERBB3mut*) may
89 provide an alternative mechanism for upregulation of HER2 activity that is therapeutically
90 tractable using second generation HER2 tyrosine kinase inhibitors such as neratinib [13, 14].

91

92 The epidermal growth factor (EGF) family of receptor tyrosine kinases (HER1-4) are activated
93 by ligand-dependent homo-/heterodimerisation and regulate cellular proliferation and tumor
94 progression [15, 16]. In *ERBB2* amplified cells, the oncogenic effect of HER2 is mediated by

95 heterodimerisation with HER3 in a ligand-independent manner [17]. Thus HER3 is necessary
96 for HER2 oncogenic activity, and both HER2 and HER3 are therapeutic targets in *ERBB2*
97 amplified breast cancer [18]. Mutations in *ERBB2* cluster in the tyrosine kinase and extra-
98 cellular domains of HER2 and exert their oncogenic effects by activating tyrosine kinase
99 activity or increasing HER2 dimerisation, respectively [13]. *In vitro* studies of HER2 activity in
100 cell line and xenograft models identified 13 mutations (listed in the Methods) that enhanced
101 proliferation and/or demonstrated growth inhibition with the irreversible HER2/EGFR
102 tyrosine kinase inhibitor, neratinib [13]. Mutations in HER3, a critical binding partner for
103 HER2, have been shown to promote ligand (EGF)-independent transformation of breast
104 epithelial cells only in the presence of kinase-active HER2 [19]. This indicates that known
105 oncogenic mutations in *ERBB3*, e.g. G284R and E928G, may also be therapeutically targetable
106 *via* HER2 inhibition [19].

107

108 *ERBB2* mutations have previously been linked with worse prognosis in *CDH1*-altered ILC: a
109 study of ILC cases in the TCGA dataset ($N=169$) found that *ERBB2* mutations ($N=6$) occurred
110 exclusively in *CDH1*-altered tumors ($N=100$) [20]. However, prognostic data on the 6
111 *ERBB2*mut cases were limited to 2 patient events for both disease-free survival and OS
112 analyses. The overall rate of *ERBB2*mut in this study of primary ILC was 3.6% ($N=6$ out of 169).
113 A study of relapsed *CDH1*-mutated ILC found *ERBB2*mut in 18% of cases ($N=4$ of 22),
114 suggesting further acquisition of *ERBB2* mutations in *CDH1*-altered ILC due to the selective
115 pressure of treatment [21].

116

117 To demonstrate the potential clinical benefit of targeting low frequency somatic mutations,
118 prognostic analyses using large clinical datasets are required. In the MA12 trial comprising

119 328 premenopausal patients with ER+ primary breast cancer of all histological subtypes, non-
120 silent *ERBB2* mutations occurred in 5.2% of patients ($N=17$) and were adversely prognostic of
121 OS ($p=0.0114$) [22]. A study of 5,605 cases of relapsed breast cancer found *ERBB2*mut in 2.4%
122 of cases ($N=138$), of which 20% ($N=27$ of 138) were in ILC tumors [23]. However, neither of
123 these larger studies stratified clinical outcome by histological subtype.

124

125 An ILC-specific study of 630 cases of primary ILC found *ERBB2* and *ERBB3* mutations in 5.1%
126 and 3.6% of tumors respectively [24]. Comparison of cases of ER+, HER2- ILC from the same
127 study ($N=371$) with cases of ER+, HER2- IDC from TCGA ($N=338$) indicated significant
128 enrichment of both *ERBB2* and *ERBB3* mutations in ILC (4.3% and 3.5% in ILC vs. 1.5% and
129 0.6% in IDC) [24]. The study reported limited statistical evidence of a time-dependent effect
130 of *ERBB2* mutational status associated with short-term breast cancer-specific survival.
131 However, confirmation in datasets including patients with long-term follow-up and
132 decoupling of activating from silent mutations is needed.

133

134 We hypothesized that low-frequency somatic mutations in *ERBB2* and *ERBB3* are enriched in
135 ER+, *ERBB2* non-amplified primary ILC cases and may have demonstrable prognostic effect.
136 We tested these hypotheses by mining a combined dataset of the three largest primary breast
137 cancer series with data on tumor *ERBB2* and *ERBB3* mutational status, gene expression,
138 clinicopathological features and patient survival outcomes. Our overall goal was to determine
139 the association between targetable *ERBB2/3* mutations and survival in ILC, and thereby
140 provide evidence for a clinically actionable strategy to improve outcomes for patients with
141 ILC.

142

143 **Patients and methods**

144 **Patients and outcome measures**

145 Genomic and clinical outcome data associated with tumor samples from patients with
146 primary breast cancer in TCGA 2015 ($N=817$), METABRIC 2012 and 2016 ($N=2,509$) and MSK-
147 IMPACT 2018 ($N=918$) were accessed online *via* CBioportal [25-27]. From these datasets, ER+
148 and HER2- cases of stage I-III ILC and IDC with both clinical outcome and mutational data
149 called from next-generation sequencing (NGS) analyses were selected ($N=1,580$). Cases of
150 mixed or non-ILC/IDC histology, ER-negative/undetermined, HER2+/undetermined,
151 carcinoma-*in-situ*, and stage 4/undetermined were excluded. For TCGA and MSK datasets,
152 HER2 status was determined by IHC (positive/negative) or, where IHC was indeterminate, by
153 ISH assessment of *ERBB2* amplification, in line with standard clinical practice [11]. For
154 METABRIC cases, HER2 status was determined using the Affymetrix SNP6 copy number
155 inference pipeline.

156

157 The primary outcome measure, available in all datasets, was OS. Variables included *ERBB2*
158 and *ERBB3* mutational status. *ERBB2*mut status was subcategorized as oncogenic or
159 uncharacterized by cross-reference with existing data that identified mutations targetable by
160 HER2-inhibition: G309A/E, S310F, L755S, del755-759, S760A, D769H, D769Y, V777L, P780ins,
161 V842I, R896C [13]. Cases were denominated onc*ERBB2*mut if tumors harbored at least one
162 oncogenic *ERBB2* mutation.

163

164 Clinical and NGS mutation data were integrated with clinicopathological features including
165 histological subtype, lymph node (LN) status and tumor size and grade. Normalized gene

166 expression data were publicly available for METABRIC (Illumina HT12 microarray) and TCGA
 167 (RNA-seq) datasets. Baseline clinicopathological characteristics are summarized in Table 1.
 168

		METABRIC (N=702)		TCGA (N=330)		MSK-IMPACT (N=548)		TOTAL (N=1580)	
		N	%	N	%	N	%	N	%
Histology*	ILC	76	10.8	100	30.3	103	18.8	279	17.7
	IDC	626	89.2	230	69.7	445	81.2	1301	82.3
Age	<50 yrs	125	17.8	87	26.4	190	34.7	402	25.4
	≥50 yrs	577	82.2	243	73.6	358	65.3	1178	74.6
Menopause	Pre-	125	17.8	90	27.2	234	42.7	449	28.5
	Post-	577	82.2	219	66.4	309	56.4	1105	69.9
	unknown	0	0	21	6.4	5	0.9	26	1.6
Stage*	I	238	33.9	67	49.3	270	49.3	575	36.4
	II	418	59.5	188	34.3	188	34.3	794	50.3
	III	46	6.6	75	16.4	90	16.4	211	13.4
Tumor size	<20mm	215	30.6	96	29.1	330	60.2	641	40.6
	≥20mm	487	69.4	234	70.9	218	39.8	939	59.4
Tumor grade	1	61	8.7	49	14.8	51	9.3	161	10.2
	2	344	49.0	190	57.6	191	34.9	725	45.9
	3	268	38.2	81	24.5	288	52.6	637	40.3
	unknown	29	4.1	10	3.0	18	3.3	57	3.6
Follow-up	<5 yrs	137	19.5	261	82.8	454	82.8	852	53.9
	5-10 yrs	195	27.8	61	11.7	64	11.7	320	20.3
	≥10 yrs	370	52.7	8	5.5	30	5.5	408	25.8
Status	Alive**	310	44.2	299	90.1	494	90.1	1103	69.8
	Deceased	392	55.8	31	9.9	54	9.9	477	20.3

169 **Table 1:** Baseline clinicopathological characteristics of the combined cohort. *Cases of ILC/IDC
170 histology, stage I-III, ER+ and HER2- status with clinical outcome and mutational data were
171 selected *via* CBioportal. **at last follow-up.

172

173 **Statistical analysis**

174 For analysis of binary somatic mutation status, combined cohort analysis was performed.

175 Enrichment for cases with mutations in candidate genes (*ERBB2/3mut*) by histological
176 subtype (ILC vs. IDC) was determined by χ^2 test for association between categorical variables.

177 For ILC and IDC separately, Kaplan-Meier (KM) survival curves stratified by mutation status
178 were compared using Logrank and generalized Wilcoxon tests. Multivariate analysis of OS was
179 performed using a Cox-regression model. Covariates included tumor grade, size (<20mm or
180 \geq 20mm) and LN status (positive or negative). Tumor grade was classified as low (grade 1-2)
181 or high (grade 3).

182

183 To derive a novel gene expression signature of HER2 activity that accounted for the effect of
184 potentially targetable *ERBB2mut* in *ERBB2* non-amplified tumors, we applied a weighted
185 average difference (WAD) method to gene expression data in cases from the METABRIC 2012
186 ($N=1,980$) and TCGA 2015 ($N=817$) cohorts [25, 28, 29]. Gene expression in *ERBB2mut* cases
187 ($N=38$, selected by *ERBB2* non-amplified status and patient age>50) was compared with that
188 in a similar number of *ERBB2* wild-type cases ($N=79$, selected by *ERBB2* non-amplified status,
189 grade>1, stage>I, patient age>50). This was repeated for onc*ERBB2mut* cases ($N=23$) using
190 the same comparator and selection criteria. To incorporate the effect of HER2 activity *via*
191 *ERBB2* amplification, the overlap of differentially expressed genes DEGs shared by both
192 comparisons (*ERBB2mut* and onc*ERBB2mut* vs. *ERBB2* wild-type) with DEGs from a further

193 comparison of *ERBB2* amplified ($N=247$) versus non-amplified ($N=1,733$) cases in METABRIC
194 was calculated. Finally, to incorporate the downstream phenotype (HER2 status), the overlap
195 of this list with DEGs from a comparison of clinical HER2+ versus HER2- cases in TCGA was
196 calculated.

197

198 Multiple gene expression signatures of HER2 activity have been derived using cell line models
199 and patient tumors [30-33]. We compared our novel gene signature with the HER2 activity
200 signature established by Desmedt *et al* [31] with respect to its ability to detect potentially
201 targetable *ERBB2*mut cases in our ILC/IDC dataset. This was achieved by multivariate
202 regression modeling of response to neratinib for each gene signature using breast cancer cell
203 line pharmacogenomic data from the BROAD Institute, accessed online *via* the CellMinerCDB
204 portal [34, 35]. The Pearson coefficient for each significantly correlated signature gene was
205 used to calculate normalized signature scores for each METABRIC case in the current dataset.
206 To validate the prognostic effect of *ERBB2* in ILC, cases were then stratified by gene signature
207 score (upper vs. lower quartiles) and the signatures compared by histological subtype using a
208 Cox regression model of 10-year OS.

209

210

211

212

213

214

215

216 **Results**

217 **Clinicopathological landscape of the combined cohort**

218 To detect prognostic effects of low frequency mutations and add to the existing body of
219 literature on ILC-specific mutational drivers, a combined cohort of three public datasets was
220 collated. We first evaluated the implications of combining cases of ILC and IDC from
221 potentially disparate datasets.

222

223 In the combined cohort, long-term follow-up of patients is dominated by the largest dataset
224 (METABRIC, N=702: mean follow-up 133 months) whilst early events are enriched by the two
225 smaller datasets (TCGA and MSK, N=878: mean follow-up 33 months). Compared to the
226 smaller datasets with shorter follow-up (TCGA and MSK), patients in METABRIC were more
227 likely to be over 50 years of age and have T1 tumours (<20mm diameter), but no significant
228 difference in grade or LN status was found (see Table 2). As the principle skew in the combined
229 dataset is towards longer follow-up in METABRIC cases, OS analysis was limited to 10 years –
230 thus providing a clinically meaningful endpoint for all patients, irrespective of age.

231

232

233

234

235

236

237

238

		METABRIC (N=702)		TCGA AND MSK (N=878)		χ^2 -test
		N	%	N	%	
Age	<50 years	125	17.8	277	31.5	$p < 0.00001$
	≥ 50 years	577	82.2	601	68.5	
Tumor size	<20mm	215	30.6	426	48.5	$p < 0.00001$
	≥ 20 mm	487	69.4	452	51.5	
Tumor grade	1-2	405	60.2	481	56.6	$p = 0.158$
	3	268	39.8	369	43.4	
LN status	Negative	370	53.9	489	55.8	$p = 0.457$
	Positive	316	46.1	387	44.2	

239 **Table 2:** Comparison of long vs. short follow-up cohorts: METABRIC (largest dataset, long
 240 follow-up) vs. TCGA and MSK (combined smaller datasets, short follow-up). Significant
 241 difference was found with respect to age and tumor size, but not tumor grade or LN status.
 242 Table excludes “unknown” cases for each variable.

243

244 ***ERBB2* mutations are enriched in primary ILC and cluster in the HER2 tyrosine kinase domain**

245 We next assessed the prevalence of *ERBB2/3*mut in our dataset of ILC and IDC (N=1,580).
 246 Overall prevalence of *ERBB2*mut was 2.2% (N=34). *ERBB2*mut was enriched in ILC, with
 247 prevalence of 5.7% (N=16) vs. 1.4% in IDC (N=18) ($p < 0.0001$). In contrast, prevalence of
 248 *ERBB3*mut was lower (1.6% overall, N=26), and there was no enrichment of *ERBB3*mut in ILC
 249 vs. IDC (1.1%, N=3 vs. 1.8%, N=23; $p = 0.604$). Due to the small number of cases in ILC, further
 250 analysis of *ERBB3*mut was not performed.

251

252 In ILC, *ERBB2*mut clustered in the tyrosine kinase domain of HER2 (15 out of 16 cases; 93.8%).
 253 Of these, the majority have been characterized as oncogenic and potentially targetable using

254 neratinib (onc*ERBB2*mut, 11 out of 15 kinase domain mutations in ILC; 73.3%). In IDC, all
255 kinase domain mutations were known oncogenic ($N=8$) and non-characterized mutations
256 were distributed evenly across protein domains (see Figure 1). The most frequently occurring
257 *ERBB3*mut coded for E928G, which lies in the tyrosine kinase domain of HER3 ($N=7$ out of 26;
258 27%).

259

260 ***ERBB2* mutation is an adverse prognostic indicator of survival for patients with primary ILC**

261 Using our meta-cohort of 1,580 cases, we next assessed *ERBB2*mut as a prognostic marker of
262 OS in ILC ($N=279$) and IDC ($N=1,301$). Median duration of patient follow-up was 50 months
263 (range 0 – 351 months). In patients with ILC, median OS was significantly shorter if tumors
264 were *ERBB2*mut positive (inclusive of oncogenic and uncharacterized mutations) *versus*
265 *ERBB2*mut negative (66 vs. 211 months, $p=0.0001$). In contrast, there was no significant
266 difference in OS for *ERBB2*mut cases of IDC (159 vs. 166 months, $p=0.733$) (see Figure 2).

267

268 *ERBB3*mut status was not a significant prognostic indicator of OS in cases of IDC (HR=1.46,
269 95% CI 0.65 – 3.28; $p=0.359$). There were no events (patient deaths) at ten years of follow-up
270 in the three cases of *ERBB3*mut ILC.

271

272 **Targetable *ERBB2* mutation status is an independent adverse prognostic marker of 10-year** 273 **overall survival in ER+, *ERBB2* non-amplified ILC**

274 To test the effect of therapeutically actionable *ERBB2* mutations on a clinically relevant
275 endpoint, we stratified 10-year OS by onc*ERBB2*mut status. In ER+, *ERBB2* non-amplified ILC,
276 onc*ERBB2*mut status was prognostic of 10-year OS independently of LN status, tumor grade
277 and size (HR 3.65, 95% CI 1.21 – 11.00; $p=0.021$, see Figure 3).

278

279 Unselected *ERBB2*mut status (including oncogenic and uncharacterized mutations) was also
280 adversely prognostic of 10-year OS in univariate analysis (HR 3.66, 95% CI 1.54 – 8.73;
281 $p=0.003$), but in contrast to characterized *oncERBB2*mut this prognostic effect was not
282 independent of LN stage, tumor grade and / or size.

283

284 **A novel gene signature of targetable HER2 activity using *ERBB2*mut status**

285 Since existing gene signatures of HER activity derive from HER2 status (by IHC and/or ISH) we
286 generated a novel gene signature incorporating DEGs in *ERBB2*mut and *oncERBB2*mut cases,
287 as described in the Methods and outlined in Figure 4. A list of up and down-regulated *ERBB2*
288 “mutant” DEGs ($N=20$) was generated by combining the overlap between DEGs for METABRIC
289 *ERBB2* amplified versus non-amplified, *ERBB2* mutated versus wild-type (*ERBB2*mut and
290 *oncERBB2*mut separately) and TCGA HER2+ versus HER2- (Figure 4A, and see Additional file 1
291 for supplemental tables S1-3 with all gene lists).

292

293 As shown in Figure 4B, *ERBB2*mut and *oncERBB2*mut clustered in the upper quartile of the
294 novel gene signature score in cases of ER+, HER2- ILC/IDC from METABRIC. In contrast,
295 *ERBB2*mut and *oncERBB2*mut cases did not cluster when scored by the established gene
296 signature of HER2 pathway activation but were evenly distributed across the cohort. Both the
297 novel and established HER2 pathway gene signatures were predictive of response to neratinib
298 in multivariate analysis using a pharmacogenomic breast cancer cell line model (Pearson
299 $R=0.9$, $p<1e^{-14}$ for both; see Additional file 2 for the original plots from the CellMinerCDB
300 online portal). The regression model used incorporates drug response and gene expression
301 data from $N=36$ breast cancer cell lines [35, 36]. Our findings using this model verify that the

302 novel signature derived from DEGs in *ERBB2*mut versus wild-type cases reflects oncogenic
303 HER2 activity due to *ERBB2*mut. This suggests that a gene signature may have value in
304 predicting the presence of *ERBB2*mut in HER2- breast cancer. Finally, Figure 4C demonstrates
305 that the novel gene signature score (stratified into upper vs. lower quartiles) was adversely
306 associated with 10-year OS (HR=2.3, 95% CI 1.04–5.05; $p=0.040$), thus providing mRNA level
307 validation of *ERBB2*mut as an adverse prognostic marker.

308

309

310 Discussion

311 In this study, we mined clinical and NGS data from the largest clinical cohorts with data in the
312 public domain to test whether there was an ILC-specific association of *ERBB2* mutations with
313 OS. In our meta-cohort of ER+, *ERBB2* non-amplified cases of ILC and IDC, we found that
314 *ERBB2* mutations are enriched ILC, cluster in the functional kinase domain of HER2, and
315 robustly associate with adverse clinical outcomes – independently of known prognostic
316 clinicopathological features including LN status and tumor grade. In contrast, there was no
317 ILC-specific enrichment of *ERBB3*mut.

318

319 Compared to the largest previous study with data on *ERBB2* mutational status in primary ILC
320 ($N=371$) [24], the current study found higher frequency of *ERBB2*mut (5.7% vs. 4.3%) but
321 lower frequency of *ERBB3*mut (1.1% vs. 3.5%). Hazard proportionality in the study by
322 Desmedt *et al* 2016 suggested that *ERBB2* mutational status had a time-dependent effect
323 associated with short-term risk of breast cancer relapse [24]. The current study adds data on
324 the prognostic effect of specific *ERBB2* mutations that were previously shown in cell line, *in*
325 *vitro* and xenograft studies to respond to existing clinical compounds such as trastuzumab or

326 neratinib [13]. We demonstrated that the status of these therapeutically targetable *ERBB2*
327 mutations is an independent adverse prognostic marker of overall survival in a large cohort
328 of patients with ILC.

329
330 Findings from the current study imply that targeted sequencing of ILC tumors for *ERBB2*
331 mutations may be an actionable and viable strategy to improve outcomes for patients with
332 primary ILC by providing a biomarker for HER-targeted therapy in the adjuvant setting. Based
333 on worldwide ILC incidence of 250,000 [1-4], 10% rate of ER- and/or HER2- status in ILC, and
334 5% rate of potentially targetable *ERBB2*mut in ER+, *ERBB2* non-amplified ILC, we
335 conservatively estimate that 11,250 additional ILC patients per year who have *ERBB2*-
336 mutated tumors may benefit from HER2-targeting therapy.

337
338 Existing clinical studies of HER2-targeting therapy are underway in advanced breast cancer.
339 For example, neratinib (alone or in combination with fulvestrant) was tested in phase II study
340 for patients with advanced, *ERBB2* non-amplified, *ERBB2*mut breast cancer (clinical trial
341 registration number NCT01670877). Of the 16 patients who had evaluable tumor responses,
342 *N*=2 had partial response by RECIST criteria (13%). Out of 14 patients with tumors harboring
343 *oncERBB2*mut, *N*=5 (36%) derived clinical benefit, defined as stable disease or
344 partial/complete response by RECIST criteria [37]. Given that the patients in this trial had
345 received multiple prior lines of treatment, it is conceivable that patients with *ERBB2* non-
346 amplified primary tumors harboring *oncERBB2*mut will also derive benefit from (neo)adjuvant
347 HER2-targeted therapy such as neratinib.

348

349 In the current cohort of primary ER+, HER2- ILC and IDC, *ERBB2* mutations in ILC tumors
350 clustered almost exclusively in the active HER2 tyrosine kinase domain ($N=15$ of 16). Our
351 retrospective clinical outcome data indicates an ILC-specific adverse survival association for
352 patients with *ERBB2* non-amplified tumors harboring potentially targetable *ERBB2* mutations.
353 This primary finding was validated by existing mRNA data *via* a novel gene signature of HER2
354 activity, which linked *ERBB2*mut with response to neratinib. Subject to independent external
355 validation, a gene signature of HER2 activity could have clinical application as a biomarker of
356 response to HER2-targeted therapy and may be more cost-effective than targeted NGS in the
357 detection of *ERBB2*mut in primary ILC.

358

359 Taken together, these new findings suggest that focusing early phase studies of HER2-
360 inhibition (e.g. with neratinib) in patients with *ERBB2* non-amplified, *ERBB2*mut primary ILC
361 may be an effective strategy to demonstrate feasibility and clinical benefit in the
362 (neo)adjuvant setting.

363

364 Limitations of the current study include cohort size, quality of clinical outcome data and bias
365 in long-term follow-up towards cases from the largest dataset (METABRIC). Statistical power
366 is limited by the small number of patient deaths in cases of ILC and IDC with *ERBB2*mut ($N=7$
367 and 8 respectively). However, the clear deviation of the KM plot of OS in ILC stratified by
368 *ERBB2*mut (Figure 2) is reflected in a low type I error of one in 10,000 ($p=0.0001$). In contrast,
369 there was no ILC-specific enrichment of *ERBB3*mut in our database. However, in IDC there
370 was enrichment of the known oncogenic HER3 kinase domain mutation, E928G. To
371 demonstrate the prognostic effect of mutations affecting HER3, larger datasets will be
372 required.

373

374 The effect of incomplete clinical outcome data on the primary endpoint (OS) is difficult to
375 quantify but is accounted for statistically in Cox regression modelling by censoring patients at
376 last follow-up. Patients in METABRIC presented at older age, which is associated with more
377 indolent, less biologically aggressive breast cancer [38]. This implies an underestimation of
378 the adverse impact of *ERBB2* mutations on 10-year survival outcomes in the current study.
379 Since the METABRIC dataset did not have complete HER2 IHC or reverse-phase protein array
380 data, it was not possible to associate activating *ERBB2*mut directly with downstream HER2
381 protein expression. Instead, we generated a gene signature incorporating clinical HER2 status
382 to validate our findings.

383

384 **Conclusions**

385 Targetable kinase domain *ERBB2* mutations are enriched in primary ILC and their detection
386 by targeted sequencing or validated gene signature surrogate may provide an actionable
387 strategy to improve patient outcomes. Biomarker-led clinical trials of adjuvant HER2-targeted
388 therapy to treat breast tumors with activating *ERBB2* mutations are warranted in HER2-
389 primary ILC.

390

391 **List of abbreviations**

392	CI	Confidence interval
393	DEG	Differentially expressed gene
394	EGF	Epidermal growth factor
395	ER	Estrogen receptor alpha

396	ERBB2/3mut	<i>ERBB2/3</i> mutant
397	HER2/3	Human epidermal growth factor receptor 2 / 3
398	HR	Hazard ratio
399	IDC	Invasive ductal carcinoma of the breast
400	IHC	Immunohistochemistry
401	ILC	Invasive lobular carcinoma of the breast
402	ISH	<i>In situ</i> hybridization
403	KM	Kaplein-Meier
404	LN	Lymph node
405	NGS	Next generation sequencing
406	oncERBB2mut	Oncogenic <i>ERBB2</i> mutant
407	OS	Overall survival
408	RECIST	Response evaluation criteria in solid tumors
409	WAD	Weighted average difference

410

411 **Declarations**

412 **Ethics approval and consent to participate**

413 Ethics approvals and consent for each cohort (METABRIC, TCGA and MSK-IMPACT) was
414 obtained in the original studies.

415

416 **Consent for publication**

417 Not applicable

418

419 **Availability of data and materials**

420 The datasets generated and/or analysed during the current study are available *via* the
421 CBioportal online portal (<http://www.cbioportal.org>) and in the Genome-Phenome Archives
422 EGAS00000000098 and phs000178 (<https://ega-archive.org/studies/EGAS00000000098> and
423 <https://ega-archive.org/studies/phs000178>).

424

425 **Competing interests**

426 The authors declare that they have no competing interests.

427

428 **Funding**

429 The corresponding author (SJJ) is funded by the Wellcome Trust (Royal Academy of Medical
430 Sciences grant number AAM 127669) and the National Institute for Health Research UK. Dr
431 Oesterreich's work on ILC is supported by the Breast Cancer Research Foundation and a
432 Komen Scholar awards (SAC160073).

433

434 **Authors' contributions**

435 SK, DMH, WZ, SO and SJJ conceived, designed and supervised the study. SK, MA, CM, KB, NPM
436 and SJJ analysed and interpreted the data. SK, WZ, SO and SJJ prepared the manuscript. ARG,
437 IOE and EAR had supervisory and administrative roles. SJ, TF and KS had consultative roles in
438 study design. All authors were involved in manuscript preparation and reviewing for
439 submission, gave their final approval and agreed to be accountable for all aspects of the work.

440

441 **Acknowledgements**

442 Not applicable

443 **Figure legends**

444 **Figure 1:** Rate of *ERBB2* and *ERBB3* mutations and their spatial distribution on HER2/3 in ILC
445 and IDC. *ERBB2*mut were found to be **(A)** enriched in ILC (yellow bar) vs. IDC (blue bar) and
446 **(B)** clustered in the tyrosine kinase domain of HER2; **(C)** *ERBB3*mut occurred at lower
447 frequency, with the high-frequency outlier in IDC coding for known oncogenic HER3 kinase
448 domain alteration E928G ($N=6$). Y-axes show the number of cases harboring at least one
449 *ERBB2/3* mutation at a specific amino acid (aa) of HER2/3, shown along the x-axes. Yellow-
450 filled circles indicate *oncERBB2*mut. Extracellular domains of HER2/3: Receptor L, Furin-like
451 and Growth Factor Receptor IV; intracellular: tyrosine protein kinase. $*p<0.001$; n/s = not
452 significant.

453
454 **Figure 2:** OS by *ERBB2* mutational status in ILC (left) and IDC (right). Gray line indicates *ERBB2*
455 wild-type cases; blue line indicates cases with at least one *ERBB2*mut.

456
457 **Figure 3.** Univariate and multivariate analyses of 10-year OS in $N=279$ cases of ER+, *ERBB2*
458 non-amplified ILC. Gray square dot indicates hazard ratio (HR) in univariate analysis and the
459 gray bar indicates the 95% confidence interval (CI). Significant prognostic variables in
460 univariate analyses (where 95% CI does not span HR=1) are included in multivariate analysis,
461 shown in blue. For each variable, cases with unknown values are excluded from the analysis.

462
463 **Figure 4:** A novel gene signature of HER2 activity incorporating *ERBB2*mut, *ERBB2*
464 amplification and clinical HER2 status. **(A)** Generation of a 20-gene signature of HER2 activity.
465 The upper Venn diagrams show the overlap between the top 500 DEGs by WAD score for
466 METABRIC amplified vs. non-amplified, *ERBB2* mutant vs. wild-type, and TCGA HER2+ vs.

467 HER2-. Up-regulated DEGs are shaded blue and down-regulated DEGs yellow. *ERBB2*mut and
468 *oncERBB2*mut vs. wild-type are analysed separately, and the overlap combined in the lower
469 Venn diagrams. **(B)** Comparison with an established 24-gene signature of HER pathway
470 activation [31], using a gene signature (genesig) score derived from multivariate analysis of
471 response to neratinib in breast cancer cell lines. Cases with *ERBB2*mut (gray lines) and
472 *oncERBB2*mut (blue lines) clustered in the upper quartile of normalized genesig scores for the
473 novel signature but not the established signature. **(C)** 10-year OS analysis of cases in the
474 current study stratified by histological subtype and novel genesig score (upper vs. lower
475 quartiles) indicates that *ERBB2*mut-associated DEGs are prognostic in ILC but not IDC. GE =
476 geneset enrichment.

477

478 **Additional files**

479 **Additional file 1.** Supplemental tables: **(S1)** list of genes in the novel ($N=20$) and established
480 (Desmedt *et al* [31], $N=24$) gene signatures of HER2 activity; **(S2 and S3)** top 500 DEGs by WAD
481 score stratified into genes that are up-regulated (S2) and down-regulated (S3) for METABRIC
482 *ERBB2* amplified vs. non-amplified (*ERBB2*amp), METABRIC *ERBB2*mut vs. wild-type
483 (*ERBB2*mut) and *oncERBB2*mut vs. wild-type (*oncERBB2*mut), and TCGA HER2+ vs. HER2-
484 (TCGA HER2+).

485

486 **Additional file 2.** Supplemental figure showing the correlation between the observed
487 response of breast cancer cell lines ($N=36$) to neratinib and the response predicted by
488 expression of (A) novel ($N=20$) and (B) established (Desmedt *et al* [31], $N=24$) gene panels.
489 Plots were generated using CellMinerCDB online portal and cell line data from the BROAD
490 Institute [34, 35].

491 **References**

- 492 1. Martinez V, Azzopardi JG. Invasive lobular carcinoma of the breast: incidence and variants.
493 Histopathology. 1979;3(6):467-88.
- 494 2. Li CI, Anderson BO, Daling JR, Moe RE. Trends in incidence rates of invasive lobular and ductal
495 breast carcinoma. JAMA. 2003;289(11):1421-4.
- 496 3. Li CI, Daling JR. Changes in breast cancer incidence rates in the United States by histologic
497 subtype and race/ethnicity, 1995 to 2004. Cancer Epidemiol Biomarkers Prev. 2007;16(12):2773-80.
- 498 4. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, et al. Estimating
499 the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer.
500 2019;144(8):1941-53.
- 501 5. Pestalozzi BC, Zahrieh D, Mallon E, Gusterson BA, Price KN, Gelber RD, et al. Distinct clinical
502 and prognostic features of infiltrating lobular carcinoma of the breast: combined results of 15
503 International Breast Cancer Study Group clinical trials. J Clin Oncol. 2008;26(18):3006-14.
- 504 6. Rakha EA, El-Sayed ME, Powe DG, Green AR, Habashy H, Grainge MJ, et al. Invasive lobular
505 carcinoma of the breast: response to hormonal therapy and outcomes. Eur J Cancer. 2008;44(1):73-
506 83.
- 507 7. Metzger Filho O, Giobbie-Hurder A, Mallon E, Gusterson B, Viale G, Winer EP, et al. Relative
508 Effectiveness of Letrozole Compared With Tamoxifen for Patients With Lobular Carcinoma in the BIG
509 1-98 Trial. J Clin Oncol. 2015;33(25):2772-9.
- 510 8. Sikora MJ, Cooper KL, Bahreini A, Luthra S, Wang G, Chandran UR, et al. Invasive lobular
511 carcinoma cell lines are characterized by unique estrogen-mediated gene expression patterns and
512 altered tamoxifen response. Cancer Res. 2014;74(5):1463-74.
- 513 9. NICE. Early and locally advanced breast cancer: diagnosis and management. Online: National
514 Institute for Health and Care Excellence; 2018 [NICE Pathway]. Available from:
515 <https://www.nice.org.uk/guidance/ng101>.

- 516 10. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, et al. Early breast
517 cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up dagger. *Ann Oncol.*
518 2019;30(8):1194-220.
- 519 11. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human
520 Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical
521 Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J Clin Oncol.*
522 2018;36(20):2105-22.
- 523 12. Denduluri N, Chavez-MacGregor M, Telli ML, Eisen A, Graff SL, Hassett MJ, et al. Selection of
524 Optimal Adjuvant Chemotherapy and Targeted Therapy for Early Breast Cancer: ASCO Clinical Practice
525 Guideline Focused Update. *J Clin Oncol.* 2018;36(23):2433-43.
- 526 13. Bose R, Kavuri SM, Searleman AC, Shen W, Shen D, Koboldt DC, et al. Activating HER2
527 mutations in HER2 gene amplification negative breast cancer. *Cancer Discov.* 2013;3(2):224-37.
- 528 14. Hanker A, Koch J, Ye D, Sliwoski G, Sheehan J, Kinch L, et al. Abstract PD3-05: Co-occurring
529 gain-of-function mutations in HER2 and HER3 cooperate to enhance HER2/HER3 binding, HER-
530 dependent signaling, and breast cancer growth. *Cancer Research.* 2019;79(4 Supplement):PD3-05-
531 PD3-.
- 532 15. Eroglu Z, Tagawa T, Somlo G. Human epidermal growth factor receptor family-targeted
533 therapies in the treatment of HER2-overexpressing breast cancer. *Oncologist.* 2014;19(2):135-50.
- 534 16. Hynes NE, MacDonald G. ErbB receptors and signaling pathways in cancer. *Curr Opin Cell Biol.*
535 2009;21(2):177-84.
- 536 17. Junttila TT, Akita RW, Parsons K, Fields C, Lewis Phillips GD, Friedman LS, et al. Ligand-
537 independent HER2/HER3/PI3K complex is disrupted by trastuzumab and is effectively inhibited by the
538 PI3K inhibitor GDC-0941. *Cancer Cell.* 2009;15(5):429-40.
- 539 18. Mishra R, Patel H, Alanazi S, Yuan L, Garrett JT. HER3 signaling and targeted therapy in cancer.
540 *Oncol Rev.* 2018;12(1):355.

- 541 19. Jaiswal BS, Kljavin NM, Stawiski EW, Chan E, Parikh C, Durinck S, et al. Oncogenic ERBB3
542 mutations in human cancers. *Cancer Cell*. 2013;23(5):603-17.
- 543 20. Ping Z, Siegal GP, Harada S, Eltoum IE, Youssef M, Shen T, et al. ERBB2 mutation is associated
544 with a worse prognosis in patients with CDH1 altered invasive lobular cancer of the breast. *Oncotarget*.
545 2016;7(49):80655-63.
- 546 21. Ross JS, Wang K, Sheehan CE, Boguniewicz AB, Otto G, Downing SR, et al. Relapsed classic E-
547 cadherin (CDH1)-mutated invasive lobular breast cancer shows a high frequency of HER2 (ERBB2) gene
548 mutations. *Clin Cancer Res*. 2013;19(10):2668-76.
- 549 22. Griffith OL, Spies NC, Anurag M, Griffith M, Luo J, Tu D, et al. The prognostic effects of somatic
550 mutations in ER-positive breast cancer. *Nat Commun*. 2018;9(1):3476.
- 551 23. Ross JS, Gay LM, Wang K, Ali SM, Chumsri S, Elvin JA, et al. Nonamplification ERBB2 genomic
552 alterations in 5605 cases of recurrent and metastatic breast cancer: An emerging opportunity for anti-
553 HER2 targeted therapies. *Cancer*. 2016;122(17):2654-62.
- 554 24. Desmedt C, Zoppoli G, Gudem G, Pruneri G, Larsimont D, Fornili M, et al. Genomic
555 Characterization of Primary Invasive Lobular Breast Cancer. *J Clin Oncol*. 2016;34(16):1872-81.
- 556 25. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive
557 Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*. 2015;163(2):506-19.
- 558 26. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, et al. The somatic
559 mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat*
560 *Commun*. 2016;7:11479.
- 561 27. Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, et al. The Genomic Landscape of
562 Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell*. 2018;34(3):427-38 e6.
- 563 28. Kadota K, Nakai Y, Shimizu K. A weighted average difference method for detecting
564 differentially expressed genes from microarray data. *Algorithms Mol Biol*. 2008;3:8.

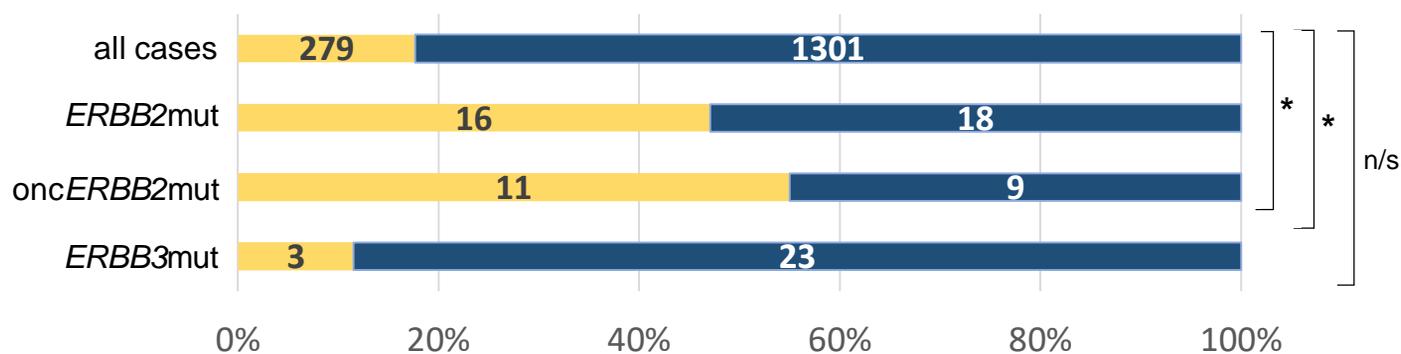
- 565 29. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and
566 transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*.
567 2012;486(7403):346-52.
- 568 30. Ferrari A, Vincent-Salomon A, Pivot X, Sertier AS, Thomas E, Tonon L, et al. A whole-genome
569 sequence and transcriptome perspective on HER2-positive breast cancers. *Nat Commun*.
570 2016;7:12222.
- 571 31. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, et al. Biological
572 processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin*
573 *Cancer Res*. 2008;14(16):5158-65.
- 574 32. Sareyeldin RM, Gupta I, Al-Hashimi I, Al-Thawadi HA, Al Farsi HF, Vranic S, et al. Gene
575 Expression and miRNAs Profiling: Function and Regulation in Human Epidermal Growth Factor
576 Receptor 2 (HER2)-Positive Breast Cancer. *Cancers (Basel)*. 2019;11(5).
- 577 33. Creighton CJ, Hilger AM, Murthy S, Rae JM, Chinnaiyan AM, El-Ashry D. Activation of mitogen-
578 activated protein kinase in estrogen receptor alpha-positive breast cancer cells in vitro induces an in
579 vivo molecular phenotype of estrogen receptor alpha-negative human breast tumors. *Cancer Res*.
580 2006;66(7):3903-11.
- 581 34. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical
582 sensitivity and basal gene expression reveals mechanism of action. *Nat Chem Biol*. 2016;12(2):109-16.
- 583 35. Rajapakse VN, Luna A, Yamade M, Loman L, Varma S, Sunshine M, et al. CellMinerCDB for
584 Integrative Cross-Database Genomics and Pharmacogenomics Analyses of Cancer Cell Lines. *iScience*.
585 2018;10:247-64.
- 586 36. Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, et al. An interactive resource
587 to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*.
588 2013;154(5):1151-61.

589 37. Ma CX, Bose R, Gao F, Freedman RA, Pegram MD, Blackwell K, et al. Phase II trial of neratinib
590 for HER2 mutated, non-amplified metastatic breast cancer (HER2(mut) MBC). J Clin Oncol.
591 2016;34(15).

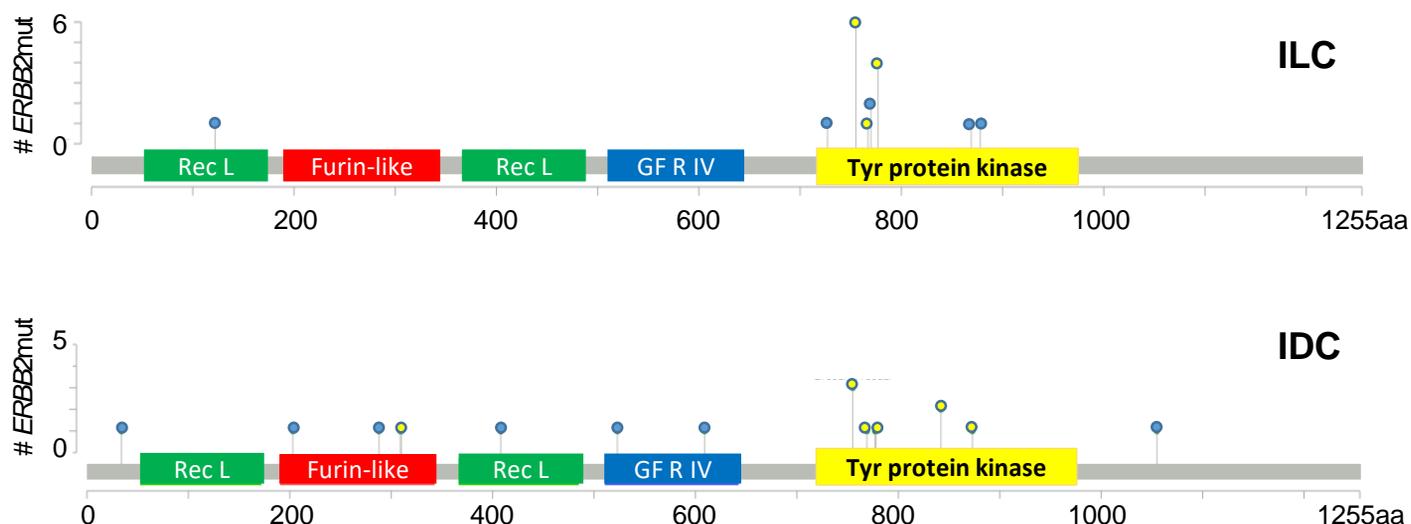
592 38. Syed BM, Green AR, Paish EC, Soria D, Garibaldi J, Morgan L, et al. Biology of primary breast
593 cancer in older women treated by surgery: with correlation with long-term clinical outcome and
594 comparison with their younger counterparts. Br J Cancer. 2013;108(5):1042-51.

595

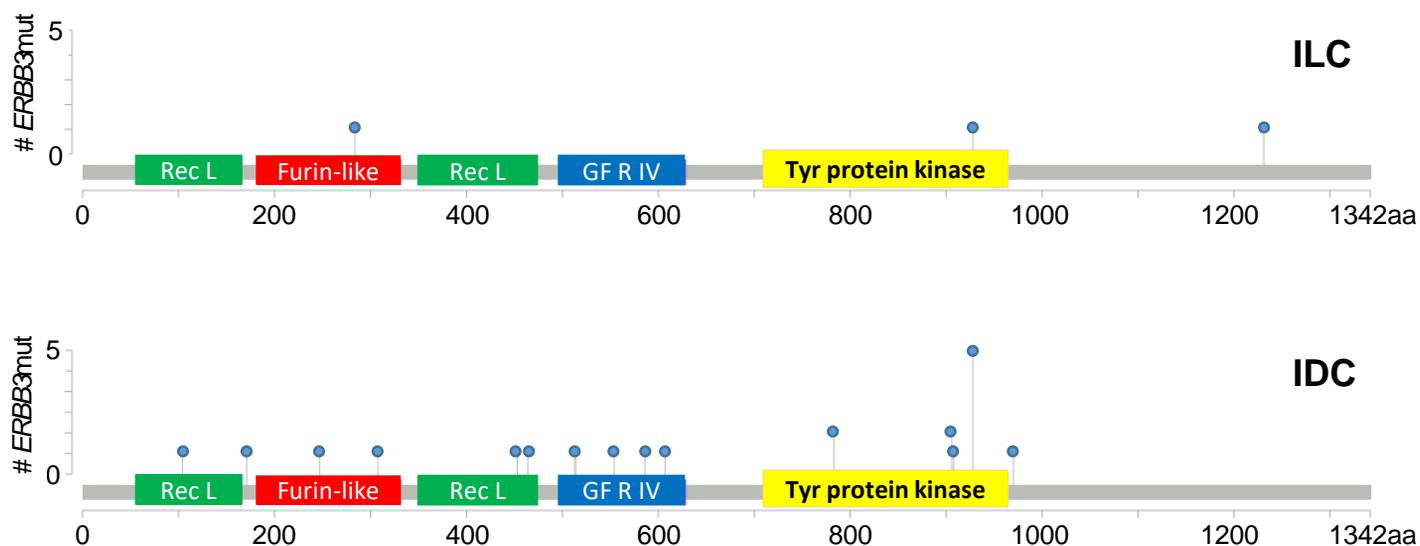
(A) Proportion of cases by histology: ILC IDC



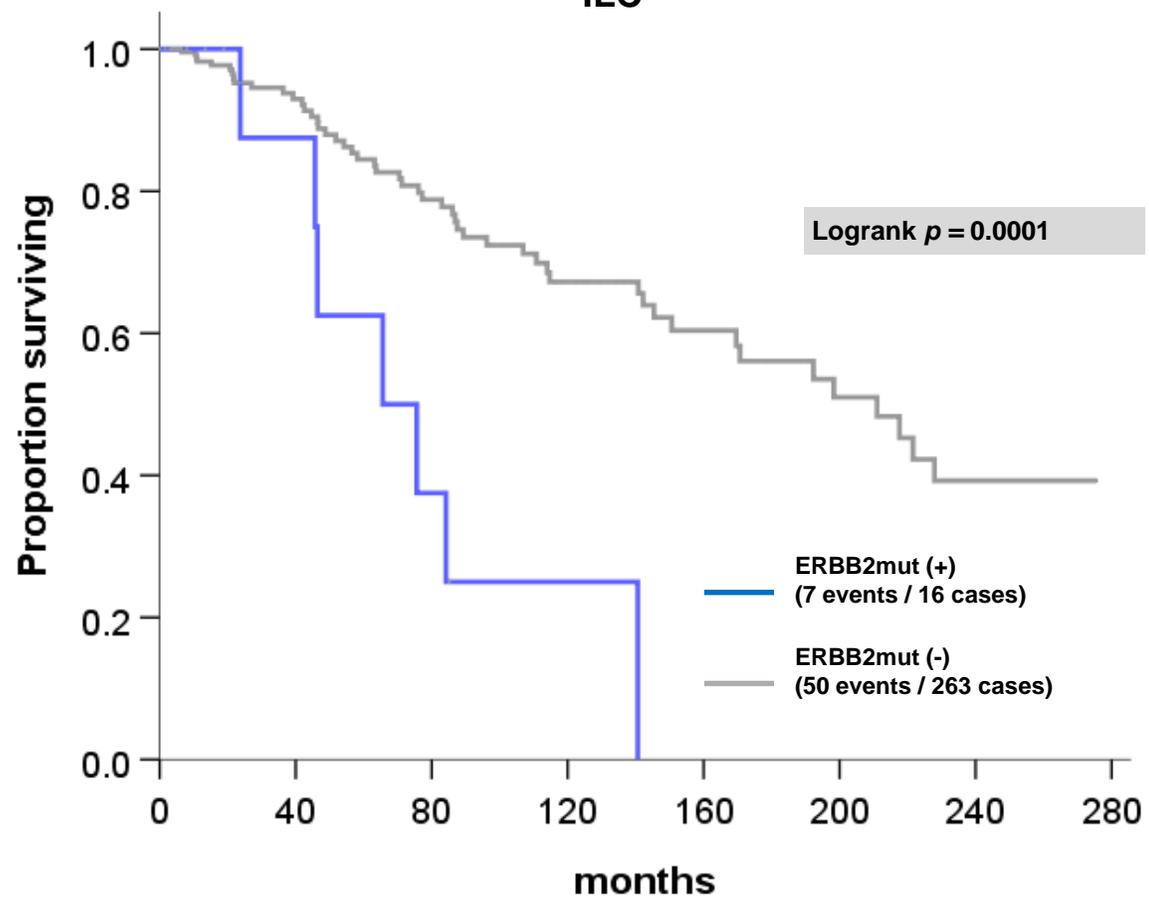
(B) ERBB2mut



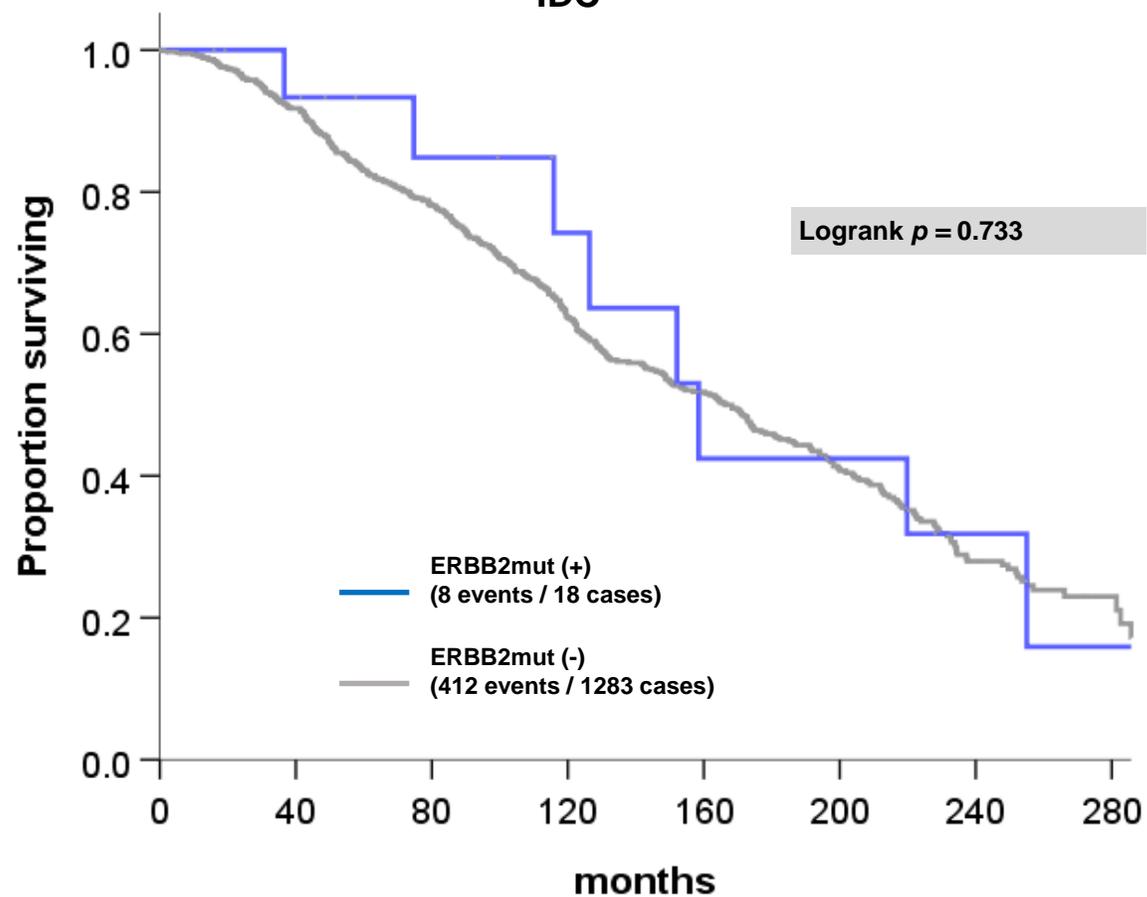
(C) ERBB3mut

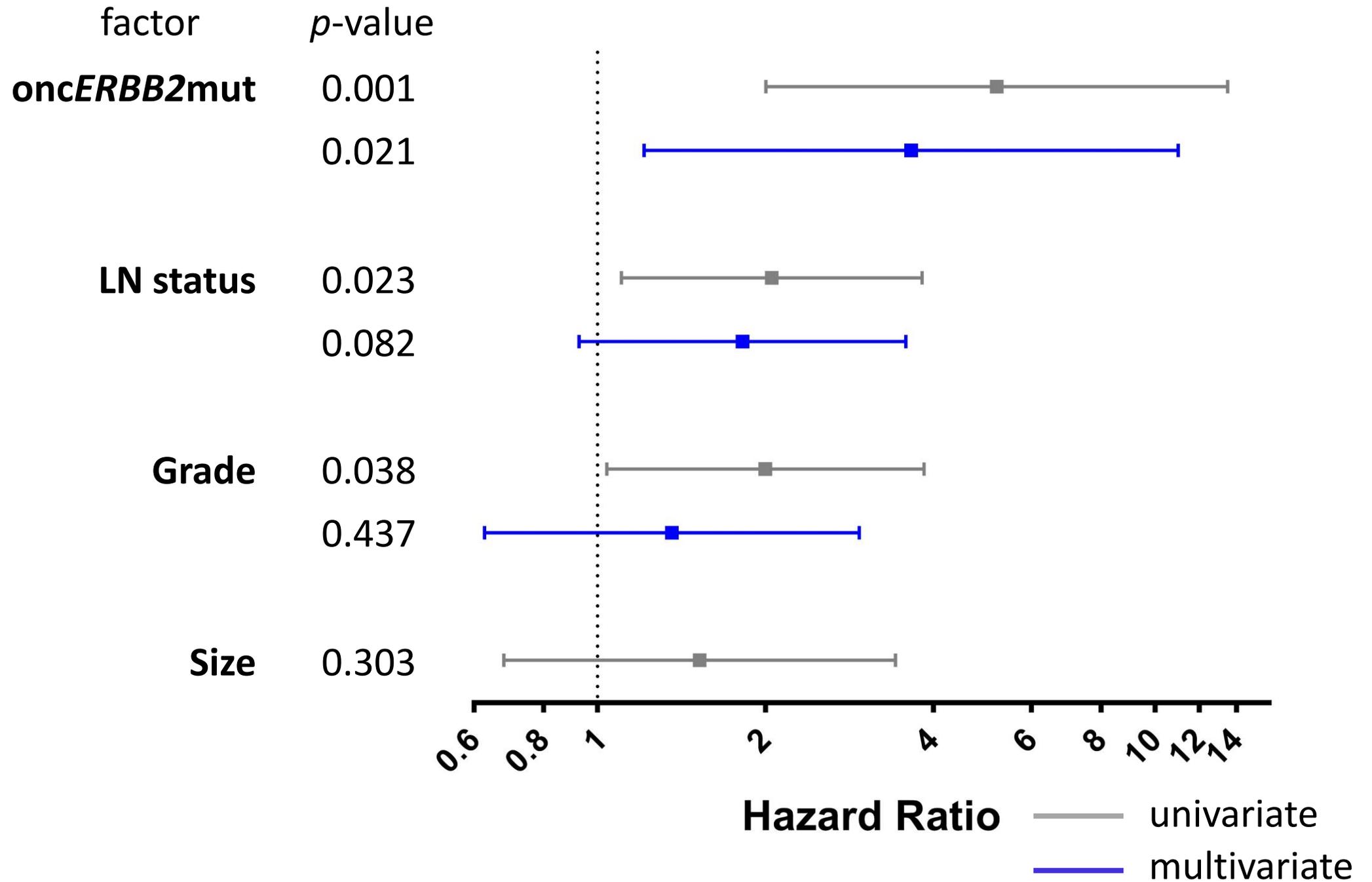


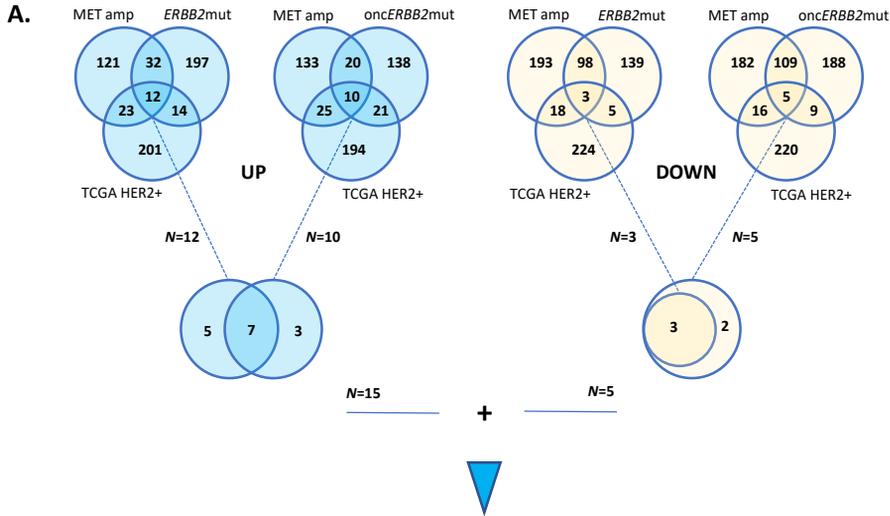
ILC



IDC







Novel *ERBB2mut* gene signature 20

24 Established HER2 pathway activation signature

