



Lessons learned from a candidate gene study investigating aromatase inhibitor treatment outcome in breast cancer



A list of authors and their affiliations appears at the end of the paper

The role of germline genetics in adjuvant aromatase inhibitor (AI) treatment efficacy in ER-positive breast cancer is poorly understood. We employed a two-stage candidate gene approach to examine associations between survival endpoints and common germline variants in 753 endocrine resistance-related genes. For a discovery cohort, we screened the Breast Cancer Association Consortium database ($n \geq 90,000$ cases) and retrieved 2789 AI-treated patients. Cox model-based analysis revealed 125 variants associated with overall, distant relapse-free, and relapse-free survival (p -value $\leq 1E-04$). In validation analysis using five independent cohorts ($n = 8857$), none of the six selected candidates representing major linkage blocks at *CELA2B/CASP9*, *NR112/GSK3B*, *LRP1B*, and *MIR143HG (CARMN)* were validated. We discuss potential reasons for the failed validation and replication of published findings, including study/treatment heterogeneity and other limitations inherent to genomic treatment outcome studies. For the future, we envision prospective longitudinal studies with sufficiently long follow-up and endpoints that reflect the dynamic nature of endocrine resistance.

Breast cancer is the most frequent cancer and leading cause of cancer death in women¹ affecting both pre- and postmenopausal women. Most breast cancers (75%) express the estrogen receptor α (ER α) known to fuel tumor growth through estrogen signaling. Its presence renders early-stage breast cancers (i.e., tumors that have not spread beyond breast and nearby lymph nodes) amenable to endocrine therapy as the key therapeutic approach². Valid adjuvant options include the selective estrogen receptor modulator tamoxifen that competes with estrogen for the binding to ER α to block signaling, and the 3rd generation aromatase inhibitors (AIs) anastrozole, letrozole, and exemestane that remove estrogen from the signaling cascade by blocking its synthesis³. AIs evolved as the standard-of-care for postmenopausal women⁴ and have become an option for premenopausal women together with concurrent ovarian suppression⁵. Their mechanism of action in lowering estrogen levels is particularly important for postmenopausal women, as AIs target the aromatase enzyme (CYP19A1) in peripheral nonglandular tissues (and breast tumor) to block the conversion of testosterone and androstenedione to 17 β -estradiol and estrone, respectively³. Current guidelines recommend 5 to 10 years of treatment to maximize risk-reduction². During this period and beyond, women continue to be at risk of recurrence and death that extends for decades following treatment⁶. Therefore, additional biomarkers for the proper selection of drug treatment regimens are a high research priority.

Inherited germline variants account for the inter-patient variability of pharmacokinetic and pharmacodynamic drug effects. In the case of AIs, putative predictors include polymorphisms at drug metabolizing enzymes such as CYP3A4 and UGT1A4 for anastrozole, CYP2A6 for letrozole, as well as CYP3A4 and UGT2B17 for exemestane, and associations have been reported with regards to systemic AI concentrations, systemic estrogen concentrations, treatment efficacy, and toxicities⁷. However, despite compelling biological and pharmacological rationales, the evidence of linking the magnitude of estrogen suppression and AI efficacy with the patients' constitutional background remains a challenge. For example, comparative analyses of patients with anastrozole and exemestane (MA.27 trial) as well as letrozole treatment (PreFace trial) linked insufficient estrogen suppression with a 2.2-fold increased risk of early breast cancer events (p -value = 5E-04)⁸. A corresponding pharmacogenomic Genome-Wide Association Study (GWAS) identified a common variant (rs6981827) in the human CUB And Sushi multiple domains 1 gene (*CSMD1*, alias *PPP1R24*)⁹ to be associated with the changes in estrogen levels under anastrozole, and a weakly linked expression Quantitative Trait Locus (eQTL) single-nucleotide polymorphism (SNP rs6990851) that was associated with breast cancer-free interval in anastrozole- and exemestane-treated patients. Moreover, the risk of breast cancer recurrence in patients treated with anastrozole and exemestane has been linked to SNPs in the long non-coding RNA *MIR2052HG*¹⁰. Both findings were underpinned by mechanistic studies in

✉ e-mail: reiner.hoppe@ikp-stuttgart.de

that the *CSMD1* genotype correlated with *CYP19A1* expression in an anastrozole-dependent fashion through transcription regulation⁹. In the case of *MIR2052HG*, variant genotypes exhibited increased ER α binding to estrogen response elements relative to wild-type genotypes, thereby downregulating ER α transcription in vitro, which may affect the risk of breast cancer recurrence in women treated with AI¹⁰.

Despite important insights, the identification of relevant AI outcome predictors remains a challenge given the scarcity of suitable study and replication cohorts for early breast cancer. Association studies in the breast cancer endocrine treatment setting are particularly challenging, as the lengthy 5-year treatment and further 5 to 10 year follow-up period during which recurrence or death events may occur, limit the availability of such cohorts. In the current work we performed a two-stage approach, utilizing the database of the Breast Cancer Association Consortium (BCAC) with its comprehensive clinical, epidemiologic, and genetic data set^{11–13} for discovery, and independent clinical cohorts (MA.27, PreFace, KARMA/pKARMA, MCBCS, IKP211) for validation analysis to explore potential associations between common germline variants and survival endpoints in AI-treated patients. The aim of our study was to identify putative pharmacogenetic/-genomic polymorphisms in genes, that based on their suspected role in endocrine resistance and converging/interconnected mechanisms may contribute to AI treatment outcome in early-stage ER-positive breast cancer.

Results

Candidate gene association (CGA) study in AI-treated patients (discovery analysis - stage 1)

Candidate gene association analysis was performed in 2789 AI-treated breast cancer patients selected from the BCAC database (version 12) (Fig. 1). Details on patients, their clinical data and tumor characteristics are given in Table 1. Median follow-up time was 7.2 years, and event rates within a 15-year follow-up period were 14.4% (overall survival, OS), 15.2% (distant relapse-free survival, DRFS), and 16.8% (relapse-free survival, RFS). Analyses were performed using genotype data of 753 selected genes (Supplementary Data 1) related to endocrine resistance and converging interconnected mechanisms. Subgroup analyses included 1335 patients treated with anastrozole (47.9%) and 689 patients treated with letrozole (24.7%). No subgroup analysis was performed for patients treated with exemestane due to low case numbers (Fig. 1).

Primary endpoints OS, DRFS, and RFS. All AI-treated patients: Unadjusted analysis revealed 47, 65, and 54 variants with an association p -value $\leq 1E-04$ for OS, DRFS, and RFS, respectively (Supplementary Data 2; minor-allele frequency (MAF) $\geq 5\%$; 125 non-overlapping variants in total). Their corresponding chromosomal regions are highlighted in the Manhattan plots of Fig. 2a and include the following gene regions: For OS, top associated variants mapped to the *CELA2B/CASP9* ($n = 16$ variants) and *NR1I2/GSK3B* loci ($n = 30$). For DRFS, variants in the *CELA2B/CASP9* region were also among the top hits ($n = 52$), including 37 variants that annotate to the intronic and 5' region of *CASP9*. Other DRFS top hits included variants in *LRP1B* ($n = 8$) and the *CARMN/MIR143HG* locus ($n = 2$). For RFS, top hits comprised variants at *CELA2B/CASP9* ($n = 16$; all among DRFS top hits), *LRP1B* ($n = 31$; including the 8 variants associated with DRFS), and *CARMN/MIR143HG* ($n = 2$) gene regions. Similar results were observed for the top associated variants in analyses adjusted for patient characteristics (age, menopausal status, and family history of breast cancer in a first degree relative), tumor characteristics (size, grade, nodal status, stage, and progesterone receptor status), treatment (AI, adjuvant chemotherapy, trastuzumab, and radiation), and surgery type (Supplementary Data 3).

For the top hits, we scrutinized eQTLs to support the selection of candidate variants to be further pursued in validation analysis. No survival eQTLs were revealed in PanCanQTL analysis, which however was not focused on AI-treated patients. Yet, PanCanQTL and FUMA (Functional Mapping and Annotation) tools identified the top variants of the *CELA2B/*

CASP9 region as cis-eQTLs associated with *CASP9* expression in breast cancer, breast mammary tissue as well as numerous other normal tissues and cancer entities. *GSK3B* top hits were revealed as cis-eQTLs in breast mammary tissue, non-breast tumor entities, and various normal tissues. Furthermore, cis-eQTLs were identified for *LRP1B* and *CARMN/MIR143HG* in non-breast tissues. We then used KM Plotter to investigate whether expression levels of genes identified in the eQTL databases were correlated with OS in ER-positive, endocrine treated breast cancer patients. Here, high *CASP9*, *GSK3B*, and *LRP1B* expression levels were associated with better survival. For *GSK3B*, an opposite effect was observed when patients with chemotherapy were excluded. Moreover, stratification by molecular subtype demonstrated that high *GSK3B* expression was associated with a better outcome in Luminal B but worse outcome in Luminal A patients (Supplementary Fig. 1–3 in the Supplementary Information).

Anastrozole-treated patients: We identified $n = 2$ (OS), $n = 4$ (DRFS), and $n = 8$ (RFS) top associated variants in unadjusted analysis (Supplementary Data 4; p -value $\leq 1E-04$ and MAF $\geq 10\%$). Their corresponding chromosomal regions are indicated in Fig. 2b: For OS, top associated variants were located in an intergenic region 5' of the *ERN1* gene ($n = 2$). DRFS top associated variants included singletons at *ASAH2B*, *KCNH5*, *CACNB2*, and in the *FGFBP2/PROM1* locus. Associated variants for RFS mapped to *CACNB2* ($n = 4$) as well as to *TNFRSF10B* ($n = 2$), *CACNA1C*, and *LRP1B* (both singletons). Note that the single *LRP1B* hit (rs77397926; MAF $\sim 40\%$) is not linked to the *LRP1B* variants identified in the all AI analysis.

Letrozole-treated patients: Unadjusted analysis showed 22, 9, and 33 top associated variants for OS, DRFS, and RFS, respectively (Supplementary Data 5; p -value $\leq 1E-04$ and MAF $\geq 10\%$). Their corresponding chromosomal regions are highlighted in Fig. 2c: For OS, top associated variants mapped to the *CACNA2D3* ($n = 4$), *ABCC2* ($n = 4$), and *PLCE1/NOC3L* ($n = 12$) loci. Linked top variants in the latter gene region were also identified in DRFS ($n = 2$) and RFS ($n = 22$) analyses, however they showed rather low MAFs (iCOGS: $\leq 8.5\%$, OncoArray: $\leq 11.4\%$). In addition, top associated variants in DRFS and RFS analyses were identified at the *CELA2B/CASP9* locus ($n = 3$ and 5, respectively), all of which were also among the DRFS top hits in the all AI analyses. Further RFS associated top variants were located at the *ESR1* ($n = 4$) and *KCNJ11/NCR3LGI* ($n = 2$) loci.

Secondary endpoints breast cancer-specific survival (BCSS), distant recurrence-free interval (DRFI), and recurrence-free interval (RFI). Unadjusted analyses of all AI-treated patients for the secondary endpoints BCSS, DRFI, and RFI revealed 14, 3, and 3 variants with association p -value $\leq 1E-04$, respectively (Supplementary Fig. 4 in the Supplementary Information, Supplementary Data 6; MAF $\geq 5\%$). Of the 14 BCSS top hits, 13 were tightly linked variants in *CACNA1D* and the remainder was located in the *ERN1* gene region. Top variants in DRFI analysis comprised 3 variants with association p -value $\leq 1E-04$ (2 variants in the *FGFBP2/PROM1* gene region and a singleton in the *KCNK9* gene), and another $n = 85$ with p -value $\leq 5E-04$. These included further 9 variants in the *FGFBP2/PROM1* gene region, top hits in the *CASP9/CELA2B* locus ($n = 17$) and *LRP1B* ($n = 7$) as well as variants in *ESRGG* ($n = 6$), *NEDD9* ($n = 7$), and *VCAN* ($n = 3$). Top RFI hits comprised 67 variants with p -value $\leq 5E-04$, including 16 variants in *LRP1B* and 11 variants in *VCAN* (including one SNP with p -value $\leq 1E-04$), as well as singletons each in *DIXDC1* and *KCNK9* with p -value $\leq 1E-04$.

Selection of candidate variants for validation

The primary objective of our study was the validation of selected top variants identified in the OS, DRFS, and RFS discovery analyses of all AI-treated patients, using independent AI treatment cohorts. Candidate selection was based on a p -value-based filtering strategy (association p -value $\leq 1E-04$) and a power/sample size estimation approach that accounts for “winner’s curse bias”. In total, six SNPs were selected for primary analysis, including rs3820071 (DRFS, RFS) and rs6685648 (DRFS) in the *CELA2B/CASP9* locus, rs10496860 (DRFS, RFS) in *LRP1B*, rs3107669 (OS) in *GSK3B*, as well as rs353298 (DRFS, RFS) and rs353296 (RFS) in the *CARMN/MIR143HG*

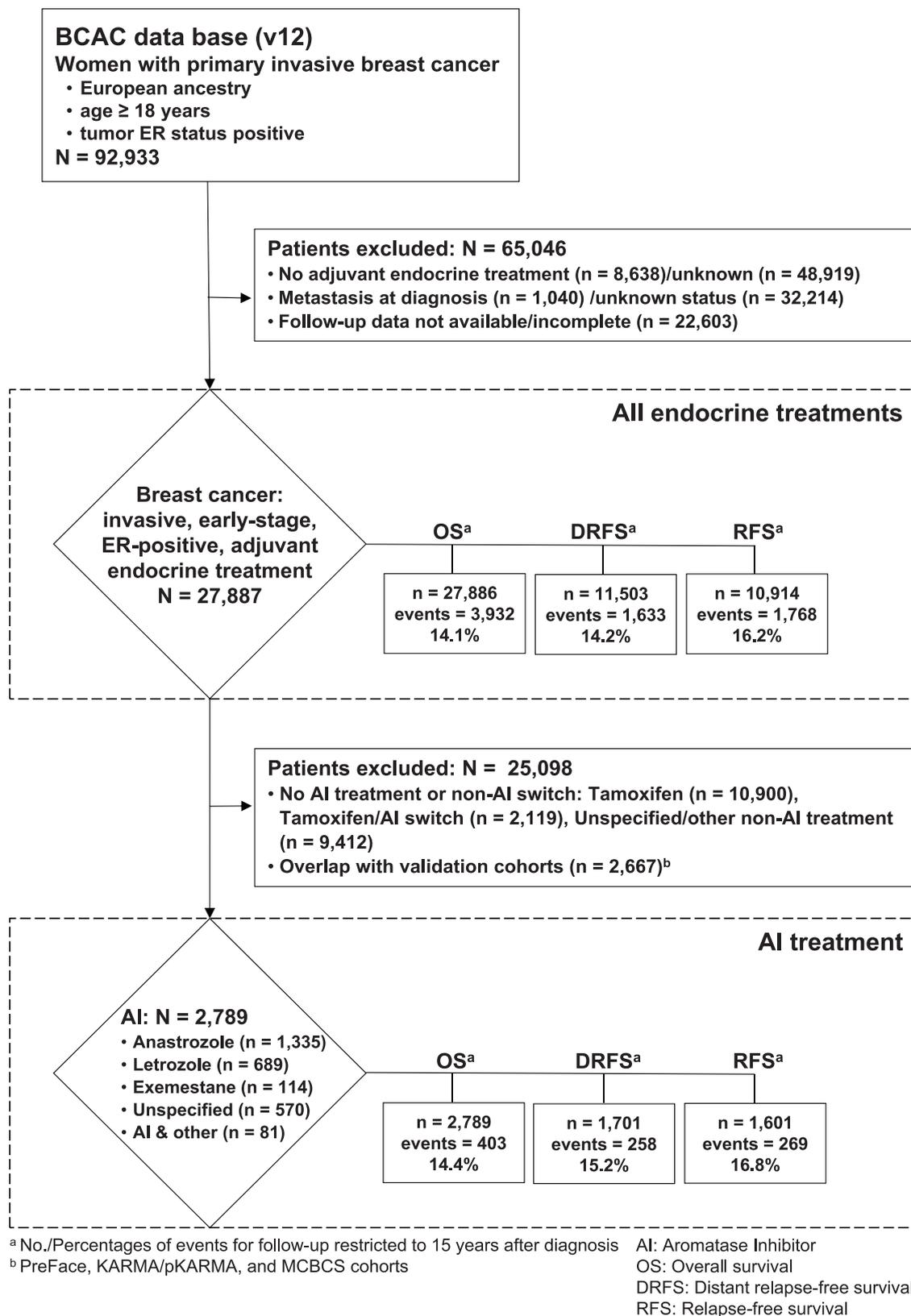


Fig. 1 | Flow diagram of BCAC patients with primary invasive ER-positive breast cancer, early-stage disease characteristics and AI therapy. The BCAC database (version 12) comprised 92,933 patients with a confirmed diagnosis of invasive ER-positive breast cancer who are of European descent (based on genotype data) and older than or equal to 18 years of age. Of these, 27,887 patients had early-stage disease with documented adjuvant endocrine therapy (AI, Tamoxifen, switch

therapy, or unspecified) and complete follow-up data. The subgroup of AI-treated patients analysed in this study comprised 2789 women of which 1335 had received anastrozole, 689 letrozole, and 114 exemestane. For 570 patients, AI treatment was not further specified and 81 patients had received an additional endocrine treatment. AI Aromatase Inhibitor, OS Overall survival, DRFS Distant relapse free survival, RFS Relapse free survival.

Table 1 | Characteristics of breast cancer cohorts included in AI candidate gene association analysis

	Discovery cohort	Validation cohorts				
	BCAC AI ^a	MA.27	PreFace	KARMA/ pKARMA	MCBCS	IKP211
No. of patients, <i>n</i>	2789	4408	2761	643	392	353
Follow-up (years), median [IQR ^b]	7.2 [3.9,10.5]	4.1 [3.9,5.0]	5.3 [4.7,5.7]	10.8 [7.8,12.2]	10.1 [7.0,11.7]	5.5 [4.6,10.1]
No. of deaths (any cause), <i>n</i> (%)	403 (14.4)	231 (5.2)	113 (4.1)	116 (18.0)	57 (14.5)	42 (11.9)
No. of breast cancer-specific deaths, <i>n</i> (%)	203 (7.3)	96 (2.2)	45 (1.6)	44 (6.8)	12 (3.1)	25 (7.2)
No. of distant relapses, <i>n</i> (%)	172 (6.2)	146 (3.3)	107 (3.9)	57 (8.9)	22 (5.6)	36 (10.2)
No. of relapses ^c , <i>n</i> (%)	233 (8.4)	238 (5.4)	155 (5.6)	78 (12.1)	28 (7.1)	42 (11.9)
Characteristics						
AI treatment, <i>n</i> (%)						
Anastrozole	1335 (47.9)	2195 (49.8)	–	–	335 (85.5)	143 (40.5)
Letrozole	689 (24.7)	–	2761 (100)	–	38 (9.7)	174 (49.3)
Exemestane	114 (4.1)	2213 (50.2)	–	–	19 (4.8)	21 (5.9)
AI & other hormonal therapy ^d	81 (2.9)	–	–	–	–	–
Unspecified/AI-to-AI switches	570 (20.4)	–	–	643 (100)	–	15 (4.2)
Age at diagnosis/randomization (years), median [IQR ^e]	61 [54,67]	64 [58,71]	64 [58,69]	64 [58,69]	63 [57,69]	64 [59,69]
Year of diagnosis/randomization , median [IQR ^e]	2006 [2004,2009]	2005 [2005,2005]	2009 [2009,2010]	2007 [2006,2008]	2007 [2005,2008]	2008 [2007,2009]
Menopausal status, <i>n</i> (%)						
Pre-/perimenopausal	335 (12.0)	–	–	64 (10.0)	55 (14.0)	–
Postmenopausal	1766 (63.3)	4408 (100)	2761 (100)	474 (73.7)	337 (85.6)	344 (97.5)
Missing	688 (24.7)	–	–	105 (16.3)	–	9 (2.5)
ER status, <i>n</i> (%)						
Negative	–	24 (0.5)	28 (1.0)	–	–	3 (0.8)
Positive	2638 (94.6)	4384 (99.5)	2727 (98.8)	639 (99.4)	392 (100)	349 (98.9)
Missing	151 (5.4)	–	6 (0.2)	4 (0.6)	–	1 (0.3)
PR status, <i>n</i> (%)						
Negative	524 (18.8)	729 (16.5)	383 (13.9)	170 (26.4)	60 (15.3)	48 (13.6)
Positive	1784 (64.0)	3587 (81.4)	2374 (86.0)	453 (70.5)	332 (84.7)	303 (85.8)
Missing	481 (17.2)	92 (2.1)	4 (0.1)	20 (3.1)	–	2 (0.6)
HER2 status, <i>n</i> (%)						
Negative	1836 (65.8)	–	2174 ^g (78.7)	291 (45.3)	328 (83.7)	266 ^g (75.4)
Positive/Equivocal	421 (15.1)	–	531 ^g (19.2)	59 (9.2)	23 (5.9)	78 ^g (22.1)
Missing	532 (19.1)	4408 (100)	56 (2.1)	293 (45.6)	41 (10.5)	9 (2.5)
Tumor size, <i>n</i> (%)						
≤2 cm	1327 (47.6)	3242 (73.5) ^f	1768 (64.0)	363 (56.5)	264 (67.3)	204 (57.8)
2–5 cm	1066 (38.2)	1046 (23.7) ^f	847 (30.7)	209 (32.5)	104 (26.5)	127 (36.0)
>5 cm	165 (5.9)	104 (2.4) ^f	115 (4.2)	14 (2.2)	21 (5.4)	18 (5.1)
Missing	231 (8.3)	16 (0.4) ^f	31 (1.1)	57 (8.9)	3 (0.8)	4 (1.1)
Nodal status, <i>n</i> (%)						
Negative	1397 (50.1)	3232 (73.3)	1963 (71.1)	275 (42.8)	256 (65.3)	237 (67.1)
Positive	1306 (46.8)	1107 (25.1)	779 (28.2)	309 (48.1)	133 (33.9)	108 (30.6)
Missing	86 (3.1)	69 (1.6)	19 (0.7)	59 (9.2)	3 (0.8)	8 (2.3)
M, <i>n</i> (%)						
M0	1544 (55.4)	4408 (100)	2682 (97.1)	643 (100.0)	392 (100)	340 (96.3)
MX/Missing	1245 (44.6)	–	79 (2.9)	–	–	13 (3.7)
Tumor grade, <i>n</i> (%)						
G1	445 (16.0)	–	520 (18.8)	80 (12.4)	132 (33.7)	59 (16.7)
G2	1470 (52.7)	–	1776 (64.3)	335 (52.1)	193 (49.2)	223 (63.2)
G3/4	778 (27.9)	–	461 (16.7)	194 (30.2)	64 (16.3)	69 (19.5)
GX/Missing	96 (3.4)	4408 (100)	4 (0.1)	34 (5.3)	3 (0.8)	2 (0.6)

Table 1 (continued) | Characteristics of breast cancer cohorts included in AI candidate gene association analysis

	Discovery cohort	Validation cohorts				
	BCAC AI ^a	MA.27	PreFace	KARMA/ pKARMA	MCBCS	IKP211
Surgery, n (%)						
No surgery	21 (0.8)	–	–	–	–	–
Breast saving	1011 (36.2)	3000 (68.1)	2137 (77.4)	361 (56.1)	235 (59.9)	254 (72.0)
Mastectomy (with or w/o axillary)	705 (25.3)	1408 (31.9)	588 (21.3)	281 (43.7)	156 (39.8)	99 (28.0)
Type Unknown	770 (27.6)	–	–	–	–	–
Missing	282 (10.1)	–	36 (1.3)	1 (0.2)	1 (0.3)	–
Radiation, n (%)						
No	509 (18.3)	–	352 (12.7)	126 (19.6)	151 (38.5)	61 (17.3)
Yes	2158 (77.4)	–	2393 (86.7)	516 (80.2)	241 (61.5)	286 (81.0)
Missing	122 (4.4)	4408 (100)	16 (0.6)	1 (0.2)	–	6 (1.7)
Neoadjuvant chemotherapy, n (%)						
No	2372 (85.0)	–	2559 (92.7)	597 (92.8)	392 (100)	325 (92.1)
Yes	165 (5.9)	–	190 (6.9)	44 (6.8)	–	25 (7.1)
Missing	252 (9.0)	4408 (100)	12 (0.4)	2 (0.3)	–	3 (0.8)
Adjuvant chemotherapy, n (%)						
No	1433 (51.4)	3096 (70.2)	1842 (66.7)	354 (55.1)	274 (69.9)	251 (71.1)
Yes	1215 (43.6)	1312 (29.8)	900 (32.6)	283 (44.0)	118 (30.1)	99 (28.0)
Missing	141 (5.1)	–	19 (0.7)	6 (0.9)	–	3 (0.8)
Trastuzumab, n (%)						
No	1461 (52.4)	840 (19.1)	730 (26.4)	329 (51.2)	364 (92.9)	–
Yes	182 (6.5)	34 (0.8)	169 (6.1)	44 (6.8)	28 (7.1)	–
Missing	1146 (41.1)	3534 (80.2)	1862 (67.4)	270 (42.0)	–	353 (100)

^aThe BCAC AI discovery cohort is composed of eligible patients from 25 individual BCAC studies.

^bInterquartile Range.

^cDistant relapse, locoregional relapse, ipsilateral BC or contralateral BC.

^dE.g.: aromatase inhibitor plus LHRH, megestrol, faslodex or provera.

^eCombining HER2 IHC 0/1+ as negative and 2+ /3+ as positive/equivocal.

^fDerived from T1-T3 and TX (no cases with T4 in MA.27 data set).

gene region (Table 2). The investigation of the candidates for the respective remaining survival endpoints as well as in the anastrozole and letrozole treatment subgroups was part of our secondary objectives. Further secondary objectives included the examination of top hits additionally identified in the anastrozole and letrozole treatment subgroup analyses as well as in the analyses of the secondary endpoints BCSS, DRFI, and RFI. For the former, we selected rs11012983 (*CACNB2*; anastrozole-treated patients) and rs827423 (*ESR1*; letrozole-treated patients; Supplementary Table 1a in the Supplementary Information) applying similar filter criteria as for the primary analyses. Selected BCSS, DRFI, and RFI candidates covered variants in *CACNA1D* (rs2276836), *ESRRG* (rs830321), *NEDD9* (rs4713432), and *VCAN* (rs251124 and rs2541311), where we relaxed the *p*-value criterion to $p \leq 5E-04$ (Supplementary Table 1b in the Supplementary Information).

Validation analysis of selected candidates (stage 2)

The five validation cohorts comprised 8557 AI-treated patients (2673 patients treated with anastrozole, 2973 patients with letrozole, 2253 patients with exemestane, and 658 patients with unspecified AI treatment or switch regimens). Details on clinical and histopathological characteristics as well as treatment regimens are given in Table 1. Median follow-up times for MA.27, PreFace, and IKP211 were 4.1, 5.3, and 5.5 years, and exceed 10 years for KARMA/pKARMA and MCBCS, respectively. Validation analyses of the six primary candidates were performed similar to the discovery analysis. Concerning our primary objective, we observed effect sizes (per-allele HR) up to 1.2; however, we did not identify any significant associations in the analysis of all AI-treated patients (Table 3, Fig. 3a–c, Supplementary Fig. 5 in

the Supplementary Information). Lowest *p*-values were observed for the *CARMN* variants rs353298 (DRFS and RFS, *p*-value = 0.280 and 0.108) and rs353296 (RFS, *p* = 0.084), however with opposite effect directions compared to the discovery analysis. This trend was consistent for three cohorts (PreFace, MCBCS, and IKP211), but not for MA.27. Particularly for rs353296, effects were heterogeneous across the validation cohorts ($I^2 > 52\%$, heterogeneity *p*-value < 0.1; Fig. 3c). Considerable study heterogeneity was also observed for *GSK3B* variant rs3107669 ($I^2 = 49.7\%$, heterogeneity *p*-value = 0.078; Fig. 3a, upper part). A detailed examination of rs3107669 revealed that the exclusion of one cohort, i.e., MA.27, considerably reduced heterogeneity, and the corresponding hazard ratio (HR) was shifted towards the discovery finding (HR = 1.17, 95% confidence interval (CI) = [1.00, 1.37], meta-analysis *p*-value = 0.045, $I^2 = 0\%$, heterogeneity *p*-value = 0.321; Fig. 3a, lower part). All other variants showed homogeneous effects in the analysis of all AI-treated patients ($I^2 < 25\%$, heterogeneity *p*-value > 0.2; Supplementary Fig. 5 in the Supplementary Information). Secondary adjusted analyses as well as sensitivity analyses did not reveal substantially different results (Supplementary Table 2 in the Supplementary Information).

Similar results for the six primary candidates were also observed in treatment subgroups analyses. No significant associations were obtained in the unadjusted analysis of anastrozole-treated patients ($n = 2673$; MA.27, MCBCS, IKP211; Supplementary Fig. 6 in the Supplementary Information). The corresponding letrozole analysis ($n = 2973$; PreFace, MCBCS, IKP211) revealed a low association *p*-value for the *CARMN* variant rs353296 with RFS and no study heterogeneity, but again an opposite effect direction was

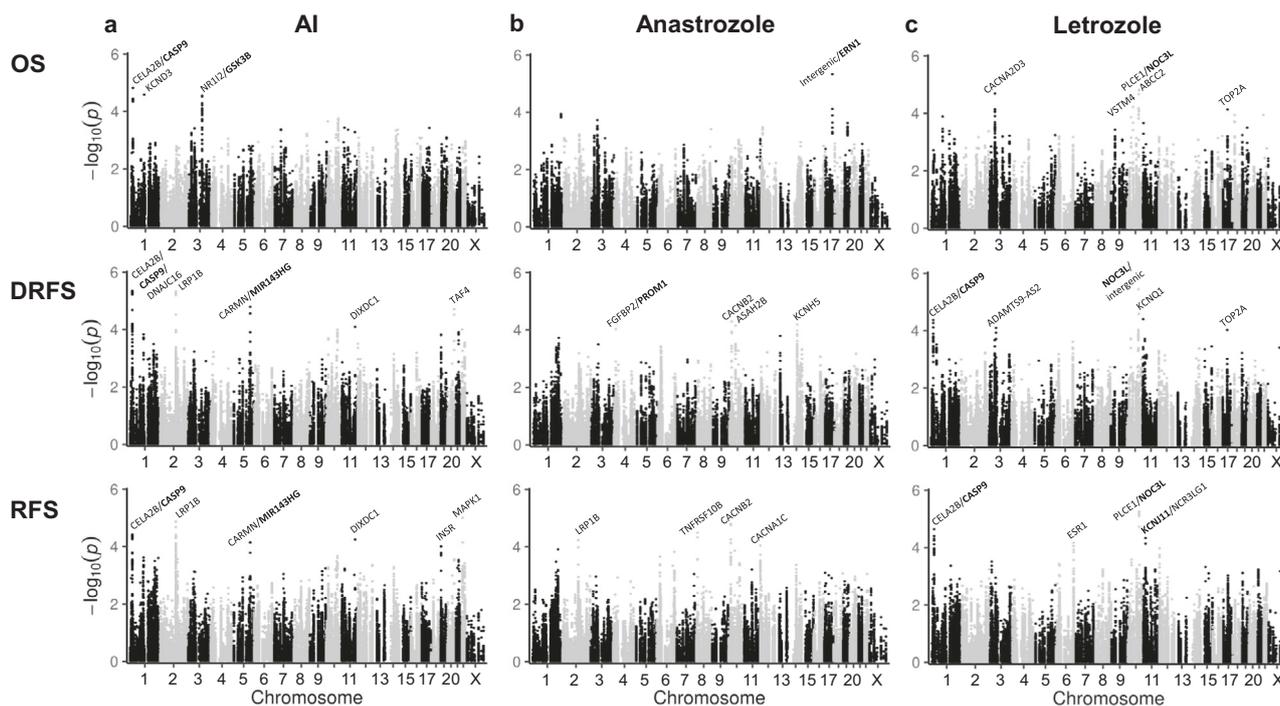


Fig. 2 | Candidate gene association study of 15-year overall, distant relapse-free, and relapse-free survival in AI-treated BCAC patients (discovery cohort). The Manhattan plots display the results of unadjusted analyses (stratified by country and corrected for genetic principal components). The y-axis shows the $-\log_{10}$ meta-analysis p -values and the x-axis the chromosomal positions of the variants. Variants are located within 753 selected candidate genes or surrounding gene region (defined as ± 10 kb from gene start/end). For variants with association p -value $\leq 1E-04$, the corresponding gene symbols are displayed on top (“intergenic” in case no gene was

annotated). Genes marked in bold were part of the 753 candidate genes. Analysis of (a) patients treated with any AI ($n = 2789$) and approximately 170,000 variants with $MAF \geq 0.05$, (b) patients treated with anastrozole ($n = 1335$) and approximately 138,000 variants with $MAF \geq 0.1$, and (c) patients treated with letrozole ($n = 689$) and approximately 138,000 variants with $MAF \geq 0.1$. Survival endpoints include OS (overall survival), DRFS (distant relapse-free survival), and RFS (relapse-free survival).

Table 2 | Results of meta-analyses in AI-treated BCAC patients (discovery cohort) for the six primary candidate variants

rs ID	chr:position ^a	ref/effect allele	EAF ^b (%)	OS		DRFS		RFS	
				HR ^c [95% CI]	P ^d	HR ^c [95% CI]	P ^d	HR ^c [95% CI]	P ^d
rs3820071	1:15808767	G/A	26.2	1.38 [1.18,1.60]	3.7E-05	1.57 [1.29,1.90]	4.6E-06	1.49 [1.23,1.80]	4.0E-05
rs6685648	1:15825195	T/C	31.6	1.27 [1.10,1.47]	1.3E-03	1.44 [1.20,1.73]	9.0E-05	1.39 [1.16,1.67]	<i>4.0E-04</i>
rs10496860	2:141671250	T/C	11.9	1.32 [1.03,1.70]	2.8E-02	1.86 [1.43,2.44]	4.9E-06	1.73 [1.33,2.24]	4.5E-05
rs3107669	3:119567101	C/A	60.3	1.38 [1.19,1.60]	2.9E-05	1.19 [0.99,1.44]	6.1E-02	1.12 [0.93,1.36]	2.2E-01
rs353298	5:148801379	A/G	41.3	0.78 [0.67,0.92]	2.3E-03	0.63 [0.51,0.78]	1.6E-05	0.66 [0.54,0.81]	7.1E-05
rs353296	5:148801826	G/T	34.3	1.23 [1.06,1.43]	6.1E-03	1.43 [1.18,1.73]	<i>2.7E-04</i>	1.46 [1.21,1.77]	7.4E-05

^a Position according to hg19.

^b Effect allele frequency (%).

^c Per-allele hazard ratio.

^d P -values $\leq 1E-04$ are marked in bold, p -values between $1E-03$ and $1E-04$ in italic.

observed when compared to the discovery finding (HR [95% CI] = 0.69 [0.54, 0.89], meta-analysis p -value = 0.003, $I^2 = 0\%$, heterogeneity p -value = 0.452). Opposite effect directions were also observed for the other endpoints (OS, DRFS) and for rs353298. The effect direction for GSK3B variant rs3107669 in the letrozole analysis matched that of the discovery finding, however the effect was insignificant and heterogeneous in the validation cohorts (OS: HR [95% CI] = 1.42 [0.79, 2.56], meta-analysis p -value = 0.237, $I^2 = 65.3\%$, heterogeneity p -value = 0.056; Supplementary Fig. 7 in the Supplementary Information). Further secondary objectives in our validation analysis were concerned with the examination of top hits identified in the treatment subgroup analyses as well as in the analysis for the additional survival endpoints BCSS, DRFI, and RFI considering breast-cancer specific death. The CACNB2 variant rs11012983 did not show a

significant effect in the anastrozole analysis (all p -values > 0.1 , Supplementary Fig. 6 in the Supplementary Information). In the letrozole analysis, the ESR1-SNP rs827423 showed a similar trend in discovery and validation cohorts; however this trend was not significant (RFS: HR [95% CI] = 0.87 [0.70, 1.08], meta-analysis p -value = 0.208, $I^2 = 3.9\%$, heterogeneity p -value = 0.402; Supplementary Fig. 7 in the Supplementary Information). Examination of the five BCSS, DRFI, and RFI candidates also did not reveal any significant results in the validation studies (all p -values > 0.1) (Supplementary Fig. 8 in the Supplementary Information).

Discussion

This study addressed the identification of potential pharmacogenomic predictors for AI treatment outcome in patients with hormone receptor-

Table 3 | Results of meta-analyses in the validation cohorts for the six primary candidate variants

rs ID	chr:position ^a	ref/effect allele	EAF ^b range (%)	OS		DRFS		RFS	
				HR ^c [95% CI]	P ^d	HR ^c [95% CI]	P ^d	HR ^c [95% CI]	P ^d
rs3820071	1:15808767	G/A	23.7-27.3	1.00 [0.88,1.14]	0.990	1.02 [0.89,1.18]	0.749	1.03 [0.91,1.17]	0.614
rs6685648	1:15825195	T/C	30.9-32.6	0.95 [0.84,1.08]	0.450	1.00 [0.88,1.14]	0.975	0.99 [0.85,1.14]	0.854
rs10496860	2:141671250	T/C	10.3-12.5	0.96 [0.79,1.16]	0.656	0.98 [0.80,1.19]	0.811	0.91 [0.77,1.09]	0.307
rs3107669	3:119567101	C/A	57.6-59.7	1.09 [0.91,1.31]	0.344	1.13 [0.87,1.46]	0.359	1.07 [0.87,1.31]	0.527
rs353298	5:148801379	A/G	39.6-41.6	1.12 [0.98,1.27]	<i>0.091</i>	1.09 [0.93,1.28]	0.280	1.11 [0.98,1.27]	0.108
rs353296	5:148801826	G/T	34.9-35.9	0.87 [0.71,1.08]	0.205	0.82 [0.64,1.04]	0.101	0.83 [0.67,1.03]	<i>0.084</i>

^a Position according to hg19.

^b Effect allele frequency (%).

^c Per-allele hazard ratio.

^d P-values ≤ 0.1 are marked in italic.

positive early breast cancer. To do so, we leveraged one of the largest clinical, epidemiologic, and genetic breast cancer data sets available from BCAC to perform hypothesis-driven pharmacogenomic analyses, using OS, DRFS, and RFS as the main survival endpoints. Our two-stage candidate gene approach focused on patients that met the clinical definition for adjuvant AI treatment as reflected in the favorable event rates in both discovery and validation cohorts (Table 1). In this study, we analyzed approximately 170,000 genetic variants located in 753 gene regions that have been selected based on their proposed roles in breast cancer endocrine resistance and/or related and intersecting pathways, in which ER constitutes the principal signal transduction pathway¹⁴⁻¹⁸.

Results from the discovery analysis of 2789 AI-treated patients highlighted 125 top associated loci for the main endpoints OS, DRFS, and RFS (p -value ≤ 1E-04) including variants located in the regions of *CASP9* (Chr. 1), *LRP1B* (Chr. 2), *GSK3B* (Chr. 3), and *CARMN* (Chr. 5), but not in any known pharmacogenes. Published candidates from survival GWAS or CGA studies of ER-positive, early-stage, AI-treated breast cancer patients were not significant in our analysis^{9,10,19,20}. Rather, OS associations were particularly evident for *CASP9* and *GSK3B*, with multiple closely linked variants ($r^2 > 0.9$), respectively. For DRFS and RFS, multiple associations were again observed for *CASP9*, but also for *LRP1B* (Fig. 2, Supplementary Data 2). Supportive evidence for the top variants' potential modulatory role in breast cancer outcome came from queries on eQTL in tumor and normal tissue. For *CASP9*, encoding a tumor suppressor caspase and recognized as a hub gene in inflammatory breast cancer²¹, the variants were associated with its mRNA expression. High *CASP9* expression in turn showed a favorable role for OS in patients with ER-positive breast tumors based on RNA-seq data²², which is in line with results of a gene-chip-based analysis²¹. This remained statistically significant when the analysis was limited to patients with a good prognosis (i.e., without chemotherapy; Supplementary Fig. 1 in the Supplementary Information). Similar associations have been observed for the neighboring genes *CELA2B* and *DNAJC16*, highlighting a putative role of this locus in the survival of early-stage breast cancer patients.

Likewise, for *GSK3B*, a unique serin/threonine kinase that acts as a signaling node at the intersection of multiple cancer-related pathways such as NFκB, Wnt/β-catenin, MEK/ERK, PI3K/Akt and Notch²³, top variants were identified as cis-eQTLs that impact on gene expression in mammary tissue. While *GSK3B* is subject to bidirectional regulation contingent on phosphorylation at Y216 (activation) and S9 (inhibition), a dual role as tumor suppressor and promoter has been noted²⁴. The tumor-suppressive role is mainly attributed to its activated state resulting in the downregulation of proto-oncogenic genes (e.g., *MYC*, *CCND1*) mediated by the blockage of β-catenin signaling via its phosphorylation and subsequent proteasomal degradation and/or direct regulation at protein level (e.g., Cyclin D1)²⁴. Notably, Cyclin D1 is among the most commonly overexpressed proteins in breast cancer and a combinatory role of Cyclin D1 and *GSK3B* expression in survival has been suggested with high *GSK3B* levels increasing the risk for distant relapse and low levels being more favorable²⁵. With regard to gene

expression, we observed that high *GSK3B* levels showed a protective effect on the OS of patients with ER-positive breast tumors, while an inverse effect was revealed in the subgroup of patients with Luminal A (but not Luminal B) tumors, stressing *GSK3B*'s putative dual role in the biology of these distinct molecular tumor subtypes.

Although our discovery analysis and in silico resources support a putative role of *CASP9* and *GSK3B* in breast cancer and AI treatment outcome, none of the six primary candidate variants have been validated in the independent breast cancer AI treatment cohorts. A reason for the failed replication could be that these are false positives of our discovery analysis. On the other hand, we may have missed polymorphisms with low frequency or weaker effects on the survival endpoints. Yet, the objective of the first stage of the analysis was the identification of a candidate set that would most likely include the top true survival-associated polymorphisms, which would then be analyzed in independent validation cohorts with sufficient power to correct for multiple testing and winner's curse.

While the strength of our study is its two-stage design that meets the gold standard of genetic association studies²⁶, and its large size of more than 11,300 AI-treated breast cancer patients, it is at the same time subject to several limitations intrinsic to large drug treatment outcome related studies^{27,28}. These comprise differences in study design or clinical characteristics such as age at diagnosis, histopathological parameters, treatment regimens, as well as length of follow-up, that all may contribute to heterogeneity among study cohorts. Moreover, studies could potentially suffer from underreporting of events, and commonly investigated endpoints may be limited surrogates of endocrine treatment efficacy in pharmacogenomic investigations. In terms of heterogeneity, we noticed considerable differences for clinical characteristics and treatment regimens across the validation cohorts. While adjusted analyses did not reveal substantial differences compared to unadjusted analysis, we observed especially for *GSK3B* rs3107669 that the exclusion of MA.27 considerably reduced heterogeneity, with the corresponding HR shifting in the direction of our discovery results (OS: HR [95% CI] = 1.17 [1.00, 1.37], p -value = 0.045, $I^2 = 0\%$, heterogeneity p -value = 0.321; Fig. 3a). Thus, we cannot rule out that heterogeneity across the study cohorts has hampered, at least to some extent, our validation effort. In this context, it should be noted that genetic variants with similar effects across different AI treatment regimens have been described in the literature, but even opposite effect directions were reported. For example, *MIR2052HG* variants were associated with breast cancer-free interval (BCFI) in the combined anastrozole and exemestane treatment group of the MA.27 trial¹⁰, however in anastrozole- and exemestane-only analyses opposite BCFI effects of multiple other genetic variants were identified²⁹. Yet, the previously described variants assessed in our discovery analyses were not strongly associated, neither in all AI-treated patients nor in anastrozole-only treated patients (all unadjusted p -values > 0.04). This may be due to the use of other endpoints (OS, DRFS, RFS) and the limited sample size of our discovery cohort, particularly with regard to anastrozole-only ($n = 1335$) and letrozole-only ($n = 689$) analyses. Of note, our subgroup

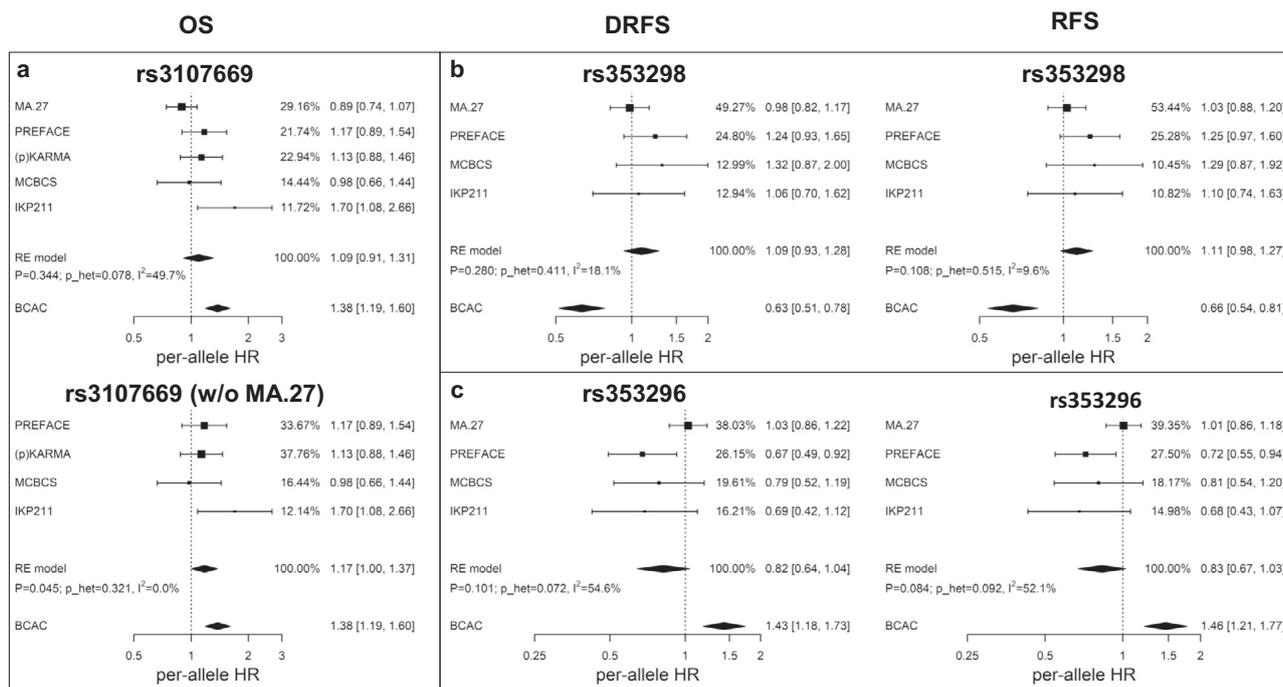


Fig. 3 | Results of meta-analyses for selected candidate variants in validation cohorts and BCAC patients (discovery cohort). Forest plots depicting per-allele hazard ratios (x-axis) and corresponding confidence intervals for (a) rs3107669, (b) rs353298, and (c) rs353296 in eligible validation cohorts. The size of the squares reflects the study size. The KARMA/pKARMA studies were omitted in DRFS and

RFS analyses due to delayed study entry of nearly all 643 patients (> 0.5 years after diagnosis). Diamonds represent results of random effects (RE) meta-analyses in validation cohorts and of the discovery analyses in BCAC patients, respectively. Endpoints are OS (overall survival), DRFS (distant relapse-free survival), and RFS (relapse-free survival).

analyses revealed only few additional candidates as compared to the all-AI analysis ($n = 2789$), a reason why we chose the latter as primary analysis.

It is important to view our findings within the context of reported challenges and uncertainties inherent to treatment outcome association studies that may limit their value^{27,28}. We support the view that better-quality cohort appropriateness is key for the detection of treatment-related effects and appreciate that efficacy (survival) endpoints such as OS, DFS, or RFS differ from phenotype endpoints (e.g. disease risk), as they are liable to additional confounders related to patients' drug treatment history (e.g. type, duration, adherence, etc). Thus, it will be advisable to control for these in the future to warrant detectability of small treatment-related effects.

A particular concern may relate to the specific drug action inherent to endocrine treatment, that currently is not well-captured by standard descriptive variables and endpoints. In contrast to cancer treatments that act via cytotoxic effects²⁸, the efficacy of endocrine treatment including AI is subject to a cytostatic effect that persists over time. At a glance, a minimum of 5-year treatment is required to keep micrometastases in the women's body at check, which is conveyed by a carry-over effect beyond the actual drug intake until resistance occurs^{4,6,16}. Clinical events such as relapse and death follow an evolving micrometastatic phenotype via selection and expansion of a resistant subpopulation of tumor cells and reprogramming of the tumor cells' biological pathways. Accordingly, the tumor cell itself potentially holds promise for a more precise efficacy endpoint than survival. Despite laboratory proof of concept of an adapting molecular landscape during AI treatment¹⁴⁻¹⁸, and clinical proof of concept that this adaptive process is a critical determinant of endocrine treatment outcomes³⁰, the notion of an application of the dynamic tumor molecular landscape as a measurable AI-related endpoint may be premature. Yet, innovation on micrometastasis detection and manipulation are on the rise^{31,32} and it may therefore be reasonable to speculate that proper techniques for the assessment of quantifiable endpoints may become available.

Although we did not meet our primary objective, our work generated important findings from which we conclude that efforts towards further standardization of endocrine treatment pharmacogenomic studies for ER-

positive early-stage breast cancer need to be intensified. A lesson learned is to widen the view towards innovative approaches to better capture treatment efficacy. The latter requires revisiting the way we define informative endpoints, which can only be achieved via controlled longitudinal clinical trials with sufficiently long follow-up assessing drug-related parameters including adherence that would then be complemented by the monitoring of measurable molecular tumor cell parameters.

Methods

Breast Cancer Patients

Patient characteristics, clinical data (e.g., treatment, follow-up, and survival endpoints), and genotype data (e.g., treatment, follow-up, and survival endpoints) have been retrieved from established data repositories including hospital-based and case-control studies, as well as prospective clinical trials and observational studies for the investigation of an association between common germline variants and patients' outcome upon AI treatment. Study subjects include patients that had received an AI, i.e. anastrozole, letrozole, or exemestane for the treatment of their primary invasive early breast cancer that was ERα and/or progesterone receptor (PR) positive. Patient origin, size of discovery cohort (stage 1) and validation cohorts (stage 2), as well as extraction procedures are briefly summarized as follows with more details provided in Supplementary Data 7 and 8, respectively.

Discovery cohort (stage 1). We used data from selected breast cancer patients from studies participating in the Breast Cancer Association Consortium (BCAC) stored in the BCAC database at the University of Cambridge, UK. Data have been retrieved from database version 12 (07/2019) that in addition to clinical, epidemiological, and genotype data includes information about histopathology, survival, and treatment pooled and harmonized at the Netherlands Cancer Institute prior to integration^{12,13}. Participating studies were approved by respective ethics committees, and patients' informed consent was obtained from all patients (Supplementary Data 7). Female breast cancer patients have been first selected based on a diagnosis of primary invasive breast cancer,

European ancestry (based on genotype data), age at diagnosis ≥ 18 years, and ER-positive tumor status ($n = 92,933$; Fig. 1). To match the study sample to the definition of an intended 5-year adjuvant endocrine treatment setting, patients not having received endocrine treatment or with unknown treatment status, metastatic disease at diagnosis or unknown metastatic status, as well as unknown or incomplete follow-up information on survival endpoints have been subsequently excluded ($n = 65,046$ patients; Fig. 1). The remaining 27,887 patients (30%) included 2789 AI-treated patients that served as the discovery cohort in this current investigation. Figure 1 illustrates the selection process, and Supplementary Data 7 provides details of the 25 contributing BCAC studies from Europe, USA, and Australia. Among the AI-treated patients, the majority had received anastrozole ($n = 1335$; 47.9%) or letrozole ($n = 689$; 24.7%) in line with the then in force standard of care in their countries of origin, with fewer patients having received exemestane ($n = 114$; 4.1%). For a large subgroup, AI treatment was not further specified ($n = 570$; 20.4%) or patients in addition to AI had received additional endocrine treatment ($n = 81$; 2.9%). The median follow-up of the 2789 AI-treated patients was 7.2 years (Interquartile Range (IQR): 3.9 to 10.5 years, Table 1). Note, patients treated with tamoxifen (Fig. 1) are subject to separate analyses and are not included in this current work.

Validation cohorts (stage 2). Contributing cohorts were selected based on availability and study design to match the aforementioned criteria for adjuvant AI endocrine treatment. The total cohort comprised 8557 AI-treated patients from five independent studies including AI outcome-related prospective randomized and observational clinical trials (MA.27³³, PreFace³⁴, IKP211^{35,36}) as well as case-control and cohort studies (KARMA/pKARMA³⁷, MCBCS³⁸). Overall, there were 2673 patients treated with anastrozole (MA.27, MCBCS, IKP211), 2973 patients treated with letrozole (PreFace, MCBCS, IKP211), and 2253 patients treated with exemestane (MA.27, MCBCS, IKP211), as well as 658 patients with unspecified AI treatment or switch regimens. Median follow-up times for MA.27, PreFace, and IKP211 were 4.1, 5.3, and 5.5 years, and exceeded 10 years for KARMA/pKARMA and MCBCS, respectively (Table 1). Details of each individual study regarding study design, recruitment, type of AI treatment, and follow-up are given below and in Supplementary Data 8.

MA.27: The MA.27 trial (ClinicalTrials.gov identifier NCT00066573) is a phase III cooperative group study, designed as multicenter, multinational, randomized open-label trial. The trial included postmenopausal women with histologically confirmed and completely resected stage I-III breast cancer (AJCC Version 6) that was ER α and/or PR positive. Patients were randomized to 5 years of anastrozole or exemestane for the testing of superior efficacy of the latter with regard to event-free survival³³. Following approval by health regulatory authorities and the centers' regulatory approval boards, the study enrolled 7575 women, among whom North American patients provided a blood sample and gave consent for genetic testing. Research was performed after approval by local institutional review boards in accordance with assurances filed with, and approved by the Department of Health and Human Services¹⁰. Our validation study used baseline and clinical as well as relevant genetic data of 4408 patients, of whom 2195 patients had received anastrozole and 2213 patients exemestane. Median follow-up time was 4.1 years (IQR: 3.9-5.0 years; Table 1).

PreFace: The PreFace study (ClinicalTrials.gov identifier NCT01908556) is a single-armed, multicenter, open-label, phase IV trial for the evaluation of predictive factors of the effectivity of AI therapy. The study recruited 3483 postmenopausal early-stage breast cancer patients with a hormone receptor-positive breast cancer (without distant metastases) across 220 study sites in Germany. Patients were assigned to treatment with 2.5 mg daily letrozole after the completion of standard therapies (surgery, chemotherapy, and radiotherapy) in accordance with national guidelines. Treatment was continued until disease recurrence or up to a maximum of 5 years³⁴. All patients provided blood samples for genetic testing of putative associations between polymorphic loci and survival as well as quality of life. The study was approved by Ethics

Committee of the Medical Faculty of the University of Erlangen-Nuremberg (Number 25_2008) and all other relevant ethics committees. Our validation study included baseline, clinical, and relevant genetic data of 2761 letrozole-treated patients. Median follow-up time was 5.3 years (IQR: 4.7-5.7 years; Table 1). Patient characteristics selected for this current study did not differ from the patient characteristics of the main study³⁴.

KARMA/pKARMA: The Karolinska Mammography project (KARMA) is a prospective cohort study of women attending mammography screening or clinical mammography at four hospitals in Sweden³⁷ for the assessment of epidemiological and molecular breast cancer risks. Between 10/2010 and 03/2013, 70,877 women were enrolled, including 3000 incident breast cancer cases diagnosed at cohort entry. All study participants gave informed consent and ethics approval was obtained from the ethical board at the Karolinska Institute. Approximately 53% of the recruited women were postmenopausal. The cohort is matched to relevant Swedish medical quality registries for biannual follow-up. There is an estimated annual rate of 250 new breast cancer diagnoses. Follow-ups recorded at the national registries include information on the use of medication, secondary diagnoses, and death events. Similarly, pKARMA is a case-control study with incident breast cancer cases recruited between 01/2001 and 12/2008 in the Stockholm/Gotland area by the Stockholm Breast Cancer Registry. Recording of survival data follows that of the KARMA study. Our validation study included altogether 643 AI treated patients from both cohorts with a median follow-up time of 10.8 years (IQR: 7.8-12.2 years). Specific information on the type of AI treatment was not available (Table 1).

MCBCS: Incident breast cancer cases from 6 US states (MN, WI, IA, IL, ND, SD) were recruited within the framework of a hospital-based case-control study for questionnaire-based, molecular risk factor analyses at the Division of Medical Oncology at the Mayo Clinic in Rochester, MN (02/2001-06/2005)³⁸. Cases were diagnosed within the previous six months and no prior history of cancer (except non-melanoma skin cancer), and clinical and treatment data as well as follow-up and survival data were retrieved by medical record abstraction during annual follow-up surveys. Our validation study used data from 392 patients that were treated with an AI. Of these, 335 received anastrozole, 38 letrozole, and 19 exemestane. Median follow-up time was 10.1 years (IQR: 7.0-11.7 years; Table 1).

IKP 211 Study: The IKP211 study (German clinical trials registry identifier DRKS00000605) is a prospective multi-center observational breast cancer study for the investigation of endocrine treatment efficacy and toxicity. The study recruited 1286 postmenopausal patients with a diagnosis of early-stage invasive breast cancer that was ER and/or PR positive. Patients received 5 years of adjuvant treatment, either with tamoxifen or an AI or switch from tamoxifen to AI or vice versa in line with local clinical practice and national guidelines. Patients were recruited at 37 breast centers and hospitals. Most patients were followed up during a 10 year study extension period for the documentation of recurrence and survival. All patients provided blood and plasma samples for genetic testing and drug metabolite measurements^{35,36}. Ethics approvals were obtained from the Ethics Committee of the University of Tübingen, as well as respective local committees of the participating centers. Follow-up was recorded for 1159 patients, of whom 550 patients were treated with an AI (48%) and the remainder with tamoxifen (292 patients; 19%) or switch therapy (317 patients; 27%). AI treated patients in most cases received letrozole ($n = 418$). Our validation study used clinical and relevant genetic data of 353 AI treated patients that were composed of 143 anastrozole-, 174 letrozole-, and 21 exemestane-treated patients, as well as 15 patients with AI-to-AI switch therapy. Note, 210 letrozole-treated patients were not considered in this current validation due to an overlap with the PreFace study that was conducted in parallel. Median follow-up time was 5.5 years (IQR: 4.6-10.1 years; Table 1).

Ethics statement

The study was performed in accordance with the Declaration of Helsinki. All individual studies included in the analyses were approved by the appropriate institutional ethical review boards following their national

guidelines for informed consent; the details are provided in Supplementary Data 7 and 8.

Genotyping, quality control, and imputation of genetic variants

Discovery cohort. Genotypes were established as part of the Collaborative Oncological Gene-Environment Study (COGS) and the OncoArray project using the Illumina iSelect array (211,155 genetic variants) and the Illumina OncoArray-500K BeadChip (533,631 genetic variants), respectively^{11,39}. The complete list of genotyped SNPs located in the 753 candidate gene regions is provided in Supplementary Data 1. Non-genotyped variants were imputed based on the 1000 Genomes Project Phase 3 release (October 2014) as reference panel. More details on genotyping and imputation of genetic variants as well as quality control are described elsewhere^{12,39,40}. Our main discovery analysis in 2789 patients considered approximately 170,000 variants in 753 candidate gene regions that had a MAF \geq 5% and were typed or imputed with imputation $r^2 \geq$ 0.8 in at least one of the datasets (iCOGS and OncoArray). Due to lower sample sizes, the anastrozole ($n = 1335$) and letrozole ($n = 689$) subgroup analyses have been restricted to variants with MAF \geq 10%, resulting in approximately 138,000 variants.

Validation cohorts

MA.27 samples. Genotyping was performed on three different platforms: Illumina Human610 Quad BeadChip, Illumina Human OmniExpress, and Illumina Human OmniExpressExome. Imputation was performed across the three platforms separately. Details on genotyping and imputation of genetic variants as well as quality control are described in Ingle et al.¹⁰ and literature cited therein. PreFace samples: Genotyping was also established as part of the OncoArray project using the Illumina OncoArray-500K BeadChip and the same imputation methods as described above for the discovery cohort. KARMA/pKARMA and MCBCS samples: Genotyping data (including typed and imputed variants) were retrieved from BCAC database version 12. IKP211 samples: Genotyping was performed on the Illumina Infinium OmniExpress 12/24 BeadChips platform (Illumina, Singapore). Quality control and imputation were performed similar to Khor et al.⁴¹. In brief, samples with call rates $<$ 95% or showing extremes of heterozygosity as well as variants with call rates $<$ 90%, MAF $<$ 1%, or Hardy Weinberg Equilibrium p -value $<$ $1E-06$ and non-autosomal markers were excluded. The McCarthy Group imputation preparation tool (<https://www.well.ox.ac.uk/~wrayner/tools/index.html>, version 4.2.11) was applied for further quality control and filtering, and SHAPEITv2⁴² as well as IMPUTE v2.3.2⁴³ were used for pre-phasing and genotype imputation, respectively, using the 1000 Genomes Phase 3 reference panel.

Selection of candidate genes

Candidate genes have been selected based on their putative roles in distinct biological pathways of endocrine resistance and converging/intersecting pathways. They include on the one hand genes associated with disease progression and breast cancer relapse during adjuvant endocrine treatment and thereafter. On the other hand, based on in vitro models, they include genes involved in the dynamic reconfiguration of the tumor cells' survival program during long-term estrogen deprivation, with ER constituting the principal signal transduction pathway for replication and death¹⁴⁻¹⁸. This knowledge framework served as a source for the selection of 753 genes investigated in this current CGA study (Supplementary Data 1). In brief, selected genes operate within the following context: endocrine resistance ($n = 217$), microRNAs of endocrine resistance, their targets and biosynthesis ($n = 262$), cancer-related channels, particularly potassium channels in endocrine resistance ($n = 144$), cholesterol biosynthesis and transport (sterol pathway; $n = 52$), CYP3A4 drug metabolism and ER signaling ($n = 53$), as well as other cell adhesion, surface, and transmembrane proteins including receptors and stem cell markers that may contribute to endocrine resistance ($n = 25$). Variants have been retrieved from 10 kb upstream to 10 kb downstream of the chromosomal positions of the 753 candidate genes. Detailed information is given in Supplementary Data 1.

Survival endpoints

Outcome measures followed the standardized definitions for efficacy endpoints in adjuvant breast cancer trials as described by the STEEP system⁴⁴. They are defined as follows, with deviations from STEEP indicated in italics. Time-to-event was calculated from the date of diagnosis for all cohorts except the MA.27 trial, for which the date of randomization was used instead³³.

Primary endpoints

Overall survival (OS). Time from diagnosis/randomization to death of any cause (death from breast cancer, death from nonbreast cancer cause, or death from unknown cause). Distant relapse-free survival (DRFS): Time from diagnosis/randomization to time of first documented distant recurrence or death of any cause. *Relapse-free survival (RFS)*: Time from diagnosis/randomization to first documented recurrence (invasive ipsilateral breast tumor recurrence, *invasive contralateral breast cancer*, local/regional invasive recurrence or distant recurrence) or death of any cause.

Secondary endpoints

Breast cancer-specific survival (BCSS). Time from diagnosis/randomization to death from breast cancer. Distant recurrence-free interval (DRFI): Time from diagnosis/randomization to first documented distant recurrence or death from breast cancer. Recurrence-free interval (RFI): Time from diagnosis/randomization to first documented recurrence (invasive ipsilateral breast tumor recurrence, *invasive contralateral breast cancer*, local/regional invasive recurrence or distant recurrence) or death from breast cancer.

For each outcome measure, patients who did not experience any of the defined events or were lost to follow-up were censored at the date of their last contact or at 15 years after diagnosis, whichever came first. Of the 2789 AI-treated BCAC patients, 1479 (53.0%) had a delayed study entry (for 563 patients within 0.5 years after diagnosis, and for 916 patients more than 0.5 years after diagnosis). Complete DRFS data were available for 1887 patients (67.7%). Of these, 708 (37.5%) had a delayed study entry that was within 0.5 years after diagnosis for 522 patients and more than 0.5 years for 186 patients. Data for RFS were complete for 1760 of the 1887 patients (93.3%). Of these, 656 (37.3%) had a delayed study entry, that was within 0.5 years after diagnosis for 497 patients and more than 0.5 years for 159 patients.

Statistical analysis

Throughout the manuscript, statistical significance was defined as p -value $<$ 0.05. Unless otherwise stated, all statistical tests were two-tailed.

Discovery analysis (stage 1). For each SNP and each outcome measure, survival analysis was performed using Cox proportional hazards regression and an additive genetic model. For OS and BCSS, delayed entry Cox models were applied where time at risk was defined as starting from the date of diagnosis or date of study entry in the case of a delayed entry. For all other survival endpoints, patients entering a study more than 0.5 years after diagnosis were excluded and delayed entry was not considered. Corresponding sensitivity analyses were performed with all available patients and follow-up time, both starting from date of diagnosis and date of study entry. Analyses were conducted for all AI-treated patients as well as the subgroups of anastrozole- and letrozole-treated patients using the statistical software R (version 4.2.0; <https://www.r-project.org>) and additional packages *rms*_6.7.1⁴⁵ and *metafor*_4.4-0⁴⁶.

All Cox models were stratified by country and included 10 ancestry informative principal components to adjust for population structure. OncoArray and iCOGS datasets were analyzed separately and results were subsequently combined using random-effects meta-analysis. Here, standard errors of the HR estimates were recomputed based on the likelihood ratio test statistic, as previously described¹². For top hits identified in the unadjusted analyses (meta-analysis p -value \leq $1E-04$), we conducted secondary analyses with adjustment for different combinations of patient characteristics (age, body mass index (BMI), menopausal status, family history of breast cancer in a first degree relative), tumor characteristics (size, nodal

status, grade, stage, ER/PR status), therapy (type of AI treatment, adjuvant/neoadjuvant chemotherapy, radiation) as well as surgery and study type.

PancanQTL was examined to see whether top variants of the unadjusted analyses have been identified as cis-expression quantitative trait loci (cis-eQTL) or survival QTLs in the TCGA breast cancer cohort or other tumor entities⁴⁷. In addition, the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS) tool⁴⁸ was applied to check for cis-eQTLs in normal tissues including breast mammary tissue, using an eQTL p -value threshold of $1E-03$. Genes identified in eQTL databases were further examined in KM Plotter for associations of their RNA-seq expression levels with overall survival in ER-positive, endocrine treated breast cancer patients with or without exclusion of patients that had received chemotherapy²².

Selection of candidate variants for validation analysis (stage 2). The selection of candidate variants was based on the results of the discovery analysis (stage 1). For the primary endpoints (OS, DRFS, RFS), we first filtered for meta-analysis p -value $\leq 1E-04$ (Supplementary Data 2, 4, 5). For the secondary endpoints (BCSS, DRFI, RFI) with less pronounced effects, a cutoff of $p \leq 5E-04$ was applied (Supplementary Data 6). Selection of candidate variants based on low p -values introduces optimistic bias into the corresponding regression estimates, a phenomenon which is also known as winner's curse bias or Beavis effect. To address this problem, we applied a bootstrap-based resampling approach for survival analysis to the discovery cohort⁴⁹. We used one set of bootstrap samples ($n = 500$) to obtain bias-reduced HR estimates, and a second set of bootstrap samples ($n = 100$) to empirically estimate variance and covariance of within-sample and out-of-sample estimates for each variant. Subsequently, the power/sample size calculation for the final selection of the primary validation candidates was based on the winner's curse-corrected estimates and a one-sided score test⁵⁰. Given the number of events of the primary endpoints in the validation cohorts, the MAFs of the variants in 1000 G phase3 EUR, and a Bonferroni-corrected alpha level of 0.56% ($= 5\%/9$), the power was 80% to detect six selected primary candidate variants: rs3820071 (DRFS, RFS), rs6685648 (DRFS), rs10496860 (DRFS, RFS), rs3107669 (OS), rs353298 (DRFS, RFS), and rs353296 (RFS). Furthermore, as secondary candidates we selected rs11012983 and rs827423 from the anastrozole- and letrozole-only analyses, respectively, as well as rs830321, rs2276836, rs251124, rs2541311, and rs4713432 from the analyses of the secondary endpoints.

Validation analysis (stage 2). Survival analysis in the validation cohorts was performed similar to the discovery analysis. For each validation cohort, Cox proportional hazards regression was applied to examine the associations between candidate variants and outcome measures in an additive genetic model. For OS and BCSS, delayed entry Cox models were used. For all other endpoints, patients entering a study more than 0.5 years after diagnosis were excluded from the analysis. For MA.27, Cox models were stratified by country (US, Canada) and included the previously described top seven eigenvectors of the SNP correlation matrix¹⁰. Here, survival analysis was performed separately for each of the three platforms and results were combined in a fixed-effects meta-analysis. KARMA and pKARMA patients were analyzed together, where Cox models were stratified for study. For all other validation cohorts, unstratified Cox models were applied. As described above, standard errors of the HR estimates were recomputed based on the likelihood ratio test statistic. Results of the five validation cohorts were combined in a random-effects meta-analysis. Since nearly all 643 KARMA/pKARMA patients had a delayed study entry (> 0.5 years after diagnosis), these were only included in the OS and BCSS analysis. For all other survival endpoints, only MA.27, PreFace, MCBCS, and IKP211 were used for validation analysis.

In secondary adjusted analyses, we included nodal status and adjuvant chemotherapy as strata in the Cox models¹⁰ with and without

additional adjustment for tumor size, ER/PR status, type of AI treatment (MA.27, MCBCS, and IKP211), as well as ECOG performance score and bisphosphonate use (MA.27 only). Additional stratification factors included trastuzumab as well as celecoxib and aspirin use for MA.27³³. In corresponding sensitivity analyses, stratification factors were alternatively included as covariates in the Cox models. Further sensitivity analyses considered (additional) adjustment for different combinations of age or menopausal status, BMI, family history of breast cancer in a first degree relative, tumor grade, tumor stage, neoadjuvant chemotherapy, radiation, and/or surgery. All analyses were similarly performed in the subgroup of patients treated with anastrozole and letrozole, respectively.

Data availability

The datasets analyzed in the current study cannot be made publicly available in full due to restraints imposed by ethics committees for the protection of patient privacy and confidentiality, as well as ownership of the contributing consortia and institutions. A subset of the BCAC data that support the findings of this study is publicly available via dbGaP (www.ncbi.nlm.nih.gov/gap; accession number phs001265.v1.p1). Requests for BCAC data can be made to the BCAC Data Access Coordination Committee (DACC) (<https://ccge.medschl.cam.ac.uk/breast-cancer-association-consortium-bcac>). Requests for other data should be directed to the corresponding author for further approval of the involved institutions. Summary results of the discovery analysis are available at <https://ccge.medschl.cam.ac.uk/breast-cancer-association-consortium-bcac>; summary results of the validation analysis are included in the manuscript and its supplement.

Code availability

This study did not generate original codes. All software and algorithms used in this study are publicly available and listed in the Method section. R codes used to analyze data and generate figures are available via the corresponding author upon reasonable request. The code that supports the findings of this study is available via the corresponding author on reasonable request.

Received: 24 July 2024; Accepted: 4 February 2025;

Published online: 19 February 2025

References

- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Maximov, P. Y. et al. A unifying biology of sex steroid-induced apoptosis in prostate and breast cancers. *Endocr. Relat. Cancer* **25**, R83–R113 (2018).
- Smith, I. E. & Dowsett, M. Aromatase inhibitors in breast cancer. *N. Engl. J. Med.* **348**, 2431–2442 (2003).
- Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet* **386**, 1341–1352 (2015).
- Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Aromatase inhibitors versus tamoxifen in premenopausal women with oestrogen receptor-positive early-stage breast cancer treated with ovarian suppression: a patient-level meta-analysis of 7030 women from four randomised trials. *Lancet Oncol.* **23**, 382–392 (2022).
- Pan, H. et al. 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *N. Engl. J. Med.* **377**, 1836–1846 (2017).
- Hertz, D. L., Henry, N. L. & Rae, J. M. Germline genetic predictors of aromatase inhibitor concentrations, estrogen suppression and drug efficacy and toxicity in breast cancer patients. *Pharmacogenomics* **18**, 481–499 (2017).
- Ingle, J. N. et al. Anastrozole has an Association between Degree of Estrogen Suppression and Outcomes in Early Breast Cancer and is a

- Ligand for Estrogen Receptor α . *Clin. Cancer Res.* **26**, 2986–2996 (2020).
9. Cairns, J. et al. Pharmacogenomics of aromatase inhibitors in postmenopausal breast cancer and additional mechanisms of anastrozole action. *JCI Insight* **5**, e137571, <https://doi.org/10.1172/jci.insight.137571> (2020).
 10. Ingle, J. N. et al. Genetic Polymorphisms in the Long Noncoding RNA MIR2052HG Offer a Pharmacogenomic Basis for the Response of Breast Cancer Patients to Aromatase Inhibitor Therapy. *Cancer Res.* **76**, 7012–7023 (2016).
 11. Michailidou, K. et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45**, 353–361 (2013).
 12. Escala-Garcia, M. et al. Genome-wide association study of germline variants and breast cancer-specific mortality. *Br. J. Cancer* **120**, 647–657 (2019).
 13. Morra, A. et al. Association of germline genetic variants with breast cancer-specific survival in patient subgroups defined by clinic-pathological variables related to tumor biology and type of systemic treatment. *Breast Cancer Res.* **23**, 86 (2021).
 14. Dixon, J. M. Endocrine Resistance in Breast Cancer. *N. J. Sci.* **2014**, 390618, <https://doi.org/10.1155/2014/390618> (2014).
 15. Clarke, R., Tyson, J. J. & Dixon, J. M. Endocrine resistance in breast cancer—An overview and update. *Mol. Cell. Endocrinol.* **418**, 220–234 (2015).
 16. Jordan, V. C. The new biology of estrogen-induced apoptosis applied to treat and prevent breast cancer. *Endocr. Relat. Cancer* **22**, R1–R31 (2015).
 17. Ma, C. X., Reinert, T., Chmielewska, I. & Ellis, M. J. Mechanisms of aromatase inhibitor resistance. *Nat. Rev. Cancer* **15**, 261–275 (2015).
 18. Fan, P., Maximov, P. Y., Curpan, R. F., Abderrahman, B. & Jordan, V. C. The molecular, cellular and clinical consequences of targeting the estrogen receptor following estrogen deprivation therapy. *Mol. Cell. Endocrinol.* **418**, 245–263 (2015).
 19. Houtsma, D. et al. The variant T allele of PvuII in ESR1 gene is a prognostic marker in early breast cancer survival. *Sci. Rep.* **11**, 3249 (2021).
 20. Conte, B. et al. SNP of Aromatase Predict Long-term Survival and Aromatase Inhibitor Toxicity in Patients with Early Breast Cancer: A Biomarker Analysis of the GIM4 and GIM5 Trials. *Clin. Cancer Res.* **29**, 5217–5226 (2023).
 21. Zhang, M., Wu, K., Wang, M., Bai, F. & Chen, H. CASP9 As a Prognostic Biomarker and Promising Drug Target Plays a Pivotal Role in Inflammatory Breast Cancer. *Int. J. Anal. Chem.* **2022**, 1043445, <https://doi.org/10.1155/2022/1043445> (2022).
 22. Györfy, B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput. Struct. Biotechnol. J.* **19**, 4101–4109 (2021).
 23. Walz, A. et al. Molecular Pathways: Revisiting Glycogen Synthase Kinase-3 β as a Target for the Treatment of Cancer. *Clin. Cancer Res.* **23**, 1891–1897 (2017).
 24. He, R., Du, S., Lei, T., Xie, X. & Wang, Y. Glycogen synthase kinase 3 β in tumorigenesis and oncotherapy (Review). *Oncol. Rep.* **44**, 2373–2385 (2020).
 25. Quintayo, M. A. et al. GSK3 β and cyclin D1 expression predicts outcome in early breast cancer patients. *Breast Cancer Res. Treat.* **136**, 161–168 (2012).
 26. Uffelmann, E. et al. Genome-wide association studies. *Nat. Rev. Methods Prim.* **1**, 59, <https://doi.org/10.1038/s43586-021-00056-9> (2021).
 27. McInnes, G., Yee, S. W., Pershad, Y. & Altman, R. B. Genomewide Association Studies in Pharmacogenomics. *Clin. Pharmacol. Ther.* **110**, 637–648 (2021).
 28. Pirmohamed, M. Pharmacogenomics: current status and future perspectives. *Nat. Rev. Genet.* **24**, 350–362 (2023).
 29. Cairns, J. et al. Interaction Between SNP Genotype and Efficacy of Anastrozole and Exemestane in Early-Stage Breast Cancer. *Clin. Pharmacol. Ther.* **110**, 1038–1049 (2021).
 30. Ellis, M. J. et al. Lower-dose vs high-dose oral estradiol therapy of hormone receptor-positive, aromatase inhibitor-resistant advanced breast cancer: a phase 2 randomized study. *JAMA* **302**, 774–780 (2009).
 31. Mao, X. et al. Emerging Technologies for the Detection of Cancer Micrometastasis. *Technol. Cancer Res. Treat.* **21**, 15330338221100355, <https://doi.org/10.1177/15330338221100355> (2022).
 32. Cescon, D. W. et al. Therapeutic Targeting of Minimal Residual Disease to Prevent Late Recurrence in Hormone-Receptor Positive Breast Cancer: Challenges and New Approaches. *Front. Oncol.* **11**, 667397, <https://doi.org/10.3389/fonc.2021.667397> (2022).
 33. Goss, P. E. et al. Exemestane versus anastrozole in postmenopausal women with early breast cancer: NCIC CTG MA.27—a randomized controlled phase III trial. *J. Clin. Oncol.* **31**, 1398–1404 (2013).
 34. Hack, C. C. et al. Long-term Follow-up and Safety of Patients after an Upfront Therapy with Letrozole for Early Breast Cancer in Routine Clinical Care - The PreFace Study. *Geburtshilfe Frauenheilkd.* **84**, 185–195 (2024).
 35. Mürdter, T. E. et al. Activity levels of tamoxifen metabolites at the estrogen receptor and the impact of genetic polymorphisms of phase I and II enzymes on their concentration levels in plasma. *Clin. Pharmacol. Ther.* **89**, 708–717 (2011).
 36. Schroth, W. et al. Gene Expression Signatures of BRCAness and Tumor Inflammation Define Subgroups of Early-Stage Hormone Receptor-Positive Breast Cancer Patients. *Clin. Cancer Res.* **26**, 6523–6534 (2020).
 37. Gabrielson, M. et al. Cohort Profile: The Karolinska Mammography Project for Risk Prediction of Breast Cancer (KARMA). *Int. J. Epidemiol.* **46**, 1740–1741g (2017).
 38. Olson, J. E. et al. A comprehensive examination of CYP19 variation and risk of breast cancer using two haplotype-tagging approaches. *Breast Cancer Res. Treat.* **102**, 237–247 (2007).
 39. Amos, C. I. et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol. Biomark. Prev.* **26**, 126–135 (2017).
 40. Guo, Q. et al. Identification of novel genetic markers of breast cancer survival. *J. Natl Cancer Inst.* **107**, djv081 (2015).
 41. Khor, C. C. et al. Cross-Ancestry Genome-Wide Association Study Defines the Extended CYP2D6 Locus as the Principal Genetic Determinant of Endoxifen Plasma Concentrations. *Clin. Pharmacol. Ther.* **113**, 712–723 (2023).
 42. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
 43. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457–470 (2011).
 44. Hudis, C. A. et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J. Clin. Oncol.* **25**, 2127–2132 (2007).
 45. Harrell, F. E. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis (Springer International Publishing: Cham, Switzerland, 2nd ed., 2015).
 46. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
 47. Gong, J. et al. PanCanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res.* **46**, D971–D976 (2018).
 48. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

49. Poirier, J. G. et al. Resampling to Address the Winner's Curse in Genetic Association Analysis of Time to Event. *Genet Epidemiol.* **39**, 518–528 (2015).
50. Hsieh, F. Y. & Lavori, P. W. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Control. Clin. Trials* **21**, 552–560 (2000).

Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. This work was supported by Cancer Research UK grant: PPRPGM-Nov20\100002 and by core funding from the NIHR Cambridge Biomedical Research Centre (NIHR203312) [*]. *The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Additional funding for BCAC is provided by the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health, the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and the PERSPECTIVE I&I project, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation. The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. Genotyping of the OncoArray was funded by the NIH Grant U19 CA148065, and Cancer Research UK Grant C1287/A16563 and the PERSPECTIVE project supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (grant GPH-129344) and, the Ministère de l'Économie, Science et Innovation du Québec through Genome Québec and the PSRSIIRI-701 grant, and the Quebec Breast Cancer Foundation. Funding for iCOGS came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, and Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. Detailed acknowledgements of individual studies with funding information are provided in Supplementary Note 1 in the Supplementary Information.

Author contributions

Conception and study design: R.H., S.W., W.-Y.L., H.B. Provision of data: R.H., S.W., W.-Y.L., K.M., M.K.B., R.K., Q.W., J.D., M.L., K.R.K., M.P.G., L.W., J.C., R.W., L.S., B.E.C., L.H., M.R., M.W.B., W.H., N.L.L., S.M.A., W.S., B.C., C.C.K., M.A., A.C.A., T.B., J.E.C., J.C.C., NBCS Collaborators, T.D., D.M.E., J.D.F., M.G.D., J.A.G.-S., M.G., C.C.H., U.H., S.H., M.J.H., H.H., ABCTB Investigators, E.M.J., Y.-D.K., V.N.K., S.L., S.M., D.M., H.N., P.N., N.O., T.-W.P.-S., K.P., M.U.R., A.R., E.S., E.J.S., W.J.T., I.T., C.W., R.W., A.M.D., J.S., P.H., P.D.P.P., M.S., F.J.C., K.C., P.A.F., D.F.E., M.K.S., J.N.I., H.B. Data base support: M.K.B., R.K., Q.W., J.D., M.L., K.M., M.K.S., D.F.E., K.R.K., B.E.C., J.N.I., L.S., L.H., P.A.F., W.H., P.H., K.C., N.L.L., S.M.A., F.J.C., W.S., S.W., M.S., H.B. Statistical analyses: S.W., K.M., K.R.K., B.E.C., L.H. Interpretation of results: R.H., S.W., H.B., J.N.I., K.R.K., M.P.G., M.K.S. Manuscript writing: The original draft of the manuscript was written by R.H., S.W., and H.B., and all authors contributed to the critical revision and editing

of the final version of the manuscript for publication. All authors read and approved the final manuscript.

Competing interests

Peter A. Fasching reports personal fees from Novartis, Pfizer, Daiichi-Sankyo, AstraZeneca, Eisai, Merck Sharp & Dohme, Lilly, Pierre Fabre, SeaGen, Roche, Hexal, Agendia, and Gilead and grants from Biotech and Cepheid. Carolin C. Hack received Honoraria from AstraZeneca, Daiichi Sankyo, Eisai, Novartis, Pfizer, Roche, Gilead and MSD as well as support for attending meetings from Daiichi Sankyo. Matthias Schwab received funding from Agena Bioscience GmbH, HepaRegenix GmbH, Robert Bosch GmbH, and CORAT Therapeutics GmbH, as well as honoraria for oral presentations at academically organized congresses and meetings; he received payment for expert testimony from the Research Impact Fund Committee (RIF), Research Grant Council (RCG), Hong Kong, and from the German Federal Ministry of Education and Research (BMBF); he is (section-) editor for *Pharmacogenetics and Genomics*, *Drug Research*, and *Genome Medicine*. Diana M. Eccles has received research support from AstraZeneca. Matthew P. Goetz reports consulting fees to Mayo Clinic from ARC Therapeutics, AstraZeneca, Biotheranostics, Blueprint Medicines, Lilly, Novartis, Rna Diagnostics, Sanofi Genzyme, Seattle Genetics, Sermonix, Engage Health Media, Laekna and TerSera Therapeutics/Ampity Health; grant funding to Mayo Clinic from Lilly, Pfizer, Sermonix, Loxo, AstraZeneca and ATOSSA Therapeutics; personal fees for CME activities from Research to Practice, Clinical Education Alliance, Medscape, and MJH Life Sciences; personal fees serving as a panelist for a panel discussion from Total Health Conferencing and personal fees for serving as a moderator for Curio Science; and travel support from Lilly. All other authors report no conflict of interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41523-025-00733-y>.

Correspondence and requests for materials should be addressed to Reiner Hoppe.

Reprints and permissions 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.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Reiner Hoppe^{1,2,75}✉, Stefan Winter^{1,2,75}, Wing-Yee Lo³, Kyriaki Michailidou^{4,5}, Manjeet K. Bolla⁵, Renske Keeman⁶, Qin Wang⁵, Joe Dennis⁵, Michael Lush⁵, Krishna R. Kalari⁷, Matthew P. Goetz⁷, Liewei Wang⁸, Junmei Cairns⁸, Richard Weinshilboum⁸, Lois Shepherd⁹, Bingshu E. Chen⁹, Lothar Häberle¹⁰, Matthias Ruebner¹⁰, Matthias W. Beckmann¹⁰, Wei He¹¹, Nicole L. Larson¹², Sebastian M. Armasu¹³, Werner Schroth^{1,2}, Balram Chowbay^{14,15,16}, Chiea Chuen Khor^{17,18,19}, Mustapha Abubakar²⁰, Antonis C. Antoniou⁵, Thomas Brüning²¹, Jose E. Castelao²², Jenny Chang-Claude^{23,24}, NBCS Collaborators^{25,26,27,28,29,30,31,32,33}, Thilo Dörk³⁴, Diana M. Eccles³⁵, Jonine D. Figueroa^{20,36,37}, Manuela Gago-Dominguez³⁸, José A. García-Sáenz³⁹, Melanie Gündert^{40,41,42}, Carolin C. Hack¹⁰, Ute Hamann⁴³, Sileny Han⁴⁴, Maartje J. Hoening⁴⁵, Hanna Huebner¹⁰, ABCTB Investigators⁴⁶, Esther M. John^{47,48}, Yon-Dschun Ko⁴⁹, Vessela N. Kristensen^{25,27}, Sabine Linn^{6,50,51}, Sara Margolin^{52,53}, Dimitrios Mavroudis⁵⁴, Heli Nevanlinna⁵⁵, Patrick Neven⁵⁶, Nadia Obi^{57,58}, Tjong-Won Park-Simon³⁴, Katri Pylkäs^{59,60}, Muhammad U. Rashid^{43,61}, Atocha Romero⁶², Emmanouil Saloustros⁶³, Elinor J. Sawyer⁶⁴, William J. Tapper³⁵, Ian Tomlinson⁶⁵, Camilla Wendt⁵³, Robert Winqvist⁶⁹, Alison M. Dunning⁶⁶, Jacques Simard⁶⁷, Per Hall^{11,52}, Paul D. P. Pharoah⁶⁸, Matthias Schwab^{1,2,69,70,71}, Fergus J. Couch⁷², Kamila Czene¹¹, Peter A. Fasching¹⁰, Douglas F. Easton^{5,66}, Marjanka K. Schmidt^{6,73,74}, James N. Ingle⁷ & Hiltrud Brauch^{1,2,70,71}

¹Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany. ²University of Tübingen, Tübingen, Germany. ³Department of Clinical Pathology, Centre for Cancer Research, University of Melbourne, Melbourne, Victoria, Australia. ⁴Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus. ⁵Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ⁶Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, the Netherlands. ⁷Department of Oncology, Mayo Clinic, Rochester, MN, USA. ⁸Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA. ⁹Canadian Cancer Trials Group, Queen's University, Kingston, Ontario, Canada. ¹⁰Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg, University Hospital Erlangen, Erlangen, Germany. ¹¹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ¹²Department of Quantitative Health Sciences, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA. ¹³Department of Quantitative Health Sciences, Division of Clinical Trials and Biostatistics, Mayo Clinic, Rochester, MN, USA. ¹⁴Laboratory of Clinical Pharmacology, Division of Cellular & Molecular Research, National Cancer Centre Singapore, Singapore, Singapore. ¹⁵Centre for Clinician Scientist Development, Duke-NUS Medical School, Singapore, Singapore. ¹⁶Singapore Immunology Network, Agency for Science, Technology & Research (A*STAR), Singapore, Singapore. ¹⁷Genome Institute of Singapore (GIS), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. ¹⁸Singapore Eye Research Institute, Singapore, Singapore. ¹⁹Ophthalmology and Visual Sciences Academic Clinical Program, Duke-NUS Medical School, Singapore, Singapore. ²⁰Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA. ²¹Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum, Bochum, Germany. ²²Oncology and Genetics Unit, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS) Foundation, Complejo Hospitalario Universitario de Santiago, SERGAS, Vigo, Spain. ²³Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ²⁴Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ²⁵Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway. ²⁶Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital-Radiumhospitalet, Oslo, Norway. ²⁷Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway. ²⁸Department of Research, Vestre Viken Hospital, Drammen, Norway. ²⁹Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital - Radiumhospitalet, Oslo, Norway. ³⁰Department of Oncology, Division of Surgery and Cancer and Transplantation Medicine, Oslo University Hospital-Radiumhospitalet, Oslo, Norway. ³¹Department of Oncology, Akershus University Hospital, Lørenskog, Norway. ³²Oslo Breast Cancer Research Consortium, Oslo University Hospital, Oslo, Norway. ³³Department of Community Medicine, The Arctic University of Norway, Tromsø, Norway. ³⁴Gynaecology Research Unit, Hannover Medical School, Hannover, Germany. ³⁵Faculty of Medicine, University of Southampton, Southampton, UK. ³⁶Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, UK. ³⁷Cancer Research UK Edinburgh Centre, The University of Edinburgh, Edinburgh, UK. ³⁸Cancer Genetics and Epidemiology Group, Genomic Medicine Group, Fundación Pública Galega de Medicina Xenómica, Fundación Instituto de Investigación Sanitaria de Santiago de Compostela (FIDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain. ³⁹Medical Oncology Department, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Centro Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain. ⁴⁰Molecular Epidemiology Group, C080, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴¹Molecular Biology of Breast Cancer, University Womens Clinic Heidelberg, University of Heidelberg, Heidelberg, Germany. ⁴²Institute of Diabetes Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. ⁴³Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁴⁴Leuven Multidisciplinary Breast Center, Department of Obstetrics and Gynecology, University Hospitals Leuven, Leuven Cancer Institute, Leuven, Belgium. ⁴⁵Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands. ⁴⁶Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia. ⁴⁷Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA. ⁴⁸Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. ⁴⁹Department of Internal Medicine, Johanniter GmbH Bonn, Johanniter Krankenhaus, Bonn, Germany. ⁵⁰Department of Medical Oncology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. ⁵¹Department of Pathology, University Medical Center Utrecht, Utrecht, the Netherlands. ⁵²Department of Oncology, Södersjukhuset, Stockholm, Sweden. ⁵³Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden. ⁵⁴Department of Medical Oncology, University Hospital of Heraklion, Heraklion, Greece. ⁵⁵Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. ⁵⁶Leuven Multidisciplinary Breast Center, Department of Oncology, Leuven Cancer Institute, University Hospitals Leuven, Leuven, Belgium. ⁵⁷Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ⁵⁸Institute for Occupational and Maritime Medicine (ZfAM), University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ⁵⁹Laboratory of Cancer Genetics and Tumor Biology, Translational Medicine Research Unit, Biocenter Oulu, Medical Research Centre, University of Oulu, Oulu, Finland. ⁶⁰Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland. ⁶¹Department of Basic Sciences, Shaikat Khanum Memorial Cancer Hospital and Research Centre (SKMCH & RC), Lahore, Pakistan. ⁶²Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid, Spain. ⁶³Department of Oncology, University Hospital of Larissa, Larissa, Greece. ⁶⁴School of Cancer & Pharmaceutical Sciences, Comprehensive Cancer Centre, Guy's Campus, King's College London, London, UK. ⁶⁵Department of Oncology, University of Oxford, Oxford, UK. ⁶⁶Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. ⁶⁷Genomics Center, Centre Hospitalier Universitaire de Québec—Université Laval Research Center, Québec City, Québec, Canada. ⁶⁸Department of Computational Biomedicine, Cedars-Sinai Medical Center, West Hollywood, CA, USA.

⁶⁹Departments of Clinical Pharmacology, Pharmacy and Biochemistry, University of Tübingen, Tübingen, Germany. ⁷⁰Cluster of Excellence iFIT (EXC2180) “Image-Guided and Functionally Instructed Tumor Therapies”, University of Tübingen, Tübingen, Germany. ⁷¹German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Tübingen, Tübingen, Germany. ⁷²Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. ⁷³Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. ⁷⁴Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands. ⁷⁵These authors contributed equally: Reiner Hoppe, Stefan Winter. ✉ e-mail: reiner.hoppe@ikp-stuttgart.de