

Dynamics of breast–cancer relapse reveal late–recurring ER–positive genomic subgroups

Oscar M. Rueda¹, Stephen–John Sammut^{1,13}, Jose A. Seoane^{2,3,4,13}, Suet–Feung Chin¹, Jennifer L. Caswell–Jin², Maurizio Callari¹, Rajbir Batra¹, Bernard Pereira¹, Alejandra Bruna¹, H. Raza Ali¹, Elena Provenzano^{5,6}, Bin Liu¹, Michelle Parisien⁷, Cheryl Gillett⁸, Steven McKinney⁹, Andrew R. Green¹⁰, Leigh Murphy⁷, Arnie Purushotham⁸, Ian O. Ellis¹⁰, Paul D. Pharoah^{1,5,6,11}, Cristina Rueda¹², Samuel Aparicio⁹, Carlos Caldas^{1,5,6*} & Christina Curtis^{2,3,4*}

The rates and routes of lethal systemic spread in breast cancer are poorly understood owing to a lack of molecularly characterized patient cohorts with long-term, detailed follow-up data. Long-term follow-up is especially important for those with oestrogen-receptor (ER)-positive breast cancers, which can recur up to two decades after initial diagnosis^{1–6}. It is therefore essential to identify patients who have a high risk of late relapse^{7–9}. Here we present a statistical framework that models distinct disease stages (locregional recurrence, distant recurrence, breast-cancer-related death and death from other causes) and competing risks of mortality from breast cancer, while yielding individual risk-of-recurrence predictions. We apply this model to 3,240 patients with breast cancer, including 1,980 for whom molecular data are available, and delineate spatiotemporal patterns of relapse across different categories of molecular information (namely immunohistochemical subtypes; PAM50 subtypes, which are based on gene-expression patterns^{10,11}; and integrative or IntClust subtypes, which are based on patterns of genomic copy-number alterations and gene expression^{12,13}). We identify four late-recurring integrative subtypes, comprising about one quarter (26%) of tumours that are both positive for ER and negative for human epidermal growth factor receptor 2, each with characteristic tumour-driving alterations in genomic copy number and a high risk of recurrence (mean 47–62%) up to 20 years after diagnosis. We also define a subgroup of triple-negative breast cancers in which cancer rarely recurs after five years, and a separate subgroup in which patients remain at risk. Use of the integrative subtypes improves the prediction of late, distant relapse beyond what is possible with clinical covariates (nodal status, tumour size, tumour grade and immunohistochemical subtype). These findings highlight opportunities for improved patient stratification and biomarker-driven clinical trials.

Breast cancer is a multistate disease with clinically relevant intermediate end points, such as locoregional recurrence and distant recurrence¹⁴. A patient's prognosis can differ considerably depending on when and where a relapse occurs, time since surgery, and time since locoregional or distant recurrence^{15,16}. These events are associated, and individual analyses of disease-free survival (DFS) or overall survival alone cannot fully capture patterns of recurrence associated with differential prognosis. In addition, most survival analyses use disease-specific death (DSD) as the primary end point, and censor deaths from other causes. However, when competing risks of mortality occur, this approach induces bias¹⁷. This is particularly problematic for breast cancer, where ER-positive patients experience higher mortality from

nonmalignant causes owing to their increased age at diagnosis relative to ER-negative patients.

We evaluated the extent of such bias on breast-cancer survival estimates by analysing 3,240 patients who had been diagnosed between 1977 and 2005, and for whom there was a median clinical follow-up of 14 years (referred to as the 'full dataset'; see Extended Data Fig. 1, Supplementary Table 1 and Methods). We compared the naive cumulative incidence for DSD (computed as one minus the survival probability), stratified by ER status and considering only cancer-related deaths (Extended Data Fig. 2a), relative to estimates with the proper cumulative incidence functions accounting for different causes of death (Extended Data Fig. 2b). These comparisons indicate that the incidence of DSD is overestimated for ER-positive tumours relative to ER-negative tumours (0.46 versus 0.37 at 20 years) owing to the increased age of diagnosis (median 63.9 versus 53.0 years; $P < 1 \times 10^{-6}$; Extended Data Fig. 2c). Moreover, because the baseline survival functions for these subgroups are distinct, their differences cannot be adequately summarized with a single parameter in a Cox proportional hazards model.

To overcome these limitations, we developed a nonhomogenous (semi)-Markov-chain model that accounts for different disease states (locregional recurrence and distant recurrence) and time scales (time since surgery or locoregional or distant recurrence), as well as competing risks of mortality and distinct baseline hazards across molecular subgroups, thereby enabling individual risk-of-relapse predictions (see Fig. 1a and Methods). The model also incorporates clinical variables known to influence breast-cancer survival^{18,19}, including age, tumour grade, tumour size and number of tumour-positive ('positive') lymph nodes (all measured at diagnosis). We refer to this as the base clinical model, into which information on molecular subtype can be incorporated.

We fitted this multistate model to the full dataset, and recorded the hazards of moving through distinct states and the number of transitions between each pair of states (Supplementary Table 2 and Methods). As expected, most cancer-related deaths (83% in ER-positive and 87% in ER-negative tumours) occurred after distant metastasis. The remainder of the cases probably reflect undetected recurrences or death due to other malignancies. Age at diagnosis was associated with a transition to death by other causes ($P < 1 \times 10^{-6}$). Examination of the log hazard ratios and 95% confidence intervals for all other variables indicated that their effect decreased with disease progression (Extended Data Fig. 2d). That is, clinical variables related to the primary tumour were more prognostic for earlier transitions than for

¹Cancer Research UK Cambridge Institute and Department of Oncology, Li Ka Shing Centre, University of Cambridge, Cambridge, UK. ²Department of Medicine, Division of Oncology, Stanford University School of Medicine, Stanford, CA, USA. ³Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. ⁴Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. ⁵Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust, Cambridge, UK. ⁶NIHR Cambridge Biomedical Research Centre and Cambridge Experimental Cancer Medicine Centre, Cambridge University Hospital NHS Foundation Trust, Cambridge, UK. ⁷Research Institute in Oncology and Hematology, Winnipeg, Manitoba, Canada. ⁸NIHR Comprehensive Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and Research Oncology, Cancer Division, King's College London, London, UK. ⁹Department of Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia, Canada. ¹⁰Division of Cancer and Stem Cells, School of Medicine, University of Nottingham and Nottingham University Hospital NHS Trust, Nottingham, UK. ¹¹Strangeways Research Laboratory, University of Cambridge, Cambridge, UK. ¹²Departamento de Estadística e Investigación Operativa, Universidad de Valladolid, Valladolid, Spain. ¹³These authors contributed equally: Stephen–John Sammut, Jose A. Seoane. *e-mail: Carlos.Caldas@cruk.cam.ac.uk; cncurtis@stanford.edu

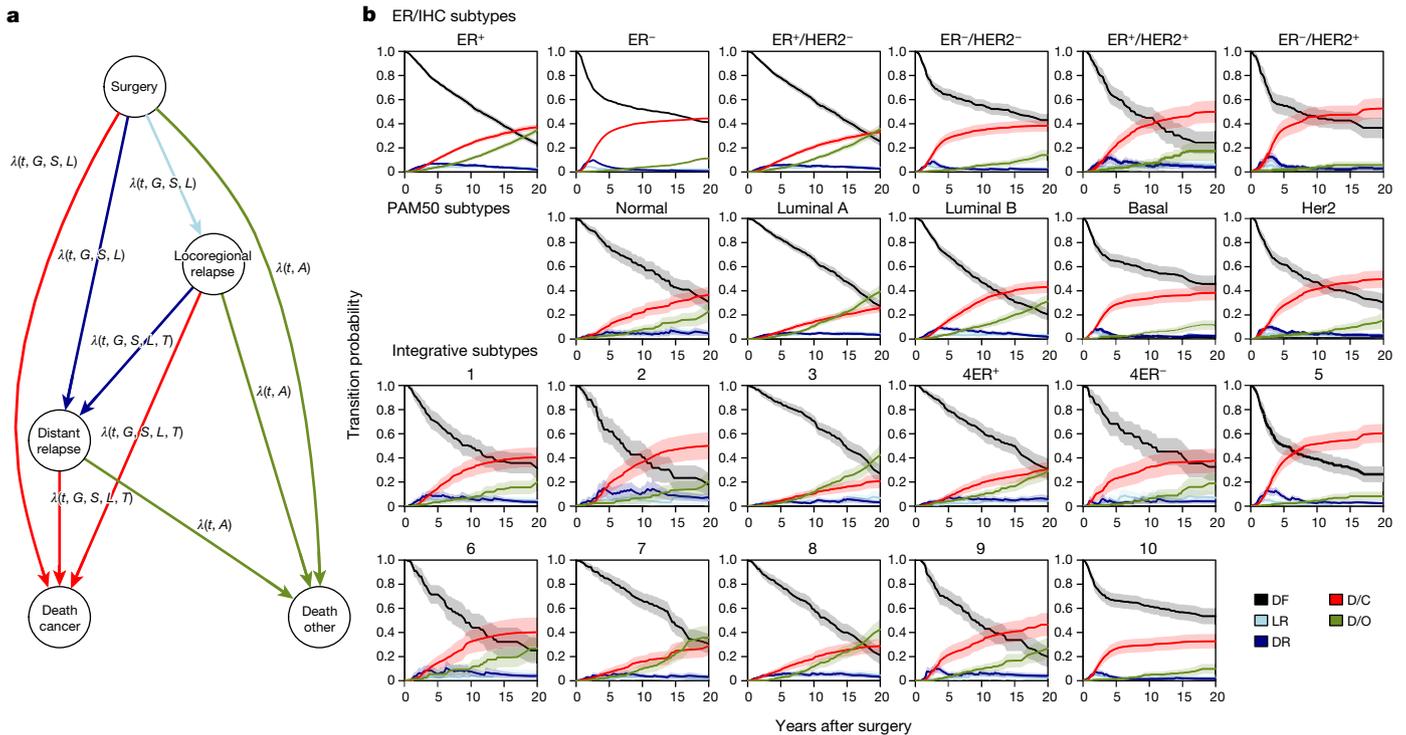


Fig. 1 | A multistate model of breast-cancer relapse enables individual risk-of-relapse predictions throughout disease progression. **a**, Graphical representation of the model. Nodes represent possible states and arcs represent possible transitions between states, where parameters that have an effect on the hazard are indicated. *A*, patient age; *G*, tumour grade; *L*, number of lymph nodes; *S*, tumour size; *t*, time since state entry; *T*, time since surgery. $\lambda()$ is the hazard function. **b**, Subtype-specific risks of relapse at the time of diagnosis. Transition probabilities from surgery to

other states (DF, disease-free; LR, locoregional relapse; DR, distant relapse; D/C, cancer-specific death; D/O, death by other causes) are shown for individual average patients across the indicated breast-cancer subtypes. Subtypes were defined on the basis of ER status using the full dataset, and for IHC, PAM50 and integrative (IntClust) subtypes using the molecular dataset. 95% confidence bands (shaded areas) were computed using bootstrapping (see Methods).

later transitions. However, several tumour characteristics informed the risk of progression from locoregional to distant recurrence, and from distant recurrence to death. In ER-positive disease, higher tumour grade, number of positive lymph nodes and tumour size all increased the risk of progression to a later state. A longer time between surgery and locoregional or distant recurrence decreased the risk of transition to a later state, and was more pronounced in ER-negative disease. We confirmed that our models were well-calibrated, concordant with the established tool PREDICT¹⁸, and that they performed comparably in external datasets (see Extended Data Figs. 1, 3, Methods and Supplementary Information).

A notable feature of our multistate model is that hazard rates can be transformed into transition probabilities that represent the probability of moving from one state into another after a given time. To evaluate patterns of recurrence across established breast-cancer molecular subgroups, we turned to the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) molecular dataset. This dataset is based on 1,980 patients (Extended Data Fig. 1), and includes assignments to: immunohistochemistry (IHC) subtypes (namely ER⁺/HER2⁺, ER⁺/HER2⁻, ER⁻/HER2⁺ and ER⁻/HER2⁻, where HER2 is human epidermal growth factor receptor 2); 5 intrinsic gene-expression subtypes (otherwise known as PAM50 subtypes)¹¹; and the 11 Integrative Cluster (IntClust) subtypes, which are characterized by distinct copy-number and gene-expression profiles^{12,13} (Supplementary Table 3). We computed the baseline transition probabilities from surgery, locoregional recurrence or distant recurrence at various time intervals (2, 5, 10, 15 and 20 years) and the corresponding standard errors of prediction (s.e.) for average individuals in each subgroup (using the full dataset for comparisons by ER status, and the molecular dataset for all others; Supplementary Table 4).

After surgery, state transitions differed substantially across the various subtypes (Fig. 1b). For example, the transition probabilities

post-surgery reveal different change points for ER-positive versus ER-negative disease. ER-negative patients had a higher risk of distant recurrence and death from cancer in the first five years, after which their risk decreased considerably. By contrast, ER-positive patients had a smaller but longer risk period during the first ten years, and the risk increased at a lower rate. Among ER-negative patients, the PAM50 ‘basal-like’ subgroup was nearly indistinguishable from the ER⁻/HER2⁻ subgroup, with most cancer-related deaths occurring in the first five years—similar to HER2⁺ patients (before the widespread use of trastuzumab). By contrast, the three predominantly ER-negative IntClust subgroups (IntClust4ER⁻, IntClust5 and IntClust10) exhibited substantial differences in their recurrence trajectories. As expected, IntClust5 (HER2⁺ enriched) generally had poor prognosis at 5 years (transition probability to relapse/cancer-related death 0.48; s.e. = 0.04), with the risk increasing to 0.65 (s.e. = 0.04) at 20 years. For IntClust10 (‘basal-like’ enriched), the first 5 years after surgery largely defined patient outcomes: the probability of relapse or cancer-related death at 5 years was 0.33 (s.e. = 0.03), and this rose after 20 years only to 0.37 (s.e. = 0.04) for an average patient. This pattern was distinct from that seen for IntClust4ER⁻ patients, who exhibited a persistent and increasing risk of relapse or cancer-related death with a probability of 0.30 (0.05) at 5 years and 0.49 (0.05) after 20 years.

The distinction between IntClust4ER⁻ and IntClust10 is also apparent when examining the average probabilities of relapse among all patients across the IntClust subtypes after surgery or after being disease-free for five and ten years (Fig. 2a). Indeed, through the course of the disease, the risk of relapse changed considerably across the integrative subtypes, and to a lesser extent across the IHC and PAM50 subtypes (Fig. 2a and Extended Data Fig. 4). Moreover, the probabilities of distant recurrence or cancer-related death among ER⁻/HER2⁻ patients who were disease-free at five years after diagnosis revealed low (IntClust10) and high (IntClust4ER⁻) risks for late-relapse

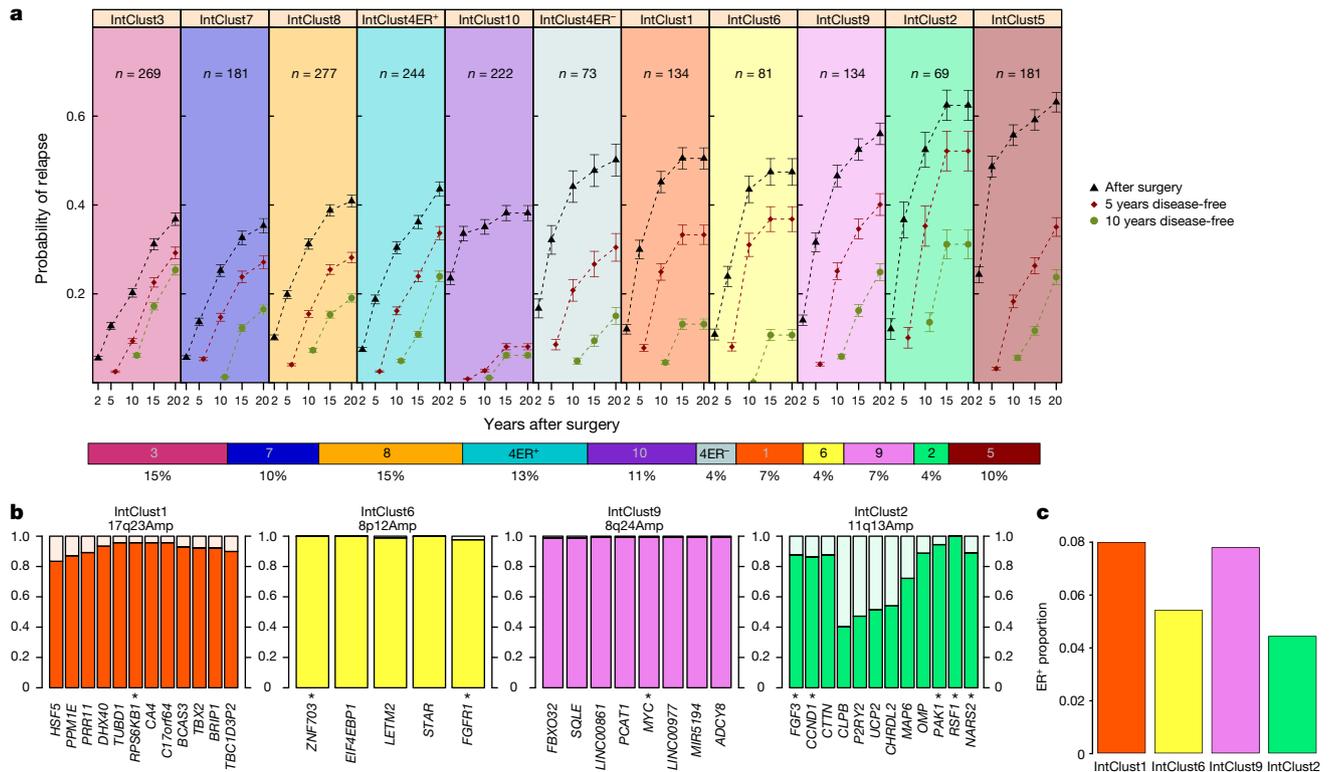


Fig. 2 | The integrative breast-cancer subtypes exhibit distinct patterns of relapse. **a**, Mean probabilities of relapse after surgery and after 5 and 10 disease-free years for the patients in each of the 11 integrative (IntClust) subtypes, ordered by increasing risk of relapse. IntClust3, IntClust7, IntClust8 and IntClust4ER⁺ represent lower-risk ER-positive subtypes; IntClust10 and IntClust4ER⁻ represent TNBC subtypes with variable relapse patterns; IntClust1, IntClust6, IntClust9 and IntClust2 represent late-relapsing ER-positive subtypes; and IntClust5 represents HER2-positive tumours before widespread use of trastuzumab. Error bars show

95% confidence intervals. The lower coloured bar shows the prevalence of each integrative subtype in the breast-cancer population. **b**, Frequencies of copy-number amplifications in specific IntClust subtypes (IntClust1, IntClust6, IntClust9 and IntClust2). Putative driver genes are indicated by asterisks. The amplified chromosomal regions are indicated above each chart. **c**, Proportion of ER-positive tumours that belong to the four late-relapsing IntClust subtypes. This analysis was done with the molecular dataset.

triple-negative breast-cancer (TNBC) subgroups, whereas IHC (and PAM50) subtypes homogenized this risk (Extended Data Fig. 5).

Marked differences were also apparent among ER-positive patients, with patients with IntClust3, IntClust7, IntClust8 and IntClust4ER⁺ subtypes exhibiting a better prognosis, whereas patients with IntClust1, IntClust2, IntClust6 and IntClust9 subtypes exhibited late-recurring cancer with a poor prognosis (Fig. 2a). These latter four subgroups had an exceedingly high risk of relapse, with mean probabilities ranging from 0.47 to 0.62 up to 20 years after surgery. The IntClust2 subtype exhibited the worst prognosis, with a probability of relapse (0.62; s.e.m. = 0.02) second only to that of IntClust5. Collectively, these subgroups comprise 26% of ER-positive cases (Fig. 2b, c) and thus define the minority of patients who may benefit from extended monitoring and treatment given the chronic nature of their disease^{5,6}.

Importantly, the four 'high risk of relapse' subgroups were enriched in characteristic genomic-copy-number alterations, which represent the likely drivers of each subgroup (Fig. 2b). For example, IntClust2 tumours are defined by amplification and concomitant overexpression of multiple oncogenes on chromosome 11q13, including *CCND1*, *FGF3*, *EMSY*, *PAK1* and *RSF1* (refs 20–22). IntClust2 accounts for 4.5% of ER-positive cases, 96% of which have *RSF1* amplification, compared with 0–22% in other subgroups. IntClust6 (5.5% of ER-positive tumours) is characterized by focal amplification of *ZNF703* (ref. 23) and *FGFR1* (ref. 24) on chromosome 8p12 (100% of IntClust6 cases versus 2–21% of others). IntClust1 (8% of ER-positive tumours) exhibited amplification of chromosome 17q23 in a region spanning the mTOR effector *RPS6KB1* (also known as *S6K1*)²⁵, which was gained or amplified in 96% and 70% of cases, respectively (versus amplification

in 0–25% of other subtypes). IntClust9 accounted for another 8% of ER-positive cases and was characterized by amplification of the *MYC* oncogene at chromosome 8q24, with amplification in 89% of these tumours (versus 3–42% of other groups). Thus the late-recurring ER-positive subgroups are defined by genomic drivers, several of which are viable therapeutic targets^{25–27}.

Similar differences in the probability of late, distant relapse were seen in the subset of patients whose tumours were ER⁺/HER2⁻ (Fig. 3a, b and Extended Data Fig. 4a–f)—a group in which late relapse and strategies to target this, such as extended endocrine therapy, represent critical clinical challenges. In particular, the probabilities of distant recurrence or cancer-related death reveal a significant risk for IntClust subtypes 1, 2, 6 and 9 (relative to IntClust3) that varied over time. Moreover, the risk was not fully captured by a model that included IHC subtype together with clinical variables (age, tumour size, grade, number of positive lymph nodes and time since surgery) that have been shown to dictate distant-relapse outcomes even after a long disease-free interval⁵ (Fig. 3a). We therefore assessed whether the integrative subtypes provide information about a patient's risk of late distant relapse above and beyond what could be inferred optimally from standard clinical information. We found that the model including clinical variables combined with IHC subtype provided substantial information about the probability of distant relapse in ER⁺/HER2⁻ patients who were relapse free at five years. The concordance index (*C*-index) predicting the risk of distant relapse was 0.63 (confidence interval 0.58–0.68) at 10 years, 0.62 (0.58–0.67) at 15 years, and 0.61 (0.57–0.66) at 20 years (Fig. 3c). However, including the IntClust subtypes significantly improved the predictive value: the *C*-index was 0.70 (confidence interval 0.64–0.75;

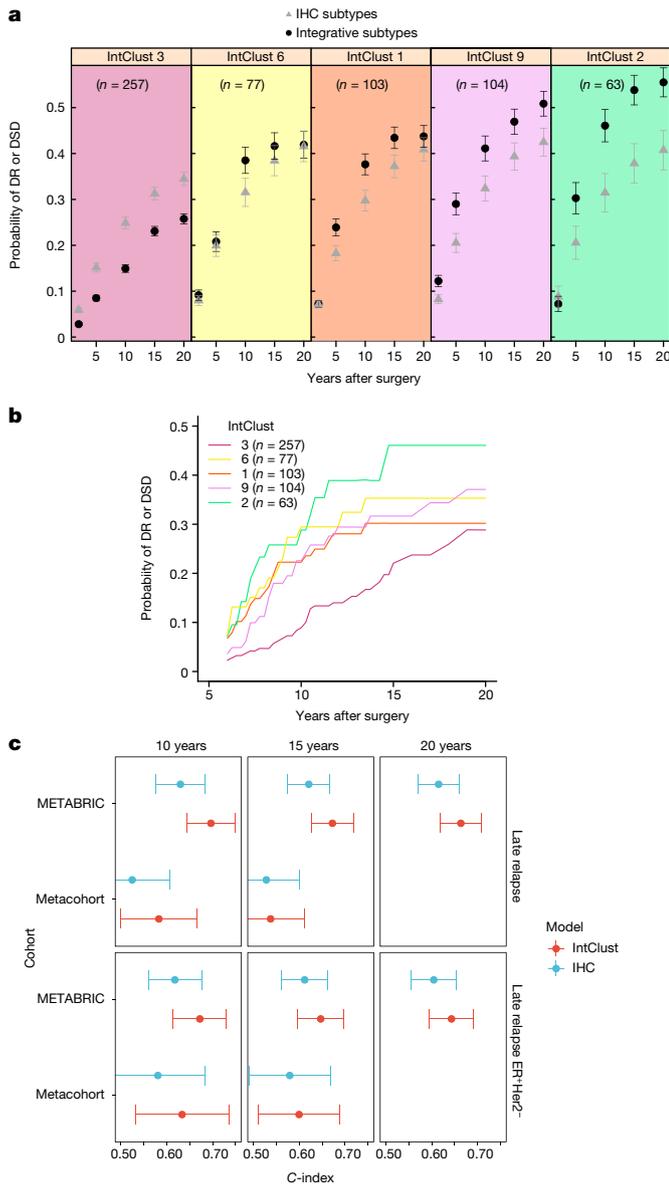


Fig. 3 | The integrative subtypes improve prediction of late, distant recurrence in ER⁺/HER2⁻ breast cancer beyond clinical covariates.

a, Probabilities of distant relapse or DSD among ER⁺/HER2⁻ patients reveal a substantial risk for IntClust subtypes 1, 2, 6 and 9 relative to IntClust3; this risk varies over time and is not captured by the standard clinical model. Data points represent average probabilities and error bars 95% confidence intervals. **b**, Average probabilities of distant relapse or DSD for ER⁺/HER2⁻ patients in the four late-relapsing subgroups relative to IntClust3 for patients who were relapse free at five years after diagnosis. **c**, Evaluation of the utility of the IHC model relative to the IntClust model for predicting late distant relapse in ER⁺/HER2⁻ patients who were relapse free at five years. C-indices are shown for both models at different time intervals in the METABRIC cohort (total $n = 1,337$; ER⁺/HER2⁻ $n = 1,013$) and the external validation metacohort (total $n = 1,080$; ER⁺/HER2⁻ $n = 739$). Error bars represent 95% confidence intervals. This analysis was done with the molecular dataset.

improvement over the clinical model, $P = 0.00011$) at 10 years, 0.67 (0.63–0.72; $P = 0.0016$) at 15 years and 0.66 (0.62–0.71; $P = 0.0017$) at 20 years. These trends were recapitulated in an external validation cohort, despite the smaller sample size and shorter follow-up times (prohibiting analyses at 20 years) (Fig. 3c and Extended Data Fig. 3e). Thus, information about the dynamics of late relapse that is provided by integrative subtype could not be inferred from standard clinical variables, including IHC subtype.

We next turned to the subset of patients who experienced a locoregional recurrence. Such a relapse is commonly treated with curative intent, and is thought to be a high-risk event that is associated with increased rates (45–80%) of distant relapse²⁸. The transition probabilities after locoregional recurrence varied substantially depending on the pathological features of the primary tumour at diagnosis and the molecular subtype, highlighting opportunities for intervention (Extended Data Figs. 6, 7 and Supplementary Tables 2, 3). By contrast, after the initial distant relapse, all subgroups exhibited a high probability of cancer-related death, although the median times differed (Extended Data Fig. 8 and Supplementary Tables 2, 3).

Unique to our cohort is a subset of 618 patients (out of the 1,079 from the full dataset who relapsed) for whom a complete description of all recurrences is available (this is the recurrent-event dataset). This enables a detailed analysis of the rates and routes of distant metastases and their lethality. These data reveal the varied time course over which metastases occurred and indicate that no sites of metastasis are exclusive to ER-positive or ER-negative disease (Extended Data Fig. 9a). Moreover, multiple distant metastases were common, even among subgroups with a favourable prognosis (Extended Data Fig. 9b). We next examined the cumulative incidence and number of metastases at different organ sites stratified by ER status (Fig. 4a). ER-negative cases harboured substantially more visceral disease than did ER-positive cases (for example, brain/meningeal, 27% versus 11%; pulmonary, 50% versus 41%). As previously reported^{29,30}, bone metastases were more common in ER-positive than in ER-negative cases (71% versus 43%), but the cumulative incidence was similar. Thus, the higher proportions observed in ER-positive disease appear not to reflect site-specific tropism: rather, bone metastases take a long time to develop, and ER-negative patients tend to die of other metastases first. In addition, ER-positive tumours more commonly presented with a first metastasis in the bone (76% versus 61%). Similar comparisons stratified by IHC, PAM50 and IntClust subtypes reveal additional variability (Extended Data Fig. 10). Striking differences in the rates of distant metastasis are also evident: ER-negative disease was characterized by a rapid series of relapses early after diagnosis, while most ER-positive patients suffered just one early relapse (commonly bone metastases), and if a second relapse occurred, the probability of additional relapses increased (Fig. 4b and Methods). Thus, after distant recurrence, subtype continues to dictate the rate of subsequent metastases, underscoring the importance of tumour biology. Both the number and the site of relapses influenced the risk of death after recurrence, with brain metastasis being most predictive. Risk estimates (Fig. 4c) were comparable between ER-positive and ER-negative tumours, suggesting that the impact of the site of metastasis on progression to death is similar.

In summary, by leveraging a cohort of 3,240 patients—including 1,980 from METABRIC, for whom detailed molecular characterization and recurrence data are available—we have delineated the spatio-temporal dynamics of breast-cancer relapse at a high resolution. Our analyses are based on a multistate statistical model that yields individual risk-of-relapse estimates, using tumour features, clinical, pathological and molecular covariates, and disease chronology, and is available via a web application (<https://caldaslab.cruk.cam.ac.uk/brcarepred>). In contrast to existing models used to calculate the benefits of adjuvant therapy at diagnosis, such as PREDICT¹⁸, our research tool can be used to assess how a patient's risk of recurrence changes throughout follow-up. Learning whether specific treatments change the outcomes of different integrative subtypes is important and will require analysis of randomized clinical trial cohorts.

By classifying breast tumours into the 11 integrative subtypes, important differences in recurrence rates have become apparent that were obscured in the IHC and PAM50 subtypes. Among TNBC patients, the IntClust10 cluster remains largely relapse free after five years, whereas IntClust4ER⁻ patients continue to be at a substantial risk of recurrence. Among ER⁺/HER2⁻ patients, IntClust subtypes 1, 2, 6 and 9 have a markedly increased risk of distant relapse up to 20 years after diagnosis, and together account for around one quarter of all ER-positive tumours

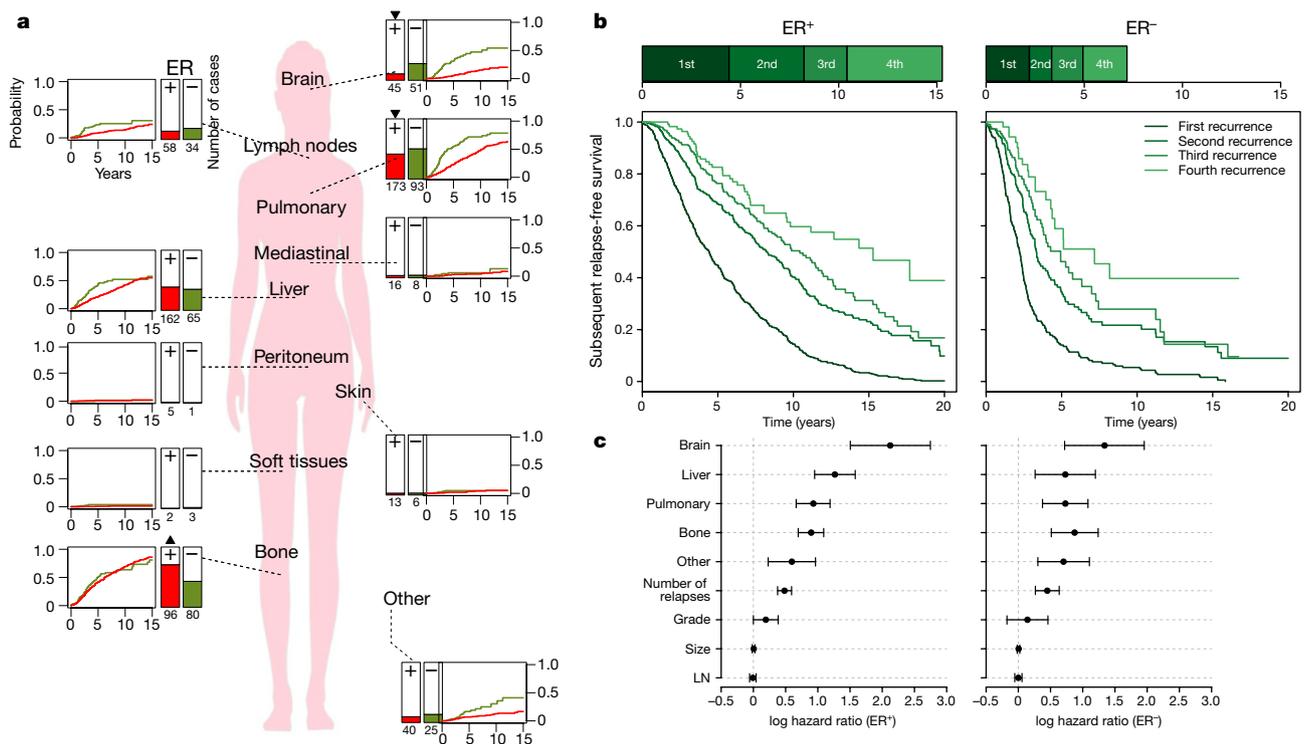


Fig. 4 | Organ-specific patterns and timing of distant relapse in ER-positive and ER-negative patients. **a**, Percentages of patients (bar plots) and cumulative incidence (lines; 1 – Kaplan–Meier estimates) for each site of metastasis in ER-positive and ER-negative cases. Total numbers are also shown. Upright triangles indicate significant positive differences and inverted triangles indicate significant negative differences in that group with respect to the overall mean (see Methods; P values were computed on the basis of logistic regression and are as follows: bone, 2.283×10^{-10} ; brain/meningeal, 3.983×10^{-7} ; pulmonary, 0.0456). **b**, Relapse-free survival curves for sequential recurrences in ER-negative ($n = 186$) and

ER-positive ($n = 419$) patients, computed using a conditional Prentice, Williams and Peterson (PWP) model. Each curve shows the probability of not having another relapse for individuals who have had a previous relapse. The top bar shows the median time (in years) until the n th relapse. **c**, log hazard ratios of DSD, with 95% confidence intervals, from the time-dependent Cox model for distant relapse in ER-negative ($n = 179$) and ER-positive ($n = 410$) patients. This analysis was done with the recurrent-event dataset. Female silhouette is from the public-domain human body diagrams at https://commons.wikimedia.org/wiki/Human_body_diagrams.

and the vast majority of late recurrences. Moreover, the integrative subtypes markedly improved the prediction of distant recurrence after five years in ER⁺/HER2⁻ patients. Our findings thus address one of the contemporary challenges in breast oncology, namely identification of the subset of ER-positive patients who have a high risk of recurrence and tumour biomarkers that are more predictive of recurrence than are standard clinical covariates^{7,8}. Integrative subtyping may help to determine whether women who are relapse free five years after diagnosis might benefit from extended endocrine therapy or other interventions to improve late outcomes. Critically, the four late-recurring ER-positive subgroups are enriched for genomic-copy-number driver alterations that can be therapeutically targeted^{24–27}, paving the way for new treatment strategies for these high-risk patient populations.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-019-1007-8>.

Received: 4 July 2018; Accepted: 31 January 2019;
Published online 13 March 2019.

- Blows, F. M. et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med.* **7**, e1000279 (2010).
- Davies, C. et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet* **381**, 805–816 (2013).
- Sestak, I. et al. Factors predicting late recurrence for estrogen receptor-positive breast cancer. *J. Natl Cancer Inst.* **105**, 1504–1511 (2013).

- Sgrosi, D. C. et al. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. *Lancet Oncol.* **14**, 1067–1076 (2013).
- 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).
- Dowsett, M. et al. Integration of clinical variables for the prediction of late distant recurrence in patients with estrogen receptor-positive breast cancer treated with 5 years of endocrine therapy: CTS5. *J. Clin. Oncol.* **36**, 1941–1948 (2018).
- Harris, L. N. et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology clinical practice guideline. *J. Clin. Oncol.* **34**, 1134–1150 (2016).
- Sledge, G. W. et al. Past, present, and future challenges in breast cancer treatment. *J. Clin. Oncol.* **32**, 1979–1986 (2014).
- Richman, J. & Dowsett, M. Beyond 5 years: enduring risk of recurrence in oestrogen receptor-positive breast cancer. *Nat. Rev. Clin. Oncol.* **1**, <https://doi.org/10.1038/s41571-018-0145-5> (2018).
- Perou, C. M. et al. Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Parker, J. S. et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
- Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
- Ali, H. R. et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. *Genome Biol.* **15**, 431 (2014).
- Putter, H., van der Hage, J., de Bock, G. H., Elgatal, R. & van de Velde, C. J. H. Estimation and prediction in a multi-state model for breast cancer. *Biom. J.* **48**, 366–380 (2006).
- Fisher, B. et al. Significance of ipsilateral breast tumour recurrence after lumpectomy. *Lancet* **338**, 327–331 (1991).
- Insa, A. et al. Prognostic factors predicting survival from first recurrence in patients with metastatic breast cancer: analysis of 439 patients. *Breast Cancer Res. Treat.* **56**, 67–78 (1999).
- Putter, H., Fiocco, M. & Geskus, R. B. Tutorial in biostatistics: competing risks and multi-state models. *Stat. Med.* **26**, 2389–2430 (2007).
- Wishart, G. C. et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res.* **12**, R1 (2010); erratum **12**, 401 (2010).

19. Michaelson, J. S. et al. Improved web-based calculators for predicting breast carcinoma outcomes. *Breast Cancer Res. Treat.* **128**, 827–835 (2011).
20. Ormandy, C. J., Musgrove, E. A., Hui, R., Daly, R. J. & Sutherland, R. L. Cyclin D1, EMS1 and 11q13 amplification in breast cancer. *Breast Cancer Res. Treat.* **78**, 323–335 (2003).
21. Sanchez-Garcia, F. et al. Integration of genomic data enables selective discovery of breast cancer drivers. *Cell* **159**, 1461–1475 (2014).
22. Shrestha, Y. et al. PAK1 is a breast cancer oncogene that coordinately activates MAPK and MET signaling. *Oncogene* **31**, 3397–3408 (2012).
23. Holland, D. G. et al. ZNF703 is a common luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. *EMBO Mol. Med.* **3**, 167–180 (2011).
24. Reis-Filho, J. S. et al. FGFR1 emerges as a potential therapeutic target for lobular breast carcinomas. *Clin. Cancer Res.* **12**, 6652–6662 (2006).
25. Liu, H. et al. Pharmacologic targeting of S6K1 in PTEN-deficient neoplasia. *Cell Reports* **18**, 2088–2095 (2017).
26. Delmore, J. E. et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904–917 (2011).
27. Pearson, A. et al. High-level clonal FGFR amplification and response to FGFR inhibition in a translational clinical trial. *Cancer Discov.* **6**, 838–851 (2016).
28. Wapnir, I. L. et al. A randomized clinical trial of adjuvant chemotherapy for radically resected locoregional relapse of breast cancer: IBCSG 27-02, BIG 1-02, and NSABP B-37. *Clin. Breast Cancer* **8**, 287–292 (2008).
29. Clark, G. M., Sledge, G. W. Jr, Osborne, C. K. & McGuire, W. L. Survival from first recurrence: relative importance of prognostic factors in 1,015 breast cancer patients. *J. Clin. Oncol.* **5**, 55–61 (1987).
30. Kennecke, H. et al. Metastatic behavior of breast cancer subtypes. *J. Clin. Oncol.* **28**, 3271–3277 (2010).

Acknowledgements We thank the women who participated in this study and the UK Cancer Registry. O.M.R. was supported by a Cancer Research UK (CRUK) travel grant (SWAH/047) to visit C. Curtis' laboratory. C.R. is supported by award MTM2015-71217-R. C. Caldas is supported by ECMC, NIH, the Mark Foundation for Cancer Research and Cancer Research UK Cambridge Centre (C9685/A25177). C. Curtis is supported by the National Institutes of Health through the NIH Director's Pioneer Award (DP1-CA238296), the American Association for Cancer Research and the Breast Cancer Research Foundation. This study is dedicated to J.M.W. and J.N.W.

Reviewer information Nature thanks Jeff Gerold, Martin A. Nowak, Peter Van Loo and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions O.M.R., C. Caldas and C. Curtis conceived the study. O.M.R. performed statistical analyses and implemented the model. J.A.S. compiled the validation cohort and performed statistical analyses. S.-J.S. led the annotation of clinical samples, with input from S.-F.C., M.C., R.B., B.P., A.B., H.R.A., E.P., B.L., M.P., C.G., S.M., A.R.G., L.M., A.P., I.O.E., S.A. and C. Caldas. A.R.G., L.M., A.P., I.O.E., S.A. and C. Caldas provided data. P.D.P. and C.R. provided statistical advice. C. Caldas and S.A. are METABRIC principal investigators. O.M.R., J.A.S., J.L.C.-J., C. Caldas and C. Curtis interpreted the results. O.M.R., J.L.C.-J., C. Caldas and C. Curtis wrote the manuscript, which was approved by all authors. C. Caldas and C. Curtis supervised the study.

Competing interests S.A. is founder and shareholder of Contextual Genomic and a scientific advisor to Sangamo Biosciences and Takeda Pharmaceuticals. C. Caldas is a scientific advisor to AstraZeneca-iMed and has received research funding from AstraZeneca, Servier and Genentech/Roche. C. Curtis is a scientific advisory board member and shareholder of GRAIL and consultant for GRAIL and Genentech. A patent application has been filed on aspects of the described work, entitled 'Methods of treatment based upon molecular characterization of breast cancer' (C. Curtis, C. Caldas, J.A.S. and O.M.R.).

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41586-019-1007-8>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1007-8>.

Reprints and permissions information is available at <http://www.nature.com/reprints>.

Correspondence and requests for materials should be addressed to C. Caldas or C. Curtis.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

Clinical cohort. We used data from 3,240 patients who were diagnosed between 1977 and 2005 (with a median follow-up time of 9.77 years overall, and of 14 years among patients who remain alive), derived from five tumour banks in the UK and Canada. Primary breast tumours and linked pseudo-anonymized clinical data were obtained with ethical approval from the relevant institutional review boards. The METABRIC study protocol was approved by the ethics committees at the University of Cambridge and the British Columbia Cancer Research Centre. Manual curation and basic quality control were performed on the data. Observations that had relapse times equal to zero or relapse times equal to the last observed time were shifted by 0.1 days. Local relapses that occurred after distant relapses were omitted. In total, 11 cases with stage 4 tumours and 6 benign and phylloid tumours were removed from all analyses. Last follow-up time or time of death was the final end point for all patients. Special care was taken to remove second primary tumours from the dataset. Clinical parameters, such as tumour grade, were not centrally reviewed, which can lead to variability in the estimation of their effects.

Samples were allocated to three datasets, depending on the information available. For the full dataset cohort, clinical and pathological variables are available (15,394 transitions from 3,147 patients). For a subset of 1,980 patients, we previously described an integrated genomic analysis based on gene-expression and copy-number data¹², and refer to this as the molecular dataset or METABRIC molecular dataset (9,512 transitions from 1,962 patients). For this cohort, tumours were stratified on the basis of the IHC subtypes (ER⁺/HER2⁺, ER⁺/HER2⁻, ER⁻/HER2⁺, ER⁻/HER2⁻), the intrinsic subtypes (PAM50)^{10,11} and the integrative (IntClust) subtypes^{12,13}. Finally, for a subset of patients who experienced distant metastasis (618 out of the 1,079 from the full dataset who relapsed), the date of each recurrence was available, enabling analysis of their spatiotemporal dynamics. We refer to this as the recurrent-events dataset. The three datasets are summarized in Extended Data Fig. 1a, with clinical details and basic parameters describing the intermediate end points of locoregional and distant recurrence across distinct subgroups in Supplementary Table 1. We also established an independent metacohort composed of 1,380 patients with breast cancer from eight cohorts, enabling external validation of our findings, despite their shorter median follow-up (eight years) (Extended Data Fig. 1b). We used the maximum information available to fit the models, keeping for each patient all of the transitions that had complete observations in the variables needed to estimate the hazard of those specific transitions. Therefore, the total number of cases used in each model differs owing to the differing missing values in clinical variables and molecular classification that can affect different transitions.

Model description. The general model that we fit to our datasets is a multistate model that reflects the different risks of locoregional relapse, distant relapse or disease-specific death, conditioned on the current status of the patient. Although multistate survival models for breast cancer were proposed more than 60 years ago³¹, there are few such analyses in the literature^{14,32,33}. Specifically, we used a nonhomogenous semi-Markov chain with two absorbent states (death/cancer and death/other), as shown schematically in Fig. 1. The model was stratified by molecular subtype and used a clock-reset time scale, in which the clock stops (clock-reset) when the patient enters a new state. Although there were a small number of transitions from distant to local relapse (15 ER-positive cases and 7 ER-negative cases), we omitted the local relapse in these instances as we considered it to be redundant, and allowed only transitions from local to distant relapse in our model. We also included the possibility of cancer-related death without a recurrence to account for cases where metastasis was not detected. The R packages *survival*³⁴ and *mstate*³⁵ were used to fit the data.

Several covariates were included in the model: age at state entry (diagnosis or relapse), tumour grade, tumour size and the number of positive lymph nodes, all of them as continuous variables (although, in the case of lymph nodes, all values larger than 10 lymph nodes were coded as 10, to avoid excessive influence in the estimation of the slope from extreme cases). The time from diagnosis was also included as continuous. Note that these formulations are a simplification of the modelling in our previous work¹², where age, size and lymph nodes were modelled nonlinearly through splines. We have simplified these effects to reduce the number of parameters in the model, but also, in the case of age, because its nonlinearity is relevant only when overall survival is the end point.

For the full dataset, we fit a Cox model that was stratified on ER status. The effect of age on death/other causes was modelled with a different coefficient for each transition into nonmalignant death (based on ER status), to account for differences in the age at relapse or diagnosis. 'Grade', 'size' and 'lymph nodes' were allowed to have different coefficients from the starting state to states of recurrence/cancer-related death for each ER status. Time since diagnosis had different coefficients from the starting state of relapse to states of recurrence/cancer-related death for each ER status, and time since locoregional recurrence had different coefficients from distant relapse state to cancer-related death for each ER status. The time since

locoregional recurrence was not predictive of the time to distant recurrence and therefore was not included in further analyses.

For the molecular dataset, and because of the large number of molecular subtypes, we reduced the number of parameters, constraining their values to be the same for the different molecular subtypes. On the basis of different fits and the results of likelihood ratio tests, age was allowed to have a different coefficient for transitions from surgery, locoregional relapse or distant relapse into death/other causes, although these coefficients were constrained to be the same for all molecular subtypes. Grade and lymph nodes were allowed to have one value for transitions from diagnosis and another for transitions from relapse to states of recurrence/death, identical for each molecular subtype. Size was allowed one value for transitions from diagnosis and another for transitions from locoregional relapse to states of recurrence/death, identical for each molecular subtype. Time since diagnosis had the same coefficient from the starting state of relapse to states of recurrence/death, identical for all molecular subtypes. This model was fit three times, one for each molecular classification, based on ER/HER2 status, PAM50 and Integrative Clusters; each of them stratified according to the respective molecular subgroups. We used a robust variance estimate in all models and performed likelihood ratio tests in order to reduce the number of parameters in each model (as mentioned above).

Transition probabilities for each molecular subtype. Using the model fit, we obtained the hazards for each transition for a given individual. We used these hazards to compute the corresponding transition probabilities as follows. We used a clock-reset model and defined all probabilities starting at the time of entry to the last state. All times (s and t) are also defined starting from the time of entry. Let the set of states be $\{S = \text{disease-free/after surgery}, L = \text{locoregional relapse}, D = \text{distant relapse}, C = \text{cancer-related death}, O = \text{other cause of death}\}$. We condition on the vector of clinical covariates \mathbf{x} , which includes the time from surgery (in the case of relapse, this variable has an effect on the hazards).

Transitions from distant relapse. Following previously published studies^{14,36}, we define the conditional probability of having no further event between times t and s for a patient with distant relapse at time t as:

$$S_D(s, t|\mathbf{x}) = \exp\left\{-\int_t^s (\lambda_{D,C}(u|\mathbf{x}) + \lambda_{D,O}(u|\mathbf{x}))du\right\}$$

where $\lambda_{i,j}(t|\mathbf{x})$ is the hazard of moving from state i to state j at time t with the vector of covariates \mathbf{x} (including the time from surgery or age, which must be updated after a relapse).

Then, the prediction probabilities for each path are:

$$\pi_D^C(u, t|\mathbf{x}) = \int_t^u \lambda_{D,C}(s|\mathbf{x})S_D(s, t|\mathbf{x})ds$$

$$\pi_D^O(u, t|\mathbf{x}) = \int_t^u \lambda_{D,O}(s|\mathbf{x})S_D(s, t|\mathbf{x})ds$$

$$\pi_D(u, t|\mathbf{x}) = 1 - (\pi_D^C(u, t|\mathbf{x}) + \pi_D^O(u, t|\mathbf{x}))$$

Transitions from locoregional relapse. Similarly, we obtain:

$$S_L(s, t|\mathbf{x}) = \exp\left\{-\int_t^s (\lambda_{L,D}(u|\mathbf{x}) + \lambda_{L,C}(u|\mathbf{x}) + \lambda_{L,O}(u|\mathbf{x}))du\right\}$$

$$\pi_L^{D,C}(u, t|\mathbf{x}) = \int_t^u \lambda_{L,D}(s|\mathbf{x})\pi_D^C(u-s, 0|\mathbf{x})S_L(s, t|\mathbf{x})ds$$

$$\pi_L^{D,O}(u, t|\mathbf{x}) = \int_t^u \lambda_{L,D}(s|\mathbf{x})\pi_D^O(u-s, 0|\mathbf{x})S_L(s, t|\mathbf{x})ds$$

$$\pi_L^D(u, t|\mathbf{x}) = \int_t^u \lambda_{L,D}(s|\mathbf{x})\pi_D(u-s, 0|\mathbf{x})S_L(s, t|\mathbf{x})ds$$

$$\pi_L^C(u, t|\mathbf{x}) = \int_t^u \lambda_{L,C}(s|\mathbf{x})S_L(s, t|\mathbf{x})ds$$

$$\pi_L^O(u, t|\mathbf{x}) = \int_t^u \lambda_{L,O}(s|\mathbf{x}) S_L(s, t|\mathbf{x}) ds$$

$$\pi_L(u, t|\mathbf{x}) = 1 - (\pi_L^{D,C}(u, t|\mathbf{x}) + \pi_L^{D,O}(u, t|\mathbf{x}) + \pi_L^D(u, t|\mathbf{x}) + \pi_L^C(u, t|\mathbf{x}) + \pi_L^O(u, t|\mathbf{x}))$$

Transitions after surgery.

$$S_S(s, t|\mathbf{x}) = \exp\left[-\int_t^s (\lambda_{S,L}(u|\mathbf{x}) + \lambda_{S,D}(u|\mathbf{x}) + \lambda_{S,C}(u|\mathbf{x}) + \lambda_{S,O}(u|\mathbf{x})) du\right]$$

$$\pi_S^{L,D,C}(u, t|\mathbf{x}) = \int_t^u \lambda_{S,L}(s|\mathbf{x}) \pi_L^{D,C}(u-s, 0) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^{L,D,O}(u, t|\mathbf{x}) = \int_t^u \lambda_{S,L}(s|\mathbf{x}) \pi_L^{D,O}(u-s, 0) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^{L,C}(u, t|\mathbf{x}) = \int_t^u \lambda_{S,L}(s|\mathbf{x}) \pi_L^C(u-s, 0) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^{L,O}(u, t|\mathbf{x}) = \int_t^u \lambda_{S,L}(s|\mathbf{x}) \pi_L^O(u-s, 0) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^{L,D}(u, t|\mathbf{x}) = \int_t^u \lambda_{S,L}(s|\mathbf{x}) \pi_L^D(u-s, 0) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^{D,C}(u, t|\mathbf{x}) = \int_t^u \lambda_{S,D}(s|\mathbf{x}) \pi_D^C(u-s, 0) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^{D,O}(u, t|\mathbf{x}) = \int_t^u \lambda_{S,D}(s|\mathbf{x}) \pi_D^O(u-s, 0) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^L(u, t|\mathbf{x}) = \int_t^u \lambda_{S,L}(s|\mathbf{x}) \pi_L(u-s, 0) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^D(u, t|\mathbf{x}) = \int_t^u \lambda_{S,D}(s|\mathbf{x}) \pi_D(u-s, 0) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^C(u, t|\mathbf{x}) = \int_t^u \lambda_{S,C}(s|\mathbf{x}) S_S(s, t|\mathbf{x}) ds$$

$$\pi_S^O(u, t|\mathbf{x}) = \int_t^u \lambda_{S,O}(s|\mathbf{x}) S_S(s, t|\mathbf{x}) ds$$

$\pi_S(u, t|\mathbf{x})$ can be computed as 1 minus the sum of the others.

Prediction probabilities for being in a particular state at a certain time can also be computed by summing the appropriate paths. Note that the main difficulty in computing these probabilities is updating the corresponding hazards every time a transition occurs, as they may depend on variables that change over time or after a transition to a different state. In our implementation, we tried to follow the style of the *mstate* package³⁵.

Standard errors for transition probabilities in our model. If our model was Markovian (as the clock-forward model), the transition probabilities could be easily computed through the product-integral representation³⁷ and it would also be straightforward to obtain estimates of their standard errors. However, for our clock-reset model the estimation of standard errors is complicated, so we used a semiparametric bootstrap approach to obtain such estimates³⁸. In brief, for every bootstrap replicate ($B = 100$), we sampled trajectories for each observation in our original dataset on the basis of our fitted model. These trajectories were fitted to the original model and bootstrap hazards for the original average individuals were

computed. Then, the formulas described earlier were used to obtain bootstrap transition probabilities. Because these bootstrap estimates are not likely to converge to the theoretical estimates in transitions with a small number of observed instances, we computed the standard deviation of the bootstrap estimates as an indication of the variability of these predictions for a given patient.

Transition probabilities for specific events. The transition probabilities obtained for each patient can be aggregated to obtain probabilities of visiting specific states (locoregional recurrence, distant recurrence) or specific end points. We used these probabilities in two ways: as an example of individual predictions for an average patient for each molecular subtype (based on typical or average values of each covariate), as in Supplementary Table 4b, Fig. 1b and Extended Data Figs. 6, 8, together with a confidence interval computed using the obtained probabilities ± 1.96 times the standard deviation of the bootstrap estimates described above, which represents variability around individual predictions. We also computed probabilities for all patients to show their distribution in each molecular subtype, as in Supplementary Table 4a, Figs. 2a, 3a and Extended Data Figs. 4, 5, 7. Confidence intervals computed using the mean of the probabilities ± 1.96 times the standard error of the mean represent variability around the mean in each subtype.

Sites of relapse. In the recurrent-event datasets, each patient can have several relapses. Instead of adding the site to our multistate models, we selected only patients who had a distant relapse. First, in Fig. 4a and Extended Data Fig. 10, we tested whether the proportions of relapses in each organ differed by molecular subtype. We fitted a logistic regression model with relapse as a binary variable and the sites of metastases as dependent variables. We computed simultaneous tests using the R package *multcomp*³⁹ with the Dunnett method⁴⁰. Only those proportions with $P < 0.05$ were considered significant. In the same figures, cumulative-incidence distributions for each organ were computed independently—that is, no competing risk model was fitted.

Next, we modelled recurrent distant metastases (Fig. 4b) using the PWP conditional model⁴¹. This model allows for different baseline hazards for each consecutive recurrence while keeping at risk for recurrence i only those individuals that have experienced the recurrence $i - 1$.

Finally, in Fig. 4c we fitted a Cox model with time-dependent variables to estimate the hazard of DSD for metastases in different organs. We also included in this model the clinical variables from the primary tumour (tumour grade, tumour size and number of tumour-positive lymph nodes).

Goodness-of-fit testing. Goodness-of-fit testing was performed for all models. Proportional hazards assumptions were tested with the Schoenfeld residuals versus time, using the survival package function `cox.zph()`³⁴. None of the models showed covariates that violated the assumption, except the model for sites of metastasis (ER-positive), where the number of metastases and 'other metastasis' were significant, and the model for sites of metastasis (ER-negative), where the grade and number of metastases were significant (see Supplementary Information). Visual inspection of the plots showed that the trend was roughly flat and thus the violation was not critical. As shown previously, in the model that includes ER, ER violates the proportional hazard assumption. However, this model was only used to test differences in the hazard ratios of the other covariates according to ER.

Model validation and calibration. We validated each of the models using several approaches, as outlined below.

Internal validation. We validated the global predictions of the model on all transitions using a bootstrap approach that has previously been described in detail⁴², using the *rms* R package. We used the following measures of predictive ability: (1) Somers' Dxy rank correlation (Dxy), which is $2(C - 0.5)$, where C is the C -index; (2) Nagelkerke's R2, which is the square root of the proportion of log-likelihood explained by the model, to the log-likelihood that could be explained by a 'perfect' model, with a penalty for model complexity; (3) the slope shrinkage (slope), a measure of how much the estimates are affected by extreme observations; (4) the discrimination index D , derived from the log-likelihood at the shrunken linear predictor; (5) the unreliability index U , a measure of how different the model maximum log-likelihood is from that of a model with frozen coefficients; (6) the overall quality index Q , a normalized and penalized-for-unreliability log-likelihood; and (7) the g -index (g) on the log relative hazard (linear predictor) scale (Gini's mean difference).

Each measure was computed on the training set and on 200 bootstrap test sets, estimating the optimism and the corrected indexes for predictions at 5, 10 and 15 years (see Extended Data Fig. 3a).

Internal calibration. We also use the following procedure for model calibration, as described previously⁴²: (1) interpolation of the hazard function using splines (haz method) among all the cases as a general function of the predictor variables and time; (2) computation of the predicted values for a given time point (5, 10 or 15 years); (3) computation of the differences between observed and predicted; and (4) using 200 bootstrap datasets, computation of the optimism in those differences. Extended Data Figure 3b shows a box plot of the mean absolute error of all predictions.

External calibration. As an external comparison of the predicted probabilities of our models, we used predict v.2.1¹⁸, a tool that has been validated extensively. PREDICT uses a model with several variables (including the effect of treatment) and produces estimates of the probability of cancer-specific death (C/D) and non-malignant death (O/D), as well as estimates of the effect of treatment. We compared the probabilities for these events with PREDICT using Pearson correlation (see Extended Data Fig. 3c, d).

External validation. We used two sets of external samples to validate the predictions of our models. First, we used a set of METABRIC samples that were not part of the original study, including 121 patients with copy-number data and 57 patients with expression data. Survival data are available for these patients (in fact, they are part of the full dataset, but because they had not been used to fit the IntClust model, they could be used to test the validity of the C-index on an external dataset). We classified these tumours into IntClust groups using the iC10 package¹³.

Second, we used an external dataset of 1,380 patients from 8 different cohorts and different survival information. We validated predictions of disease-specific survival, overall survival, relapse-free survival and distant-relapse-free survival. We compiled a metacohort by merging early breast-cancer cohorts for which expression data (Affymetrix array), outcome and covariates are available, including GSE19615 (DFHCC cohort⁴³), GSE42568 (Dublin cohort⁴⁴), GSE9195 (Guyt2 cohort⁴⁵), GSE45255 (IRB/JNR/NUH cohort⁴⁶), GSE11121 (Maintz cohort⁴⁷), GSE6532 (TAM cohort⁴⁵), GSE7390 (Transbig cohort⁴⁸) and GSE3494 (Upp cohort⁴⁹). Original data (raw CEL files) were downloaded and preprocessed using the rma function from the affy⁵⁰ package. The intensities were then quantile normalized and corrected for batch effects with the COMBAT function from the sva⁵¹ package. PAM50 was called using the genefu⁵² package. The ER, progesterone receptor and HER2 status were extracted from the expression with probes 205225_at, 208305_at and 216836_s_t, using a Gaussian mixture model. IntClust10 subgroups were called using the iC10 package. C-indices and summary C-indices were calculated using the survcomp⁵³ package. For the combined metacohort scores, we calculated C-scores for each individual cohort and then combined them using the function combine.est from the survcomp⁵³ package. Confidence intervals and P values for comparing C-indices were computed with the same package. Extended Data Figure 3e shows the C-indices and confidence intervals for these comparisons.

General statistical considerations. All tests were performed two-sided (except where indicated). Adjustment for multiple comparisons was done as described in the section ‘Comparison of probabilities of relapse in ER⁺ high risk integrative subtypes’ (see Supplementary Methods) and separately when comparing the proportions of metastases in each organ (Fig. 4a and Extended Data Fig. 10). All analyses were conducted in R version 3.5.1⁵⁴. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

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

Code availability

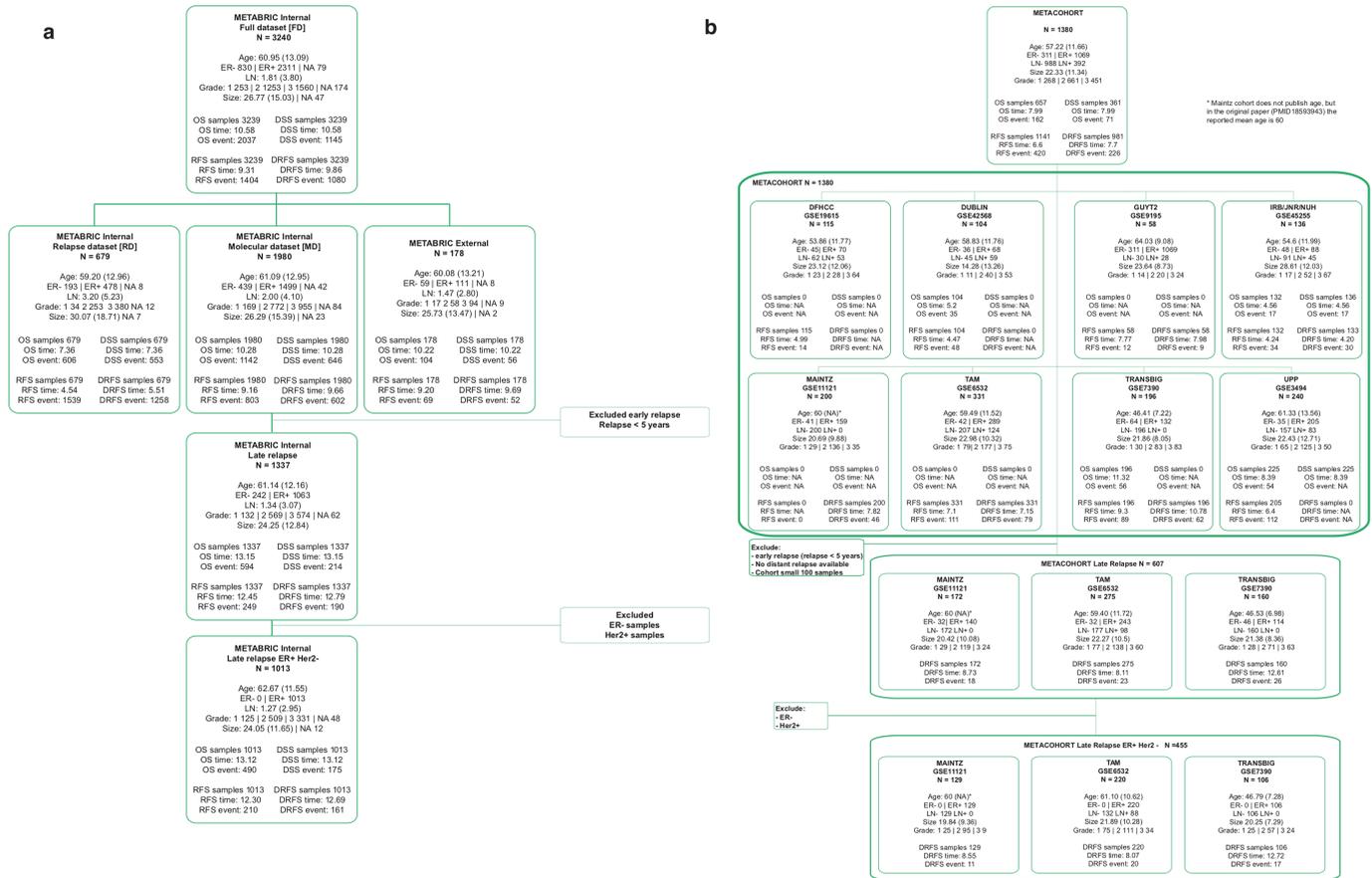
All code and scripts are available for academic use at <https://github.com/cclab-brca/brcarepred>.

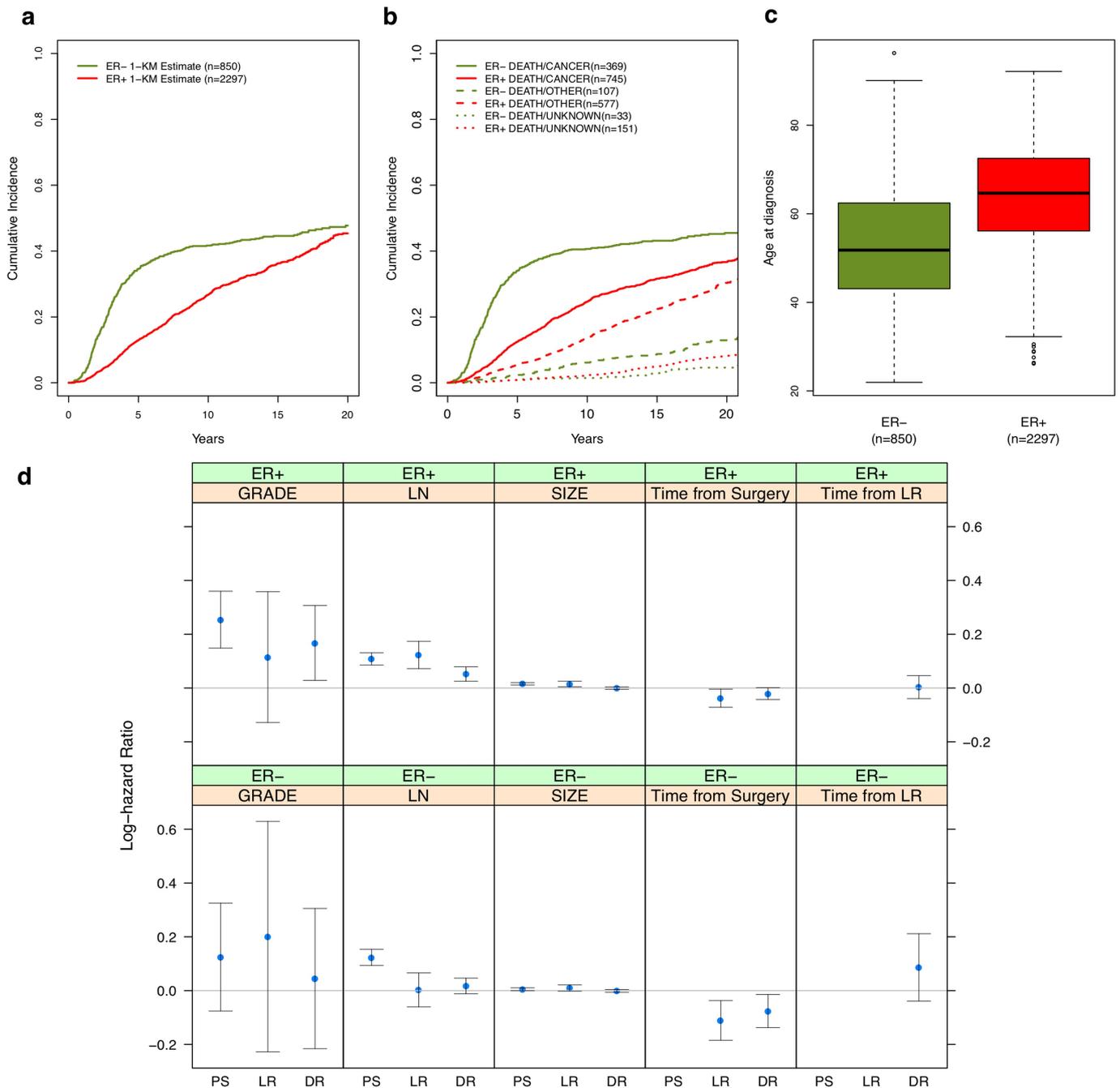
Data availability

The genomic copy number, gene-expression and molecular-subtype information has been described previously¹² and is available at the European Genome-Phenome

Archive at <https://www.ebi.ac.uk/ega/studies/EGAS00000000083>. Clinical data are available in Supplementary Tables 5–8. The breast-cancer-recurrence predictor is available as a web application for academic use at <https://caldaslab.cruk.cam.ac.uk/brcarepred>.

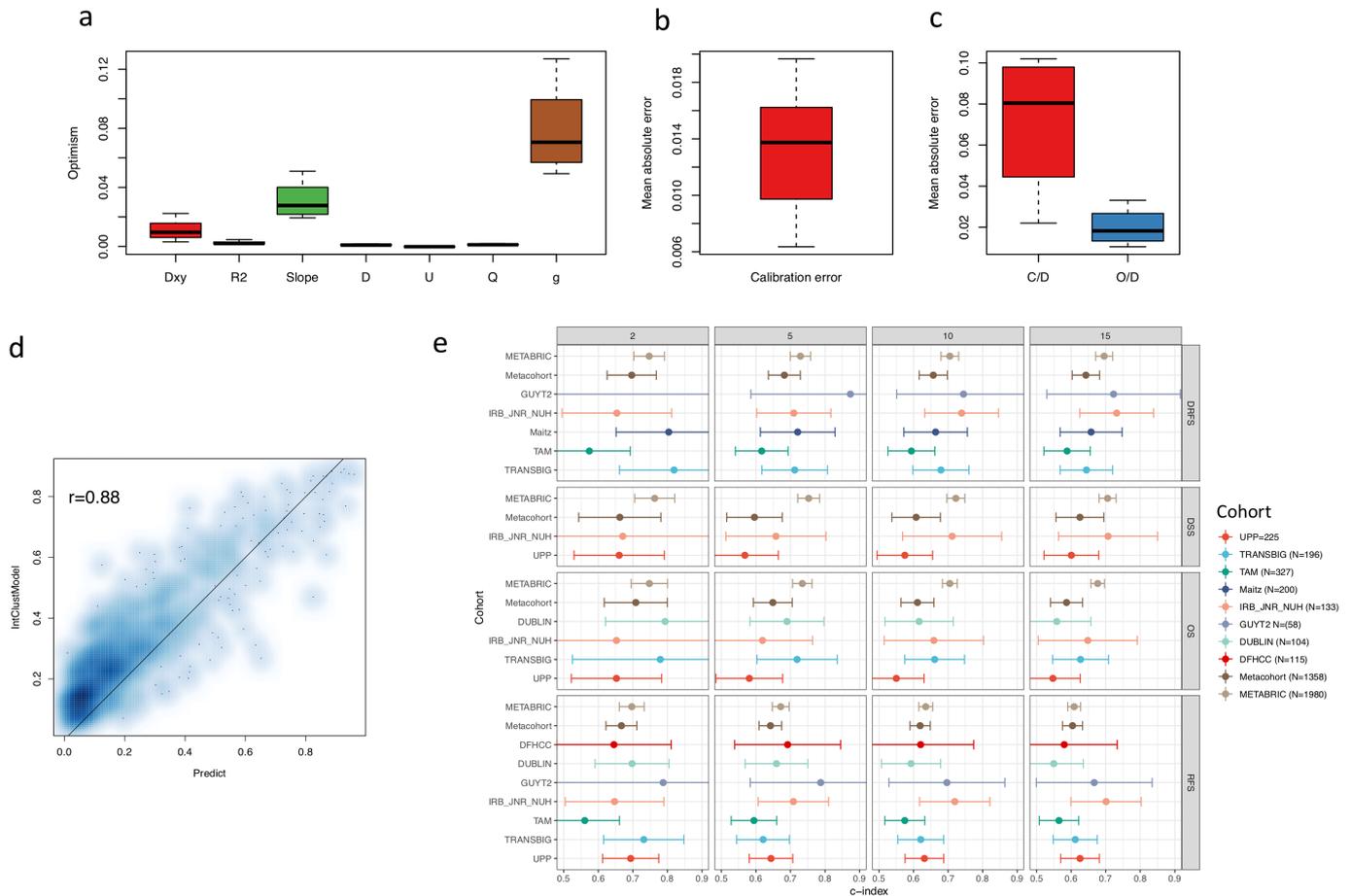
- Fix, E. & Neyman, J. A simple stochastic model of recovery, relapse, death and loss of patients. *Hum. Biol.* **23**, 205–241 (1951).
- Brøt, P. et al. Analyzing prognostic factors in breast cancer using a multistate model. *Breast Cancer Res. Treat.* **54**, 83–89 (1999).
- Meier-Hirmer, C. & Schumacher, M. Multi-state model for studying an intermediate event using time-dependent covariates: application to breast cancer. *BMC Med. Res. Methodol.* **13**, 80 (2013).
- Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model* (Springer, New York, 2000).
- de Wreede, L. C., Fiocco, M. & Putter, H. mstate: an R package for the analysis of competing risks and multi-state models. *J. Stat. Software* **38**, 1–30 (2011).
- Klein, J. P., Keiding, N. & Copelan, E. A. Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplantation patients. *Stat. Med.* **12**, 2315–2332 (1993).
- Aalen, O., Borgan, O. & Gjessing, H. *Survival and Event History Analysis—A Process Point of View* (Springer, New York, 2008).
- Fiocco, M., Putter, H. & van Houwelingen, H. C. Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Stat. Med.* **27**, 4340–4358 (2008).
- Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363 (2008).
- Dunnett, C. W. A multiple comparison procedure for comparing several treatments with a control. *J. Am. Stat. Assoc.* **50**, 1096–1121 (1955).
- Prentice, R. L., Williams, B. J. & Peterson, A. V. On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373–379 (1981).
- Harrell, F. E. J. *Regression Modeling Strategies* (Springer, 2001).
- Li, Y. et al. Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat. Med.* **16**, 214–218 (2010).
- Clarke, C. et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. *Carcinogenesis* **34**, 2300–2308 (2013).
- Loi, S. et al. Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* **9**, 239 (2008).
- Nagalla, S. et al. Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis. *Genome Biol.* **14**, R34 (2013).
- Schmidt, M. et al. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res.* **68**, 5405–5413 (2008).
- Desmedt, C. et al. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin. Cancer Res.* **13**, 3207–3214 (2007).
- Miller, L. D. et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc. Natl Acad. Sci. USA* **102**, 13550–13555 (2005); correction **102**, 17882 (2005).
- Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
- Gendoo, D. M. A. et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics* **32**, 1097–1099 (2016).
- Schröder, M. S., Culhane, A. C., Quackenbush, J. & Haibe-Kains, B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* **27**, 3206–3208 (2011).
- R Core Team. *R: A Language and Environment for Statistical Computing*. <http://www.r-project.org/> (2015).





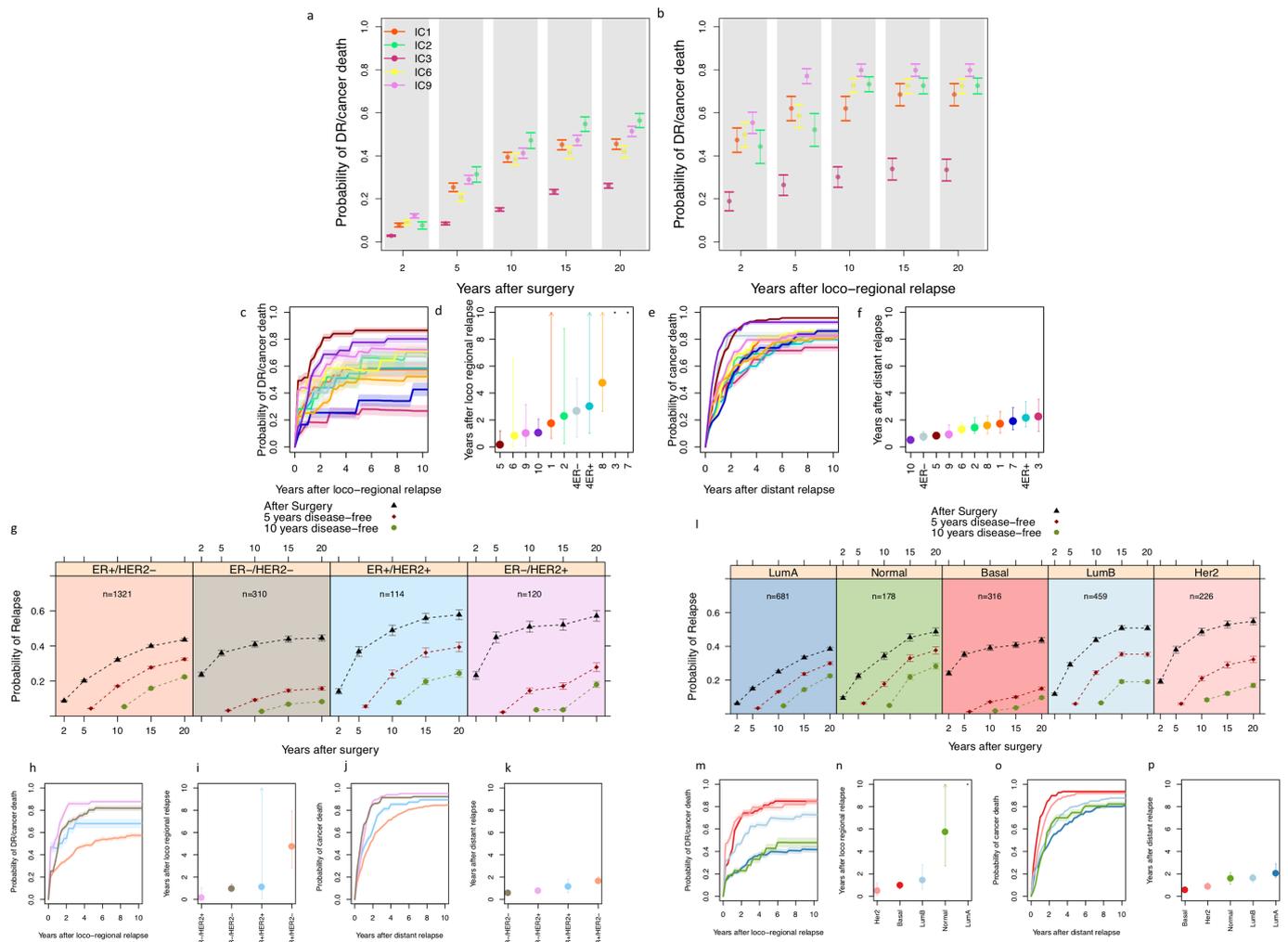
Extended Data Fig. 2 | Effect of censoring nonmalignant deaths on the estimation of disease-specific survival, and prognostic value of clinical covariates at different disease states. **a**, Cumulative incidence computed as 1 – Kaplan–Meier (KM) estimator, using only disease-specific death as an end point and censoring other types of death. **b**, Cumulative incidence computed using a competing-risk model that takes into account different causes of death. The bias of the 1 – Kaplan–Meier estimator is visible. **c**, Distribution of age at the time of diagnosis for ER-negative and ER-positive patients. The number of patients in each group is indicated in all panels. This analysis was done with the full dataset. Box plots were computed using the median of the observations (centre line). The first and

third quartiles are shown as boxes, and the whiskers extend to the ± 1.58 interquartile range divided by the square root of the sample size. Outliers are shown as dots. **d**, log hazard ratios calculated using the multistate model stratified by ER status ($n = 3,147$) for different covariates, namely grade, lymph-node (LN) status, tumour size (size), time from surgery and time from local relapse (LR). log hazard ratios are shown for different states, including post-surgery (PS; hazard ratio of progressing to relapse or DSD), locoregional recurrence (LR; hazard ratio of progressing to distant relapse or DSD) and distant recurrence (DR; hazard ratio of cancer-specific death). 95% confidence intervals are shown. This analysis was done with the full dataset.



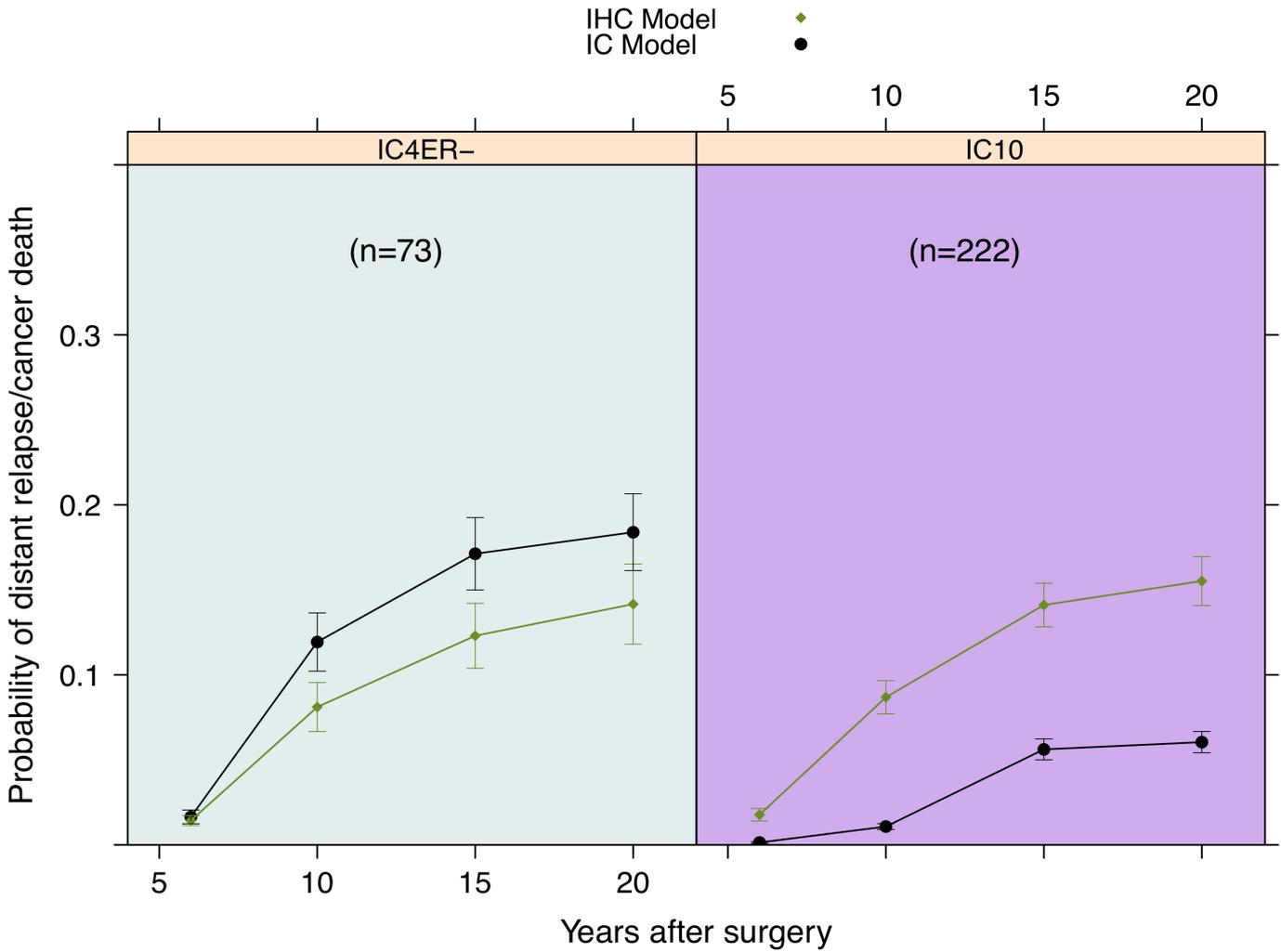
Extended Data Fig. 3 | Model calibration and validation in an external dataset. **a**, Internal validation of the global predictions of the models on all transitions using bootstrap ($n = 200$). Discriminant measures of predictive ability are shown on the x axis, as described in the Methods section ‘Model validation and calibration’. The y axis shows the optimism, that is, the difference between the training predictive ability and the test predictive ability of the discriminant measures (see Methods). **b**, Internal calibration of the global predictions of the models on all transitions using bootstrap ($n = 200$). The distribution of the mean absolute error between observed and predicted is plotted. **c**, External calibration of DSD risk and nonmalignant death risk using PREDICT 2.1 ($n = 1,841$). The distribution of the mean absolute error between the predictions of PREDICT and our model based on ER status only is plotted. **a–c**, Box plots were computed using the median of the observations (centre line).

The first and third quartiles are shown as boxes, and the whiskers extend to the ± 1.58 interquartile range divided by the square root of the sample size (see Methods). **d**, Scatter plot of the predictions of DSD risk computed by PREDICT and our model based on the IntClust subtypes only at ten years ($n = 1,841$; see Methods). The Pearson correlation is shown. **e**, Concordance index (C-index) of prediction of risk of distant relapse (DRFS), disease-specific death (disease-specific survival, DSS), death (overall survival, OS) and relapse (RFS) in the 178 withheld METABRIC samples and in a metacohort composed of eight published studies among $ER^+/HER2^-$ patients in the high-risk IntClust subtypes, where results are shown for individual cohorts and the combined metacohort (see Methods and Supplementary Information). Error bars correspond to 95% confidence intervals for the C-index. The number of patients in each group is indicated on the right.



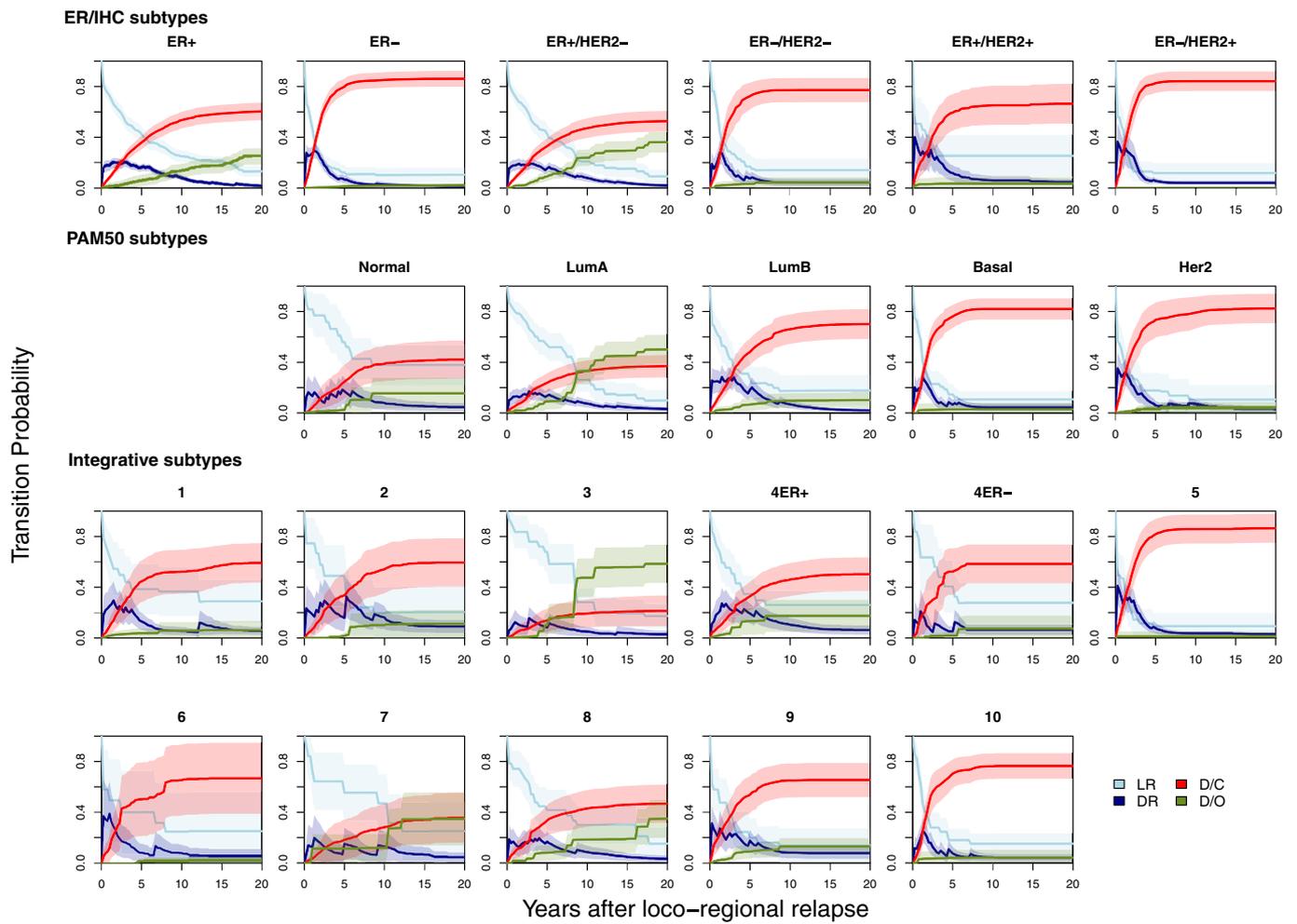
Extended Data Fig. 4 | Different subtypes have distinct probabilities of recurrence. **a**, Average probability of experiencing a distant relapse (defined as the probability of having a distant relapse at any point followed by any other transition) or cancer-related death for the high-risk ER⁺ IntClust (IC) subtypes (IC1 $n = 134$, IC6 $n = 81$, IC9 $n = 134$, IC2 $n = 69$) relative to IC3 ($n = 269$), the ER⁺ subgroup with the best prognosis. This analysis was restricted to ER⁺/HER2⁻ cases, which represent the vast majority for each of these subtypes. Error bars represent 95% confidence intervals around the mean. **b**, As for **a**, but showing the average probability of experiencing distant recurrence or cancer-related death after a local recurrence (IC1 $n = 21$, IC6 $n = 10$, IC9 $n = 21$, IC2 $n = 13$, IC3 $n = 30$). **c**, Average probability of recurrence (distant relapse or cancer-specific death) after loco-regional relapse for all patients in each of the 11 IntClust subtypes. **d**, Median time until an additional relapse (distant recurrence or cancer-specific death) after local recurrence for all patients in each of the 11 IntClust subtypes ($n = 270$). This has been computed using

a Kaplan-Meier approach with competing risks of progression and nonmalignant death. Error bars represent 95% confidence intervals around the median time. Asterisks denote situations in which the median time cannot be computed because fewer than 50% of the patients relapsed. This analysis was done with the molecular dataset. **e**, Average probability of cancer-related death after distant recurrence for all patients by subtype. **f**, As for **d**, except that the median time until cancer-specific death after distant recurrence is shown ($n = 596$). **g**, Mean probabilities of relapse after surgery and after five and ten disease-free years (see Methods and Supplementary Table 4) for the patients in each of the four IHC subtypes. Error bars represent 95% confidence intervals. The number of patients in each group is indicated. **h-k**, As for **c-f**, but for the IHC subtypes (same sample sizes). **l**, As for **g**, but for the PAM50 subtypes. The number of patients in each group is indicated. **m-p**, As for **h-k**, but for the PAM50 subtypes (with the same sample sizes, except for **p** where $n = 593$).



Extended Data Fig. 5 | The ER⁻/HER2⁻ integrative subtypes exhibit distinct risks of relapse. The probabilities of distant relapse or cancer-related death among ER⁻/HER2⁻ patients who were disease-free at five years after diagnosis reveal marked differences in the risk of relapse for TNBC IntClust subtype IC4ER⁻ versus the IC10 (basal-like enriched)

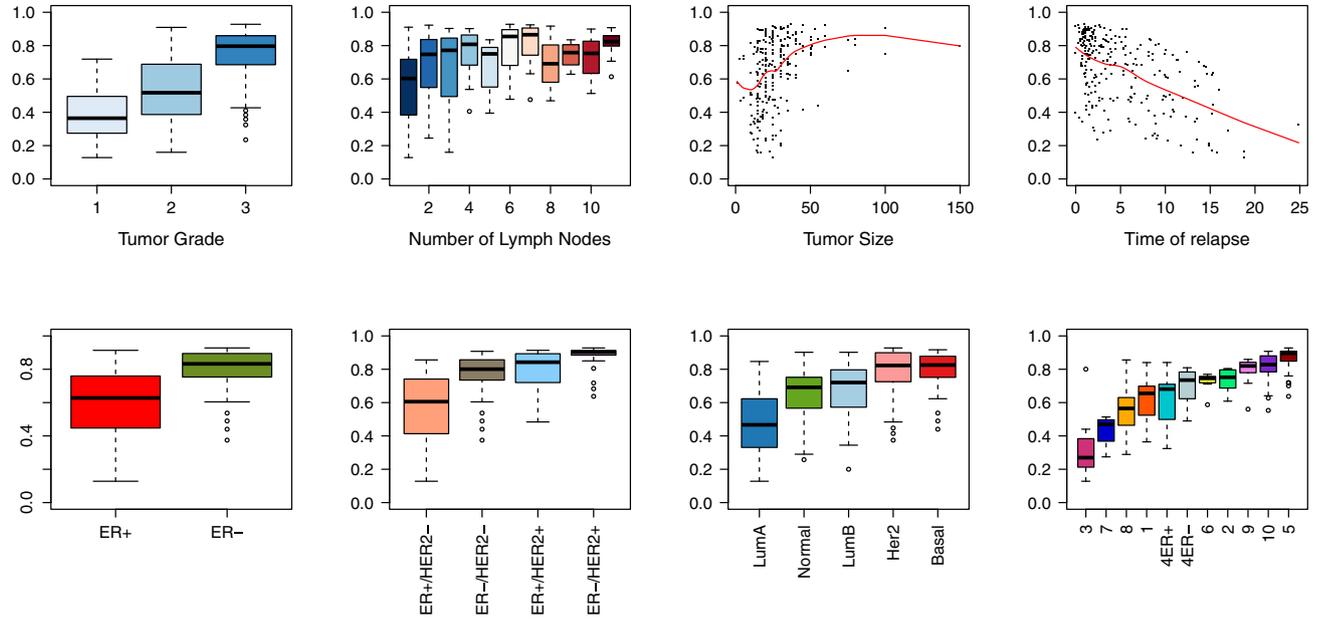
subtype. Here the base clinical model with IHC subtypes is compared with the base clinical model plus IntClust subtype information. Error bars represent 95% confidence intervals. The number of patients in each group is indicated.



Extended Data Fig. 6 | Subtype-specific risks of relapse after locoregional relapse. Transition probabilities from locoregional recurrence to other states for individual average patients, stratified on the basis of ER, IHC, PAM50 or IntClust subtype. 95% confidence bands were

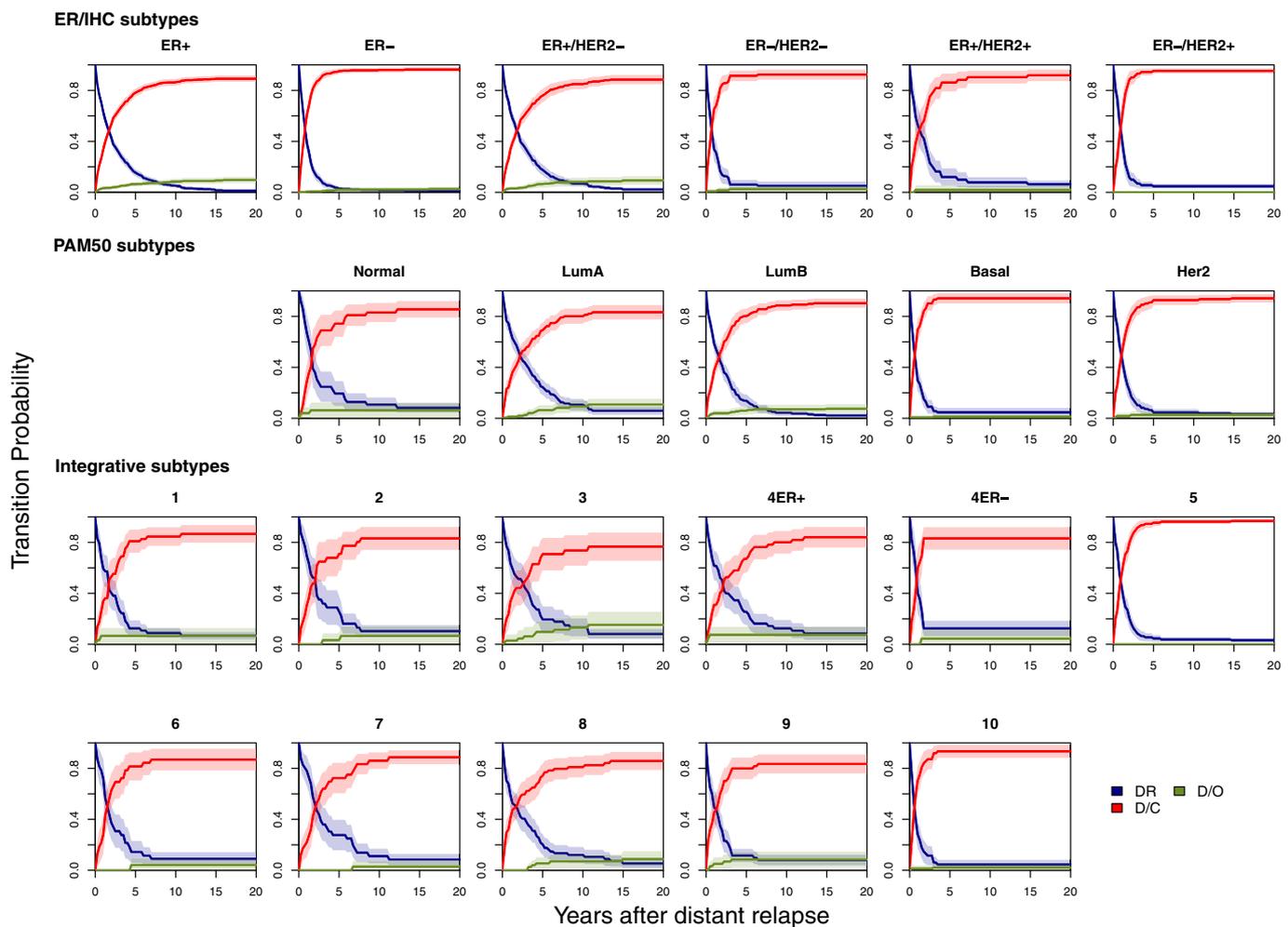
computed using bootstrap. This analysis was done with the full dataset for the comparisons between ER⁺ and ER⁻, and the molecular dataset for the remainder.

Probability of DR/Cancer Death 10 years after LR



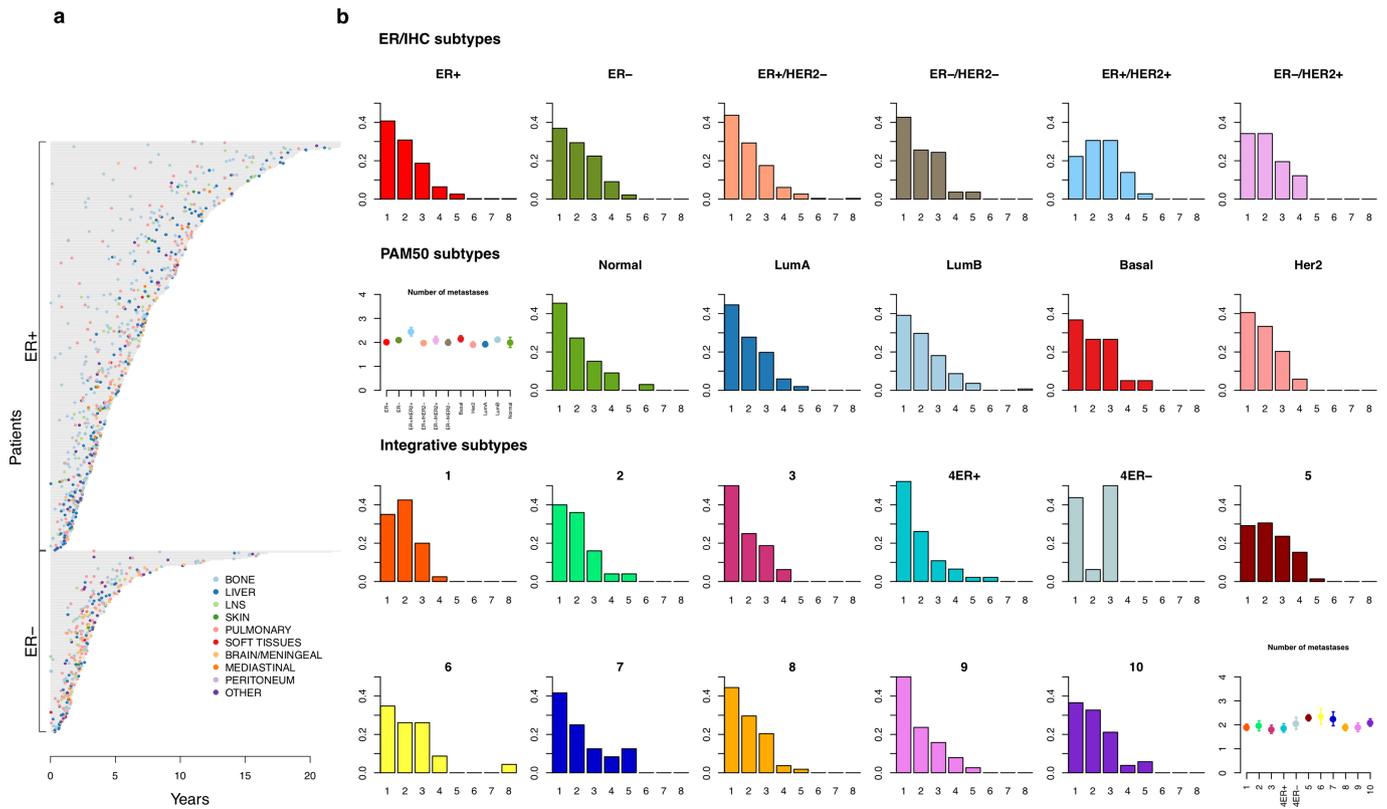
Extended Data Fig. 7 | Associations between probabilities of distant relapse ten years after locoregional relapse with clinico-pathological and molecular features of the primary tumour. For each patient that had a locoregional recurrence, the ten-year probability of having a distant relapse or cancer-related death is plotted against different variables. A loess fit is overlaid to highlight the relationship between the probability

and tumour size or time of relapse. Box plots were computed using the median of the observations (centre line). The first and third quartiles are shown as boxes, and the whiskers extend to the ± 1.58 interquartile range divided by the square root of the sample size. Outliers are shown as dots. This analysis was done with the molecular dataset and the model was stratified by IntClust subtype ($n = 257$).



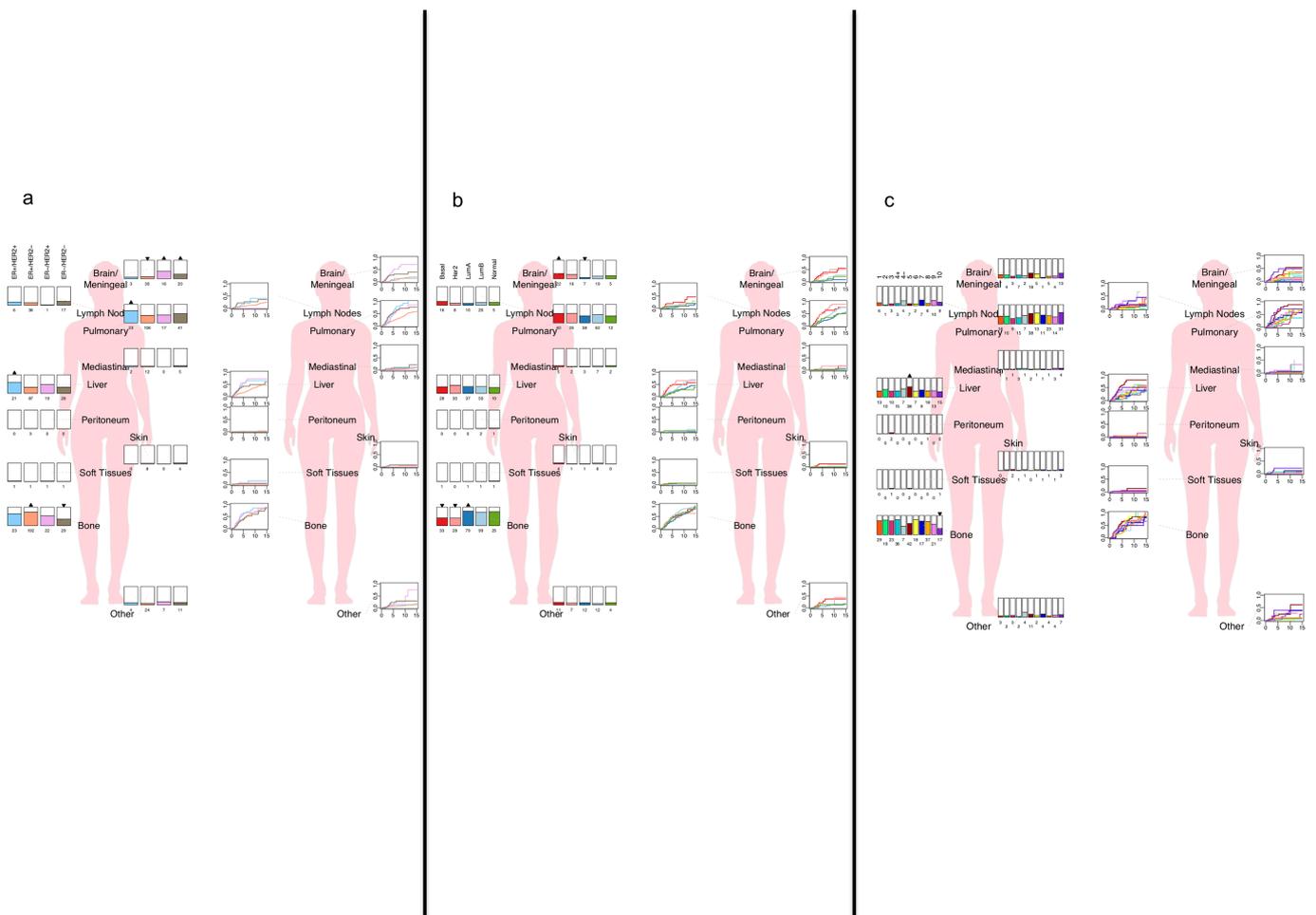
Extended Data Fig. 8 | Subtype-specific risks of cancer-related death after a distant relapse. Transition probabilities from distant relapse to other states for individual average patients stratified on the basis of ER, IHC, PAM50 or IntClust subtype. 95% confidence bands were computed

using bootstrap. This analysis was done with the full dataset for the comparisons between ER⁺ and ER⁻, and the molecular dataset for the remainder.



Extended Data Fig. 9 | Distribution of the number of relapses by molecular subtype. **a**, Times of distant recurrence for ER⁻ and ER⁺ patients ($n = 605$). Each dot represents a distant recurrence, coded by colour for different sites. **b**, Distribution of the number of distant relapses for different subtypes ($n = 609$), based on ER status (ER⁺ $n = 422$, ER⁻ $n = 187$), IHC ER/HER2 status (ER⁺/HER2⁻ $n = 263$, ER⁻/HER2⁻ $n = 82$,

ER⁺/HER2⁺ $n = 36$, ER⁻/HER2⁺ $n = 41$), PAM50 subtype (normal $n = 33$, luminal A $n = 101$, luminal B $n = 138$, basal $n = 79$, HER2 $n = 69$) and IntClust subtype (IC1 $n = 40$, IC2 $n = 25$, IC3 $n = 32$, IC4ER⁺ $n = 46$, IC4ER⁻ $n = 16$, IC5 $n = 72$, IC6 $n = 23$, IC7 $n = 24$, IC8 $n = 54$, IC9 $n = 38$, IC10 $n = 52$). ER status was imputed on the basis of expression in four samples. These analyses were done with the recurrent-events cohort.



Extended Data Fig. 10 | Site-specific patterns of relapse in the IHC, PAM50 and IntClust subtypes. **a**, Left, percentages of patients with metastases at a given site in the IHC subtypes (bar plots, total numbers also indicated). Upright triangles indicate significant positive differences in that group with respect to the overall mean and inverted triangles indicate significant negative differences in that group with respect to the overall mean using simultaneous testing of all sites (see Methods). Location of metastatic sites is not anatomically accurate. Right, cumulative

incidence functions (as $1 - \text{Kaplan-Meier estimates}$) for each site of metastasis in the IHC subtypes. The same patient can have multiple sites of metastasis. **b**, As for **a**, but for the PAM50 subtypes. **c**, As for **a**, but for the IntClust subtypes. These analyses were done with the recurrent-events cohort. Female silhouettes are from the public-domain human body diagrams at https://commons.wikimedia.org/wiki/Human_body_diagrams.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

R packages described in the text (survival, mstate, rms, survComp, affy, dva, genefu, iC10) www.r-project.org, www.bioconductor.org

Data analysis

R packages described in the text (survival, mstate, rms, survComp, affy, dva, genefu, iC10) www.r-project.org + custom code deposited in github (<https://github.com/cclab-brca/brcarepred>)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data is provided as supplementary tables with the submission.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We used all available samples from the METABRIC study (Curtis et al, 2012 Nature). To our knowledge, this is the largest breast cancer cohort with molecular data and long-term follow-up. To reduce the risk of overfitting, we only used pre-selected variables in the model and used common parameters in different subgroups.
Data exclusions	Exclusion criteria were pre-established. Missing values were queried over the original source hospital files. Those that were truly missing were excluded. Benign, stage 4 tumors, and patients for whom recurrence times were not reliable were removed.
Replication	Not applicable; these are patient data.
Randomization	METABRIC is an observational study, no randomization was performed
Blinding	No blinding was considered necessary

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	The cohort includes 3240 patients with invasive primary breast cancers derived from five tumor banks in the UK and Canada diagnosed between 1977-2005 (with a median follow-up of 14 years amongst patients who remain alive). Primary breast tumors and linked pseudo-anonymised clinical data were obtained with ethical approval from the relevant institutional review boards. A detailed description of the cohort can be found in Supplementary Table 1 and Supplementary Table 4.
Recruitment	In this observational study, women diagnosed with invasive primary breast tumors between 1977-2005 for whom clinical information could be categorically linked to fresh frozen tumor material were considered for inclusion. Further details can be found in Curtis et al (2012, Nature).