

LETTER

<https://doi.org/10.1038/s41586-019-1007-8>

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*}

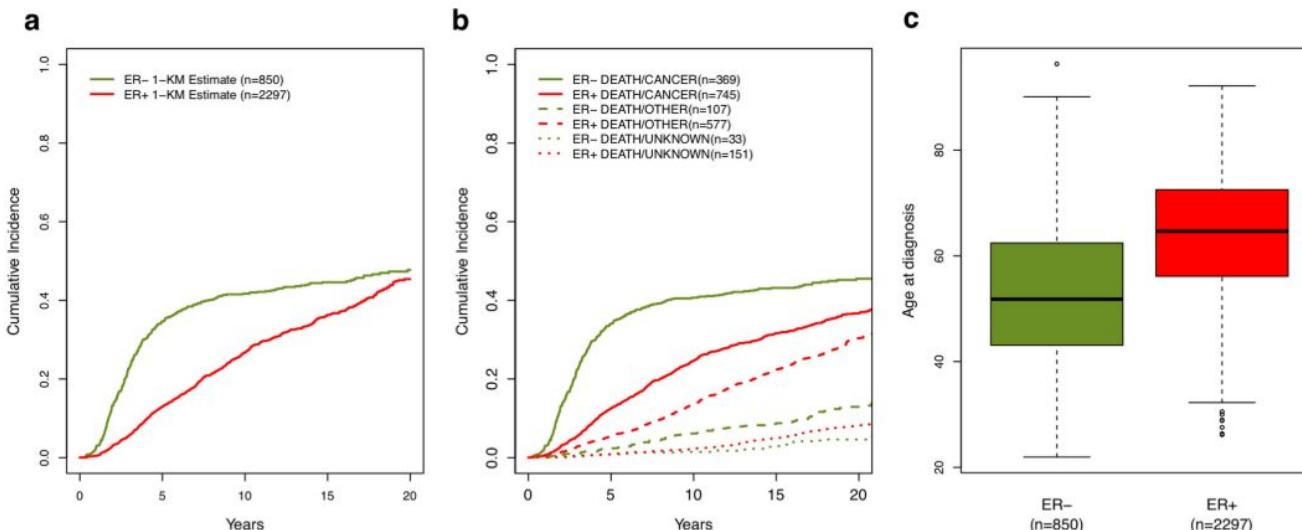
¹Cancer Research UK Cambridge Institute and Department of Oncology, Li Ka Shing Centre, University of Cambridge, Cambridge, UK. ²Department of Medicine, Division of Oncology, Stanford University School of Medicine, Stanford, CA, USA. ³Department of Genetics, Stanford University School of Medicine, Stanford, 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

Main message

- Sometimes breast cancer **relapse** occurs many years (20+) after original diagnosis
- It is important to find the group of patients that are at risk of a late relapse
 - Patients can be categorized into subgroups according to many systems
 - **ER** subtypes (ER+; ER-)
 - ER+**IHC** subtypes (ER+/HER2+; ER+/HER2-; ER-/HER2+; ER-/HER2-)
 - PAM50 subtypes (normal; luminal A; luminal B; Basal; HER2) (**gene expression**)
 - **Integrative subtypes** (1; 2; 3; 4ER-; 4ER+; 5; 6; 7; 8; 9; 10) (**copy-number & gene expression**) → allows for **more precise classification**
- **TNBC**: IntClust4ER- more likely to relapse late than IntClust10
- **ER+/HER-**: IntClusts 1, 2, 6, 9 are likely to relapse late
 - These are all enriched for genomic-copy-number driver alterations that can be therapeutically targeted

Problems with general survival approaches

- Usually disregard death of other causes (other than cancer)
- But ER+ cases are generally diagnosed at later ages
 - Patients die from other causes as well → less of them are alive
 - Incidence is overestimated ~ cancer death/(cancer death + alive)

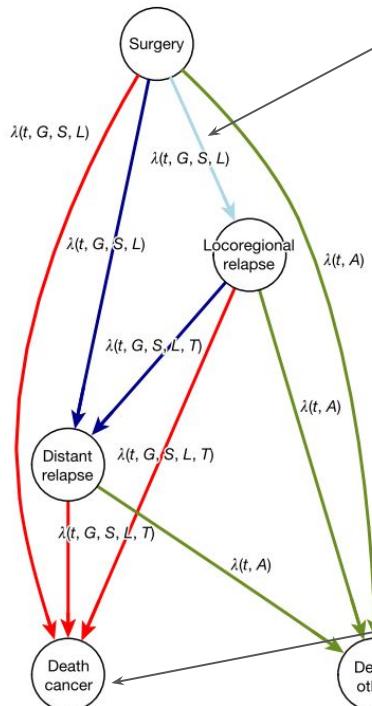


Investigated cohorts

- 3240 patients
- 5 tumor banks from UK and Canada
- Diagnosed between 1977 and 2005 (*this must introduce a giant bias though*)
- Median follow-up time: 14 years
- “full dataset”: clinical and pathological data (n = 3147)
- “molecular dataset”: clinical, pathological and molecular (copy-number, gene expression) data (n = 1962)
- “recurrent-events dataset”: patients with distant metastases (n = 618)
- + independent metacohort (n = 1380)
- “Special care was taken to remove second primary tumours from the dataset.”

Nonhomogenous (semi)-Markov-chain model

a



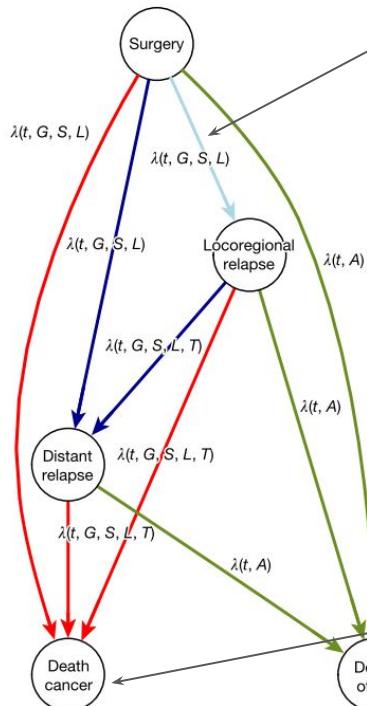
transition probabilities between distinct states, depending on

- Time from diagnosis
- Time from surgery
- Grade
- Tumor size
- Lymph node involvement
- Age
- No treatment!!!

absorbent states

Nonhomogenous (semi)-Markov-chain model

a



THIS IS NOT NEW AT ALL!

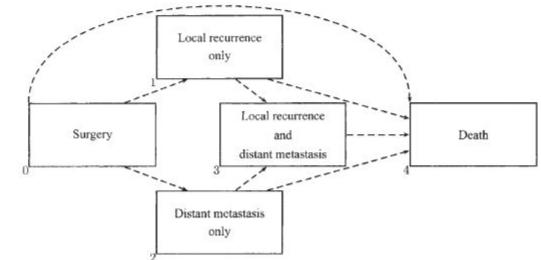


Figure 1 A graphical representation of the multi-state model. The numbers indicated below the boxes are also used to represent the states in the formulas to follow.

Putter et al. Estimation and prediction in a multi-state model for breast cancer. Biom. J. 48, 366–380 (2006).

Nonhomogenous (semi)-Markov-chain model

1. Define states (surgery, locoregional relapse, distant relapse, death of cancer, death of other causes)
2. Choose a pair of states and collect all available transitions between them (annotated with appropriate clinical data)
 - o Make sure to update age
 - o “Clock-reset method” (always set the clock to zero at the previous state)
3. Fit a multivariate Cox-model separately for all the possible transition types
 - o If necessary, stratify by subtype
4. From the fitted models we get the hazard of transitioning from state i to state j at time t , given patient parameters \mathbf{x} for a given patient as $\lambda_{i,j}(t|\mathbf{x})$
5. Use these hazards to calculate transition probabilities and probabilities of paths

Nonhomogenous (semi)-Markov-chain model

Given that a patient had a distant metastasis at time t , the probability of nothing happening until time $s > t$ is:

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

Nonhomogenous (semi)-Markov-chain model

Given that a patient had a distant metastasis at time t , the probability of nothing happening until time $s > t$ is:

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

we know these from the Cox-models!

Nonhomogenous (semi)-Markov-chain model

Given that a patient had a distant metastasis at time t , the probability of nothing happening until time $s > t$ is:

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

we know these from the Cox-models!

The probability of dying of cancer by the time $u > t$, given distant metastasis at t :

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

The probability of dying of some other cause by the time $u > t$, given distant metastasis at t :

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

The probability of not dying by the time $u > t$, given distant metastasis at t :

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

Nonhomogenous (semi)-Markov-chain model

Given that a patient had a distant metastasis at time t , the probability of nothing happening until time $s > t$ is:

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

we know these from the Cox-models!

The probability of dying of cancer by the time $u > t$, given distant metastasis at t :

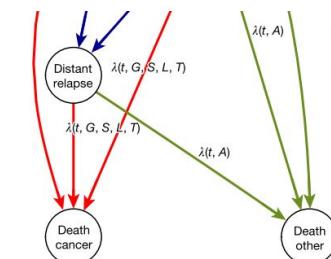
$$\pi_D^C(u, t | \mathbf{x}) = \int_t^u \lambda_{D,C}(s | \mathbf{x}) S_D(s, t | \mathbf{x}) ds$$

The probability of dying of some other cause by the time $u > t$, given distant metastasis at t :

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

The probability of not dying by the time $u > t$, given distant metastasis at t :

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

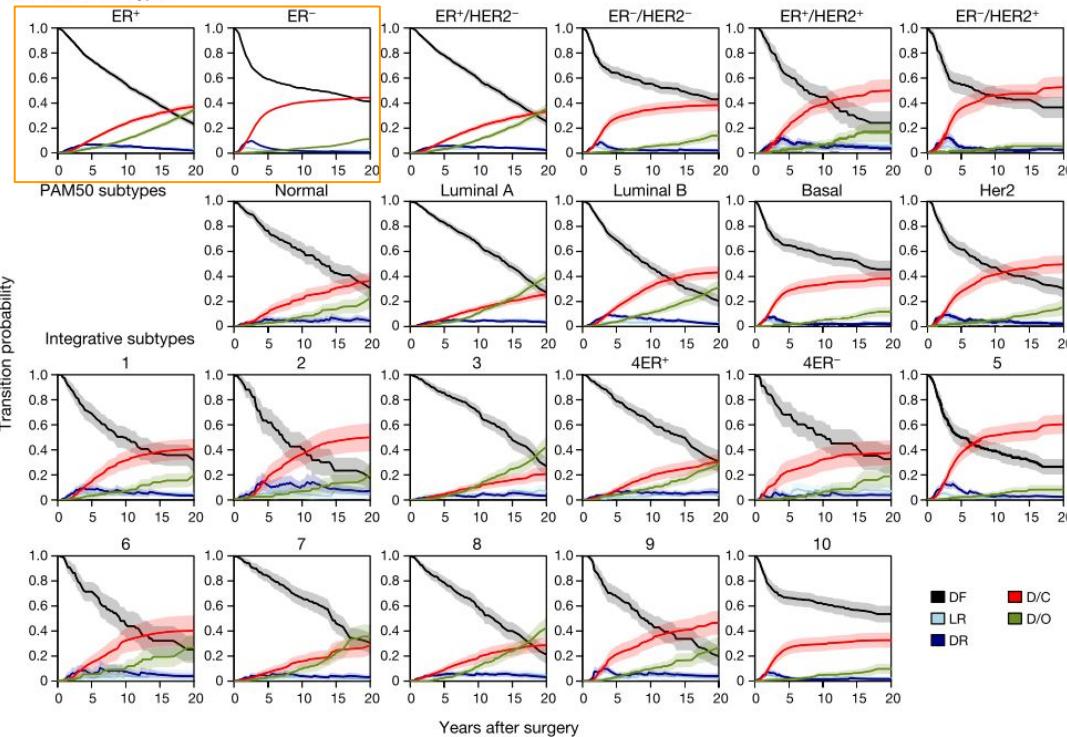


Nonhomogenous (semi)-Markov-chain model

$$\begin{aligned}
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})) \\
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
\end{aligned}$$

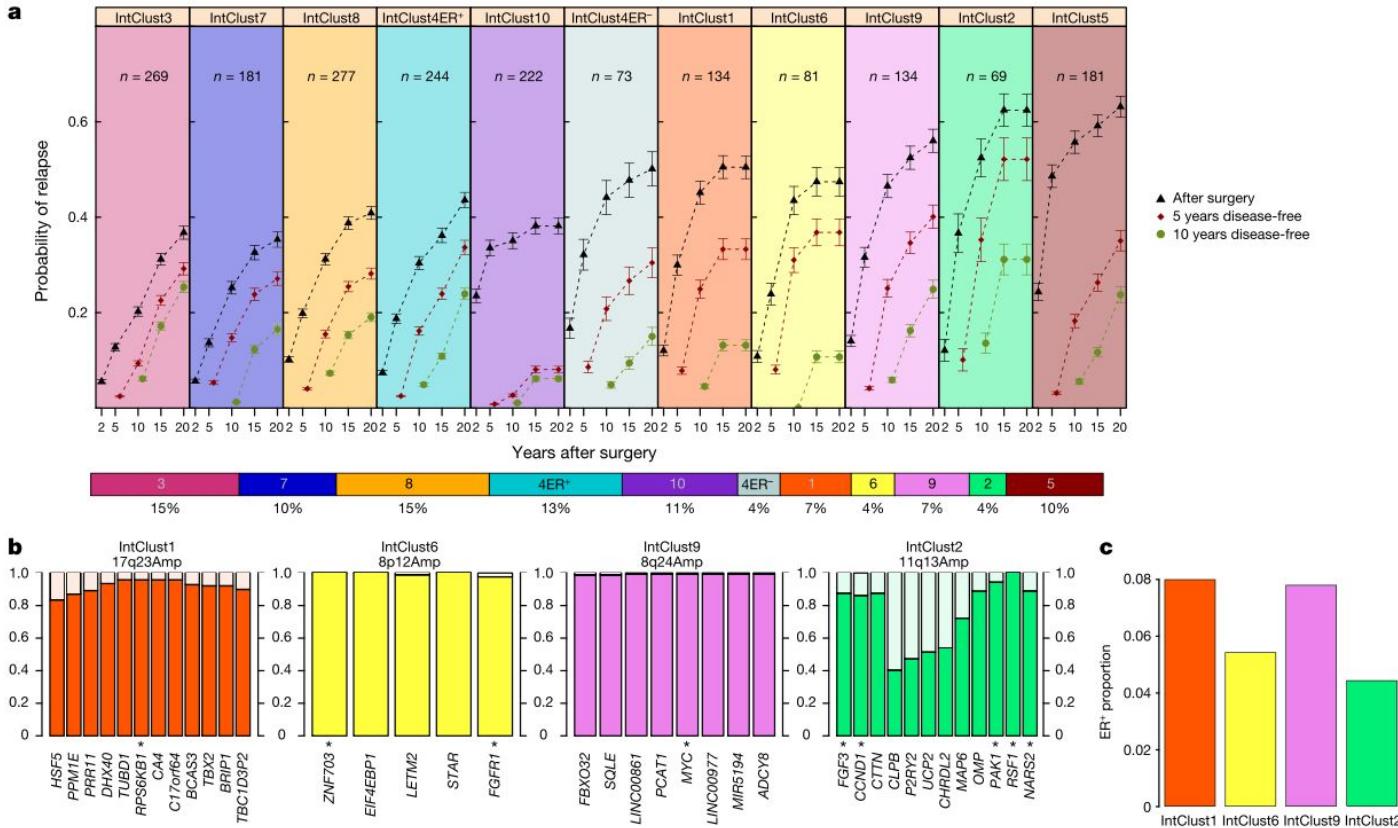
Results

b ER/IHC subtypes



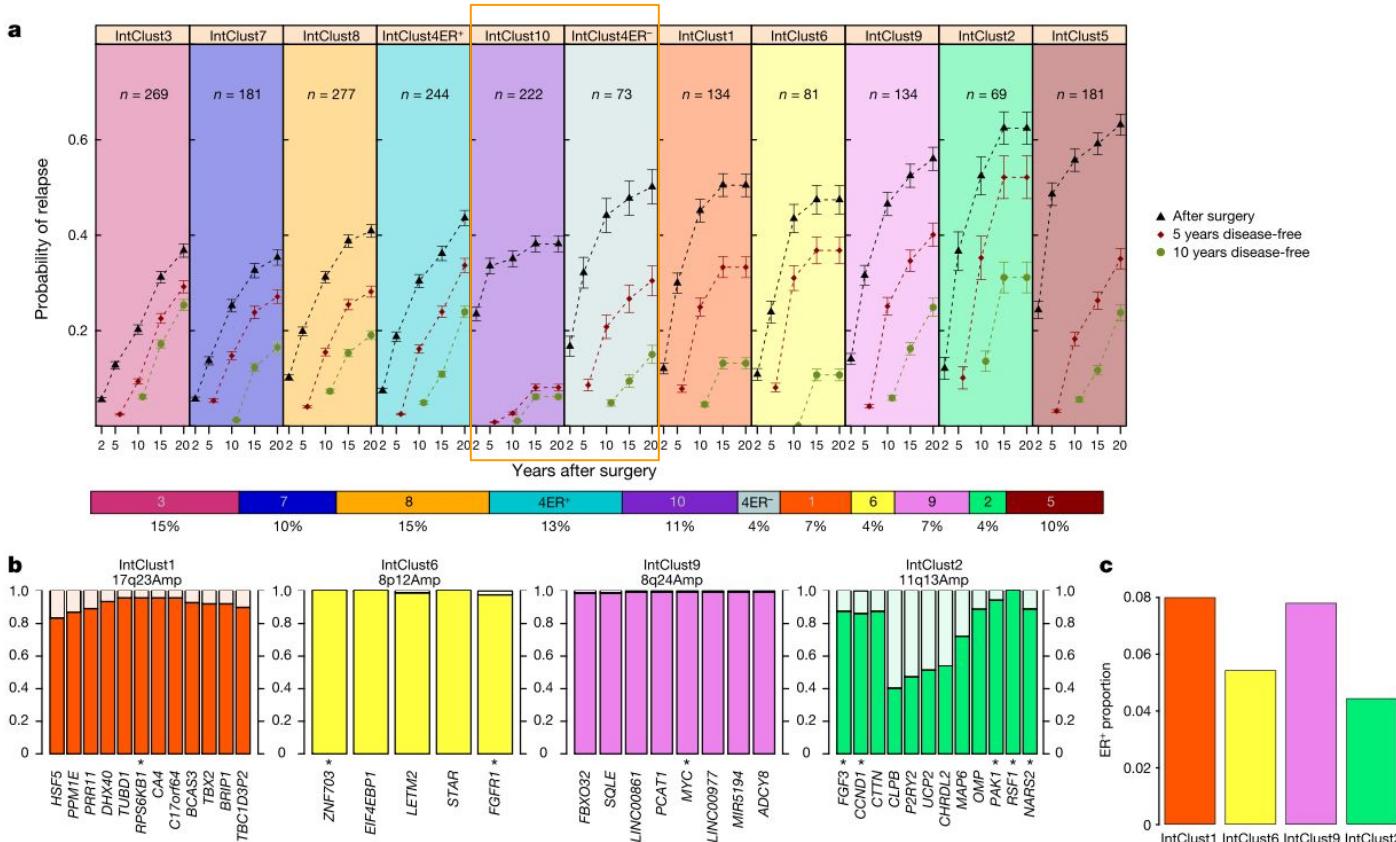
Transition probabilities are characteristically different for patients of different subtypes!

Results



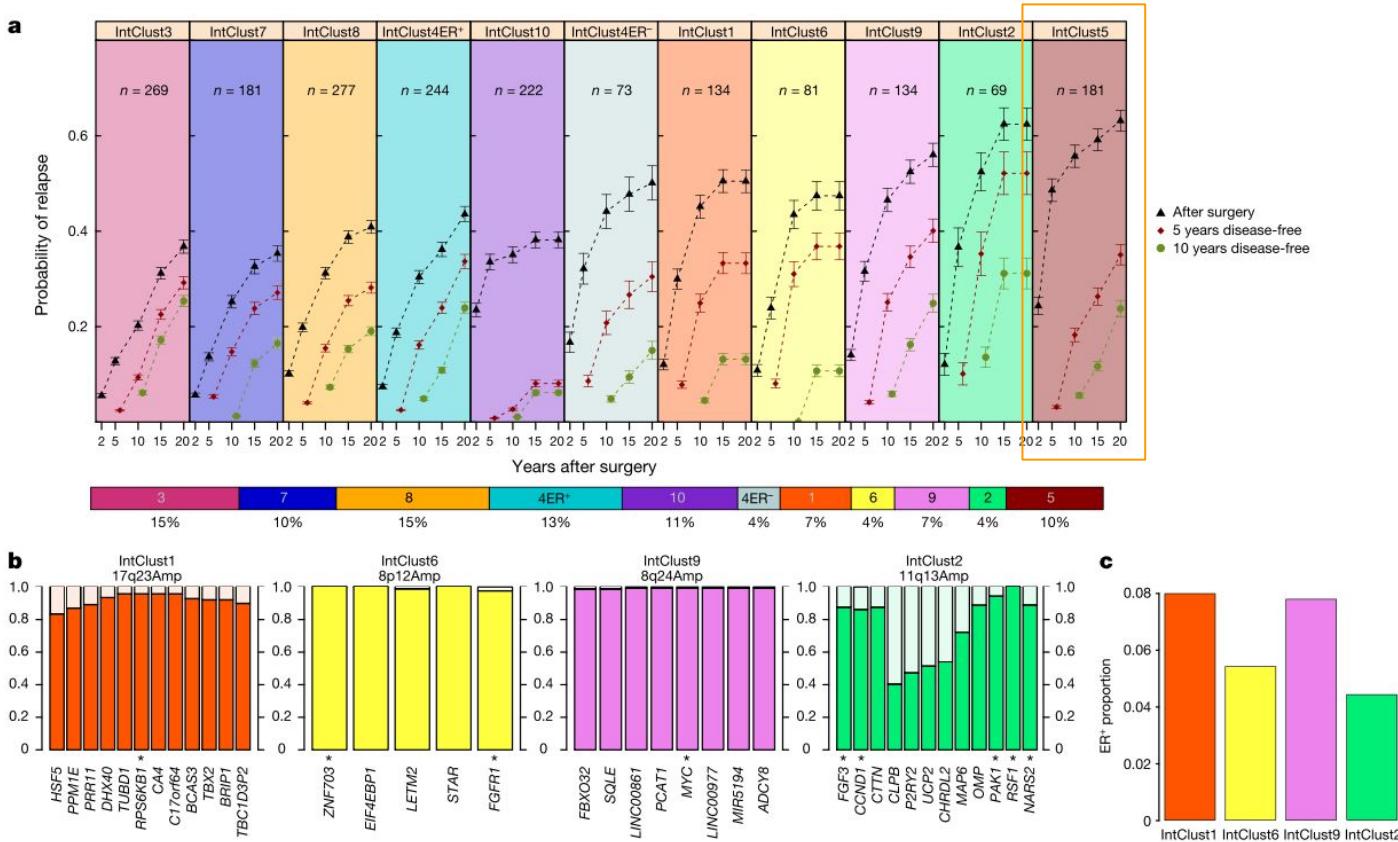
Results

both TNBC, but late relapse is more frequent for 4ER-



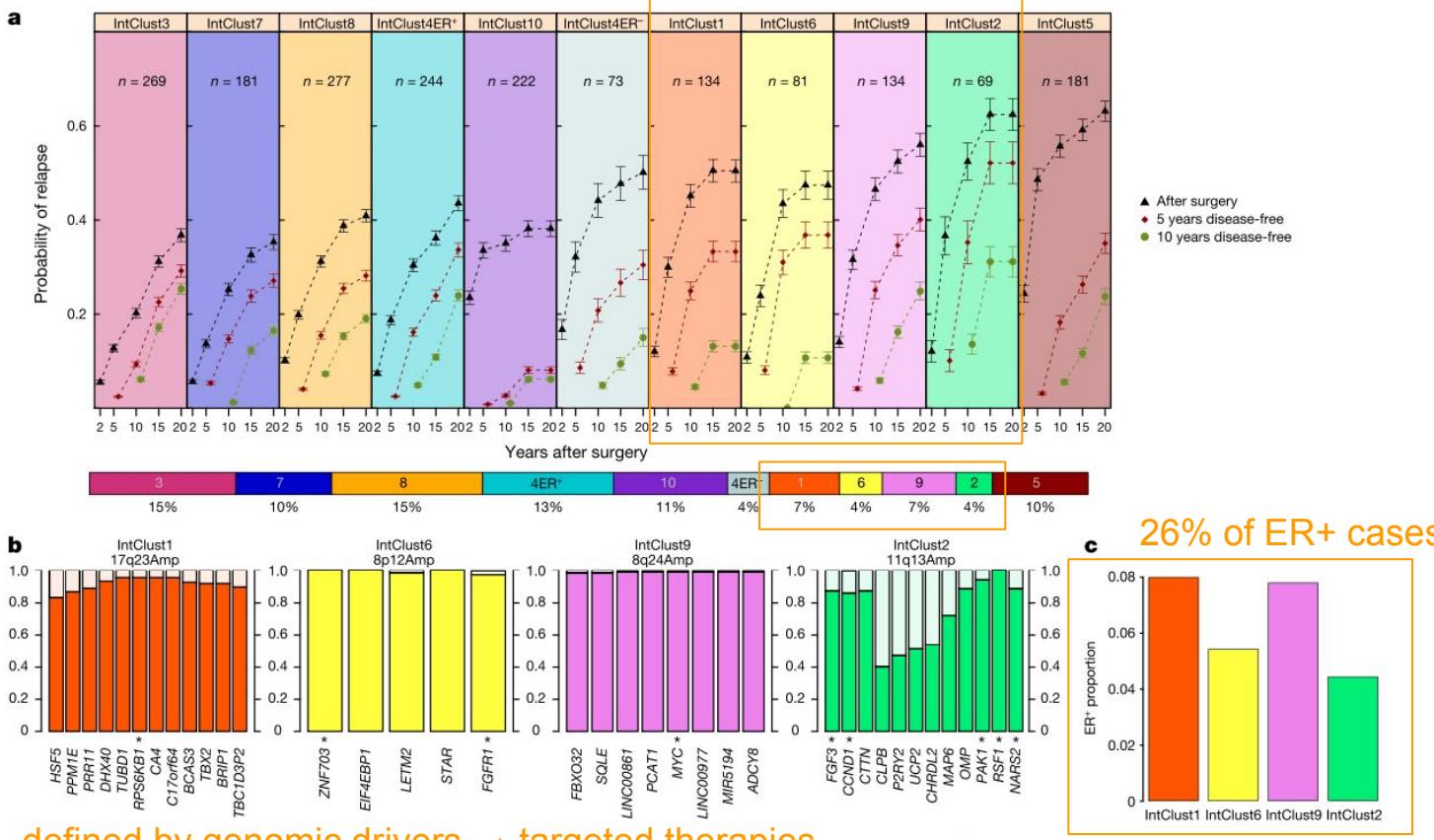
Results

HER2+ cases before use of trastuzumab



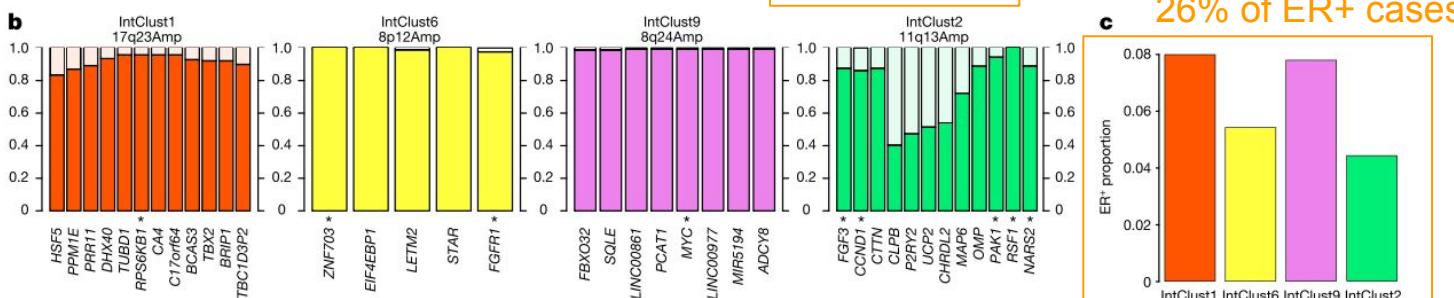
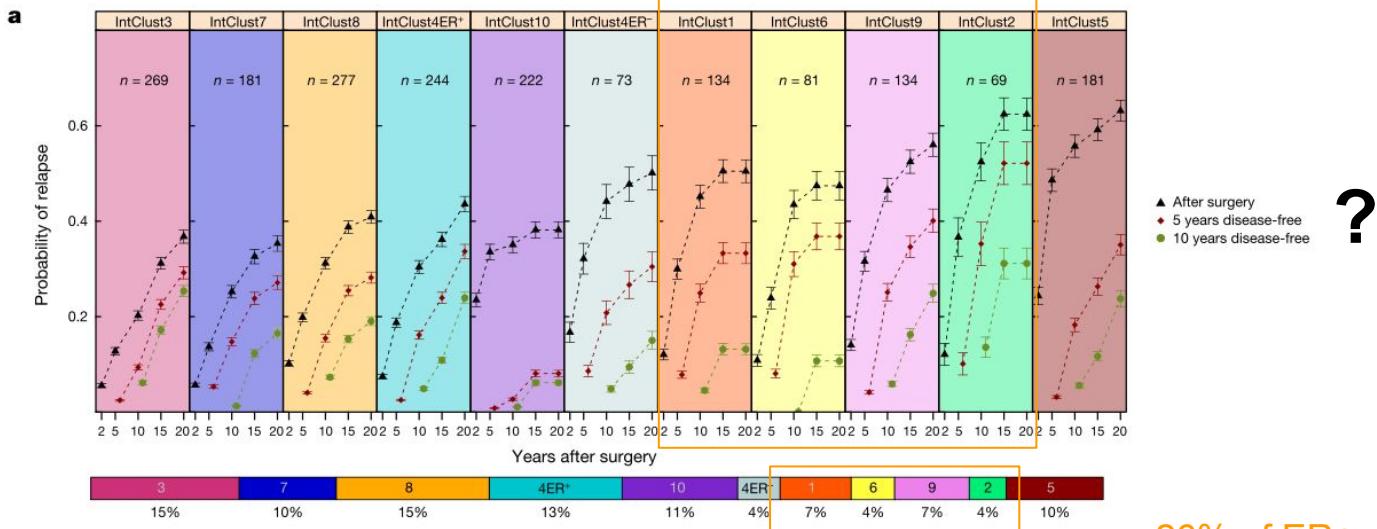
Results

late relapse subtypes



Results

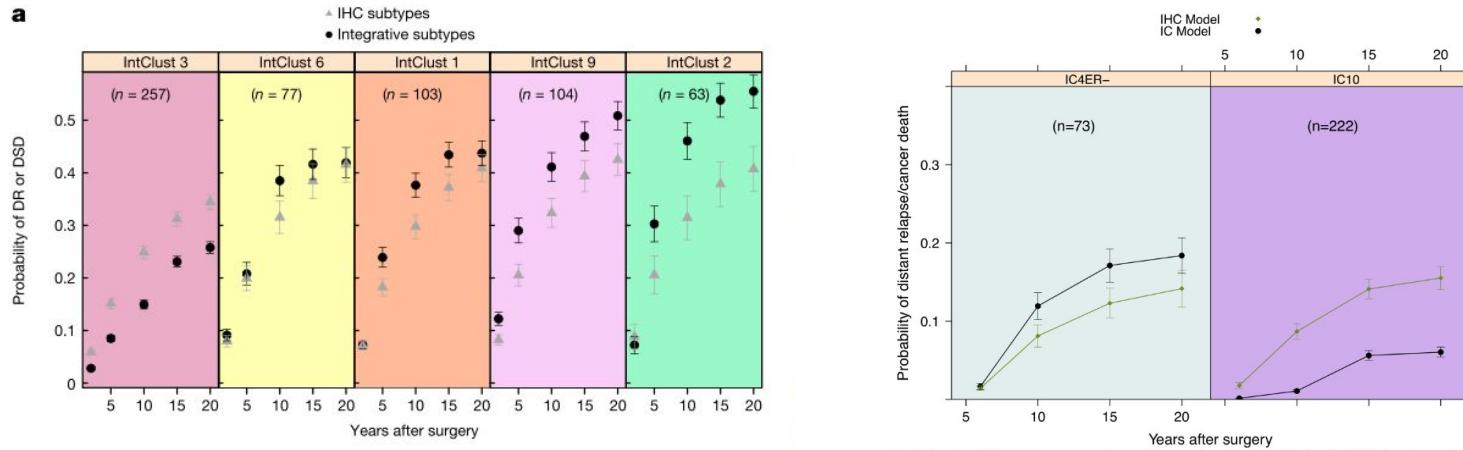
late relapse subtypes



defined by genomic drivers → targeted therapies

Results

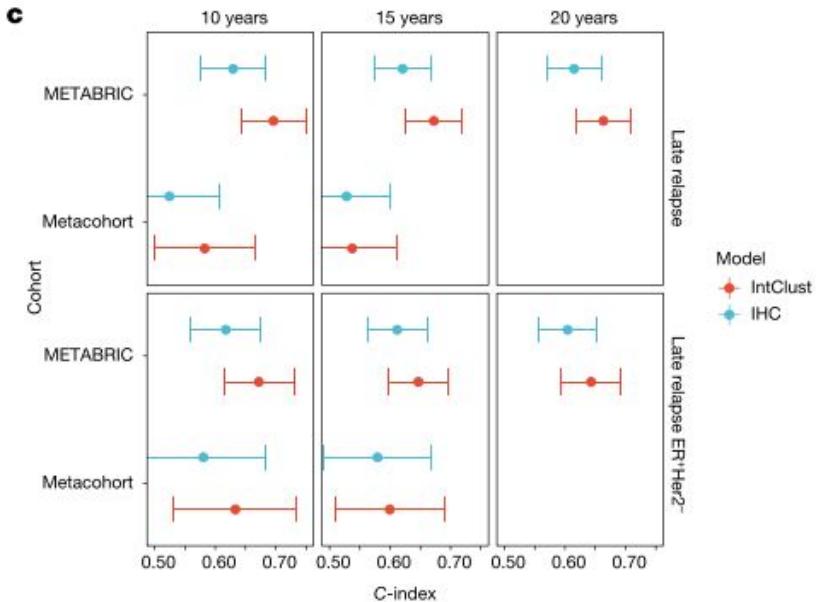
Using IntClust subtypes is **more informative** than using IHC subtypes



Differences in late relapse probabilities would not be apparent from the IHC subtype model.

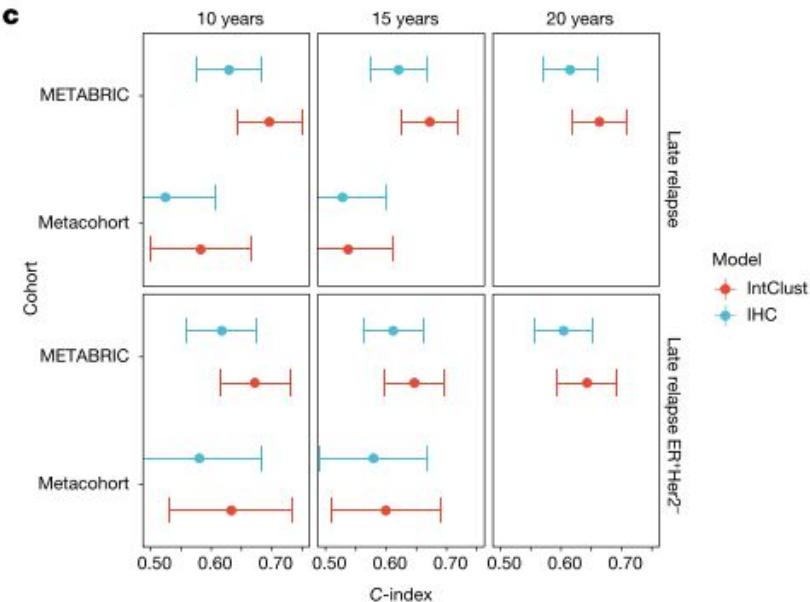
Results

Using IntClust subtypes results in more accurate models than using IHC subtypes



Results

Using IntClust subtypes results in more accurate models than using IHC subtypes

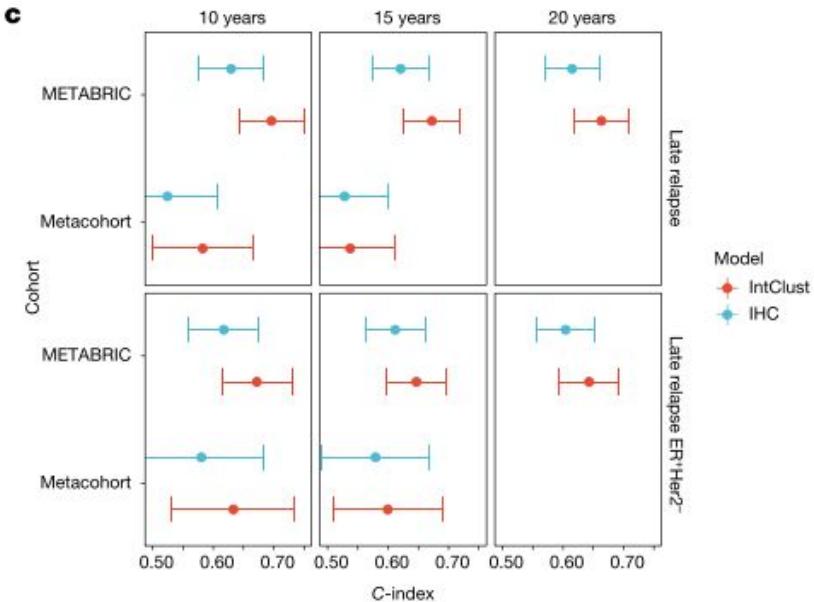


Concordance

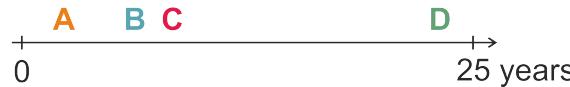
- Take all possible pairs of patients
- Compare the **true order of their transitions** to the **predicted order**
 - If same: +1
 - If not: 0
- Average for all patient pairs
- **Actual transition times are disregarded**, only order is checked

Results

Using IntClust subtypes results in more accurate models than using IHC subtypes

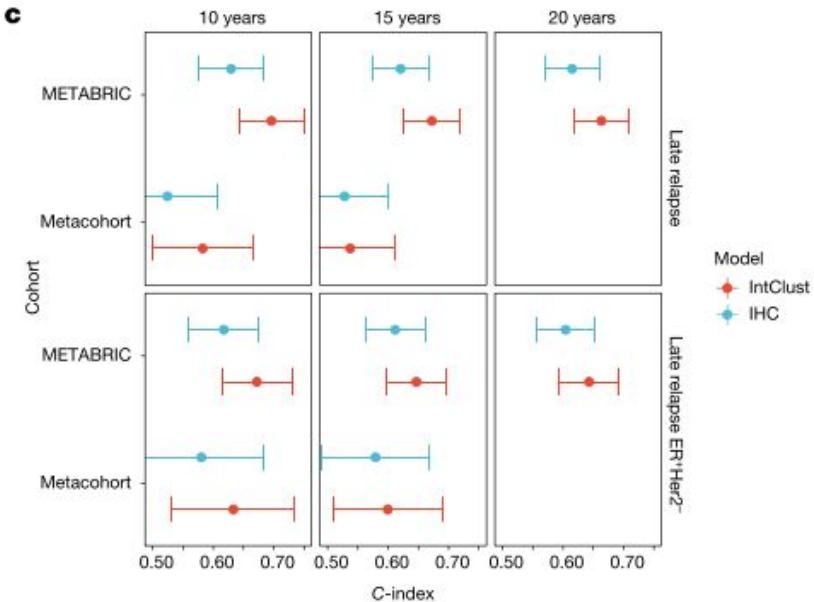


Concordance

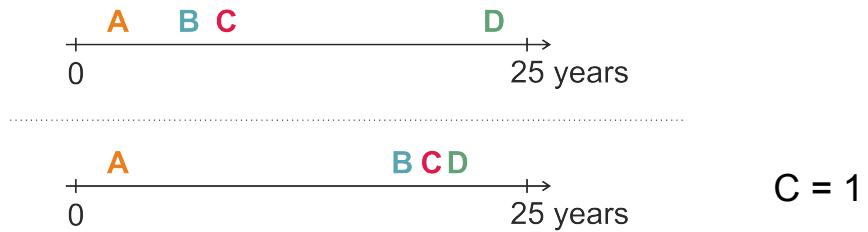


Results

Using IntClust subtypes results in more accurate models than using IHC subtypes

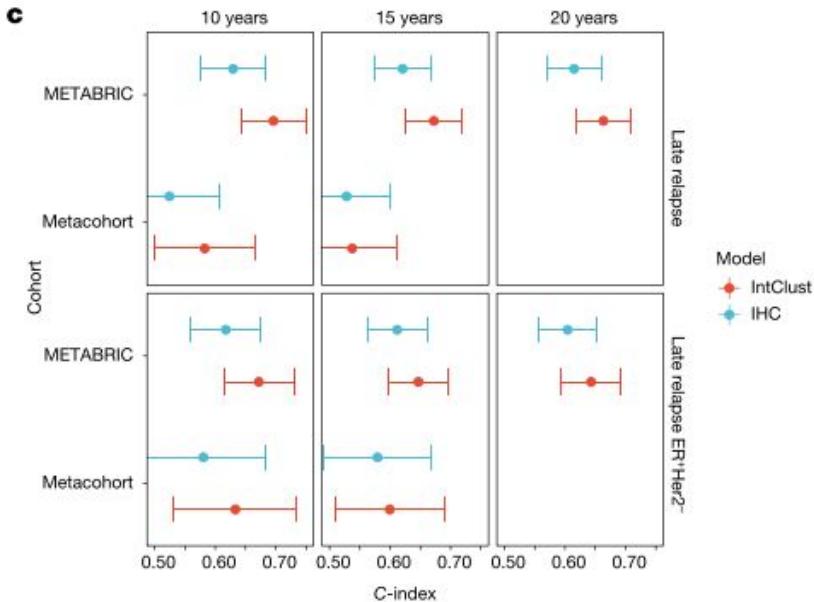


Concordance

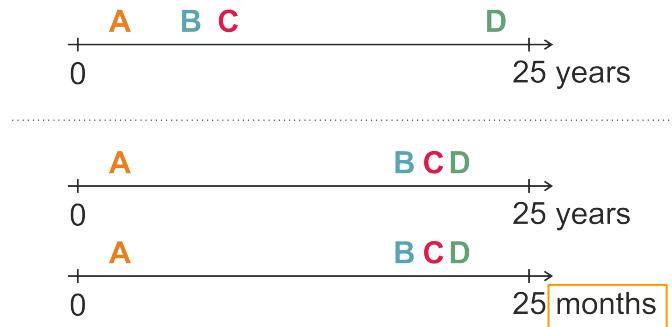


Results

Using IntClust subtypes results in more accurate models than using IHC subtypes



Concordance

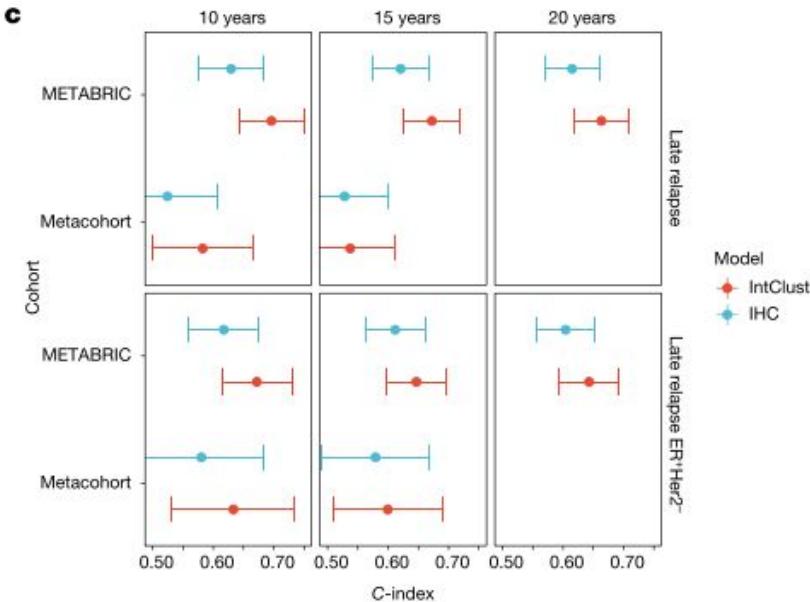


C = 1

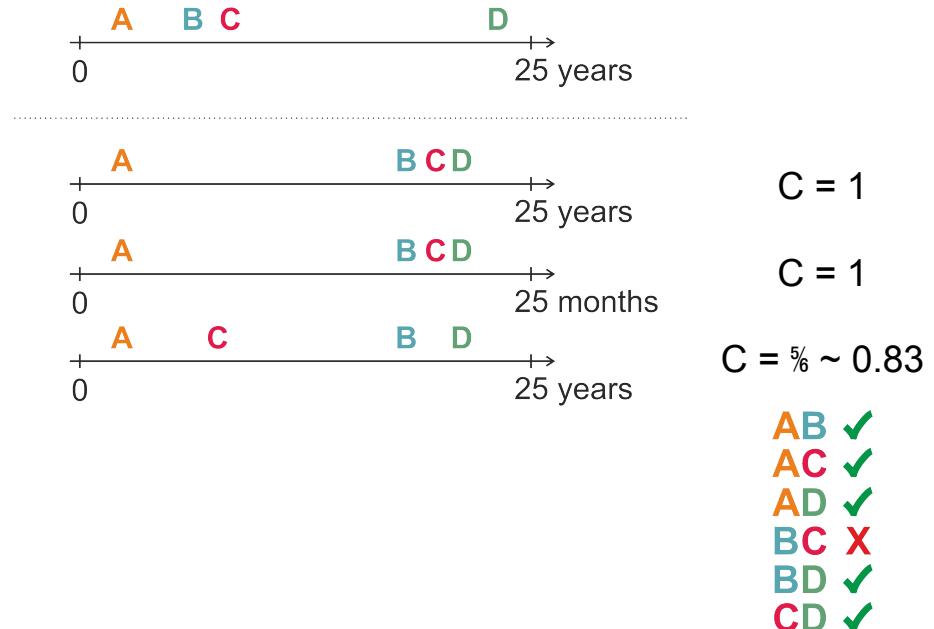
C = 1

Results

Using IntClust subtypes results in more accurate models than using IHC subtypes

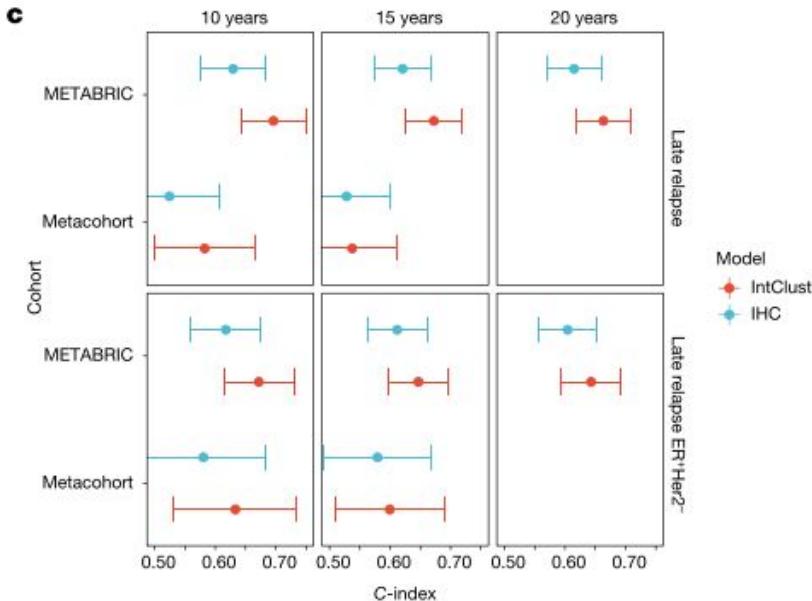


Concordance

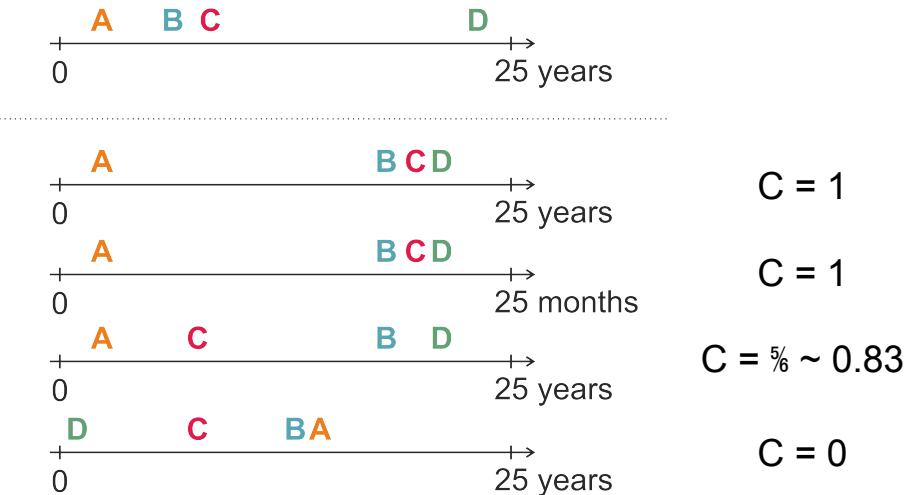


Results

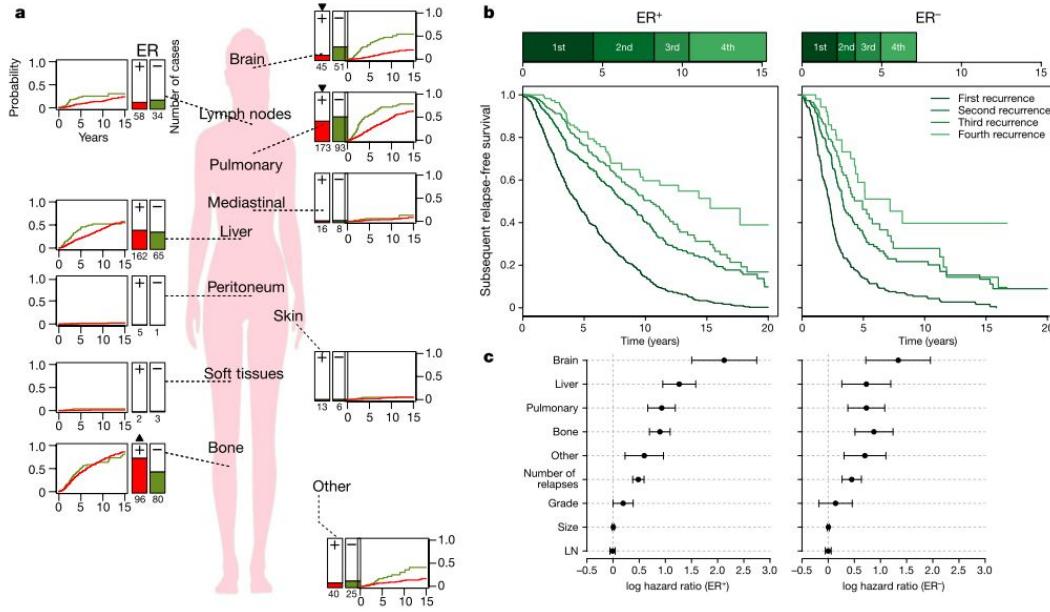
Using IntClust subtypes results in more accurate models than using IHC subtypes



Concordance



Results



- no sites of metastasis are exclusive to ER+ or ER- disease
- bone metastases take a long time to develop, and ER- patients tend to die of other metastases first
- Rates of distant mets:
 - ER- : rapid series of relapses early after diagnosis
 - ER+: just one early relapse (commonly bone), if a second occurred, the probability of additional relapses increased
- after distant recurrence, subtype continues to dictate the rate of subsequent metastases

Summary

- Used on already existing model with slight modifications
- Implementation in R already available (survival, mstate)
- Main result: **IntClust subtyping of BC cases is superior** to other subtyping methods
 - Original article:

Ali *et al.* *Genome Biology* 2014, **15**:431
<http://genomebiology.com/2014/15/8/431>



RESEARCH

Open Access

Genome-driven integrated classification of breast cancer validated in over 7,500 samples

H Raza Ali^{1,2,4†}, Oscar M Rueda¹, Suet-Feung Chin¹, Christina Curtis⁵, Mark J Dunning¹, Samuel AJR Aparicio⁶
and Carlos Caldas^{1,3,4*}

- Code available on github
- **Online prediction tool:** <https://caldaslab.cruk.cam.ac.uk/brcarepred>