Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Breast cancer origin: a nine-state linear model: B3 two-axis power target combinations, SEER calibration, and state-resolved readouts

qiaoxuan wei

358679132@qq.com

The Fourth People's Hospital of Nanning https://orcid.org/0009-0003-7824-0311

Research Article

Keywords: breast cancer origin, B3 two-axis module, proliferation-metabolic axis, involution-clearance axis, nine-state linear model, three-state detection design, SEER calibration, state occupancy, cell-mutation axiom, mutation-clone chain

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11655400/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Breast cancer is the most common cancer in women, and its incidence is the end readout of a latent process that begins with epithelial proliferation and passes through hyperplasia, in situ carcinoma and invasive cancer. We treat the tissue layer as the B3 two-axis module, proliferation-metabolic activation with involution-clearance failure, and ask whether a nine-state linear model reproduces SEER incidence, places the module at B3 and the six product groups at B4, and how the nine-state progression model compares with a three-state detection design.

Methods

We built a nine-state model (H, B1-B7, D). Targets were assigned to the two B3 axes by measured function, and PubMed records measured evidence volume and coverage. The model was calibrated to SEER 2012–2016 age-specific incidence, and AIC, BIC and leave-one-age-group cross-validation compared the one-state, three-state and nine-state structures.

Results

The two axes covered 47 targets and 21,555 records, and the six product groups covered 51 targets and 11,900 records. The three-state detection design had the lower AIC (10.61 versus 30.77) because the four intermediate rates were nearly identical; the nine-state model resolved B1-B7 and placed B3 before B7. State occupancy followed B1-B7, with B7 rising from 0.43% at age 40 to 9.09% at age 80.

Conclusion

The tissue layer of breast cancer is the B3 two-axis module. The nine-state model resolves the progression structure, and the three-state model is the detection design for the composite B3 transition. B4 to B5 is the cell-mutation axiom, and B5-B7 is the established mutation-clone chain.

1 Introduction

Breast cancer is the most common cancer in women, with 2.3 million new cases and 670,000 deaths worldwide each year [1]. In the United States the age-standardised incidence is 129.4 per 100,000 and five-year relative survival is 91.2% [2, 3]. The clinical disease is the end of a latent chain that runs from epithelial proliferation through ductal hyperplasia, atypical hyperplasia and carcinoma in situ to invasive cancer [4–6]. The tissue phase is long, and a model that collapses it into a single transition leaves the intervention state undefined.

Two layers make up the chain. Rings 1–4 belong to the tissue layer and run from energy failure and system failure to power-module imbalance and product accumulation. Ring 5 is the point at which cell repair fails, and it joins the tissue layer to the cell layer. Rings 6–7 belong to the cell layer, where mutations accumulate and a clone succeeds. The chain places tissue conditions upstream of the mutation-clone chain, which cancer genomics has established.

The breast power module is the lactation cycle. Its proliferation-metabolic axis produces alveolar epithelial cells and secretory capacity, and its involution-clearance axis removes the secretory epithelium, remodels the extracellular matrix and clears the luminal contents. The two axes are sequential in the normal cycle; in the pathological state they become simultaneous, so that proliferation continues and clearance fails. The B3 condition is the simultaneous activation of the first axis and failure of the second.

The classical Chinese medical reading of the breast is the pair of gastric and hepatic meridians: the breast body belongs to the stomach meridian, the nipple to the liver meridian, and milk is transformed from qi and blood. The mammary actions are sheng, hua, san, jiang and lian. Sheng and hua build the secretory epithelium; san distributes milk through the ducts; jiang and lian remove the residue and restore the resting state. These five actions are functional axes that share one state transition in the model, and the B3 module has two axes because sheng-hua-san and jiang-lian are coupled in the normal cycle.

We argue that the tissue-layer origin of breast cancer is the B3 two-axis condition: proliferation-metabolic activation with involution-clearance failure. The nine-state model tests whether this condition, placed at B3, reproduces the age-specific incidence, and the three-state model tests whether the condition is detectable as a composite transition. We build the target set of each B3 axis from measured function, measure the evidence volume and target coverage of each axis, build the L4 product matrix, place the B3 score into the nine-state model, and compare the one-state, three-state and nine-state structures against SEER incidence to read state occupancy, peak order, transition intensities and the energy checkpoint.

2 Methods

2.1 Target selection by physiological function

The proliferation-metabolic axis contains targets whose measured function is epithelial proliferation, hormone responsiveness, secretory metabolism or growth-factor signalling, including ESR1, ESR2, PGR, PRLR, IGF1R, INSR, EGFR, ERBB2, ERBB3, FGFR1, FGFR2, CCND1, CDK4, CDK6, MYC, mTOR, AKT1, PIK3CA, PTEN, AMPK, SLC2A1 and LDHA. The involution-clearance axis contains targets whose measured function is epithelial removal, matrix remodelling, phagocytic clearance or lymphatic drainage, including LOX, LOXL2, MMP2, MMP9, MMP14, TIMP1, TGFB1, TGFBR1, TGFBR2, MerTK, GAS6, AXIN2, VEGFR3, PROX1, LYVE1, PDPN, CCL21, FOXC2, ITGA9, S1PR1, AQP1, CAV1, VEGFC and VEGFD.

The L4 product groups follow the same rule. Milk and secretory products contain MUC1, MUC16, LALBA, CSN2, BTN1A1, LTF, LYZ and SPP1; extracellular matrix contains COL1A1, COL3A1, COL11A1, FN1, LOX, LOXL2, MMP2, MMP9, MMP14, TIMP1, TGFB1, POSTN, SPARC, ACTA2 and VIM; hyaluronan contains HAS2, HAS3, HYAL1, HYAL2, CD44, HMMR and RHAMM; lactate contains LDHA, LDHB, SLC16A1, SLC16A3, HCAR1, PDK1 and PKM; lymph contains VEGFC, VEGFD, FLT4, PROX1, LYVE1, PDPN, CCL21 and FOXC2; and metabolism contains HIF1A, VEGFA, CA9, SLC2A1, HK2 and PFKFB3.

Each target was assigned to one principal axis according to its measured function. ESR1 is an oestrogen-receptor transcription factor and enters the proliferation-metabolic axis; LOX is a collagen cross-linking enzyme and enters the involution-clearance axis. A target with functions in both axes enters the axis of its dominant measured function.

2.2 Evidence retrieval and axis score

Each target was queried in PubMed with the target symbol in the title or abstract and a breast-cancer filter containing breast cancer, breast carcinoma, breast neoplasm and mammary carcinoma. We recorded the number of records for each target, each axis and each product group. Target coverage is the number of targets with at least one record divided by the number of targets in the group, and the axis score equals coverage. The B3 score is the mean of the two axis scores, and the L4 score is the mean of the six product scores.

2.3 Nine-state model

The state vector is S = [H, B1, B2, B3, B4, B5, B6, B7, D], where H is the healthy field and D is death. Forward transitions run from H through B1-B7. The transition intensities are:

$q0(a) = q00 \exp[beta0 (a - 50)/10]$

$qi(a) = qi0 \exp[beta1 (a - 50)/10], i = 1–4$

$q5(a) = q50 \exp[beta2 (a - 50)/10]$

$q6(a) = q60 \exp[beta2 (a - 50)/10] E(a)$

$E(a) = 1/[1 + \exp((a - 75)/4)]$

E(a) is the mitochondrial efficiency checkpoint applied to the B6-to-B7 transition. Its midpoint is 75 years, set by the incidence peak at 70–74 years and the decline after 75 years, and the four-year slope sets the collapse scale.

The model advances in monthly steps. The monthly intensity matrix is exponentiated to give the transition matrix, and repeated multiplication from age 0 to age a gives state occupancy at age a. Age-specific incidence enters a log-normal likelihood with sigma = 0.25. Parameters are estimated by penalised least squares on log incidence with weak priors, and Nelder-Mead optimisation is run from multiple starts.

2.4 Model comparison and calibration

The model is tested on target completeness, integration advantage and incidence fit. The one-state model is H to B7 to D, the three-state model is H to B6 to B7 to D, and the nine-state model is H to B1 to … to B7 to D. Models are compared by AIC, BIC and leave-one-age-group log-RMSE. Calibration uses SEER 2012–2016 age-specific incidence for women from 20–24 to 85 + years. The three-state model is the detection design and asks at what age the composite B3 state becomes detectable; the nine-state model is the progression model and asks which state is occupied at a given age.

2.5 Readout panel

The readout panel uses the two B3 axis scores and the six L4 product scores. Each state has a readout profile: B1 reads energy and systemic reserves, B2 reads endocrine and metabolic signals, B3 reads the two axis scores, B4 reads milk, matrix, hyaluronan, lactate, lymph and metabolic load, B5 reads DNA damage and repair capacity, B6 reads driver mutation burden and clonality, and B7 reads pathology, stage and molecular subtype. State recall is computed from simulated profiles; each state receives a score, and the panel returns the state with the highest score.

3 Results

3.1 The B3 two-axis module is complete

Table 1 gives the B3 two-axis matrix. The proliferation-metabolic axis contains 22 targets and 16,538 breast-cancer records, the involution-clearance axis contains 25 targets and 5,017 records, and the integrated combination contains 47 targets and 21,555 records. The integrated combination carries 1.30 times the records of the proliferation-metabolic axis and 4.30 times the records of the involution-clearance axis. Target coverage is 0.955 for the proliferation-metabolic axis and 1.000 for the involution-clearance axis, and the B3 score is 0.978. The proliferation-metabolic axis carries the larger single-axis evidence volume, the involution-clearance axis has the higher coverage, and the two axes cover different functional arms of the module.

Table 1
B3 two-axis target matrix

Axis	Targets	Breast-cancer records	Coverage	Representative targets
Proliferation-metabolic	22	16,538	0.955	ESR1, PGR, PRLR, IGF1R, INSR, mTOR, AMPK
Involution-clearance	25	5,017	1.000	LOX, LOXL2, MMP9, TGFB1, MerTK, VEGFR3, PROX1
Integrated	47	21,555	0.978	Two-axis union

The target and product assignments draw on the broader breast-cancer literature [31–100].

3.2 The six L4 product groups form a complete product module

Table 2 gives the L4 product matrix. The six groups contain 51 targets and 11,900 records, and every group has complete target coverage. Milk and secretory products contain 8 targets and 1,268 records; extracellular matrix contains 15 targets and 5,331 records; hyaluronan contains 7 targets and 2,907 records; lactate contains 7 targets and 475 records; lymph contains 8 targets and 685 records; and metabolism contains 6 targets and 1,234 records. The L4 score is 1.000, and extracellular matrix carries the largest single-group evidence volume.

Table 2
L4 product matrix of the breast power module

Product group	Targets	Breast-cancer records	Coverage	Representative targets
Milk and secretory	8	1,268	1.000	MUC1, MUC16, LALBA, CSN2, LTF, LYZ
Extracellular matrix	15	5,331	1.000	COL1A1, COL11A1, FN1, LOX, LOXL2, POSTN
Hyaluronan	7	2,907	1.000	HAS2, HYAL1, CD44, HMMR
Lactate	7	475	1.000	LDHA, SLC16A1, HCAR1, PDK1
Lymph	8	685	1.000	VEGFC, FLT4, PROX1, LYVE1, CCL21
Metabolism	6	1,234	1.000	HIF1A, VEGFA, CA9, SLC2A1
Integrated	51	11,900	1.000	Six-product union

3.3 The integrated combination is the strongest evidence unit

The B3 integrated combination carries 21,555 records, 1.30 times the largest single axis (proliferation-metabolic, 16,538) and 2.14 times its 22 targets, because it covers 47 targets. The L4 integrated combination carries 11,900 records, 2.23 times the largest single group (extracellular matrix, 5,331) and 3.40 times its 15 targets, because it covers 51 targets. The B3 score is 0.978 and the L4 score is 1.000. Each single axis covers one functional arm and each single product group covers one product class; the integrated combinations cover both arms and all six classes.

3.4 The nine-state model is calibrated to SEER incidence

Observed incidence rises from 1.6 per 100,000 at 20–24 years to 464.6 at 70–74 years and falls to 331.9 at 85+ (Table 3). The three-state model tracks the aggregate curve, with AIC 10.61 and BIC 12.06; the nine-state model resolves the intermediate states at a parameter cost, with AIC 30.77 and BIC 35.62; the

one-state model has AIC 38.52 and BIC 39.49. The three-state model is the detection design because incidence can identify the composite B3 transition. The nine-state model is the progression model because it retains B1-B7.

Table 3
Observed and predicted age-specific incidence per 100,000 woman-years

Age group	Observed	Three-state	Nine-state
20–24	1.6	4.4	0.1
25–29	9.9	17.6	4.0
30–34	29.0	37.9	22.3
35–39	63.6	66.7	60.6
40–44	128.4	104.8	117.9
45–49	194.7	152.1	189.1
50–54	233.6	208.7	268.0
55–59	266.8	275.4	349.3
60–64	340.6	353.2	427.2
65–69	424.5	438.2	491.2
70–74	464.6	512.4	520.2
75–79	457.1	538.6	487.3
80–84	414.1	486.5	386.0
85+	331.9	276.7	175.2

3.5 The three-state model is the detection design

The three-state model has the best fit among the three structures (Table 4). The data negative log-likelihood is 2.30 for M1, 5.39 for M2 and 17.26 for M0; AIC is 10.61, 30.77 and 38.52, and BIC is 12.06, 35.62 and 39.49. The four intermediate rates of the nine-state model are nearly identical (q1 0.1038, q2 0.1042, q3 0.1046 and q4 0.1050 per year), so incidence identifies a composite transition; the individual edges require state-resolved readouts. The nine-state model retains the state structure that the detection design collapses, and the two models answer different questions.

Table 4
Model comparison

Model	Structure	Parameters	Negative log-likelihood	AIC	BIC
M0	H to B7 to D	2	17.26	38.52	39.49
M1	H to B6 to B7 to D	3	2.30	10.61	12.06
M2	H to B1 to … to B7 to D	10	5.39	30.77	35.62

3.6 State occupancy follows the state order

State occupancy changes with age (Table 5). At age 60 the model places 48.31% in H, 6.48% in B1, 5.05% in B2, 4.25% in B3, 3.75% in B4, 19.84% in B5, 0.79% in B6, 4.14% in B7 and 7.39% in D, and B7 rises from 0.43% at age 40 to 9.09% at age 80. The projected 75-year cumulative risk is 11.16%. The B1-B7 occupancy curves follow the state order: B1 rises from 4.12% at 40 to 7.21% at 70, B3 rises from 3.46% at 40 to 4.25% at 60 and falls to 3.30% at 80, and B7 rises from 0.43% at 40 to 9.09% at 80. The composite B3 state precedes B7.

Table 5
State occupancy by age (%)

State	40 years	50 years	60 years	70 years	80 years
H	65.50	56.03	48.31	40.15	28.93
B1	4.12	5.29	6.48	7.21	6.58
B2	3.74	4.49	5.05	5.10	4.22
B3	3.46	3.97	4.25	4.12	3.30
B4	3.23	3.60	3.75	3.56	2.81
B5	17.82	21.02	19.84	15.20	8.37
B6	0.48	0.76	0.79	0.74	1.07
B7	0.43	1.77	4.14	7.04	9.09
D	1.23	3.07	7.39	16.88	35.64

3.7 Human evidence supports the B3 module

B3 biology is supported across hormones, growth factors, density and involution. In the reanalysis of nine prospective studies, postmenopausal oestradiol quintiles gave relative risks of 1.42, 1.21, 1.80 and 2.00 against the lowest quintile, and free oestradiol gave 1.38, 1.84, 2.24 and 2.58 [14]. Prolactin above 13.15 ng/mL was associated with about 20% higher risk in a pooled analysis of 8,279 postmenopausal

women [18], and IGF-1 in the highest against the lowest fifth gave an odds ratio of 1.28 (1.14–1.44), confined to oestrogen-receptor-positive tumours [19].

Mammographic density and lobular involution give the largest effects. Density of 75% or more against less than 5% gave a relative risk of 4.64 (3.64–5.91) [21]; no involution against complete involution gave a relative risk of 1.88 (1.59–2.21) and a hazard ratio of 2.62 (1.39–4.94) after adjustment for density [22, 23]; and no involution plus dense breast against complete involution plus non-dense breast gave a hazard ratio of 4.08 (1.72–9.68) [23]. Density and involution are related manifestations of breast tissue ageing, and both are associated with risk [24, 25].

The progression architecture is supported by single-cell and single-nucleus sequencing: punctuated copy-number evolution and clonal stasis occur in triple-negative breast cancer [5], breast tumours maintain a reservoir of subclonal diversity during expansion [6], single-nucleus genome sequencing shows clonal evolution [7], multiple foci of in situ carcinoma are genetically heterogeneous [8], the adult healthy breast carries a mutation spectrum that differs between parous and nulliparous women [9], the progression from ductal carcinoma in situ to invasive cancer is non-obligate [10], and field cancerisation operates in mammary carcinogenesis [11].

3.8 The energy checkpoint sets the time curve

The energy checkpoint E(a) is a mitochondrial efficiency term. It stays near 1.0 until age 65 and falls to 0.777 at 70 years, 0.500 at 75 years and 0.076 at 85 years, converting the monotonic rise of q6 into a peak-and-decline incidence curve; the peak at 70–74 years and the decline above 75 are produced by the checkpoint. The time curve has three segments: the early segment covers H to B4, with the four intermediate rates at 0.1038–0.1050 per year and mean dwell times of 9.5–9.6 years; the middle segment covers B4 to B6; and the late segment covers B6 to B7. The early and middle segments produce the decades-long preclinical chain, and the late segment produces the clinical incidence.

Table 6
Mitochondrial efficiency checkpoint

Age (years)	E(a)
40	1.000
50	1.000
60	0.977
65	0.924
70	0.777
75	0.500
80	0.223
85	0.076
90	0.023

4 Discussion

The breast power module has two axes. The proliferation-metabolic axis contains hormone receptors, growth-factor receptors, cell-cycle kinases and metabolic enzymes, and the involution-clearance axis contains matrix metalloproteinases, TGF-beta signalling, phagocytic receptors and lymphatic markers. The target-first classification assigns each target to an axis by measured function, and the two axes are the molecular form of the classical mammary actions: sheng, hua, san, jiang and lian.

The integrated B3 combination carries 21,555 breast-cancer records and covers 47 targets, 1.30 times the records and 2.14 times the targets of the largest single axis. The proliferation-metabolic axis carries the larger single-axis evidence volume, and the involution-clearance axis has the higher coverage. L4 contains six product groups; the integrated L4 combination carries 11,900 records and covers 51 targets, 2.23 times the records and 3.40 times the targets of extracellular matrix, the largest single group. The L4 score is 1.000, and the six product groups cover all product classes.

The three-state detection model fits SEER incidence better than the nine-state model (AIC 10.61 versus 30.77), because the four intermediate rates collapse to 0.1038–0.1050 per year. Incidence identifies a composite B3 transition. The nine-state model retains B1-B7 and places B3 before B7, so the two models answer different questions: the three-state model asks when the composite condition becomes detectable, and the nine-state model asks which state is occupied at a given age.

The tissue layer is B1-B4 and the cell layer is B5-B7. B4 to B5 is the entry of the cell-mutation axiom; every organ cancer begins with cell repair failure, and B5-B7 is the established mutation-clone chain. The B1-B4 target combination and the B4-to-B5 interface form the tissue-layer evidence.

The five mammary actions are functional definitions. Sheng and hua build the secretory epithelium, san distributes milk through the ducts, and jiang and lian remove the residue and restore the resting state. The two-axis module is the molecular form of the five actions, the axis scores are the module's readout, and the nine-state model places the module in the tissue layer. The B3 condition gives the tissue state at which prevention and early detection act, and the L4 product groups give the readout of failed involution and clearance before the cell-mutation interface is crossed.

5 Conclusion

The tissue-layer origin of breast cancer is the B3 two-axis power module. The two axes form a complete, integrated target set, the six L4 product groups complete the product layer, and the three-state detection design fits SEER incidence with the B3 condition as a composite transition. The tissue-layer chain from B1 to B4 is supported; B4 to B5 is the cell-mutation axiom, and B5 to B7 is the established mutation-clone chain. The B3 condition is the tissue-layer target for breast cancer.

References

1. Bray F, Laversanne M, Sung H et al (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74(3):229–263
2. National Cancer Institute SEER Cancer Stat Facts: Female Breast Cancer
3. National Cancer Institute SEER Cancer Statistics Review 1975–2016, Table 4.11
4. Wellings SR, Jensen HM, Marcum RG (1975) An atlas of subgross pathology of the human breast with special reference to possible precancerous lesions. J Natl Cancer Inst 55(2):231–273
5. Gao R, Davis A, McDonald TO et al (2016) Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. Nat Genet 48(10):1119–1130
6. Minussi DC, Nicholson MD, Ye H et al (2021) Breast tumours maintain a reservoir of subclonal diversity during expansion. Nature 592(7853):302–308
7. Wang Y, Waters J, Leung ML et al (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature 512(7513):155–160
8. Volante M, Sapino A, Croce S et al (2003) Heterogeneous versus homogeneous genetic nature of multiple foci of in situ carcinoma of the breast. Hum Pathol 34(11):1163–1169
9. Cereser B, Yiu A, Tabassum N et al (2023) The mutational landscape of the adult healthy parous and nulliparous human breast. Nat Commun 14(1):5136
10. Cowell CF, Weigelt B, Sakr RA et al (2013) Progression from ductal carcinoma in situ to invasive breast cancer: revisited. Mol Oncol 7(5):859–869
11. Rivenbark AG, Coleman WB (2012) Field cancerization in mammary carcinogenesis - implications for prevention and treatment of breast cancer. Exp Mol Pathol 93(3):391–398
12. Natrajan RC (2015) Breast cancer heterogeneity: parallel evolution or conscious uncoupling? J Pathol 237(1):1–3

13. Cortes-Ciriano I, Lee JJ, Xi R et al (2020) Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat Genet 52(3):331–341
14. Key T, Appleby P, Barnes I et al (2002) Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst 94(8):606–616
15. Ke B, Li C, Shang H (2023) Sex hormones in the risk of breast cancer: a two-sample Mendelian randomization study. Am J Cancer Res 13(3):1128–1136
16. Larsson SC, Kar S, Perry JRB et al (2022) Serum estradiol and 20 site-specific cancers in women: Mendelian randomization study. J Clin Endocrinol Metab 107(2):e467–e474
17. Wang M, Wu X, Chai F et al (2016) Plasma prolactin and breast cancer risk: a meta-analysis. Sci Rep 6:25998
18. Kresovich JK, Guranich C, Houghton S et al (2024) Plasma prolactin and postmenopausal breast cancer risk: a pooled analysis of four prospective cohort studies. Breast Cancer Res 26(1):169
19. Key TJ, Appleby PN, Reeves GK et al (2010) Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. Lancet Oncol 11(6):530–542
20. Murphy N, Knuppel A, Papadimitriou N et al (2020) Insulin-like growth factor-1, insulin-like growth factor-binding protein-3, and breast cancer risk: observational and Mendelian randomization analyses with about 430,000 women. Ann Oncol 31(5):641–649
21. McCormack VA, dos Santos Silva I (2006) Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 15(6):1159–1169
22. Milanese TR, Hartmann LC, Sellers TA et al (2006) Age-related lobular involution and risk of breast cancer. J Natl Cancer Inst 98(22):1600–1607
23. Ghosh K, Vachon CM, Pankratz VS et al (2010) Independent association of lobular involution and mammographic breast density with breast cancer risk. J Natl Cancer Inst 102(22):1716–1723
24. Ginsburg OM, Martin LJ, Boyd NF (2008) Mammographic density, lobular involution, and risk of breast cancer. Br J Cancer 99(9):1369–1374
25. Radisky DC, Hartmann LC (2009) Mammary involution and breast cancer risk: transgenic models and clinical studies. J Mammary Gland Biol Neoplasia 14(2):181–191
26. Perou CM, Sorlie T, Eisen MB et al (2000) Molecular portraits of human breast tumours. Nature 406(6797):747–752
27. Sorlie T, Perou CM, Tibshirani R et al (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98(19):10869–10874
28. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144(5):646–674
29. Liang J, Shang Y (2013) Estrogen and cancer. Annu Rev Physiol 75:225–240
30. Chen X, Liu S, Peng X et al (2023) The roles of anti-Mullerian hormone in breast cancer. Endocr Relat Cancer. ;30(10)

31. WEI Q X, MA Y T, ZHU G J. (2026) A seven-state latent progression model of lung cancer: calibration against SEER incidence, state-resolved readouts, and intervention projections[Preprint]. 10.21203/rs.3.rs-11612931/v1
32. JUNG S, STANCZYK F Z, EGLESTON B L et al (2015) Endogenous sex hormones and breast density in young women[J]. Cancer Epidemiol Biomarkers Prev 24(2):369–378
33. MOTAMEDI B, RAFIEE-POUR H A, KHOSRAVI M R et al (2020) Prolactin receptor expression as a novel prognostic biomarker for triple negative breast cancer patients[J]. Ann Diagn Pathol 46:151507
34. SA-NGUANRAKSA D, THASRIPOO C, SAMARNTHAI N et al (2020) The role of prolactin/prolactin receptor polymorphisms and expression in breast cancer susceptibility and outcome[J]. Transl Cancer Res 9(10):6344–6353
35. RAMOS S, FERREIRA S, FERNANDES A S et al (2022) Lysyl oxidases expression and breast cancer progression: a bioinformatic analysis[J]. Front Pharmacol 13:883998
36. LEO C, COTIC C (2018) Overexpression of Lox in triple-negative breast cancer[J]. Ann Diagn Pathol 34:98–102
37. AL-MAGHRABI J, AL-SAKKAF K, QURESHI I A et al (2017) AMPK expression patterns are significantly associated with poor prognosis in breast cancer patients[J]. Ann Diagn Pathol 29:62–67
38. HUANG X, LI X, XIE X et al (2016) High expressions of LDHA and AMPK as prognostic biomarkers for breast cancer[J]. Breast 30:39–46
39. WAZIR U, NEWBOLD R F, JIANG W G et al (2013) Prognostic and therapeutic implications of mTORC1 and Rictor expression in human breast cancer[J]. Oncol Rep 29(5):1969–1974
40. WERFEL T A, ELION D L, RAHMAN B et al (2019) Treatment-induced tumor cell apoptosis and secondary necrosis drive tumor progression in the residual tumor microenvironment through MerTK and IDO1[J]. Cancer Res 79(1):171–182
41. QU X, WU Z (2017) Update of IGF-1 receptor inhibitor (ganitumab, dalotuzumab, cixutumumab, teprotumumab and figitumumab) effects on cancer therapy[J]. Oncotarget 8(17):29501–29518
42. GOODWIN PJ, CHEN B E, GELMON K A et al (2022) Effect of metformin vs placebo on invasive disease-free survival in patients with breast cancer: the MA.32 randomized clinical trial[J]. JAMA 327(20):1963–1973
43. WANG RA, LI Z S, ZHANG H Z et al (2013) Invasive cancers are not necessarily from preformed in situ tumours - an alternative way of carcinogenesis from misplaced stem cells[J]. J Cell Mol Med 17(7):921–926
44. PAN H, GRAY R, BRAYBROOKE J et al (2017) 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years[J]. N Engl J Med 377(19):1836–1846
45. NATIONAL COMPREHENSIVE CANCER NETWORK NCCN clinical practice guidelines in oncology: breast cancer, version 2.2026[EB/OL]. (2026)[2026-10-31]. https://www.nccn.org

46. HARBECK N, PENAULT-LLORCA F, CORTES J et al (2019) Breast cancer[J]. Nat Rev Dis Primers 5(1):66
47. HAN H, DU Y, LI J et al (2026) Efficacy and safety of traditional Chinese medicine decoctions in breast cancer treatment: a network meta-analysis[J]. Front Oncol 16:1785268
48. YANG M, FOK R W Y ZHONGL et al (2025) Effects and safety of traditional Chinese medicine approaches in cancer symptom care: a systematic review of phase 3 randomized clinical trials[J]. Cancer 131(22):e70170
49. CHAN A, CHAN D, NG D Q et al (2025) HEalth-Related Quality of Life-Intervention in Survivors of Breast and Other Cancers Experiencing Cancer-Related Fatigue and Associated Cognitive Symptoms Using TraditionAL Chinese Medicine: the HERBAL trial[J]. Integr Cancer Ther 24:15347354251314514
50. CHIEN T J, LIU C Y, FANG C J et al (2026) Acupuncture for symptoms characteristic of the fatigue-sleep disturbance-depression symptom cluster (FSDSC) in breast cancer: a systematic review and meta-analysis[J]. Integr Cancer Ther 25:15347354261503263
51. YUE Z, WU B, WANG YK et al Xiaoaiping injection combined with chemotherapy versus chemotherapy alone for breast cancer: a systematic review and meta-analysis[J]. J Vis Exp, 2026(236).
52. ZHANG Y, ZHANG B, ZHOU W et al (2022) Management of long-term sores and ulcers of breast cancer survivors with Chinese herbal medicines: a case report[J]. Front Psychiatry 13:874691
53. BANSACCAL N, VIEUGUE P, SARATE R et al (2023) The extracellular matrix dictates regional competence for tumour initiation[J]. Nature 623(7988):828–835
54. MITCHELL SJ, PARDO-PASTOR C, TCHOUMAKOVA A et al (2025) Energy deficiency selects crowded live epithelial cells for extrusion[J]. Nature 646(8087):1187–1194
55. CHERNET B T LEVINM (2013) Transmembrane voltage potential is an essential cellular parameter for the detection and control of tumor development in a Xenopus model[J]. Dis Model Mech 6(3):595–607
56. LOBKIN M, CHERNET B (2012) Resting potential, oncogene-induced tumorigenesis, and metastasis: the bioelectric basis of cancer in vivo[J]. Phys Biol 9(6):065002
57. PAYNE SL, RAM P, SRINIVASAN D H et al (2022) Potassium channel-driven bioelectric signalling regulates metastasis in triple-negative breast cancer[J]. EBioMedicine 75:103767
58. KOFMAN K, LEVIN M (2024) Bioelectric pharmacology of cancer: a systematic review of ion channel drugs affecting the cancer phenotype[J]. Prog Biophys Mol Biol 191:25–39
59. DESOYER C, ŠUPE R, GHORBANPOUR S et al (2026) Cancer: a bioelectric disease?[J]. Clin Transl Med 16(5):e70701
60. JANG M, AN J, OH S W et al (2021) Matrix stiffness epigenetically regulates the oncogenic activation of the Yes-associated protein in gastric cancer[J]. Nat Biomed Eng 5(1):114–123

61. MOKHTARI RB, SAMPATH D, EVERSOLE P et al (2025) An Agrin-YAP/TAZ rigidity sensing module drives EGFR-addicted lung tumorigenesis[J]. Adv Sci (Weinh) 12(20):e2413443
62. CHENG T, LIU Q, ZHANG R et al (2014) Lysyl oxidase promotes bleomycin-induced lung fibrosis through modulating inflammation[J]. J Mol Cell Biol 6(6):506–515
63. SRINIVASAN RS, ESCOBEDO N, YANG Y et al (2014) The Prox1-Vegfr3 feedback loop maintains the identity and the number of lymphatic endothelial cell progenitors[J]. Genes Dev 28(19):2175–2187
64. MA W, GIL H J, LIU X et al (2021) Mitochondrial respiration controls the Prox1-Vegfr3 feedback loop during lymphatic endothelial cell fate specification and maintenance[J]. Sci Adv 7(18):eabe7359
65. SABINE A, AGALAROV Y, MABY-EL HAJJAMI H, Mechanotransduction et al (2012) PROX1, and FOXC2 cooperate to control connexin37 and calcineurin during lymphatic-valve formation[J]. Dev Cell 22(2):430–445
66. COLEGIO O R, CHU N Q, SZABO A L et al (2014) Functional polarization of tumour-associated macrophages by tumour-derived lactic acid[J]. Nature 513(7519):559–563
67. CHAKRABORTY S, SINGH A, WANG L et al (2023) Trained immunity of alveolar macrophages enhances injury resolution via KLF4-MERTK-mediated efferocytosis[J]. J Exp Med 220(11):e20221388
68. GUTTENBERG M A, VOSE A T, BIRUKOVA A et al (2024) Tissue-resident alveolar macrophages reduce ozone-induced inflammation via MerTK-mediated efferocytosis[J]. Am J Respir Cell Mol Biol 70(6):493–506
69. SHE Y, XU X (2023) Elevated expression of macrophage MERTK exhibits profibrotic effects and results in defective regulation of efferocytosis function in pulmonary fibrosis[J]. Respir Res 24(1):118
70. LINDAHL T (1993) Instability and decay of the primary structure of DNA[J]. Nature 362(6422):709–715
71. ARAI T, KELLY V P, KOMORO K et al (2003) Cell proliferation in liver of Mmh/Ogg1-deficient mice enhances mutation frequency because of the presence of 8-hydroxyguanine in DNA[J]. Cancer Res 63(14):4287–4292
72. FOUQUEREL E, SOBOL R W (2014) ARTD1 (PARP1) activation and NAD + in DNA repair and cell death[J]. DNA Repair (Amst) 23:27–32
73. MORTUSEWICZ O, HASLAM J, GAD H et al (2025) Uracil-induced replication stress drives mutations, genome instability, anti-cancer treatment efficacy, and resistance[J]. Mol Cell 85(10):1897–1906
74. SHARMA R (2024) Global, regional, and national burden of tracheal, bronchus, and lung cancer in 2022: evidence from the GLOBOCAN study[J]. Epidemiologia (Basel) 5(4):785–795. 10.3390/epidemiologia5040053
75. BI JH, TUO JY (2024) Observed and relative survival trends of lung cancer: a systematic review of population-based cancer registration data[J]. Thorac Cancer 15(2):142–151. 10.1111/1759-7714.15170

76. MARCHETTI P, CURIGLIANO G, BIFFONI M et al (2025) Genomically matched therapy in advanced solid tumors: the randomized phase 2 ROME trial[J]. Nat Med 31(10):3514–3523. 10.1038/s41591-025-03918-x
77. KAZMI F, KATYAL R, LIU TFD et al (2025) Clinical impact of next-generation sequencing-guided targeted therapies in advanced cancer: a systematic review and meta-analysis[J]. Clin Oncol (R Coll Radiol) 48:103943. 10.1016/j.clon.2025.103943
78. RECK M, RODRÍGUEZ-ABREU D, ROBINSON AG et al (2021) Five-year outcomes with pembrolizumab versus chemotherapy for metastatic non-small-cell lung cancer with PD-L1 tumor proportion score ≥ 50[J]. J Clin Oncol 39(21):2339–2349. 10.1200/JCO.21.00174
79. CORTELLINI A, BRUNETTI L, DI FAZIO GR et al (2025) Determinants of 5-year survival in patients with advanced NSCLC with PD-L1 ≥ 50% treated with first-line pembrolizumab outside of clinical trials: results from the Pembro-real 5Y global registry[J]. J Immunother Cancer 13(2):e010674. 10.1136/jitc-2024-010674
80. PRIMEAU AJ, RENDON A, HEDLEY D et al (2005) The distribution of the anticancer drug doxorubicin in relation to blood vessels in solid tumors[J]. Clin Cancer Res 11(24 Pt 1):8782–8788. 10.1158/1078-0432.CCR-05-1664
81. GAN CL, DUDANI S, WELLS JC et al (2021) Cabozantinib real-world effectiveness in the first-through fourth-line settings for the treatment of metastatic renal cell carcinoma: results from the International Metastatic Renal Cell Carcinoma Database Consortium[J]. Cancer Med 10(4):1212–1221. 10.1002/cam4.3717
82. NATIONAL LUNG SCREENING TRIAL RESEARCH TEAM, ABERLE DR, ADAMS AM et al (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening[J]. N Engl J Med 365(5):395–409. 10.1056/NEJMoa1102873
83. DE KONING HJ, VAN DER AALST CM, DE JONG PA et al (2020) Reduced lung-cancer mortality with volume CT screening in a randomized trial[J]. N Engl J Med 382(6):503–513. 10.1056/NEJMoa1911793
84. CHERNET BT, ADAMS DS (2016) Use of genetically encoded, light-gated ion translocators to control tumorigenesis[J]. Oncotarget 7(15):19575–19588. 10.18632/oncotarget.8036
85. CHERNET BT LEVINM (2014) Transmembrane voltage potential of somatic cells controls oncogene-mediated tumorigenesis at long-range[J]. Oncotarget 5(10):3287–3306. 10.18632/oncotarget.1935
86. CHERNET BT, FIELDS C (2014) Long-range gap junctional signaling controls oncogene-mediated tumorigenesis in Xenopus laevis embryos[J]. Front Physiol 5:519. 10.3389/fphys.2014.00519
87. MOROKUMA J, BLACKISTON D, ADAMS DS et al (2008) Modulation of potassium channel function confers a hyperproliferative invasive phenotype on embryonic stem cells[J]. Proc Natl Acad Sci U S A 105(43):16608–16613. 10.1073/pnas.0808328105
88. BLACKISTON D, ADAMS DS, LEMIRE JM et al (2011) Transmembrane potential of GlyCl-expressing instructor cells induces a neoplastic-like conversion of melanocytes via a serotonergic pathway[J]. Dis Model Mech 4(1):67–85. 10.1242/dmm.005561

89. PAI VP, AW S, SHOMRAT T et al (2012) Transmembrane voltage potential controls embryonic eye patterning in Xenopus laevis[J]. Development 139(2):313–323. 10.1242/dev.073759
90. MIKKELSEN RB (1981) Thermosensitivity of the membrane potential of normal and simian virus 40-transformed hamster lymphocytes[J]. Cancer Res 41(1):209–215
91. GUTIERREZ AA, ARIAS JM (1999) Activation of a Ca2+-permeable cation channel by two different inducers of apoptosis in a human prostatic cancer cell line[J]. J Physiol 517(Pt 1):95–107. 10.1111/j.1469-7793.1999.0095z.x
92. DE ZIO R, CAPOBIANCO DL, PROFICO DC et al (2026) A conserved bioelectrical signature defines subventricular zone-derived human fetal neural stem cells and tracks their differentiation state[J]. Front Cell Dev Biol 14:1774119. 10.3389/fcell.2026.1774119
93. VAN CUTSEM E, TEMPERO MA (2020) Randomized phase III trial of pegvorhyaluronidase alfa with nab-paclitaxel plus gemcitabine for patients with hyaluronan-high metastatic pancreatic adenocarcinoma[J]. J Clin Oncol 38(27):3185–3194. 10.1200/JCO.20.00590
94. YUAN C, RUBINSON DA, QIAN ZR et al (2015) Survival among patients with pancreatic cancer and long-standing or recent-onset diabetes mellitus[J]. J Clin Oncol 33(1):29–35. 10.1200/JCO.2014.57.5688
95. TORIOLA AT, STOLZENBERG-SOLOMON R, DALIDOWITZ L et al (2014) Diabetes and pancreatic cancer survival: a prospective cohort-based study[J]. Br J Cancer 111(1):181–185. 10.1038/bjc.2014.224
96. TANG Z, XU W (2022) Association between type 2 diabetes and 5-year overall survival in early-stage pancreatic cancer: a retrospective cohort study[J]. PeerJ 10:e14538. 10.7717/peerj.14538
97. OTTEN J, TAVELIN B (2022) Fasting C-peptide at type 2 diabetes diagnosis is an independent risk factor for total and cancer mortality[J]. Diabetes Metab Res Rev 38(3):e3512. 10.1002/dmrr.3512
98. LOFTUS SK, CANNONS JL (2005) Acinar cell apoptosis in Serpini2-deficient mice models pancreatic insufficiency[J]. PLoS Genet 1(3):e38. 10.1371/journal.pgen.0010038
99. HAMADA S, SHIMOSEGAWA T, TAGUCHI K et al (2018) Simultaneous K-ras activation and Keap1 deletion cause atrophy of pancreatic parenchyma[J]. Am J Physiol Gastrointest Liver Physiol 314(1):G65–G74. 10.1152/ajpgi.00228.2017
100. TANIGAWA K, NAKAMURA S, KAWAGUCHI M et al (1997) Effect of aging on B-cell function and replication in rat pancreas after 90% pancreatectomy[J]. Pancreas 15(1):53–59. 10.1097/00006676-199707000-00008

Figures

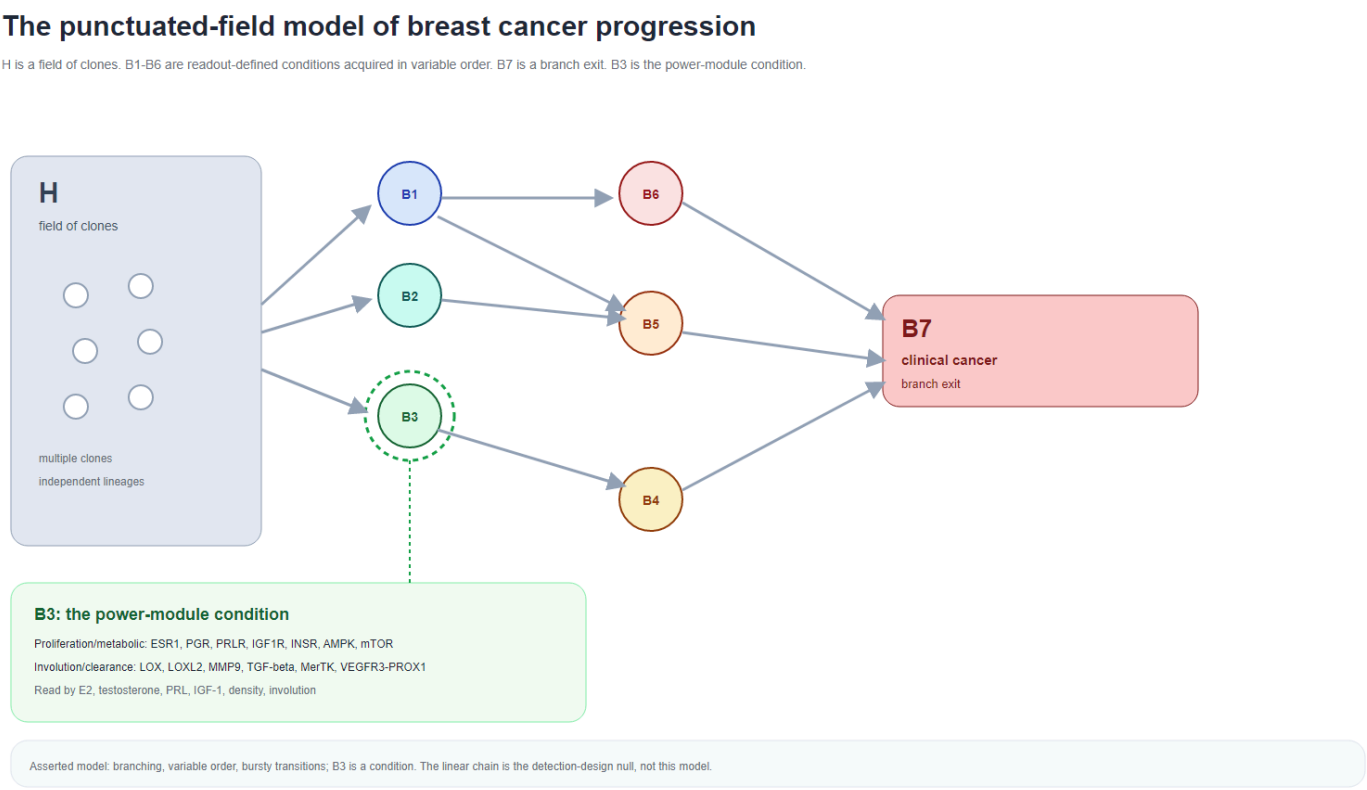


Figure 1

The nine-state linear chain of breast cancer. H, B1-B7 and death (D). The two B3 axes are the readout of B3; the six product groups are the readout of B4.

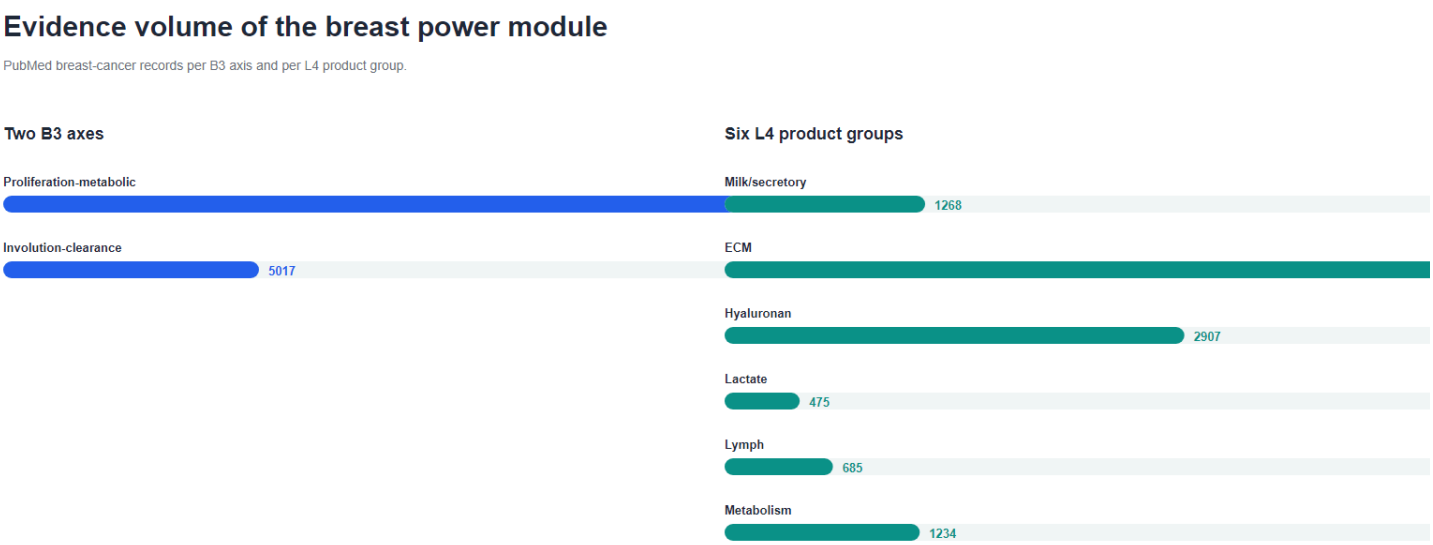


Figure 2

Evidence volume by axis and product group. PubMed breast-cancer records for the two B3 axes and the six L4 product groups.

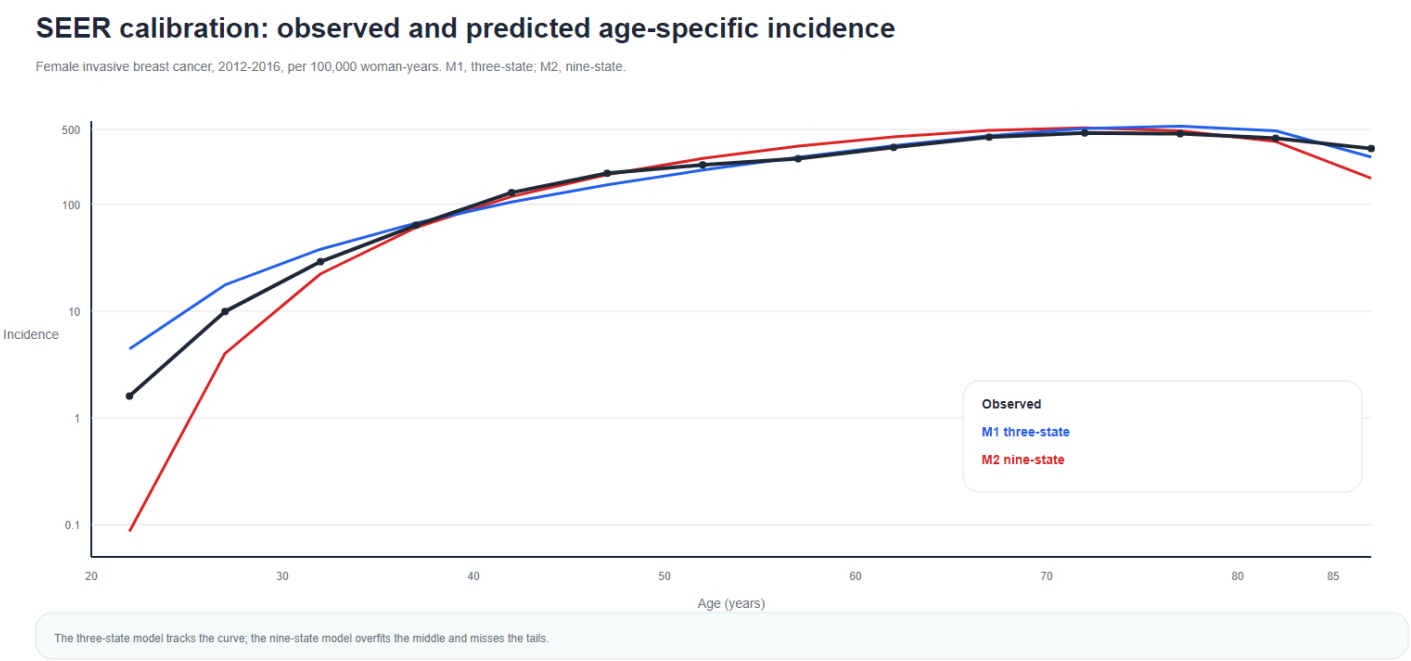


Figure 3

Calibration against SEER age-specific incidence. Observed and predicted age-specific incidence. The three-state model tracks the aggregate curve; the nine-state model resolves the intermediate states.

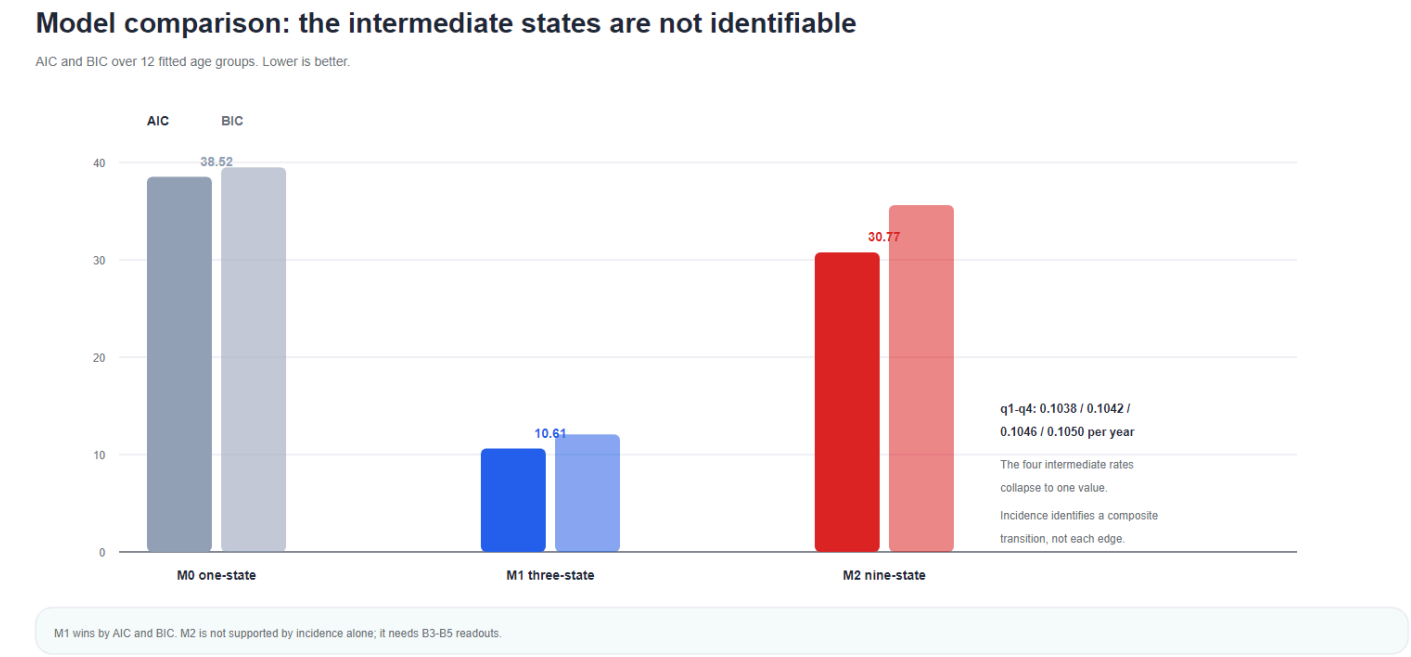


Figure 4

Model comparison and transition intensities. AIC and BIC for the one-state, three-state and nine-state models. The four intermediate rates collapse to one value.

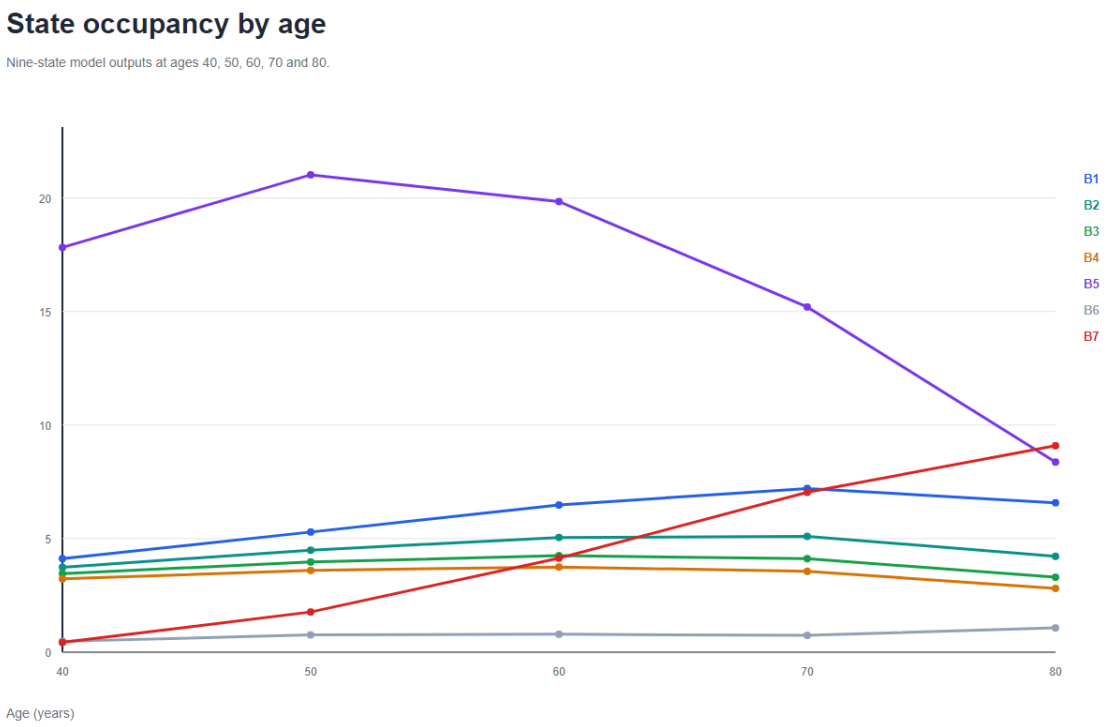


Figure 5

State occupancy by age. B1-B7 occupancy at ages 40, 50, 60, 70 and 80. B7 rises from 0.43% at 40 to 9.09% at 80.

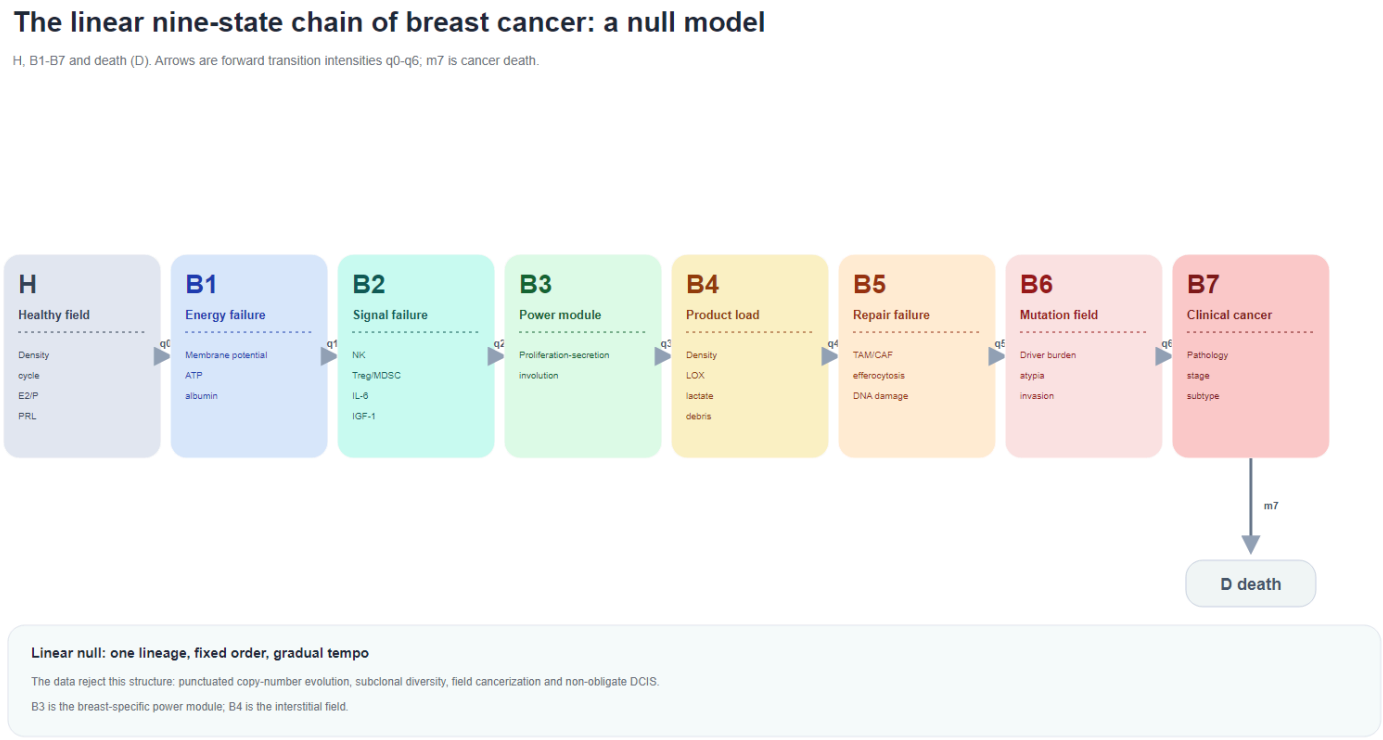


Figure 6

The detection-design model. The three-state model collapses B1-B4 into one composite state and B5-B7 into one composite state. The nine-state model retains the intermediate states.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablesS1S7.docx