Multimodal Gene Expression Deep Learning for Predicting Sentinel Lymph Node Macro-metastasis in Early Breast Cancer: Development and Validation in the SCAN-B Cohort

Supplementary materials

Table S1: Comparison of clinical characteristics between the development and independent test sets and between patients with negative and positive SLN within each set

Variables	Development set					Independent test set						
	All (n=4,625)	**Negative (n=3,880)**	**Positive (n=745)**	**P value**	**Effect size**	**All (n=2,211)**	**Negative (n=1,851)**	**Positive (n=360)**	**P value**	**Effect size**	**P value**	**Effect size**
Age, years, mean(SD)	65.0 (12.4)	65.1 (12.4)	64.3 (12.3)	0.086	-0.068^d	65.1 (12.6)	65.5 (12.4)	63.2 (13.4)	0.003	-0.182^d	0.733	-0.009^d
Menopausal status, No. (%)												
Postmenopausal	3,464 (79)	2,929 (80)	535 (77)	0.076	0.027^V	1,677 (80)	1,427 (82)	250 (74)	< 0.001	0.075^V	0.428	0.010^V
Premenopausal	895 (21)	735 (20)	160 (23)			411 (20)	321 (18)	90 (26)				
Missing	266	216	50			123	103	20				
Mode of detection, No. (%)												
Mammographic	2,500 (54)	2,152 (56)	348 (47)	< 0.001	0.065^V	1,196 (54)	1,027 (56)	169 (47)	0.003	0.064^V	0.882	0.002^V
Symptomatic	2,097 (46)	1,704 (44)	393 (53)			1011 (46)	820 (44)	191 (53)				
Missing	28	24	4			4	4	0				
Histological type, No. (%)												
ILC	631 (14)	505 (13)	126 (17)	< 0.001	0.039^V	316 (14)	252 (14)	64 (18)	0.029	0.040^V	0.241	0.014^V
NST	3,530 (76)	2,964 (76)	566 (76)			1,701 (77)	1,427 (77)	274 (76)				
Others	460 (10)	407 (11)	53 (7)			193 (9)	171 (9)	22 (6)				
Missing	4	4	0			1	1	0				
Histological grade, No. (%)												
I	842 (18)	742 (19)	100 (13)	< 0.001	0.039^V	389 (18)	347 (19)	42 (12)	0.006	0.048^V	0.672	0.008^V
II	2,423 (53)	2,003 (52)	420 (57)			1,145 (52)	944 (51)	201 (57)				
III	1,344 (29)	1,122 (29)	222 (30)			662 (30)	550 (30)	112 (32)				
Missing	16	13	3			15	10	5				
ER, %, mean(SD)	85.7 (30.2)	85.3 (30.7)	87.5 (27.8)	0.060	0.075^d	85.1 (31.8)	84.7 (32.4)	87.4 (28.0)	0.100	0.086^d	0.505	0.018^d
Missing	620	549	71			620	549	71				
PgR, %, mean(SD)	58.2 (39.1)	58.0 (39.3)	59.2 (37.7)	0.442	0.032^d	55.8 (40.0)	55.3 (40.3)	58.7 (38.1)	0.123	0.086^d	0.026	0.060^d
Missing	654	575	79			654	575	79				
HER2 status, No. (%)												
Negative	4,026 (90)	3,359 (90)	667 (90)	0.808	0.004^V	1,962 (89)	1,643 (89)	319 (89)	0.984	0.000^V	0.631	0.006^V
Positive	459 (10)	385 (10)	74 (10)			233 (11)	195 (11)	38 (11)				
Missing	140	136	4			16	13	3				
Ki67, %, mean(SD)	27.3 (19.3)	27.4 (19.6)	26.5 (17.5)	0.277	-0.048^d	28.4 (18.9)	28.6 (19.2)	27.6 (16.8)	0.315	-0.053^d	0.024	-0.062^d
Missing	1,282	1,072	210			1,282	1,072	210				
St Gallen surrogate molecular subtype, No. (%)												
HER2+	459 (13)	385 (13)	74 (12)	0.026	0.029^V	233 (11)	195 (11)	38 (11)	0.079	0.032^V	0.115	0.018^V
Luminal A	2,001 (55)	1,679 (55)	322 (53)			1,174 (54)	989 (54)	185 (53)				
Luminal B	932 (25)	757 (25)	175 (29)			588 (27)	478 (26)	110 (31)				
TNBC	275 (7)	244 (8)	31 (5)			177 (8)	158 (9)	19 (5)				
Missing	958	815	143			39	31	8				
Tumor size, mm, mean(SD)	17.8 (8.2)	17.1 (7.7)	21.8 (9.4)	< 0.001	0.573^d	17.6 (8.1)	16.8 (7.7)	21.5 (9.0)	< 0.001	0.578^d	0.286	0.027^d
Multifocality of invasive foci, No. (%)												
No	3695 (80)	3,167 (82)	528 (71)	< 0.001	0.098^V	1,751 (79)	1,528 (83)	223 (62)	< 0.001	0.189^V	0.340	0.012^V
Yes	907 (20)	695 (18)	212 (29)			457 (21)	320 (17)	137 (38)				
Missing	23	18	5			3	3	0				
LNM status, No. (%)												
Negative	3,488 (75)	3,488 (90)	0 (0)	< 0.001	0.767^V	1,635 (74)	1,635 (88)	0 (0)	< 0.001	0.743^V	0.190	0.016^V
Positive	1,137 (25)	392 (10)	745 (100)			576 (26)	216 (12)	360 (100)				
5-year distant recurrence status, No. (%)												
No relapse	2,077 (94)	1,736 (94)	341 (90)	0.005	0.060^V	1,505 (97)	1,252 (97)	253 (94)	0.019	0.059^V	< 0.001	0.068^V
Relapse	140 (6)	104 (6)	36 (10)			51 (3)	36 (3)	15 (6)				
Missing	2,408	2,040	368			655	563	92				

Comparisons were made between positive *vs.* negative patients within the development set and the independent test set, respectively. Additionally, the overall differences between the development and the independent test set were compared. P values and effect sizes were calculated. The significance level was set at $P = 0.05$, and a nontrivial effect size for continuous variables was defined as Cohen's $|d| \geq 0.50$ [1], and for categorical variables was defined as Cramer's $|V| \geq 0.30$, ≥ 0.21 and ≥ 0.17 for 1, 2, and 3 degrees of freedom, respectively [2].

Table S2: Patient characteristics by St Gallen surrogate subtypes in the development and test cohorts

Variables	Development set					Test set				
	All n=4,625	Luminal A n=2,001	Luminal B n=932	HER2+ n=459	TNBC n=275	All n=2,211	Luminal A n=1,174	Luminal B n=588	HER2+ n=233	TNBC n=177
Age, years, mean (SD)	65.0 (12.4)	65.5 (11.6)	65.4 (13.2)	62.7 (12.9)	63.0 (13.8)	65.1 (12.6)	65.8 (11.8)	64.5 (13.3)	63.0 (13.6)	66.0 (14.0)
Menopausal status, No. (%)										
Premenopausal	895 (21)	349 (19)	193 (22)	110 (26)	59 (23)	411 (20)	200 (18)	123 (22)	51 (23)	27 (16)
Postmenopausal	3,464 (79)	1,526 (81)	684 (78)	320 (74)	196 (77)	1,677 (80)	909 (82)	431 (78)	170 (77)	138 (84)
Missing	266	126	55	29	20	123	65	34	12	12
Mode of detection, No. (%)										
Symptomatic	2,097 (46)	792 (40)	479 (52)	239 (52)	169 (62)	1,011 (46)	460 (39)	304 (52)	117 (50)	110 (62)
Mammographic	2,500 (54)	1,203 (60)	451 (48)	218 (48)	105 (38)	1,196 (54)	712 (61)	284 (48)	115 (50)	66 (38)
Missing	28	6	2	2	1	4	2	0	1	1
Histological type, No. (%)										
NST	3,530 (76)	1,402 (70)	777 (83)	410 (89)	225 (82)	1,701 (77)	826 (70)	482 (82)	215 (92)	146 (82)
ILC	631 (14)	364 (18)	89 (10)	22 (5)	6 (2)	316 (14)	238 (20)	62 (11)	9 (4)	4 (2)
Others	460 (10)	235 (12)	66 (7)	27 (6)	44 (16)	193 (9)	110 (9)	43 (7)	9 (4)	27 (15)
Missing	4	0	0	0	0	1	0	1	0	0
Histological grade, No. (%)										
I	842 (18)	703 (35)	0 (0)	7 (2)	5 (2)	389 (18)	376 (32)	0 (0)	6 (3)	3 (2)
II	2,423 (53)	1,298 (65)	306 (33)	148 (32)	39 (14)	1,145 (52)	798 (68)	233 (40)	77 (33)	29 (16)
III	1,344 (29)	0 (0)	626 (67)	302 (66)	231 (84)	662 (30)	0 (0)	355 (60)	148 (64)	144 (82)
Missing	16	0	0	2	0	15	0	0	2	1
ER, %, mean (SD)	85.7 (30.2)	96.7 (7.9)	94.0 (14.0)	65.5 (41.6)	0.4 (1.3)	85.1 (31.8)	97.3 (7.6)	95.2 (11.3)	63.6 (42.4)	0.3 (1.1)
Missing	620	0	0	74	0	9	0	0	1	0
PgR, %, mean (SD)	58.2 (39.1)	69.1 (34.6)	59.1 (36.5)	33.0 (37.3)	0.2 (0.9)	55.8 (40.0)	67.9 (35.1)	58.7 (37.7)	29.6 (36.0)	0.1 (0.9)
Missing	654	0	0	86	0	14	0	0	1	0
HER2 status, No. (%)										
Negative	4,026 (90)	2,001 (100)	932 (100)	0 (0)	275 (100)	1,962 (89)	1,174 (100)	588 (100)	0 (0)	177 (100)
Positive	459 (10)	0 (0)	0 (0)	459 (100)	0 (0)	233 (11)	0 (0)	0 (0)	233 (100)	0 (0)
Missing	140	0	0	0	0	16	0	0	0	0
Ki67, %, mean (SD)	27.3 (19.3)	15.8 (7.1)	38.5 (13.5)	39.2 (18.7)	60.3 (25.4)	28.4 (18.9)	16.3 (7.5)	39.0 (13.6)	40.1 (17.0)	56.5 (24.9)
Missing	1,282	121	118	147	37	4	0	0	1	0
Tumor size, mm, mean (SD)	17.8 (8.2)	16.5 (7.9)	20.5 (8.6)	18.4 (7.9)	19.0 (7.7)	17.6 (8.1)	16.7 (8.1)	19.4 (7.8)	17.7 (8.2)	18.6 (8.2)
Multifocality of invasive foci, No. (%)										
No	3,695 (80)	1,567 (78)	742 (80)	358 (78)	245 (89)	1,751 (79)	945 (81)	437 (74)	190 (82)	151 (85)
Yes	907 (20)	433 (22)	189 (20)	99 (22)	30 (11)	457 (21)	227 (19)	150 (26)	43 (18)	26 (15)
Missing	23	1	1	2	0	3	2	1	0	0
LNM status, No. (%)										
Negative	3,488 (75)	1,527 (76)	653 (70)	345 (75)	231 (84)	1,635 (74)	884 (75)	408 (69)	170 (73)	146 (82)
Positive	1,137 (25)	474 (24)	279 (30)	114 (25)	44 (16)	576 (26)	290 (25)	180 (31)	63 (27)	31 (18)
SLNM status, No. (%)										
Negative	3,880 (84)	1,679 (84)	757 (81)	385 (84)	244 (89)	1,851 (84)	989 (84)	478 (81)	195 (84)	158 (89)
Positive	745 (16)	322 (16)	175 (19)	74 (16)	31 (11)	360 (16)	185 (16)	110 (19)	38 (16)	19 (11)
Adjuvant therapy, No. (%)										
Endocrine therapy alone	2,477 (60)	1,467 (84)	348 (38)	47 (11)	2 (1)	1,105 (57)	866 (88)	202 (35)	19 (9)	1 (1)
Chemotherapy alone	297 (7)	2 (0)	10 (1)	3 (1)	206 (98)	150 (8)	0 (0)	8 (1)	4 (2)	127 (98)
Endocrine+chemotherapy	969 (23)	268 (15)	544 (60)	6 (1)	3 (1)	501 (26)	121 (12)	366 (63)	6 (3)	2 (2)
HER2-directed therapy	395 (10)	2 (0)	4 (0)	387 (87)	0 (0)	195 (10)	0 (0)	1 (0)	194 (87)	0 (0)
Missing	487	262	26	16	64	260	187	11	10	47
Breast surgery, No. (%)										
Breast conserving surgery	3,149 (68)	1,478 (74)	611 (66)	281 (61)	184 (67)	1,570 (71)	881 (75)	400 (68)	155 (67)	109 (62)
Mastectomy	1,476 (32)	523 (26)	321 (34)	178 (39)	91 (33)	641 (29)	293 (25)	188 (32)	78 (33)	68 (38)
5-year DR status, No. (%)										
No relapse	2,077 (94)	747 (98)	344 (87)	219 (91)	93 (82)	1,505 (97)	807 (98)	410 (96)	148 (95)	108 (91)
Relapse	140 (6)	17 (2)	53 (13)	22 (9)	20 (18)	51 (3)	16 (2)	16 (4)	8 (5)	11 (9)
Missing	2,408	1,237	535	218	162	655	351	162	77	58
DRFi if no event, years, mean (SD)	6.8 (2.5)	5.9 (2.2)	6.2 (2.4)	6.9 (2.6)	6.5 (2.6)	4.9 (1.1)	4.9 (1.0)	4.9 (1.1)	4.9 (1.3)	4.8 (1.4)

St Gallen surrogate subtypes were defined according to [3]. Luminal A, luminal A-like; Luminal B, luminal B-like; HER2+, HER2-enriched; TNBC, triple negative breast cancer; NST, no special type; ILC, invasive lobular carcinoma; LNM, lymph node metastasis; SLNM, sentinel lymph node macro-metastasis; BCS, Breast conserving surgery; DRFi, distant recurrence-free interval.

Table S3: Clinical characteristics of patient groups stratified by clinical molecular subtypes and ASCO guidelines in the recalibration and hold-out splits of the independent test set

Variables	Recalibration set					Hold-out test set				
	All n=761	ER+HER2- (ASCO-omit) n=196	ER+HER2- (ASCO-SLNB) n=420	HER2+ n=84	TNBC n=45	All n=1,450	ER+HER2- (ASCO-omit) n=355	ER+HER2- (ASCO-SLNB) n=802	HER2+ n=149	TNBC n=132
Age, years, mean (SD)	63.9 (13.0)	69.1 (8.4)	62.3 (13.9)	61.0 (13.9)	63.1 (13.7)	65.7 (12.4)	69.2 (8.2)	64.3 (13.1)	64.1 (13.3)	67.0 (14.0)
Menopausal status, No. (%)										
Postmenopausal	552 (77)	196 (100)	253 (67)	60 (73)	33 (75)	1,125 (82)	355 (100)	546 (73)	110 (79)	105 (87)
Premenopausal	163 (23)	0 (0)	124 (33)	22 (27)	11 (25)	248 (18)	0 (0)	200 (27)	29 (21)	16 (13)
Missing	46	0	43	2	1	77	0	56	10	11
Mode of detection, No. (%)										
Symptomatic	333 (44)	60 (31)	194 (46)	45 (54)	25 (57)	678 (47)	113 (32)	400 (50)	72 (48)	85 (64)
Mammographic	426 (56)	136 (69)	226 (54)	38 (46)	19 (43)	770 (53)	241 (68)	401 (50)	77 (52)	47 (36)
Missing	2	0	0	1	1	2	1	1	0	0
Histological type, No. (%)										
NST	597 (79)	196 (100)	270 (64)	78 (93)	40 (89)	1,104 (76)	355 (100)	496 (62)	137 (92)	106 (80)
ILC	103 (14)	0 (0)	96 (23)	5 (6)	0 (0)	213 (15)	0 (0)	205 (26)	4 (3)	4 (3)
Others	60 (8)	0 (0)	53 (13)	1 (1)	5 (11)	133 (9)	0 (0)	101 (13)	8 (5)	22 (17)
Missing	1	0	1	0	0	0	0	0	0	0
Histological grade, No. (%)										
I	134 (18)	60 (31)	67 (16)	2 (2)	2 (4)	255 (18)	124 (35)	125 (16)	4 (3)	1 (1)
II	414 (55)	136 (69)	237 (57)	27 (32)	8 (18)	731 (51)	231 (65)	427 (54)	50 (34)	21 (16)
III	208 (28)	0 (0)	112 (27)	55 (65)	35 (78)	454 (32)	0 (0)	243 (31)	93 (63)	109 (83)
Missing	5	0	4	0	0	10	0	7	2	1
ER, %, mean (SD)	86.7 (29.4)	97.8 (7.5)	95.7 (9.5)	63.1 (42.2)	0.2 (0.9)	84.3 (32.9)	98.3 (5.2)	96.0 (10.2)	63.9 (42.6)	0.3 (1.2)
Missing	5	0	0	0	0	4	0	0	1	0
PgR, %, mean (SD)	58.4 (39.4)	64.4 (36.3)	67.3 (35.3)	32.0 (37.7)	0.0 (0.1)	54.5 (40.2)	65.5 (34.9)	63.6 (37.3)	28.3 (35.1)	0.2 (1.0)
Missing	9	0	0	0	0	5	0	0	1	0
HER2 status, No. (%)										
Negative	669 (89)	196 (100)	420 (100)	0 (0)	45 (100)	1,293 (90)	355 (100)	802 (100)	0 (0)	132 (100)
Positive	84 (11)	0 (0)	0 (0)	84 (100)	0 (0)	149 (10)	0 (0)	0 (0)	149 (100)	0 (0)
Missing	8	0	0	0	0	8	0	0	0	0
Ki67, %, mean (SD)	28.7 (19.1)	19.9 (11.2)	26.6 (16.7)	42.0 (17.1)	60.1 (23.8)	28.3 (18.7)	19.0 (10.1)	25.7 (15.3)	39.0 (16.8)	55.2 (25.2)
Missing	1	0	0	0	0	3	0	0	1	0
St Gallen surrogate molecular subtype, No. (%)										
Luminal A	413 (56)	168 (86)	245 (59)	0 (0)	0 (0)	761 (53)	308 (87)	453 (57)	0 (0)	0 (0)
Luminal B	199 (27)	28 (14)	171 (41)	0 (0)	0 (0)	389 (27)	47 (13)	342 (43)	0 (0)	0 (0)
HER2+	84 (11)	0 (0)	0 (0)	84 (100)	0 (0)	149 (10)	0 (0)	0 (0)	149 (100)	0 (0)
TNBC	45 (6)	0 (0)	0 (0)	0 (0)	45 (100)	132 (9)	0 (0)	0 (0)	0 (0)	132 (100)
Missing	20	0	4	0	0	19	0	7	0	0
Tumor size, mm, mean (SD)	17.1 (8.1)	13.0 (3.4)	18.6 (8.7)	17.7 (8.8)	21.2 (9.1)	17.9 (8.1)	13.1 (3.7)	20.1 (8.7)	17.7 (7.8)	17.8 (7.8)
Multifocality of invasive foci, No. (%)										
No	605 (80)	196 (100)	285 (68)	72 (86)	42 (93)	1,146 (79)	355 (100)	554 (69)	118 (79)	109 (83)
Yes	155 (20)	0 (0)	134 (32)	12 (14)	3 (7)	302 (21)	0 (0)	246 (31)	31 (21)	23 (17)
Missing	1	0	1	0	0	2	0	2	0	0
LNM status, No. (%)										
Negative	569 (75)	161 (82)	296 (70)	61 (73)	37 (82)	1,066 (74)	286 (81)	555 (69)	109 (73)	109 (83)
Positive	192 (25)	35 (18)	124 (30)	23 (27)	8 (18)	384 (26)	69 (19)	247 (31)	40 (27)	23 (17)
SLNM status, No. (%)										
Negative	642 (84)	180 (92)	337 (80)	71 (85)	40 (89)	1,209 (83)	315 (89)	641 (80)	124 (83)	118 (89)
Positive	119 (16)	16 (8)	83 (20)	13 (15)	5 (11)	241 (17)	40 (11)	161 (20)	25 (17)	14 (11)
Adjuvant therapy, No. (%)										
Endocrine therapy alone	383 (57)	139 (87)	232 (60)	5 (6)	0 (0)	722 (56)	250 (89)	454 (60)	14 (10)	1 (1)
Chemotherapy alone	44 (7)	0 (0)	2 (1)	2 (3)	34 (100)	106 (8)	1 (0)	5 (1)	2 (1)	93 (97)
Endocrine therapy + chemotherapy	175 (26)	20 (13)	150 (39)	4 (5)	0 (0)	326 (25)	29 (10)	291 (39)	2 (1)	2 (2)
HER2-directed therapy	67 (10)	0 (0)	0 (0)	67 (86)	0 (0)	128 (10)	0 (0)	1 (0)	127 (88)	0 (0)
Missing	92	37	36	6	11	168	75	51	4	36
Breast surgery, No. (%)										
Breast conserving surgery	532 (70)	161 (82)	280 (67)	55 (65)	27 (60)	1,038 (72)	314 (88)	535 (67)	100 (67)	82 (62)
Mastectomy	229 (30)	35 (18)	140 (33)	29 (35)	18 (40)	412 (28)	41 (12)	267 (33)	49 (33)	50 (38)
5-year DR status, No. (%)										
No relapse	583 (97)	146 (99)	332 (98)	62 (95)	30 (83)	922 (97)	224 (100)	523 (96)	86 (95)	78 (94)
Relapse	18 (3)	1 (1)	8 (2)	3 (5)	6 (17)	33 (3)	1 (0)	22 (4)	5 (5)	5 (6)
Missing	160	49	80	19	9	495	130	257	58	49
DRFi if no event, years, mean (SD)	5.1 (1.0)	4.9 (1.0)	5.1 (0.9)	5.2 (1.1)	5.1 (1.0)	4.8 (1.2)	4.8 (1.1)	4.9 (1.1)	4.7 (1.3)	4.6 (1.5)

Clinical tumor subtypes of ER+HER2-, HER2+ and TNBC were labeled according to [3]. ASCO-omit (patients recommended to omit SLNB) and ASCO-SLNB (patients recommended to undergo SLNB) were distinguished according to the 2025 ASCO guidelines. SLNM was defined as the presence of macrometastasis ($\geq$ 2 mm) on SLNB. LNM was defined as the presence of $\geq$ 0.2 mm metastasis, including both micro- and macrometastases on SLNB and/or ALND. Luminal A, luminal A-like; Luminal B, luminal B-like; HER2+, HER2-enriched; TNBC, triple negative breast cancer; NST, no special type; ILC, invasive lobular carcinoma; LNM, lymph node metastasis; SLNM, sentinel lymph node macro-metastasis; DRFi, distant recurrence-free interval.

6 Supplementary

6.1 Published cancer pathways selected for study

Table S4: Ten canonical cancer pathways curated by [4]

Pathway name	Description
Cell Cycle	Regulation of mitotic cell cycle progression
HIPPO	Regulation of cell proliferation and differentiation
MYC	Regulation of cell growth, proliferation and apoptosis
NOTCH	Regulation of cell growth and apoptosis
NRF2	Regulation of oxidative stress response
PI3K	Regulation of cell growth
RTK/RAS	Regulation of cell proliferation and survival
TGFβ	Regulation of cell proliferation and stem/progenitor phenotype
TP53	Regulation of cell proliferation, survival and apoptosis
WNT	Regulation of cell proliferation

Two curated gene lists of MutSig (Mutation Significance) and OncoKB (Oncology Knowledge Base) were included as background or benchmark genes.

Detailed lists of gene IDs is available at https://ars.els-cdn.com/content/image/1-s2.0-S0092867418303593-mmc2.xlsx

Table S5: Thirty established breast cancer pathways [5]

Reactome ID	Description
R-HSA-1640170	Cell Cycle
R-HSA-6802957	Oncogenic MAPK signaling
R-HSA-75893	TNF signaling
R-HSA-389948	PD-1 signaling
R-HSA-6806834	Signaling by MET
R-HSA-8853659	RET signaling
R-HSA-167044	Signalling to RAS
R-HSA-165159	mTOR signalling
R-HSA-4791275	Signaling by WNT in cancer
R-HSA-5358351	Signaling by Hedgehog
R-HSA-2219528	PI3K/AKT Signaling in Cancer
R-HSA-5674404	PTEN Loss of Function in Cancer
R-HSA-2644603	Signaling by NOTCH1 in Cancer
R-HSA-1980145	Signaling by NOTCH2
R-HSA-9012852	Signaling by NOTCH3
R-HSA-9013694	Signaling by NOTCH4
R-HSA-69610	p53-Independent DNA Damage Response
R-HSA-1234174	Cellular response to hypoxia
R-HSA-73893	DNA Damage Bypass
R-HSA-73942	DNA Damage Reversal
R-HSA-5685942	HDR through Homologous Recombination (HRR)
R-HSA-6796648	TP53 Regulates Transcription of DNA Repair Genes
R-HSA-1643713	Signaling by EGFR in Cancer
R-HSA-9018519	Estrogen-dependent gene expression
R-HSA-5358508	Mismatch Repair
R-HSA-109581	Apoptosis
R-HSA-170834	Signaling by TGF-beta Receptor Complex
R-HSA-354192	Integrin signaling
R-HSA-1227986	Signaling by ERBB2
R-HSA-1236394	Signaling by ERBB4

Reactome pathways gene set is available at https://download.reactome.org/95/ReactomePathways.gmt.zip

6.2 A composite score for model selection

A composite score combining the ROC AUC and PR AUC was used for model selection, formulated as $Score = (\frac{ROC\ AUC}{0.5} + \frac{PR\ AUC}{positive fraction})/2$. The baseline for ROC AUC is always 0.5 (random guessing), regardless of class distribution. The baseline for PR AUC corresponds to the positive class ratio, which changes based on dataset imbalance. Thus, the score is an average of the normalized ROC AUC and PR AUC relative to their respective baselines.

6.3 RNA-sequencing derived mutation features added no predictive benefit for SLNM compared to GEX data alone

Variant calling was performed using the pipeline described in [5], with variant filters to reduce false-positive calls resulting from either sequencing or PCR artifacts, RNA editing, or germline variants. Gene-level mutation impact index (**GeneMutImpact**) ranging from 0 to 3 was extracted using functional annotations of each gene based on Vcfanno [6]: 0 for no mutation detected, 1 for low impact, 2 for moderate impact and 3 for high impact. For patient-level mutation burden (**MutBurden**), a single value for the total burden was calculated by summing the individual mutation counts. (**GeneMutImpact**) was integrated into GEX modeling by cross-attention layers. The predictive performance of the gene-level mutation index and patient-level mutation burden was presented in Supplementary Figure S1.

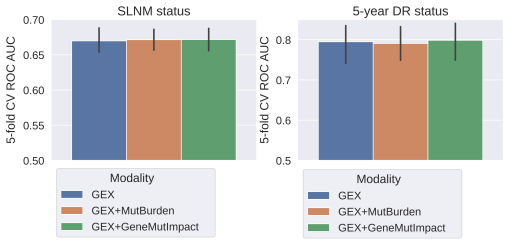
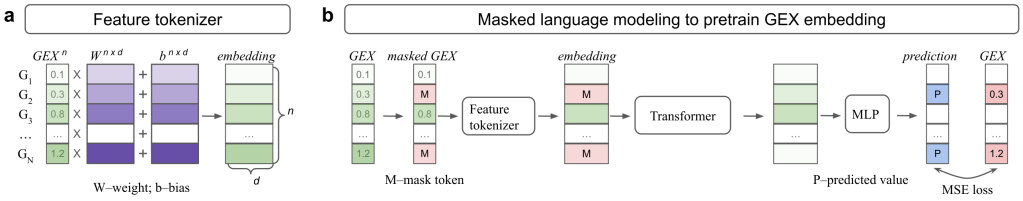


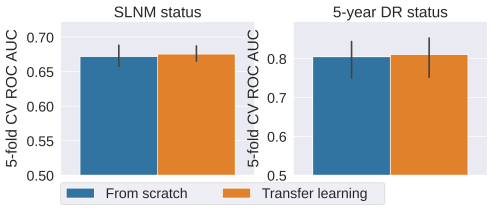
Fig. S1: 5-fold cross-validation for nodal metastasis prediction and recurrence prediction using GEX and mutation data. RNA-seq-derived mutation features did not improve SLNM prediction nor 5-year DR status prediction in the development. GEX, gene expression; MutBurden, total mutation burden at patient-level; GeneMutImpact, mutation impact index at gene-level; SLNM, sentinel lymph node macro-metastasis; DR, distant recurrence.

6.4 Pretraining on publicly available bulk RNA-seq data showed limited improvement for SLNM prediction in SCAN-B cohort

GEO expression data of 340k samples in raw counts were downloaded from ARSH4 [7]. 256k samples with $>$ 75% non-zero gene counts across the 18,536 common genes shared with the SCAN-B dataset were extracted. The gene expression data were then transformed to FPKM format and were preprocessed using the same procedures as the SCAN-B data regarding log transformation and sample-wise normalization. Next, the Transformer model was trained on the curated GEO data to predict expression values of the masked genes (as shown in Supplementary Figure S2a) for 100 epochs with a learning rate of 0.0001. Supervised finetuning was then performed in the SCAN-B cohort to predict SLNM and 5-year DR status. Results showed that transfer learning on public



(a) GEX embedding. **a**) The feature tokenizer, converting raw GEX values to meaningful embeddings, is essential for Transformer modeling. **b**) Mask language modeling by practicing a "fill-in-the-blanks" task is a self-supervised learning technique to pretrain the feature tokenizer as well as other Transformer modules. GEX, gene expression; MSE, mean squared error.



(b) Predictive performance for SLNM and 5-year DR status of GEX models pretrained on open-source data versus trained from scratch.

Fig. S2: Pretraining protocols and validation on SCAN-B data

transcriptomic data provided only a negligible benefit compared to training from scratch (Supplementary Figure S2b).

6.5 Supplementary results

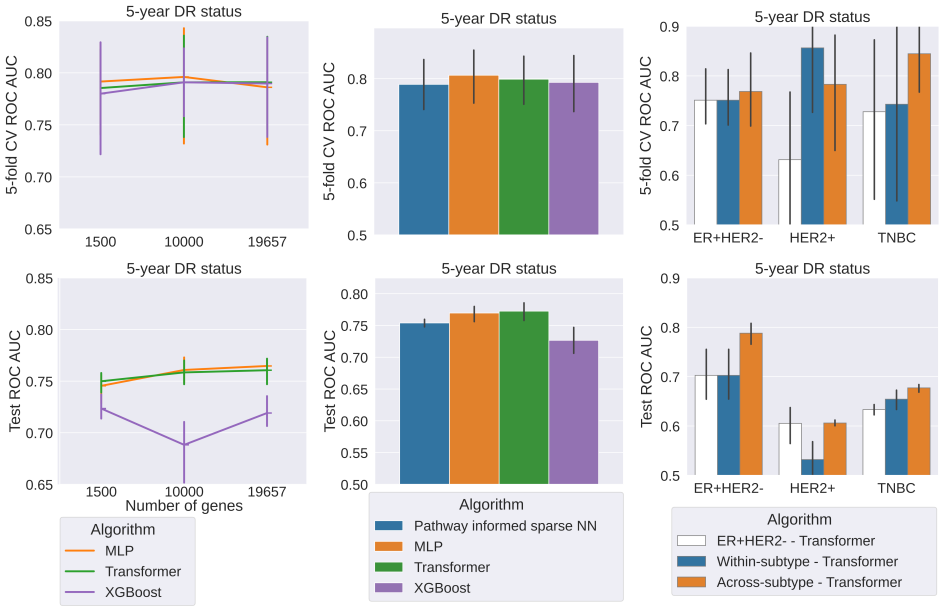


Fig. S3: 5-fold cross-validation (CV) and test performance for GEX-based distance recurrence prediction (DR). MLP and Transformer models demonstrated a higher capacity for handling a large number of gene inputs than XGBoost (the left column). MLP and Transformer exhibited enhanced ability for modeling complex correlations within GEX data (the middle column) in comparison to XGBoost and the sparse neural network(NN) incorporating pathway information. Across-subtype training improved learning for DR prediction, especially for the dominant, low-risk ER+HER2- group, by transferring knowledge from relapsed events in the high-risk groups (the right column). Compared to the test performance, the 5-fold CV showed larger error bars and a tendency towards overfitting due to the limited number of relapse events in each internal validation fold (an average of 28 cases). The error bars were calculated as the standard deviations across the 5-fold CV models.

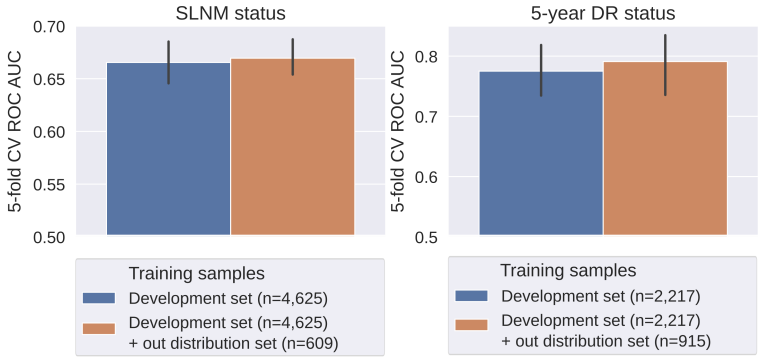
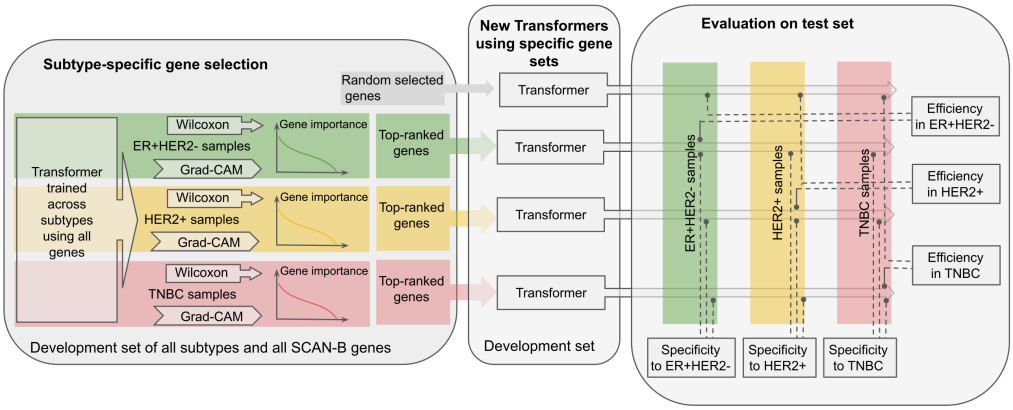
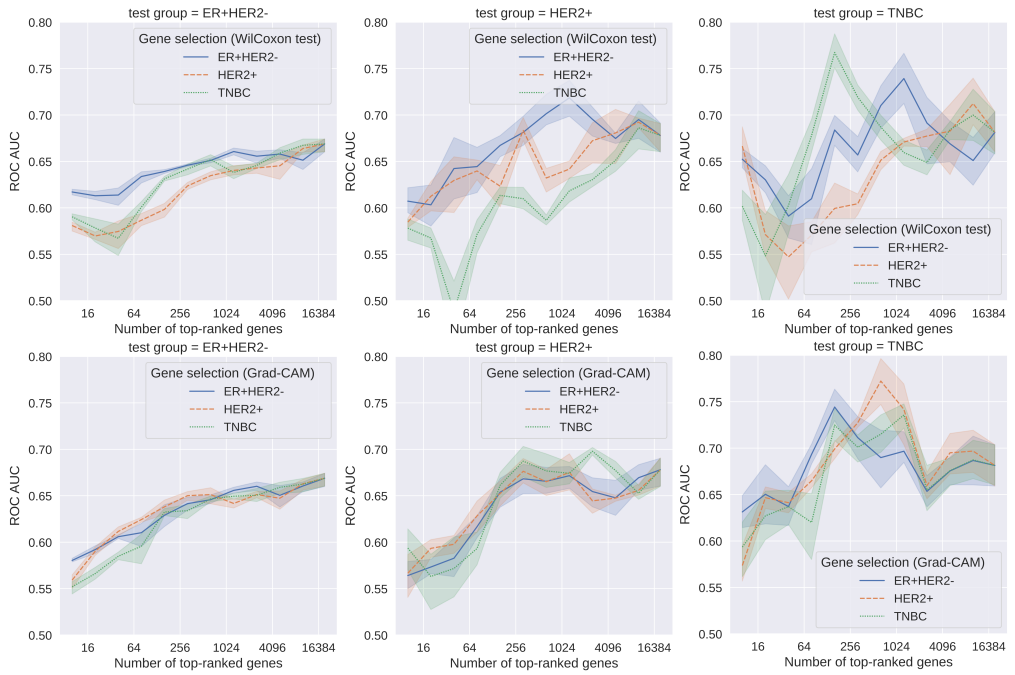


Fig. S4: Additional GEX data of out distribution patients slightly improved model performance for predicting SLNM and 5-year DR status in the development cohort. The out distribution set consisted of more advanced breast cancer patients with clinically positive nodes or with a tumor size >50 mm, who were previously excluded from the study cohort. The additional out-of-distribution data was only integrated into training and was not used for cross-validation (CV).



(a) First, all SCAN-B genes were sorted by gene importance, their association with sentinel lymph node macrometastasis quantified by either Wilcoxon test or Grad-CAM, within each subgroup. Specifically, Grad-CAM values were derived from 10 Transformer models repetitively trained on the entire development set using all SCAN-B genes. Next, varying numbers of top-ranked genes were selected in the development set based on different strategies (Wilcoxon test, Grad-CAM, or random sampling). New Transformer models were then trained separately using these gene sets on the development set. Finally, the Transformers were evaluated in the independent test set. Efficiency was assessed by comparing the predictive performance between Transformer models trained using strategically selected gene sets and those using randomly selected gene sets within each subtype. Specificity was assessed by comparing the predictive performance of Transformers trained using different gene sets within and outside subtypes. For example, the gene specificity to the TNBC subtype was measured by comparing the models trained using ER+HER2-, HER2+ specific gene sets (outside the target subtype) with those using the TNBC specific gene set (within the target subtype).



(b) High specificity was demonstrated when internal validation showed superior performance compared to the cross-validation. Overall, the Wilcoxon test-based gene selection exhibited higher specificity than the Grad-CAM-based approach. The top row shows results for selections based on the Wilcoxon rank-sum test, and the bottom row for Grad-CAM-based selection. Columns correspond to validation in each group: ER+HER2- (left), HER2+(middle) and TNBC (right).

Fig. S5: **a**) The workflow of gene selection and evaluation of the specificity and efficiency of the selected gene sets. **b**) Cross-validation within and outside subtypes to assess the specificity of gene selections.

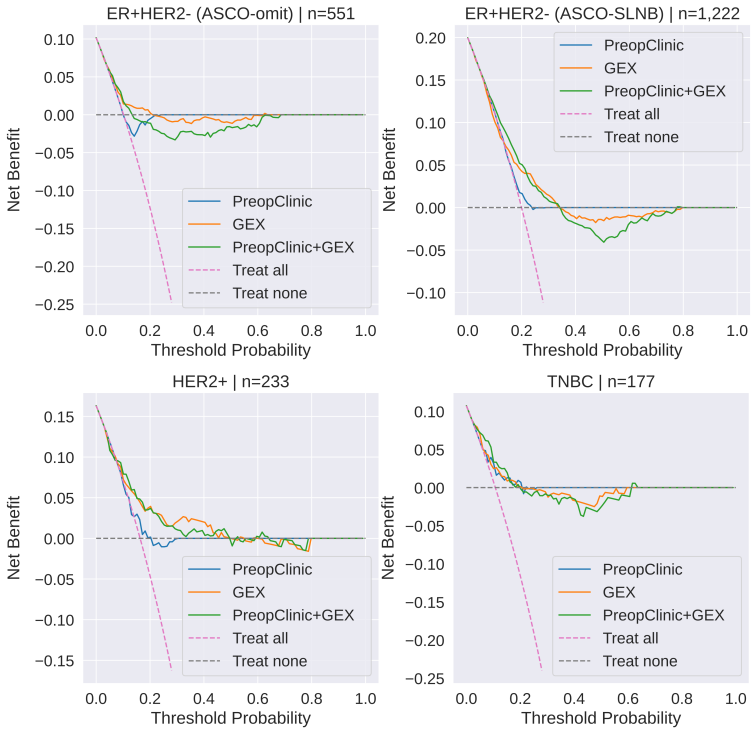


Fig. S6: Decision curves illustrating net benefit for predicting sentinel lymph node macro-metastasis by tumor subtypes in the independent test set using the developed models: PreopClinic, GEX and PreopClinic+GEX. Net benefit is a trade-off between the benefit of true positive predictions and the harm of the false positive predictions. A lower threshold probability implies a greater concern about the metastasis in axilla, while a higher threshold probability implies a greater concern about the action to be taken, such as axilla surgery. PreopClinic, preoperative clinicopathology; GEX, gene expression.

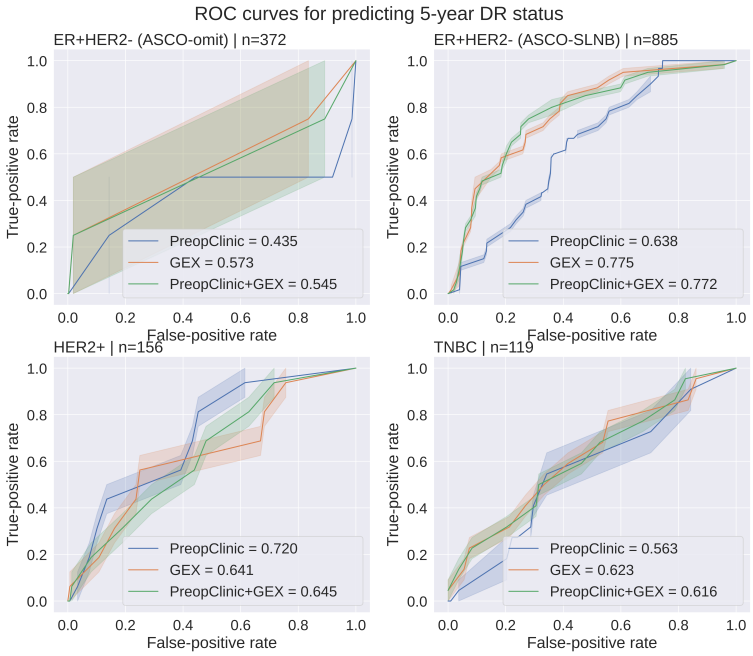


Fig. S7: Comparison of ROC curves for predicting 5-year DR in the independent test set by molecular subtypes. PreopClinic, preoperative clinicopathology; GEX, gene expression. SLNM, sentinel lymph node macro-metastasis; DR, distant recurrence.

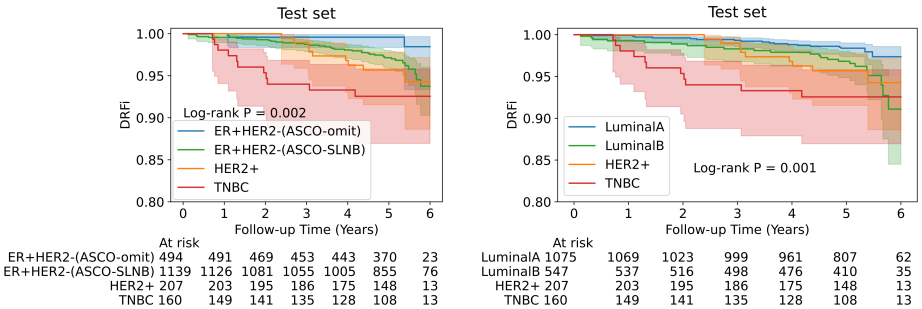
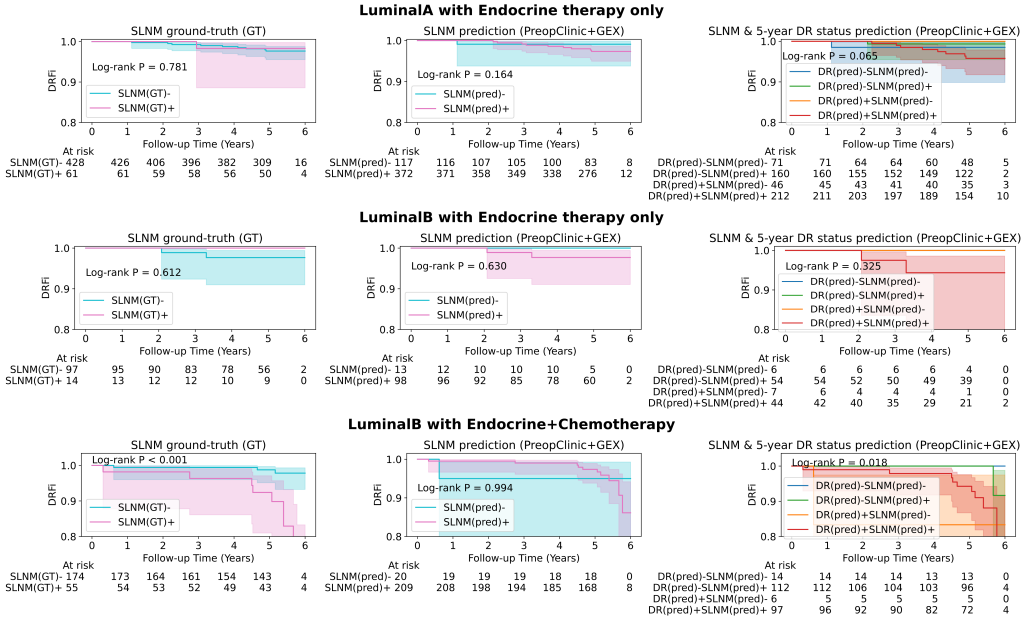


Fig. S8: Kaplan-Meier plots stratified by clinical subtypes with the endpoint of DRFi in the independent test set. Left: combination of clinical subtypes and ASCO guidelines; Right: St Gallen surrogate subtypes

(a) ER+HER2- patients divided by Luminal A- or B-like and corresponding therapy.



(b) ER+HER2- patients divided by ASCO recommendation and corresponding therapy.

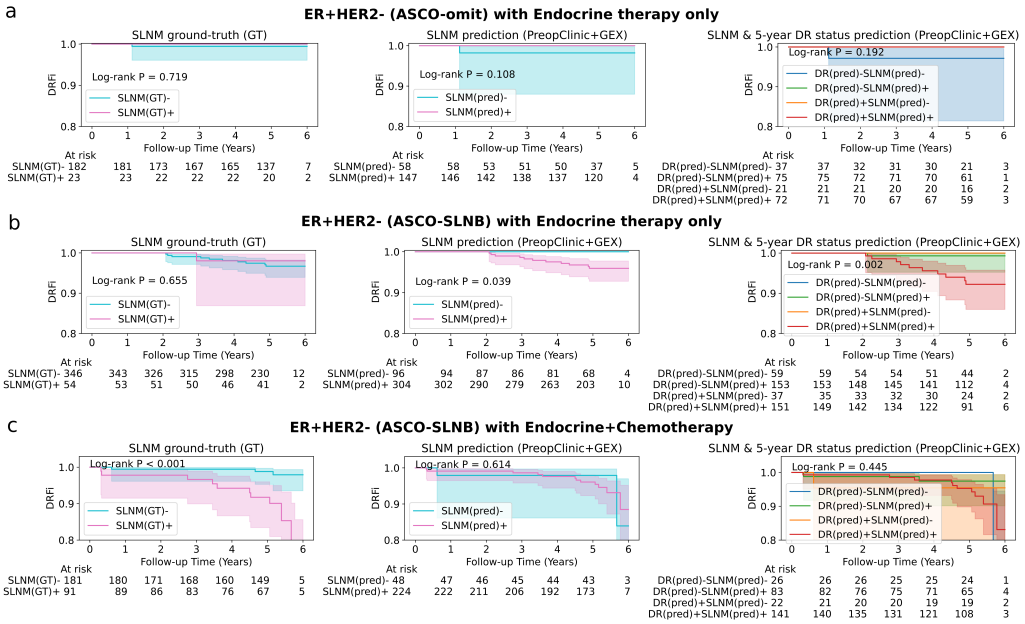


Fig. S9: Kaplan-Meier curves using DRFi as the endpoint, stratified by SLNM ground truth and GEX predictors.

(a) ER+HER2- (ASCO-SLNB) patients with endocrine therapy only

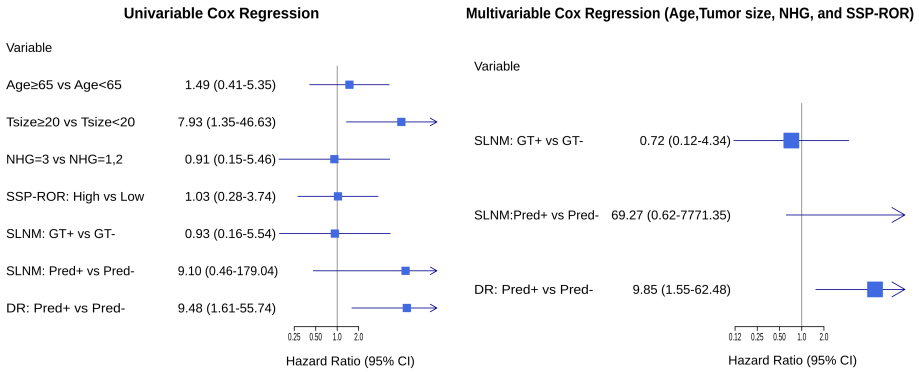


Fig. S10: Forest plots displaying hazard ratios of SLNM and DR predictions based on PreopClinic+GEX by univariable and multivariable (adjusted by age, tumor size, NHG and SSP-based ROR risk classification) Cox regression using the DRFi endpoint. SLNM, sentinel lymph node macro-metastasis; DR, distant recurrence-free; PreopClinic, preoperative clinicopathology; GEX, gene expression.

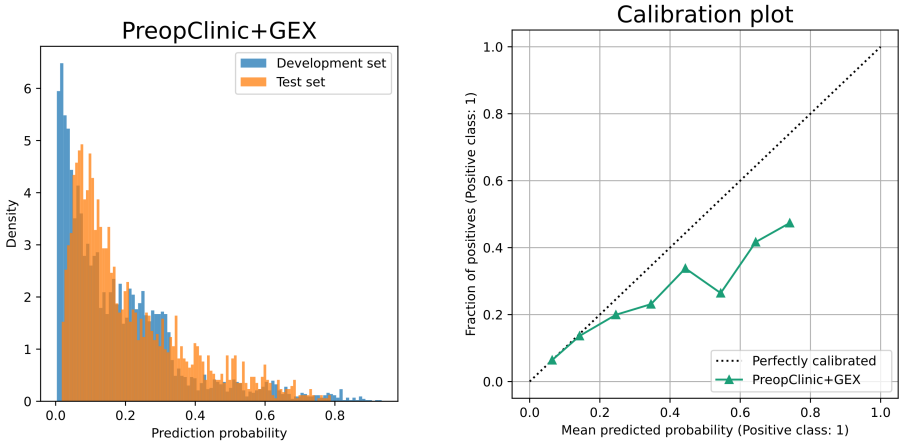


Fig. S11: Distribution plots and calibration curve of sentinel lymph node macro-metastasis (SLNM) prediction based on the combined model of PreopClinic+GEX. a) Distributions of the predicted SLNM probabilities in the development set and in the independent set. The dark orange depicts the overlap between the two distributions. b) Calibration curve of SLNM prediction. GEX data used all SCAN-B genes. GEX, gene expression analysis; PreopClinic, preoperative clinical data.

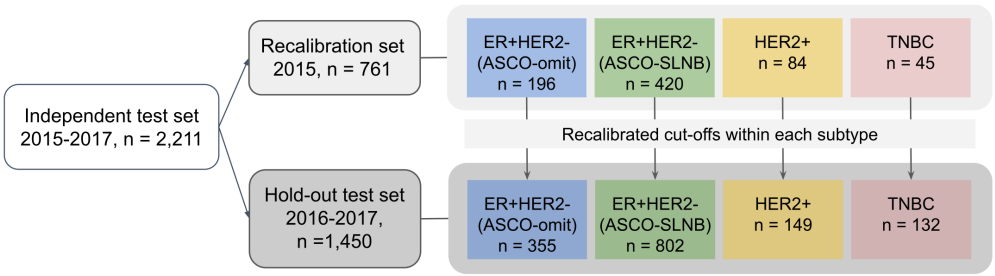


Fig. S12: Recalibration and hold-out splits of the independent test set for cut-off dependent analyses. There were 28 test samples identified as an unknown clinical subtype due to missing values. Patient characteristics of the recalibration set and hold-out test set are presented in Supplementary Table S3.

References

[1] Sawilowsky SS. New effect size rules of thumb. Journal of modern applied statistical methods. 2009;8:597–599.

[2] Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restorative dentistry & endodontics. 2017;42(2):152–155.

[3] Narbe U, Bendahl PO, Fernö M, Ingvar C, Dihge L, Rydén L. St Gallen 2019 guidelines understage the axilla in lobular breast cancer: a population-based study. British Journal of Surgery. 2021;108(12):1465–1473.

[4] Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic signaling pathways in the cancer genome atlas. Cell. 2018;173(2):321–337.

[5] Brueffer C, Gladchuk S, Winter C, Vallon-Christersson J, Hegardt C, Häkkinen J, et al. The mutational landscape of the SCAN-B real-world primary breast cancer transcriptome. EMBO molecular medicine. 2020;12(10):e12118.

[6] Pedersen BS, Layer RM, Quinlan AR. Vcfanno: fast, flexible annotation of genetic variants. Genome biology. 2016;17(1):118.

[7] Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. Nature communications. 2018;9(1):1366.

[]