



**Karolinska
Institutet**

Karolinska Institutet

<http://openarchive.ki.se>

This is a Peer Reviewed Accepted version of the following article, accepted for publication in Cancer Research.

2020-04-17

Assessment of breast cancer risk factors reveals subtype heterogeneity

Holm, Johanna; Eriksson, Louise; Ploner, Alexander; Eriksson, Mikael; Rantalainen, Mattias; Li, Jingmei; Hall, Per; Czene, Kamila

Cancer Res. 2017 Jul 1;77(13):3708-3717.

American Association for Cancer Research

<http://doi.org/10.1158/0008-5472.CAN-16-2574>

<http://hdl.handle.net/10616/47145>

If not otherwise stated by the Publisher's Terms and conditions, the manuscript is deposited under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Online-only supplementary tables and figures:

1. Supplementary Table 1. Title: Cross-validated Confusion matrix for true vs. predicted subtypes in the subset of data with PAM50 subtypes available.
Entries are percentual average cell counts across resamples.
2. Supplementary Table 2 Title: Case-only analysis with Luminal A as reference group: Genetic background risk.
3. Supplementary Table 3 Title: Case-only analysis with Luminal A as reference group: Reproductive risk factors.
4. Supplementary Table 4. Title: Case-only analysis with Luminal A as reference group: Bodyshape at age 18, Ever hormone replacement therapy (HRT) use, age at menarche, mammographic density (MD), benign breast disease (BBD).
5. Supplementary Table 5 Title: Risk of breast cancer overall and by subtype as defined by the St Gallen IHC proxy: Genetic background risk.
6. Supplementary Table 6 Title: Risk of breast cancer overall and by subtype as defined by the St Gallen IHC proxy: Reproductive risk factors.
7. Supplementary Table 7 Title: Risk for breast cancer overall and by subtype as defined by the St Gallen IHC proxy: Bodyshape at age 18, Ever hormone replacement therapy (HRT) use, age at menarche, mammographic density (MD), benign breast disease (BBD).

Supplementary table 1. Cross-validated confusion matrices for true vs. predicted subtypes in the subset of data with PAM50 subtypes available. Entries are percentual average cell counts across resamples.

Predicted subtypes	True subtypes			
By Random forest	Luminal A	Luminal B	HER2-overexpr.	Basal-like
Luminal A	47.4	10.3	2.8	0.0
Luminal B	4.7	10.7	2.5	0.5
HER2-overexpr.	1.4	2.6	6.5	1.3
Basal-like	0.0	0.0	1.3	7.9
By St Gallen				
Luminal A	31.9	4.7	0.9	0.0
Luminal B	21.7	19.1	5.5	0.9
HER2-overexpr.	0.0	0.0	4.7	0.4
Basal-like	0.4	0.0	1.7	8.1

Supplementary Table 2. Case-only analysis with Luminal A as reference group: Genetic background risk. Odds ratios with 95% confidence intervals. Adjusted for born in Sweden or not, education level and age.

Exposure		Luminal A (n, %)	Luminal B (n, %)	OR (95% CI) Luminal B	HER2- overexpressing (n, %)	OR (95% CI) HER2	Basal-like (n, %)	OR (95% CI) Basal-like
BRCA mutation, self-reported	No (Ref.)	1,505 (99%)	203 (99%)	1.00 (ref)	234 (99%)	1.00 (ref)	105 (85%)	1.00 (ref)
	Yes	17 (1%)	3 (1%)	1.05 (0.30-3.64)	3 (1%)	0.99 (0.28-3.44)	18 (15%)	11.13 (5.37-23.07)
Mother or Sister with Breast Cancer	No (Ref.)	960 (80%)	659 (79%)	1.00 (ref)	112 (82%)	1.00 (ref)	169 (80%)	1.00 (ref)
	Yes	236 (20%)	174 (21%)	1.32 (0.95-1.82)	24 (18%)	1.06 (0.76-1.97)	42 (20%)	1.15 (0.75-1.76)
Polygenic Risk Score	1 st quartile	135 (15%)	88 (15%)	1.20 (0.78-1.84)	28 (23%)	0.83 (0.55-1.25)	36 (21%)	1.19 (0.71-2.01)
	2 nd quartile (Ref.)	165 (19%)	121 (20%)	1.00 (ref)	31 (26%)	1.00 (ref)	42 (24%)	1.00 (ref)
	3 rd quartile	241 (27%)	159 (26%)	0.96 (0.63-1.44)	29 (24%)	0.74 (0.51-1.07)	46 (26%)	0.73 (0.44-1.24)
	4 th quartile	242 (39%)	236 (39%)	0.65 (0.43-0.99)	32 (27%)	0.49 (0.34-0.42)	50 (29%)	0.48 (0.28-0.81)
	Linear, per SD increase			0.82 (0.71-0.95)		0.78 (0.68-0.90)		0.66 (0.55-0.80)
Polygenic Risk Score Weighted on ER -	1 st quartile	180 (20%)	123 (23%)	1.40 (0.91-2.14)	22 (18%)	0.98 (0.64-1.49)	31 (18%)	0.76 (0.42-1.37)
	2 nd quartile (Ref.)	224 (25%)	123 (21%)	1.00 (ref)	28 (23%)	1.00 (ref)	37 (21%)	1.00 (ref)
	3 rd quartile	221 (25%)	166 (29%)	1.07 (0.70-1.64)	33 (28%)	0.99 (0.67-1.46)	43 (25%)	0.92 (0.55-1.55)
	4 th quartile	258 (29%)	192 (28%)	1.10 (0.72-1.68)	37 (31%)	1.20 (0.82-1.75)	63 (36%)	1.13 (0.68-1.86)
	Linear, per SD increase			0.93 (0.80-1.08)		1.07 (0.94-1.23)		1.11 (0.92-1.33)

Supplementary Table 3. Case-only analysis with Luminal A as reference group: Reproductive risk factors. Odds ratios with 95% confidence intervals.

Exposure		Luminal A (n, %)	Luminal B (n, %)	OR (95% CI) Luminal B	HER2- overexpressing (n, %)	OR (95% CI) HER2	Basal-like (n, %)	OR (95% CI) Basal-like
Parity ^a	Nulliparous (ref)	196 (16%)	152 (18%)	1.00 (ref)	23 (17%)	1.00 (ref)	32 (14%)	1.00 (ref)
	1-2 children	785 (63%)	513 (59%)	0.81 (0.53-1.22)	90 (65%)	1.06 (0.70-1.61)	138 (62%)	1.49 (0.85-2.62)
	>2 children	265 (21%)	205 (24%)	0.99 (0.64-1.55)	26 (19%)	1.02 (0.65-1.61)	54 (24%)	1.66 (0.90-3.07)
	Linear, per child increase			1.06 (0.94-1.20)		1.04 (0.92-1.17)		1.14 (0.97-1.33)
Age at first birth ^b	<30 (ref)	751 (27%)	499 (70%)	1.00 (ref)	74 (64%)	1.00 (ref)	131 (69%)	1.00 (ref)
	>= 30	278 (73%)	215 (30%)	1.07 (0.75-1.52)	41 (36%)	1.00 (0.72-1.39)	58 (31%)	0.92 (0.60-1.42)
Breastfeeding, parous women only ^c	Ever (ref)	999 (96%)	683 (96%)	1.00 (ref)	115 (99%)	1.00 (ref)	175 (93%)	1.00 (ref)
	Never	44 (4%)	29 (4%)	1.15 (0.54-2.47)	1 (1%)	0.57 (0.22-1.44)	13 (7%)	2.76 (1.37-5.56)
Breastfeeding, including all women ^c	Nulliparous (ref)	196 (16%)	152 (18%)	1.00 (ref)	23 (17%)	1.00 (ref)	32 (14%)	1.00 (ref)
	Parous, never breastfed	44 (4%)	29 (3%)	0.93 (0.41-2.14)	1 (1%)	0.61 (0.22-1.64)	13 (6%)	4.17 (1.79-9.73)
	Parous, breastfed >0-1.5 years	734 (59%)	466 (54%)	0.78 (0.51-1.19)	81 (58%)	1.06 (0.70-1.62)	119 (54%)	1.53 (0.86-2.72)
	Parous, breastfed >1.5 years	265 (21%)	217 (25%)	0.93 (0.56-1.53)	34 (25%)	1.06 (0.64-1.75)	56 (26%)	1.37 (0.70-2.68)

a = Parity adjusted for born in Sweden or not, age, education level, breastfeeding, age at first birth and BMI.

b = Age at first birth adjusted for born in Sweden or not, age, education level, breastfeeding, parity and BMI.

d = Breastfeeding adjusted for born in Sweden or not, age, education level, parity, age at first birth and BMI.

Supplementary Table 4. Case-only analysis with Luminal A as reference group: Ever hormone replacement therapy (HRT) use, age at menarche, mammographic density (MD), Bodyshape at age 18, benign breast disease (BBD). Odds ratios with 95% confidence intervals. SD = standard deviation.

Exposure		Luminal A (n, %)	Luminal B (n, %)	OR (95% CI) Luminal B	HER2- overexpressing (n, %)	OR (95% CI) HER2	Basal-like (n, %)	OR (95% CI) Basal-like
HRT use ^a	Never	677 (63%)	498 (66%)	1.00 (ref)	85 (71%)	1.00 (ref)	144 (72%)	1.00 (ref)
	Ever	395 (37%)	257 (34%)	0.68 (0.47-0.96)	35 (29%)	0.83 (0.60-1.15)	56 (28%)	0.74 (0.47-1.19)
Menarche ^a	Linear, per year increase	1,736	239	1.01 (0.92-1.10)	265	1.07 (0.98-1.17)	149	1.07 (0.96-1.20)
Absolute Mammographic Density ^a	Linear, per SD increase	1,240	160	1.03 (0.87-1.22)	171	0.96 (0.81-1.14)	95	0.89 (0.71-1.13)
Somatotype at age 18 ^b	Linear, Increasingly endomorph	1,739	242	1.00 (0.88-1.12)	265	0.97 (0.86-1.09)	149	1.06 (0.91-1.23)
Benign Breast Disease, Non-proliferative lesions ^c	No	1,242 (91%)	857 (93%)	1.00 (ref)	129 (90%)	1.00 (ref)	221 (94%)	1.00 (ref)
	Yes	76 (6%)	65 (7%)	1.15 (0.68-1.96)	15 (10%)	1.57 (0.99-2.48)	15 (6%)	0.78 (0.36-1.76)
Benign Breast Disease, Proliferative lesions non-atypic ^c	No	1,268 (96%)	899 (97%)	1.00 (ref)	140 (97%)	1.00 (ref)	226 (96%)	1.00 (ref)
	Yes	50 (4%)	23 (3%)	0.66 (0.28-1.53)	4 (3%)	0.88 (0.43-1.79)	10 (4%)	0.88 (0.34-2.23)

a = Adjusted for born in Sweden or not, age, education level, parity and BMI.

b = Adjusted for born in Sweden or not, age, age at menarche and education level.

c = Adjusted for born in Sweden or not, age, education level, parity and BMI.

Supplementary Table 5. Risk of breast cancer overall and by subtype as defined by the St Gallen IHC proxy: Genetic background risk. Odds ratios with 95% confidence intervals, for controls as reference group. Adjusted for born in Sweden or not, education level and age.

Exposure		Controls (n, %)	Luminal A (n, %)	OR (95% CI) Luminal A	Luminal B (n, %)	OR (95% CI) Luminal B	HER2- overexpressing (n, %)	OR (95% CI) HER2	Basal-like (n, %)	OR (95% CI) Basal-like	P heterogeneity
Mother or Sister with Breast Cancer	No (Ref.)	13,379 (86%)	960 (80%)	1.00 (ref)	659 (79%)	1.00 (ref)	112 (82%)	1.00 (ref)	169 (80%)	1.00 (ref)	
	Yes	2,083 (14%)	236 (20%)	1.55 (1.33-1.80)	174 (21%)	1.67 (1.40-1.99)	24 (18%)	1.40 (0.90-2.19)	42 (20%)	1.63 (1.16-2.30)	0.85
Polygenic Risk Score	1 st quartile	1,521 (28%)	135 (15%)	0.81 (0.63-1.04)	88 (15%)	0.73 (0.54-0.97)	28 (23%)	0.90 (0.53-1.51)	36 (21%)	0.85 (0.54-1.34)	0.86
	2 nd quartile (Ref.)	1,449 (27%)	165 (19%)	1.00 (ref)	121 (20%)	1.00 (ref)	31 (26%)	1.00 (ref)	42 (24%)	1.00 (ref)	
	3 rd quartile	1326(24%)	241 (27%)	1.66 (1.33-2.08)	159 (26%)	1.50 (1.16-1.93)	29 (24%)	1.03 (0.62-1.74)	46 (26%)	1.23 (0.80-1.89)	0.26
	4 th quartile	1,129 (21%)	242 (39%)	2.90 (2.34-3.59)	236 (39%)	2.73 (2.13-3.46)	32 (27%)	1.37 (0.83-2.27)	50 (29%)	1.62 (1.06-2.46)	0.005
	Linear, per SD increase			1.70 (1.57-1.84)		1.69 (1.54-1.85)		1.24 (1.03-1.50)		1.26 (1.08-1.57)	<0.0001
Polygenic Risk Score Weighted on ER -	1 st quartile	1,453 (27%)	180 (20%)	0.81 (0.65-1.01)	123 (23%)	1.00 (0.76-1.30)	22 (18%)	0.79 (0.45-1.38)	31 (18%)	0.83 (0.51-1.35)	0.62
	2 nd quartile (Ref.)	1,392(26%)	224 (25%)	1.00 (ref)	123 (21%)	1.00 (ref)	28 (23%)	1.00 (ref)	37 (21%)	1.00 (ref)	
	3 rd quartile	1,336 (25%)	221 (25%)	1.04 (0.84-1.48)	166 (29%)	1.42 (1.10-1.83)	33 (28%)	1.24 (0.75-2.08)	43 (25%)	1.22 (0.78-1.91)	0.23
	4 th quartile	1,244 (23%)	258 (29%)	1.32 (1.07-1.62)	192 (28%)	1.77 (1.38-2.27)	37 (31%)	1.50 (0.91-2.47)	63 (36%)	1.93 (1.27-2.93)	0.14
	Linear, per SD increase			1.21 (1.12-1.31)		1.31 (1.20-1.43)		1.34 (1.11-1.60)		1.41 (1.21-1.64)	0.20

Supplementary Table 6. Risk of breast cancer overall and by subtype as defined by the St Gallen IHC proxy: Reproductive risk factors. Odds ratios with 95% confidence intervals, for controls as reference group.

Exposure		Controls (n, %)	Luminal A (n, %)	OR (95% CI) Luminal A	Luminal B (n, %)	OR (95% CI) Luminal B	HER2- overexpressing (n, %)	OR (95% CI) HER2	Basal-like (n, %)	OR (95% CI) Basal-like	P heterogeneity
Parity ^a	Nulliparous (ref)	1,931 (12%)	196 (16%)	1.00 (ref)	152 (18%)	1.00 (ref)	23 (17%)	1.00 (ref)	32 (14%)	1.00 (ref)	
	1-2 children	10,094 (63%)	785 (63%)	0.73 (0.61-0.88)	513 (597%)	0.58 (0.47-0.72)	90 (65%)	0.61 (0.36-1.06)	138 (62%)	0.91 (0.59-1.40)	0.20
	>2 children	3,914(25%)	265 (21%)	0.62 (0.50-0.76)	205 (24%)	0.62 (0.49-0.78)	26 (19%)	0.48 (0.26-0.86)	54 (24%)	0.91 (0.57-1.44)	0.35
	Linear, per child increase			0.85 (0.81-0.90)		0.88 (0.82-0.94)		0.82 (0.69-0.97)		0.97 (0.86-1.10)	
Age at first birth ^b	<30 (ref)	9,851 (74%)	751 (27%)	1.00 (ref)	499 (70%)	1.00 (ref)	74 (64%)	1.00 (ref)	131 (69%)	1.00 (ref)	
	>= 30	3,448 (26%)	278 (73%)	1.22 (1.05-1.43)	215 (30%)	1.46 (1.22-1.45)	41 (36%)	1.52 (0.99-2.30)	58 (31%)	1.28 (0.91-1.80)	0.44
Breastfeeding ^c , parous women only	Ever (ref)	13,583 (97%)	999 (96%)	1.00 (ref)	683 (96%)	1.00 (ref)	115 (99%)	1.00 (ref)	175 (93%)	1.00 (ref)	
	Never	367 (3%)	44 (4%)	1.48 (1.06-2.07)	29 (4%)	1.52 (1.02-2.27)	1 (1%)	0.36 (0.05-2.58)	13 (7%)	3.23 (1.79-5.80)	0.06
Breastfeeding ^c , including all women	Nulliparous (ref)	1,931(12%)	196 (16%)	1.00 (ref)	152 (18%)	1.00 (ref)	23 (17%)	1.00 (ref)	32 (14%)	1.00 (ref)	
	Parous, never breastfed	367 (2%)	44 (4%)	1.09 (0.75-1.57)	29 (3%)	0.89 (0.57-1.38)	1 (1%)	0.21 (0.03-1.62)	13 (6%)	2.95 (1.47-5.90)	0.009
	Parous, breastfed >0-1.5 years	9,148 (58%)	734 (59%)	0.75 (0.62-0.91)	466 (54%)	0.58 (0.46-0.72)	81 (58%)	0.63 (0.36-1.09)	119 (54%)	0.92 (0.59-1.43)	0.15
	Parous, breastfed >1.5 years	4,435 (28%)	265 (21%)	0.64 (0.51-0.80)	217 (25%)	0.59 (0.45-0.76)	34 (25%)	0.55 (0.29-1.06)	56 (26%)	0.86 (0.51-1.44)	0.60

a = Adjusted for born in Sweden or not, age, education level, breastfeeding, age at first birth and BMI.

b = Adjusted for born in Sweden or not, age, education level, breastfeeding, parity and BMI.

c = Adjusted for born in Sweden or not, age, education level, parity, age at first birth and BMI.

Supplementary Table 7. Risk for breast cancer overall and by subtype as defined by the St Gallen IHC proxy: Bodyshape at age 18, Ever hormone replacement therapy (HRT) use, age at menarche, mammographic density (MD), benign breast disease (BBD). Odds ratios with 95% confidence intervals, for controls as reference group. SD = standard deviation.

Exposure		Controls	Luminal A (n, %)	OR (95% CI) Luminal A	Luminal B (n, %)	OR (95% CI) Luminal B	HER2- overexpressing (n, %)	OR (95% CI) HER2	Basal-like (n, %)	OR (95% CI) Basal-like	P heterogeneity
HRT use ^a	Never	10,922 (75%)	677 (63%)	1.00 (ref)	498 (66%)	1.00 (ref)	85 (71%)	1.00 (ref)	144 (72%)	1.00 (ref)	
	Ever	3,703 (25%)	395 (37%)	1.39 (1.21-1.59)	257 (34%)	1.32 (1.11-1.46)	35 (29%)	1.19 (0.78-1.82)	56 (28%)	1.24 (0.89-1.73)	0.84
Menarche ^a	Linear, per year increase	15,465	1,193	0.93 (0.89-0.97)	836	0.96 (0.91-1.00)	132	1.01 (0.90-1.14)	216	1.01 (0.92-1.10)	0.23
Absolute Mammographic Density ^a	Linear, per SD increase	14,184	869	1.71 (1.61-1.82)	569	1.69 (1.57-1.81)	88	1.64 (1.39-1.95)	137	1.63 (1.41-1.87)	0.88
Somatotype at age 18 ^b	Linear, Increasingly endomorph	15,478	1,194	0.94 (0.89-0.99)	840	0.91 (0.85-0.97)	133	0.80 (0.68-0.94)	216	1.03 (0.92-1.17)	0.06
Benign Breast Disease ^c , Non-proliferative lesions	No	14,922 (94%)	1,242 (91%)	1.00 (ref)	857 (93%)	1.00 (ref)	129 (90%)	1.00 (ref)	221 (94%)	1.00 (ref)	
	Yes	1,023 (6%)	76 (6%)	0.81 (0.62-1.04)	65 (7%)	1.16 (0.89-1.50)	15 (10%)	1.80 (1.05-3.11)	15 (6%)	1.01 (0.59-1.75)	0.03
Benign Breast Disease ^c , Proliferative lesions non-atypic	No	15,566 (98%)	1,268 (96%)	1.00 (ref)	899 (97%)	1.00 (ref)	140 (97%)	1.00 (ref)	226 (96%)	1.00 (ref)	
	Yes	379 (2%)	50 (4%)	1.80 (1.32-2.46)	23 (3%)	1.20 (0.78-1.84)	4 (3%)	1.28 (0.47-3.49)	10 (4%)	1.76 (0.90-3.46)	0.43

a = Adjusted for born in Sweden or not, age, education level, parity and BMI.

b = Adjusted for born in Sweden or not, age, age at menarche and education level.

c = Adjusted for born in Sweden or not, age, education level, parity and BMI.