Supplementary information

Lakeman et al. Clinical applicability of the Polygenic Risk Score for breast cancer risk prediction in familial cases.

Lakeman IMM, *et al. J Med Genet* 2022;0:1–10. doi: 10.1136/jmg-2022-108502

Supplementary methods

Study cohorts

HEBON

The HEBON study[1] (initiated in 1999) is an ongoing nationwide retrospective cohort study among breast cancer families with prospective follow up. Participants were invited after visiting one of the Clinical Genetic Centers in the Netherlands for breast and/or ovarian cancer counselling. Participants were asked to fill in a questionnaire about lifestyle, family history and risk factors for breast cancer. Linkage with the nationwide cancer and pathology registries is possible for follow up.

Additional selection criteria for HEBON participants included:

- At least two breast cancer cases in a family with available DNA samples
- Breast cancer diagnosis below the age of 60 years and a positive family history:
    - One first degree family member with breast cancer diagnosis below the age of 50 OR
    - Two first or second-degree family members with breast cancer diagnosis below the age of 60

ABCS-F and RBCS

The ABCS-F[2] and RBCS[3] case-cohorts included also breast cancer cases who visited the Clinical Genetic Centres of the Netherlands Cancer Institute in Amsterdam or the Erasmus Medical Center in Rotterdam, respectively. No additional selection criteria were used for ABCS-F and RBCS cases. 151 individuals from the ABCS-F study and 469 individuals from the RBCS study are included in the HEBON study as well and shown as HEBON cases in Table 1.

Quality control procedure

For the 2,179 breast cancer cases without a *BRCA1/2* pathogenic variant that were genotyped with the GSA array, quality control was performed with Plink version 1.9, which excluded 8,408 SNPs with a call rate below 95%. Another 712 SNPs were removed because of a deviation from Hardy-Weinberg equilibrium in controls at $P<1x10^{-12}$. In total, 124 individuals were excluded of which 62 individuals with a call rate below 95%, 7 individuals because they were genotypically not female or the gender was uncertain, and 17 individuals because of a sample swab. After population stratification analysis, 28 individuals were excluded because of non-European genotype (>3 SD).

Imputation pedigrees

In total, 3,492 pedigrees were collected for this study. These pedigrees consisted of 202,680 individuals (49% female) of which 12,785 individuals were affected with breast cancer. If the age of breast cancer diagnosis for a family member was not known (n=1,272), a conditional average age was estimated given the age at last

follow up of the individual and the breast cancer incidence in the Netherlands. Furthermore, for all affected individuals with breast cancer, ovarian cancer, prostate cancer or pancreatic cancer the year of birth was imputed, if this was not yet available, based on the year of birth of the closest relative (25 year difference for parents and children, average for siblings). If the age of last follow up was not known, this age was calculated based on the date of the last update of a pedigree and the year of birth.

Supplementary figures

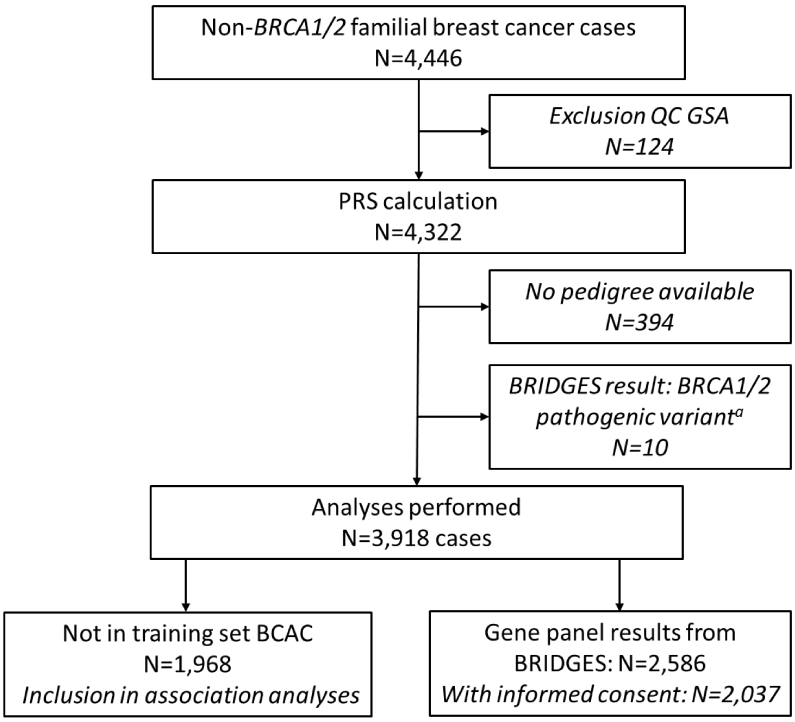

Figure S1: Flow scheme of the selection procedure

Breast cancer cases were selected from the ABCS, HEBON and RBCS studies. Details of the quality control procedure are described above. Absolute lifetime risks were calculated for all included cases (N=3,918). To exclude overlap of cases with the development dataset for the PRS_{313}[4], only 1,968 cases were included in the association analyses. For the majority of cases gene panel information was available. For cases of whom we did not have informed consent to report the clinical relevant results, only pseudo anonymized information about pathogenic variants in *ATM, CHEK2*, and *PALB2* was available (N=549). For the cases with informed consent, the number of pathogenic variants and missense variants are shown in Table S3.

[a]carriers of a pathogenic variant or family member of a carrier of a pathogenic variant in *BRCA1* or *BRCA2.*

Abbreviations: BCAC, Breast Cancer Association Consortium; BRIDGES, Breast cancer Risk after Diagnostic GEne Sequencing; PRS, Polygenic Risk Score.

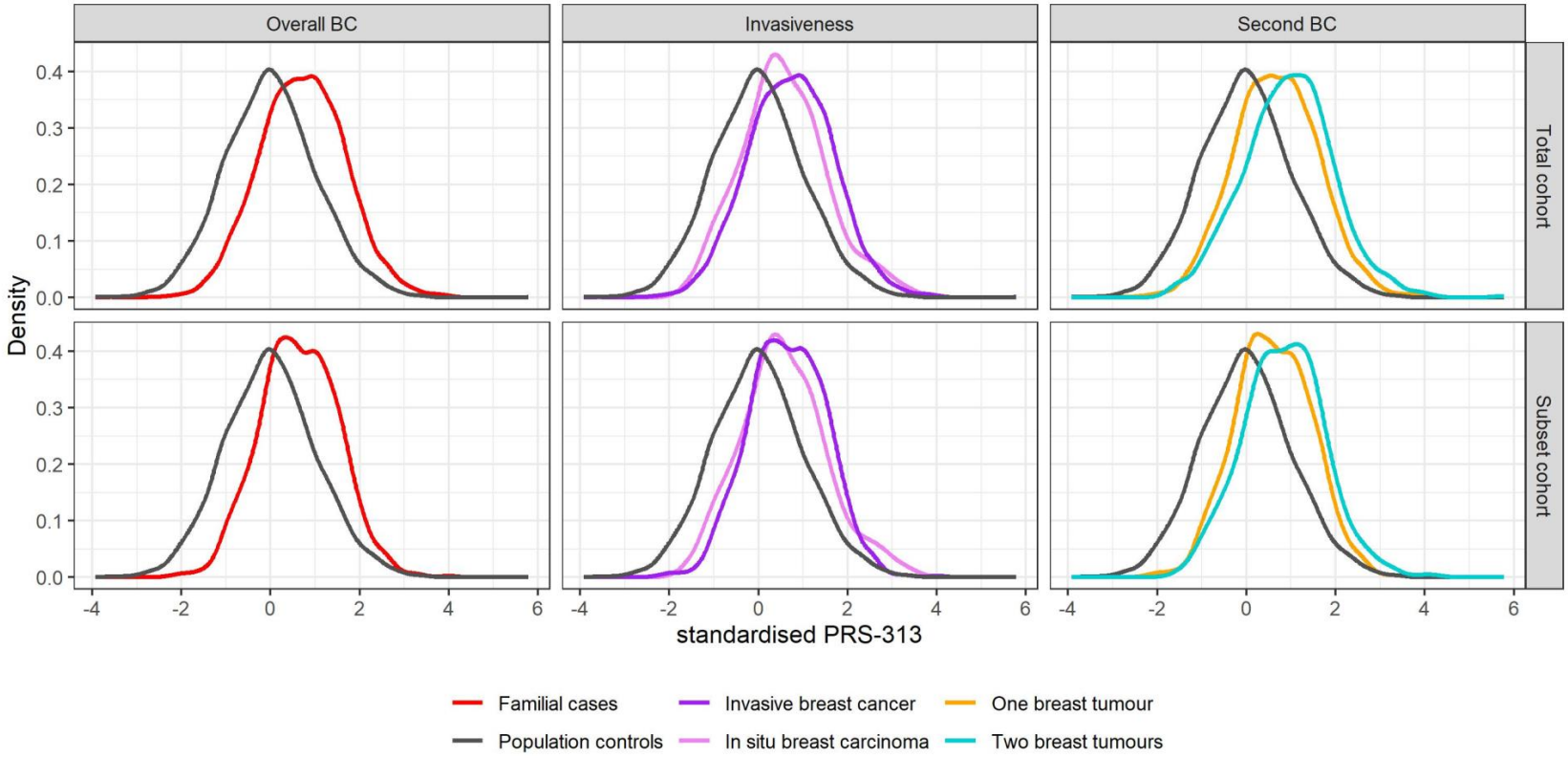

Figure S2: Density curves of the PRS_{313}

Distribution of the PRS_{313} in the included 3,474 population controls (grey line) and 3,918 and 1,968 breast cancer cases (red line) in the total and subset cohort respectively. For the invasiveness figure, 3 cases were excluded for which invasiveness for the first and/or second breast tumour was unknown. In the total cohort 3,653 and 262 cases were included with invasive (purple line) and in situ (pink line) breast cancer respectively. For the subset cohort this was 1,703 and 262. In the right figure, 719 and 327 breast cancer cases with a second breast tumour (blue line) were included in the total and subset cohort respectively.

Abbreviations: BC, Breast Cancer; PRS, Polygenic Risk Score.

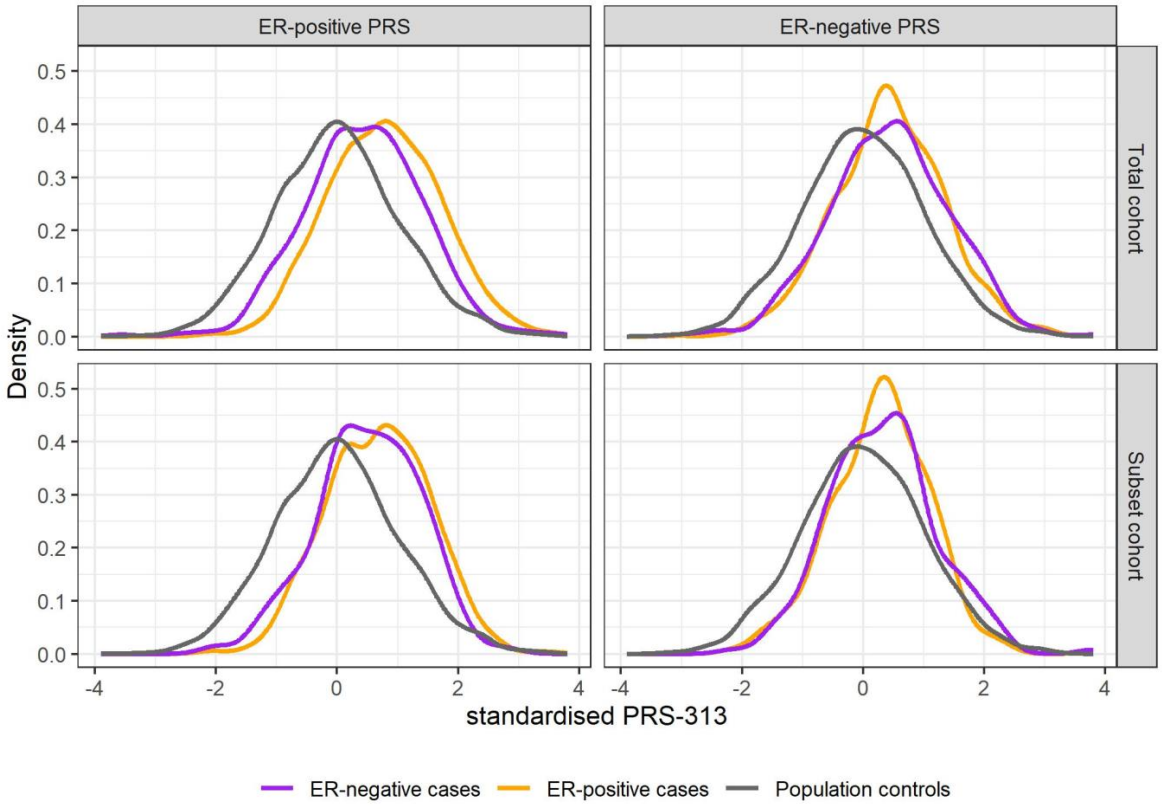

Figure S3: Density curves of the ER-positive and ER-negative PRS$_{313}$

Distribution of the ER-negative (left figures) and ER-positive (right figures) PRS$_{313}$ for cases with an ER-negative (purple line) and ER-positive (orange line) first breast tumour. As a reference, the distribution of these PRS in population controls are shown as well (grey line). In the total cohort, 1,755 and 488 breast cancer cases are included with a first ER-positve and ER-negative breast tumour respectively. For the subset cohort this was 927 and 213 respectively.

Abbreviations: ER, Estrogen Receptor; PRS, Polygenic Risk Score

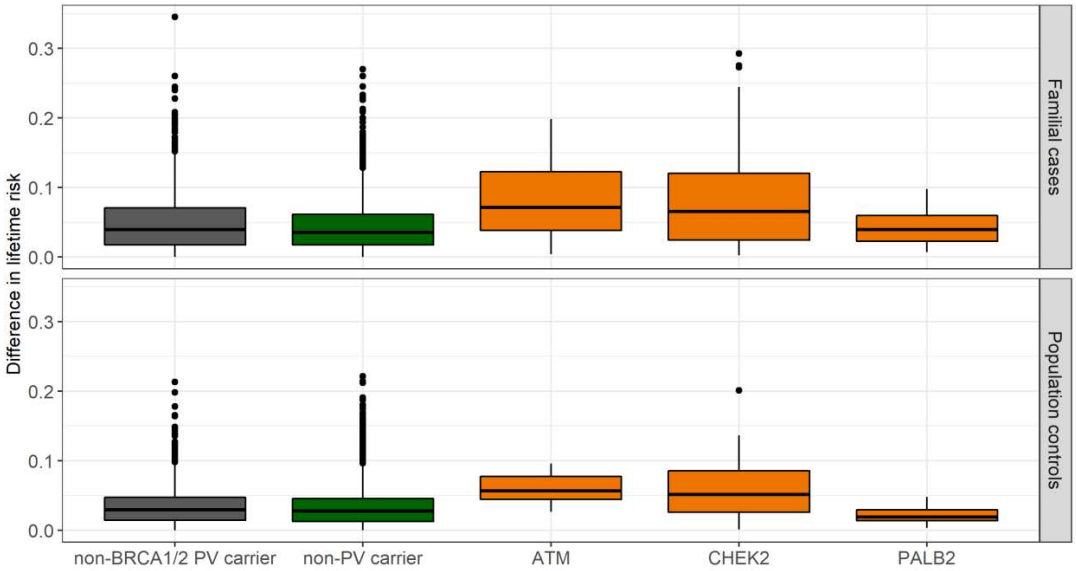

Figure S4: Difference in breast cancer lifetime risk score calculated by BOADICEA

Boxplot of the difference in breast cancer lifetime risk between the basic calculation in BOADICEA and after including the PRS_{313}. The basic calculation included birth year, gene panel results and for cases a pedigree of their family in addition. Non-carriers are the group of which we know that they do not have a pathogenic variant in *ATM, CHEK2* and *PALB2* in addition to *BRCA1/2.*

Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; PV, Pathogenic Variant.

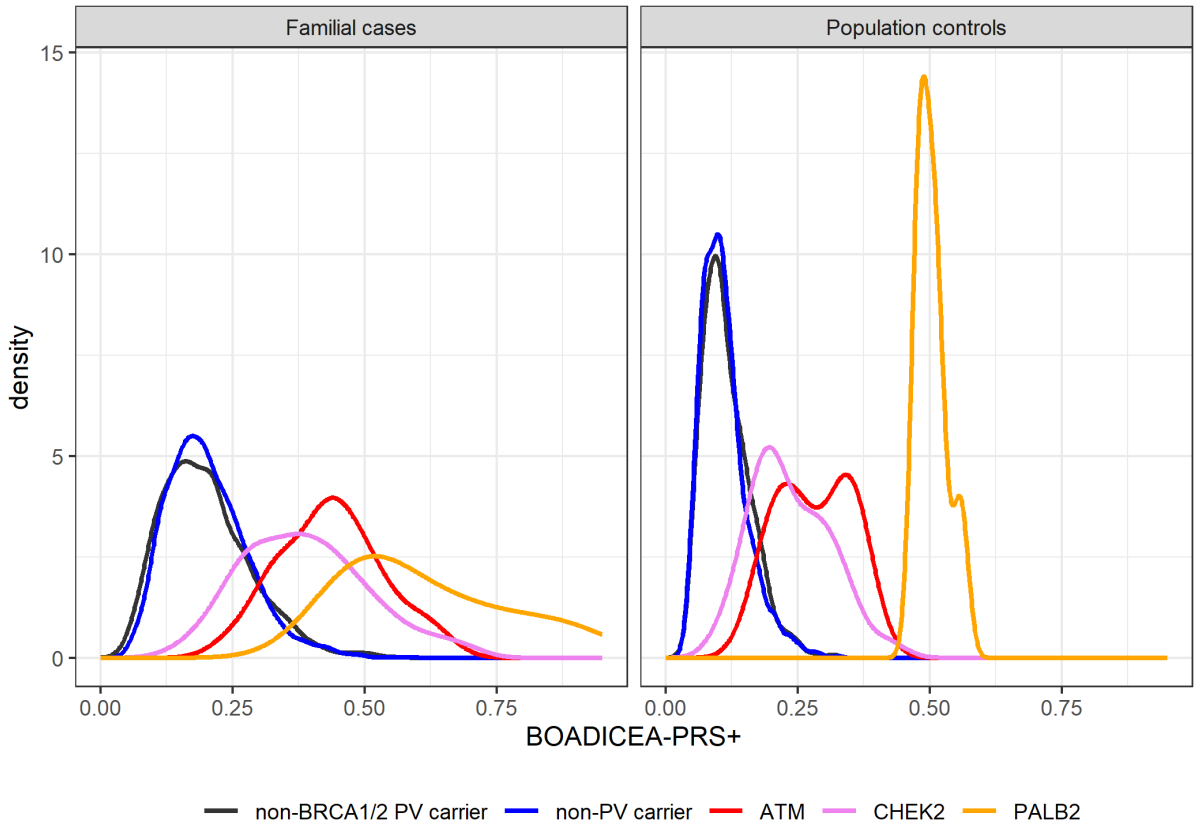

Figure S5. Distribution of breast cancer lifetime risk after including the PRS_{313}

Density plots of the distribution in breast cancer lifetime risk calculated with BOADICEA including birth cohort, gene panel results, pedigree-based family history for cases and the PRS_{313}.

Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; PV, Pathogenic Variant; PRS, Polygenic Risk Score

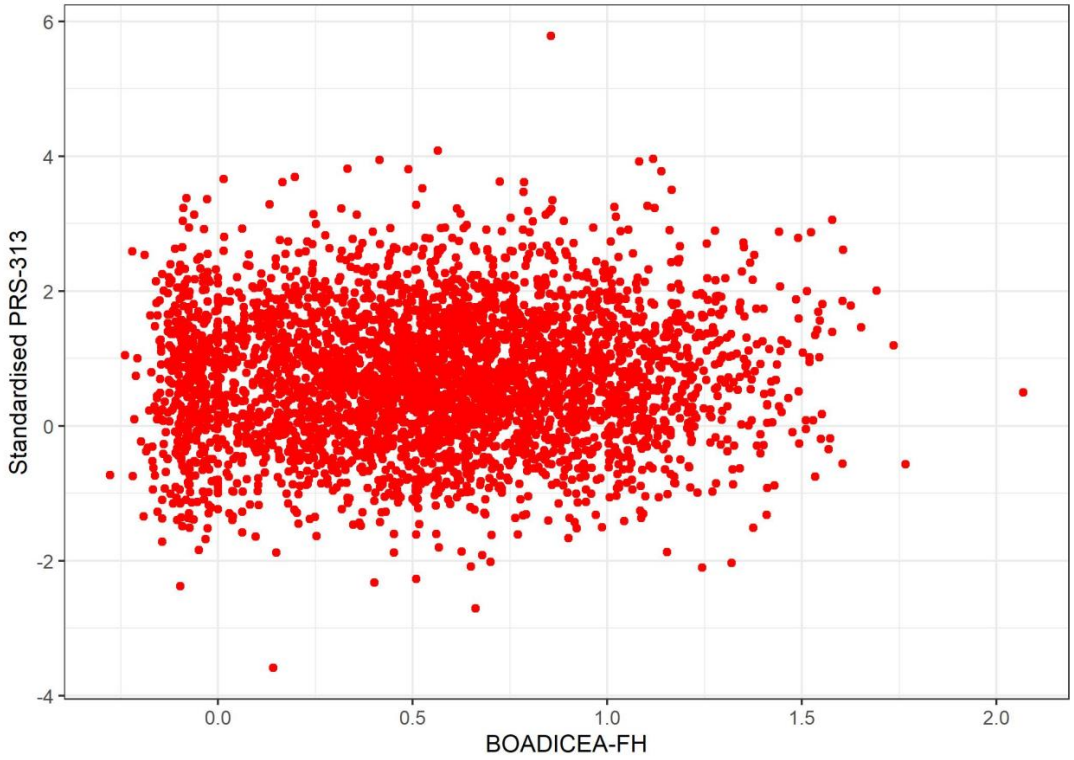

Figure S6. Correlation plot between de BOADICEA$_{FH}$ and the PRS$_{313}$

For all included breast cancer cases (N=3,918), the individual BOADICEA$_{FH}$ (polygenic load) is plotted against the PRS$_{313}$. BOADICEA$_{FH}$ was calculated with BOADICEA based on the pedigree without inclusion of the PRS$_{313}$. Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; FH, Family History; PRS, Polygenic Risk Score.

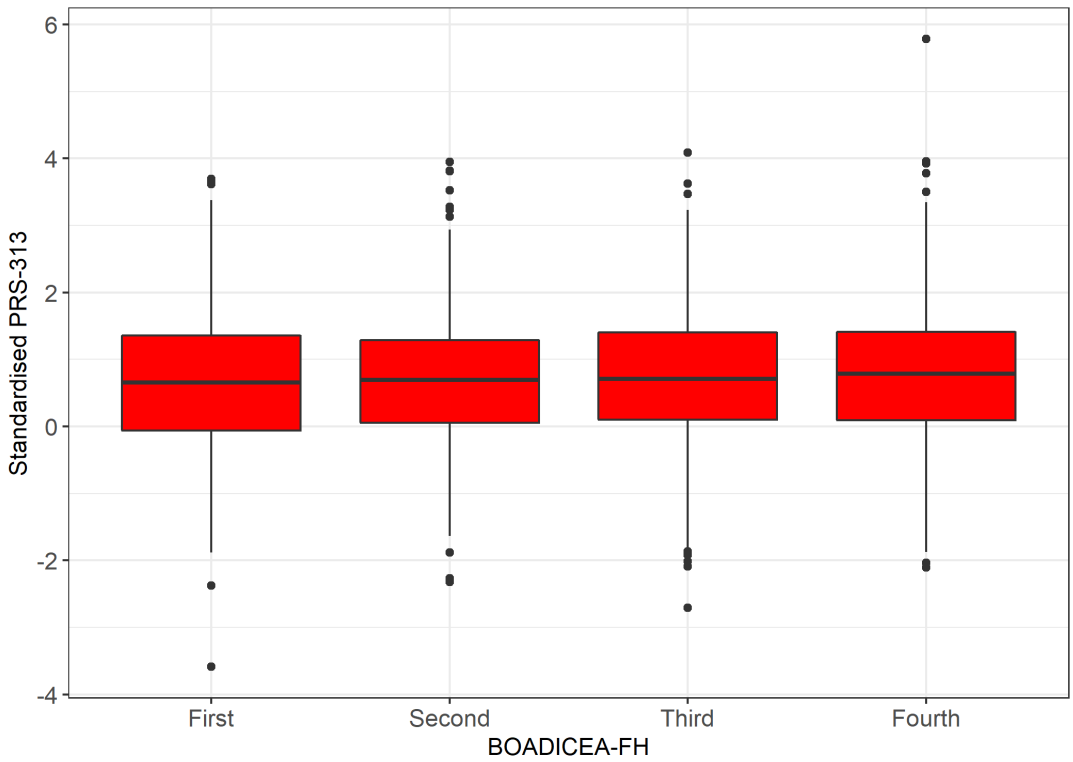

Figure S7: PRS_{313} distribution by quartiles of $BOADICEA_{FH}$

The PRS_{313} distribution for all included cases (N=3,918) separated by quartiles of the individual $BOADICEA_{FH}$ (polygenic load). $BOADICEA_{FH}$ was calculated with BOADICEA based on the pedigree without inclusion of the PRS_{313}.

Abbreviations: BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; FH, Family History; PRS, Polygenic Risk Score.

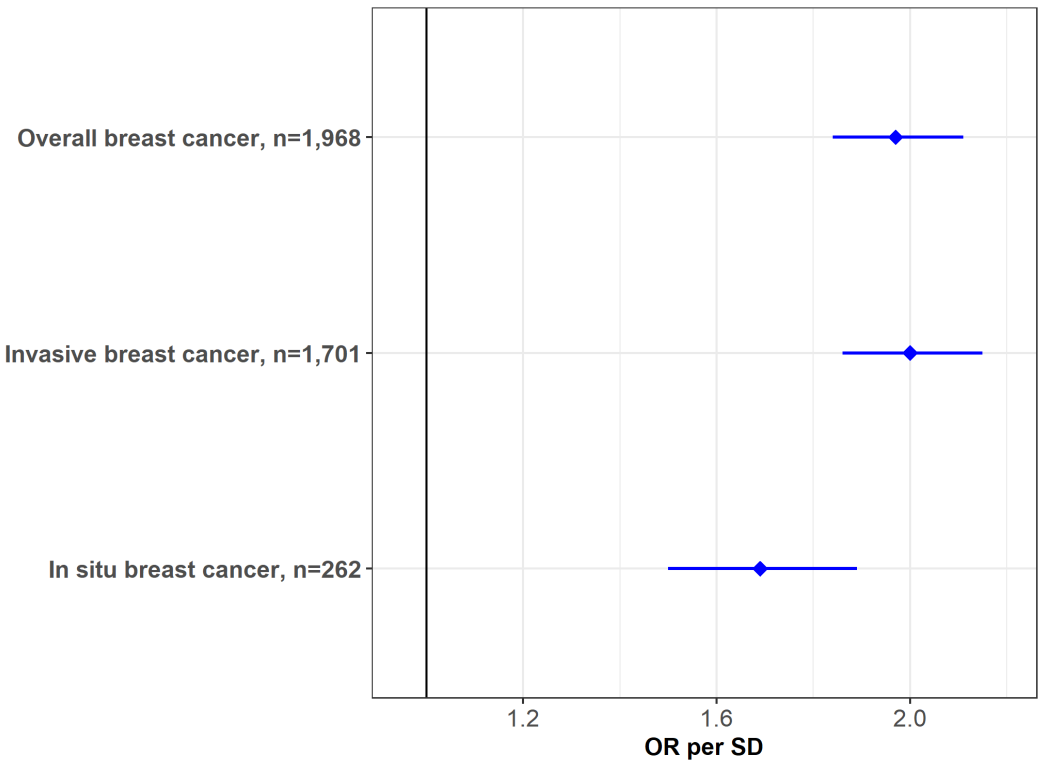

Figure S8: Association between the PRS_{313} and breast cancer

Visualisation of the effect sizes and 95% confidence intervals of the association between the PRS_{313} and breast cancer. The corresponding OR and included breast cancer cases are shown in Table 3. Abbreviations: BC, Breast Cancer; OR, Odds Ratio; PRS, Polygenic Risk Score

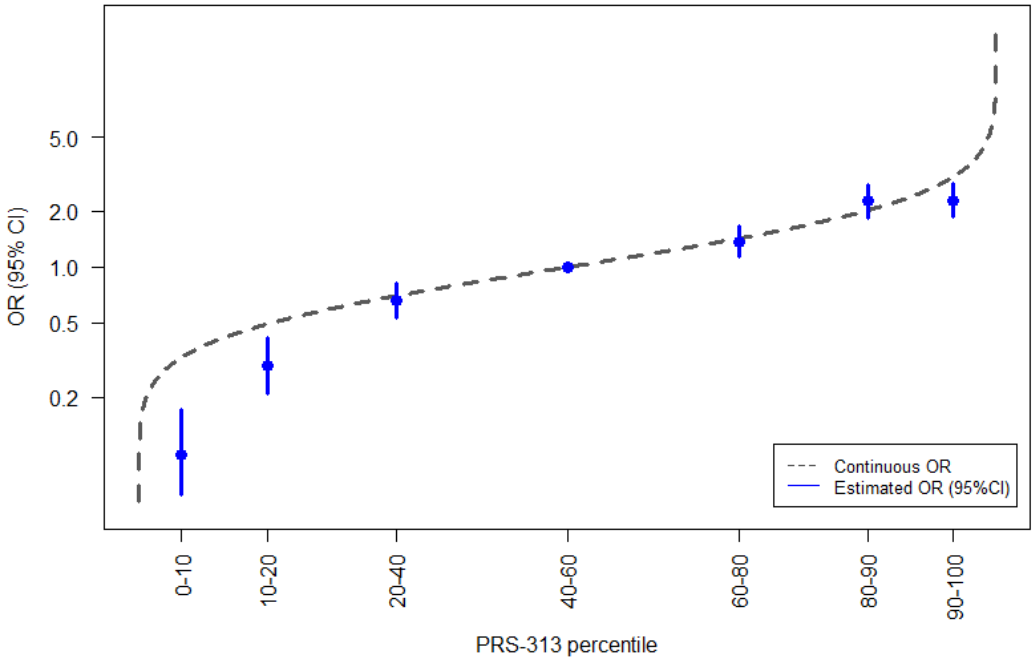

Figure S9: Association between the PRS and breast cancer by percentiles of the PRS_{313}

Plot of the effect size of the association between the continuous PRS_{313} (grey line) and breast cancer and the categorical PRS_{313} (blue dots) and breast cancer. Corresponding OR and 95% confidence intervals are shown in Table 3.

Abbreviations: CI, Confidence Interval; OR, Odds Ratio; PRS, Polygenic Risk Score.

Supplementary tables

Table S1: common low risk variants included in the PRS_{313} (large Excel file)

This table is partly published before by Mavaddat et al.[4] We added the imputation quality in this study

Table S2: Descriptives of the standardised PRS_{313}

Group	Total cohort			Family-based cases – subset[c]		
	N	Mean PRS_{313}	SD PRS_{313}	N	Mean PRS_{313}	SD PRS_{313}
All cases	3,918	0.71	0.96	1,968	0.64	0.88
Invasive cases[a]	3,653	0.73	0.96	1,703	0.65	0.86
***In situ* only cases[b]**	262	0.56	0.96	262	0.56	0.96
1 breast tumour	3,199	0.66	0.95	1,641	0.60	0.87
2 breast tumours	719	0.95	1.01	327	0.83	0.90
Population controls	3,474	0	1.03	NA	NA	NA

[a]Invasive first or second tumour

[b]no invasive first or second tumour

[c]Cases included in the association analyses which were not part of the development dataset for the PRS_{313} as described in Mavaddat et al.[4]

Abbreviations: N, Number; NA, Not Applicable; PRS, Polygenic Risk Score

Table S3: Descriptives of the standardised ER-positive and ER-negative PRS_{313}

Group	PRS	Total cohort			Family-based cases – subset[c]		
		N	Mean PRS	SD PRS	N	Mean PRS	SD PRS
ER-positive BC	ER-positive PRS	1,755	0.78	0.92	927	0.68	0.86
ER-negative BC	ER-positive PRS	488	0.43	0.98	213	0.51	0.85
ER-positive BC	ER-negative PRS	1,755	0.76	0.93	927	0.66	0.85
ER-negative BC	ER-negative PRS	488	0.46	0.97	213	0.52	0.85

[a]Invasive first or second tumour

[b]no invasive first or second tumour

[c]Cases included in the association analyses which were not part of the development dataset for the PRS_{313} as described in Mavaddat et al.[4]

Abbreviations: N, Number; NA, Not Applicable; PRS, Polygenic Risk Score

Table S4: Truncating variants in BRIDGES gene panel

Gene	Cases, N=2,037[a]		Controls, N=2,584[a]		OR	95% CI	P-value
	N	%	N	%			
ABRAXAS1	1	0.0	0	0.0	NA	NA	NA
AKT1	0	0.0	0	0.0	NA	NA	NA
ATM	**36**	**1.8**	**9**	0.3	**5.15**	**2.42-12.18**	**$1.00x10^{-06}$**
BARD1	1	0.0	1	0.0	1.27	0.02-99.55	1.00
BRCA1	NA	NA	NA	NA	NA	NA	NA
BRCA2	NA	NA	NA	NA	NA	NA	NA
BRE	0	0.0	0	0.0	NA	NA	NA
BRIP1	4	0.2	5	0.2	1.01	0.20-4.72	1.00
CDH1	0	0.0	0	0.0	NA	NA	NA
CHEK2	**131**	**6.4**	**31**	**1.2**	**5.66**	**3.78-8.70**	**$<2.00x10^{-16}$**
c.1100delC[b]	130		30				
Other	1						
EPCAM	0	0.0	2	0.1	NA	NA	NA
FANCC	5	0.2	8	0.3	0.79	0.20-2.75	0.80
FANCM	14	0.7	16	0.6	1.11	0.50-2.44	0.90
GEN1	0	0.0	0	0.0	NA	NA	NA
MEN1	0	0.0	0	0.0	NA	NA	NA
MLH1	0	0.0	0	0.0	NA	NA	NA
MRE11A	1	0.0	3	0.1	0.42	0.01-5.27	0.60
MSH2	0	0.0	2	0.1	NA	NA	NA
MSH6	1	0.0	0	0.0	NA	NA	NA
MUTYH	3	0.1	2	0.1	1.9	0.22-22.81	0.70
NBN	2	0.1	3	0.1	0.85	0.07-7.39	1,00
NF1	2	0.1	0	0.0	NA	NA	NA
PALB2	12[c]	0.6	7	0.3	2.18	0.79-6.55	0.10
PIK3CA	0	0.0	0	0.0	NA	NA	NA
PMS2	1	0.0	2	0.1	0.63	0.01-12.19	1.00
PTEN	1	0.0	1	0.0	1.27	0.02-99.55	1.00
RAD50	4	0.2	7	0.3	0.72	0.16-2.85	0.80
RAD51C	1	0.0	0	0.0	NA	NA	NA
RAD51D	5	0.2	0	0.0	NA	NA	NA
RECQL	2	0.1	3	0.1	0.85	0.07-7.39	1.00
RINT1	0	0.0	2	0.1	NA	NA	NA
STK11	0	0.0	0	0.0	NA	NA	NA
TP53	0	0.0	0	0.0	NA	NA	NA
XRCC2	0	0.0	1	0.0	NA	NA	NA
Total	227	11.1	105	4.1	-	-	-

[a]Cases and controls were included in the analyses described by Dorling et al.[5]

[b]of which 6 homozygous in cases and 1 homozygous in controls

[c]In addition to inclusion criteria for truncating variants in BRIDGES, 4 *PALB2* truncating variants in the last exon were added.

Abbreviations: CI, Confidence Interval; N, Number; NA, Not Applicable; OR, Odds Ratio.

Table S5: Missense variants in BRIDGES gene panel

Gene	Cases; N=2,038[a]		Controls, N=2,584[a]	
	Total[b]	P/LP[c]	Total[b]	P/LP[c]
ABRAXAS1	3	NA	5	NA
AKT1	2	NA	6	NA
ATM	121	5	113	4
BARD1	25	0	26	0
BRCA1	42	NA	49	NA
BRCA2	109	NA	127	NA
BRE	0	NA	0	NA
BRIP1	34	NA	41	NA
CDH1	26	NA	28	NA
CHEK2	64	8	34	2
EPCAM	9	NA	18	NA
FANCC	28	NA	23	NA
FANCM	64	NA	62	NA
GEN1	38	NA	32	NA
MEN1	4	NA	2	NA
MLH1	19	NA	21	NA
MRE11A	16	NA	19	NA
MSH2	42	NA	56	NA
MSH6	51	NA	52	NA
MUTYH	28	NA	33	NA
NBN	35	NA	23	NA
NF1	30	NA	34	NA
PALB2	23	0	23	0
PIK3CA	6	NA	10	NA
PMS2	37	NA	28	NA
PTEN	3	NA	7	NA
RAD50	50	NA	46	NA
RAD51C	9	1	9	0
RAD51D	6	0	10	0
RECQL	16	NA	20	NA
RINT1	39	NA	47	NA
STK11	0	NA	1	NA
TP53	14	4	10	0
XRCC2	6	NA	13	NA
Total	999	18	1,028	6

[a]Cases and controls were included in the analyses described by Dorling et al.[5]

[b]Total number of missense variants detected, not corrected for individuals who carry more than one missense variant in a single gene.

[c]For genes in which pathogenic variants are associated with breast cancer[5], missense variant interpretation was performed by using the ClinVar database[6].

Abbreviations: N, Number; NA, Not Applicable; P, Pathogenic; LP, Likely Pathogenic.

Table S6: Absolute change in breast cancer lifetime risk after including the PRS_{313}

	Cases			Controls		
	Min	**Mean**	**Max**	**Min**	**Mean**	**Max**
No gene-test result	0.0	5.0	34.5	0.0	3.5	21.3
Non-carriers	0.0	4.5	27.0	0.0	3.3	22.1
***ATM* PV carriers**[a]	0.4	8.0	19.8	2.6	5.9	9.6
***CHEK2* PV carriers**[a]	0.3	8.1	29.3	0.1	5.9	20.1
***PALB2* PV carriers**	0.7	4.4	9.8	0.3	2.2	4.8

[a]Two cases with both a pathogenic variant in *CHEK2* and *ATM* were excluded.

In total, 1,331 cases and 890 controls were included without a gene-test result for *PALB2*, *ATM* and *CHEK2*; 2,369 cases and 2,537 controls in the non-PV carrier group; 167 cases and 31 controls in the *CHEK2* PV carrier group; 39 cases and 9 controls in the *ATM* PV carrier group; 10 cases and 7 controls in the *PALB2* PV carrier group.

Abbreviations: Min, Minimum; Max, Maximum; PRS, Polygenic Risk Score; PV, Pathogenic Variant.

Table S7: Breast cancer lifetime risk category change based on the NCCN guideline

Group	BOADICEA Lifetime risk		No gene-test result		Non-PV carriers		*CHEK2* PV carriers[a]		*ATM* PV carriers[a]		*PALB2* PV carriers	
	Without PRS_{313}	**Including PRS_{313}**	**N**	**% change**	**N**	**% change**	**N**	**% change**	**N**	**% change**	**N**	**% change**
Cases	**<20%**	**<20%**	697	30.4	1,126	30.1	3	70.0	0	0.0	0	0.0
		>20%	305		486		7		0		0	
	>20%	**>20%**	292	11.2	605	20.1	153	2.5	39	0.0	10	0.0
		<20%	37		152		4		0		0	
	Overall change			**25.7**		**26.9**		**6.6**		**0.0**		**0.0**
	Upward change			**22.9**		**20.5**		**4.1**		**0.0**		**0.0**
Controls	**<20%**	**<20%**	851	4.4	2,419	4.7	NA		NA		NA	
		>20%	39		118							
	>20%	**>20%**	NA		NA		19	38.7	8	11.1	7	0.0
		<20%					12		1		0	
	Overall change			**4.4**		**4.7**		**38.7**		**11.1**		**0.0**
	Upward change			**4.4**		**4.7**		**0.0**		**0.0**		**0.0**

[a]Two cases with both a pathogenic variant in *CHEK2* and *ATM* were excluded.

In total, 1,331 cases and 890 controls were included without a gene-test result (no *BRCA1/2* PV); 2,369 cases and 2,537 controls in the non-PV carrier group; 167 cases and 31 controls in the *CHEK2* PV carrier group; 39 cases and 9 controls in the *ATM* PV carrier group; 10 cases and 7 controls in the *PALB2* PV carrier group.

Abbreviations: BOADICEA, the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; NCCN, the National Comprehensive Cancer Network guideline; PRS, Polygenic Risk Score; PV, Pathogenic Variant.

Table S8: Breast cancer lifetime risk category change based on the NICE guideline

Group	BOADICEA Lifetime risk		No gene-test result		Non-PV carriers		*CHEK2* PV carriers[a]		*ATM* PV carriers[a]		*PALB2* PV carriers	
	Without PRS_{313}	Including PRS_{313}	N	% change	N	% change	N	% change	N	% change	N	% change
Cases	**<17%**	**<17%**	478	38.5	699	37.1	1	0.0	NA		NA	
		>17%	299		413		0					
	17-30%	**17-30%**	332	34.3	799	31.5	34	48.5	0	100.0	NA	
		<17%	68		203		1		0			
		>30%	105		164		31		5			
	>30%	**>30%**	42	14.3	65	28.6	93	7.0	32	5.9	10	0.0
		<30%	67		26		7		2		0	
	Overall change			**36.0**		**34.0**		**23.4**		**17.9**		**0.0**
	Upward change			**29.0**		**24.4**		**18.6**		**12.8**		**0.0**
Controls	**<17%**	**<17%**	783	12.0	2,289	9.8	NA		NA		NA	
		>17%	107		248							
	17-30%	**17-30%**	NA		NA		20	35.5	5	44.4	NA	
		<17%					5		0			
		>30%					6		4			
	>30%	**>30%**	NA		NA		NA		NA		7	0.0
		<30%									0	
	Overall change			**12.0**		**9.8**		**35.5**		**44.4**		**0.0**
	Upward change			**12.0**		**9.8**		**19.4**		**44.4**		**0.0**

[a]Two cases with both a pathogenic variant in *CHEK2* and *ATM* were excluded.

In total, 1,331 cases and 890 controls were included without a gene-test result; 2,369 cases and 2,537 controls in the non-PV carrier group; 167 cases and 31 controls in the *CHEK2* PV carrier group; 39 cases and 9 controls in the *ATM* PV carrier group; 10 cases and 7 controls in the *PALB2* PV carrier group. Abbreviations: BOADICEA, the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; NICE, the National Institute for Health and Care Excellence guideline; PRS, Polygenic Risk Score; PV, Pathogenic Variant.

Table S9: Breast cancer lifetime risk by age of breast cancer diagnosis for cases based on the Dutch IKNL guideline

Group		<40 years		40-50 years		≥50 years	
	BOADICEA LTR	**Without PRS_{313}**	**Including PRS_{313}**	**Without PRS_{313}**	**Including PRS_{313}**	**Without PRS_{313}**	**Including PRS_{313}**
No gene-test result	**<20%**	403 (87%)	305 (66%)	377 (74%)	257 (50%)	222 (62%)	172 (48%)
	20-30%	58 (13%)	127 (27%)	111 (22%)	186(36%)	111 (31%)	122 (34%)
	>30%	1 (0%)	30 (6%)	24 (5%)	69 (13%)	24 (7%)	63 (17%)
Non-PV carriers	**<20%**	475 (81%)	367 (62%)	706 (65%)	557 (52%)	431 (61%)	354 (50%)
	20-30%	96 (16%)	183 (31%)	328 (30%)	395 (37%)	242 (34%)	267 (38%)
	>30%	17 (3%)	38 (6%)	44 (4%)	126 (12%)	30 (4%)	82 (12%)
***CHEK2* PV carriers[a]**	**<20%**	4 (8%)	3 (6%)	4 (5%)	1 (1%)	2 (4%)	3 (7%)
	20-30%	17 (35%)	12 (24%)	22 (30%)	11 (15%)	18 (40%)	13 (29%)
	>30%	28 (57%)	34 (69%)	47 (46%)	61 (84%)	25 (56%)	29 (64%)
***ATM* PV carriers[a]**	**<20%**	NA	NA	NA	NA	NA	NA
	20-30%	2 (20%)	1 (10%)	2 (12%)	1 (6%)	1 (8%)	0 (0%)
	>30%	8 (80%)	9 (90%)	15 (88%)	16 (94%)	11 (92%)	12 (100%)
***PALB2* PV carriers**	**<20%**	NA	NA	NA	NA	NA	NA
	20-30%	NA	NA	NA	NA	NA	NA
	>30%	4 (100%)	4 (100%)	5 (100%)	5 (100%)	1 (100%)	1 (100%)

[a]Two cases with both a pathogenic variant in *CHEK2* and *ATM* were excluded.

In total, 1,331 cases were included without a gene-test result; 2,369 cases in the non-PV carrier group; 167 cases in the *CHEK2* PV carrier group; 39 cases in the *ATM* PV carrier group; 10 cases in the *PALB2* PV carrier group.

Abbreviations: BOADICEA, the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; IKNL, Netherlands Comprehensive Cancer Organisation guideline; LTR, Life Time Risk; PRS, Polygenic Risk Score; PV, Pathogenic Variant.

References

1. van Asperen CJ, Brohet RM, Meijers-Heijboer EJ, et al. Cancer risks in BRCA2 families: estimates for sites other than breast and ovary. *Journal of medical genetics*. Sep 2005;42(9):711-9. doi:10.1136/jmg.2004.028829
2. Schmidt MK, Hogervorst F, van Hien R, et al. Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Aug 10 2016;34(23):2750-60. doi:10.1200/jco.2016.66.5844
3. Liu J, Prager-van der Smissen WJ, Schmidt MK, et al. Recurrent HOXB13 mutations in the Dutch population do not associate with increased breast cancer risk. *Sci Rep*. Jul 18 2016;6:30026. doi:10.1038/srep30026
4. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American journal of human genetics*. Jan 3 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
5. Dorling L, Carvalho S, Allen J, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *The New England journal of medicine*. Jan 20 2021;doi:10.1056/NEJMoa1913948
6. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. Jan 4 2018;46(D1):D1062-d1067. doi:10.1093/nar/gkx1153