Gene–environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study

*Ruth C Travis, Gillian K Reeves, Jane Green, Diana Bull, Sarah J Tipper, Krys Baker, Valerie Beral, Richard Peto, John Bell, Diana Zelenika, Mark Lathrop, for the Million Women Study Collaborators**

Summary

Background **Information is scarce about the combined effects on breast cancer incidence of low-penetrance genetic susceptibility polymorphisms and environmental factors (reproductive, behavioural, and anthropometric risk factors for breast cancer). To test for evidence of gene–environment interactions, we compared genotypic relative risks for breast cancer across the other risk factors in a large UK prospective study.**

Methods **We tested gene–environment interactions in 7610 women who developed breast cancer and 10 196 controls without the disease, studying the effects of 12 polymorphisms (*FGFR2*-rs2981582, *TNRC9*-rs3803662, 2q35-rs13387042, *MAP3K1*-rs889312, 8q24-rs13281615, 2p-rs4666451, 5p12-rs981782, *CASP8*-rs1045485, *LSP1*-rs3817198, 5q-rs30099, *TGFB1*-rs1982073, and *ATM*-rs1800054) in relation to prospectively collected information about ten established environmental risk factors (age at menarche, parity, age at first birth, breastfeeding, menopausal status, age at menopause, use of hormone replacement therapy, body-mass index, height, and alcohol consumption).**

Findings **After allowance for multiple testing none of the 120 comparisons yielded significant evidence of a gene–environment interaction. By contrast with previous suggestions, there was little evidence that the genotypic relative risks were affected by use of hormone replacement therapy, either overall or for oestrogen-receptor-positive disease. Only one of the 12 polymorphisms was correlated with any of the ten other risk factors: carriers of the high-risk C allele of *MAP3K1*-rs889312 were significantly shorter than non-carriers (mean height 162·4 cm [95% CI 162·1–162·7] *vs* 163·1 cm [162·9–163·2]; p=0·01 after allowance for multiple testing).**

Interpretation **Risks of breast cancer associated with low-penetrance susceptibility polymorphisms do not vary significantly with these ten established environmental risk factors.**

Funding **Cancer Research UK and the UK Medical Research Council.**

Lancet 2010; 375: 2143–51

Published Online June 2, 2010 DOI:10.1016/S0140-6736(10)60636-8

See Comment page 2123

See Department of Error page 2142

*Collaborators listed at end of paper

Cancer Epidemiology Unit (R C Travis DPhil, G K Reeves PhD, J Green MD, D Bull, S J Tipper MSc, K Baker, Prof V Beral FRS), **Clinical Trial Services Unit, Epidemiological Studies Unit** (Prof R Peto FRS), **and Division of Medical Sciences** (Prof J Bell FRS), **University of Oxford, Oxford, UK; and Centre National de Génotypage, IG, Commissariat à l'énergie Atomique, France** (D Zelenika PhD, Prof M Lathrop PhD)

Correspondence to: Dr Ruth Travis, Cancer Epidemiology Unit, Richard Doll Building, Roosevelt Drive, Oxford OX3 7LF, UK **ruth.travis@ceu.ox.ac.uk**

For the **Million Women Study website** see http://millionwomenstudy.org

Introduction

Genome-wide association studies,[1–6] together with analyses of specific candidate polymorphisms,[7,8] have identified several low-penetrance breast cancer susceptibility loci. Little is known about how the associated relative risks are affected by the established reproductive, behavioural, and anthropometric risk factors for breast cancer (here referred to collectively as environmental factors, although some, such as height, are in part genetically determined). Of the few results published so far,[9–13] some have suggested possible interactions between the effects of *FGFR2* variants and use of hormone replacement therapy (HRT), but findings are inconsistent.[11–13] Large-scale prospective data can help to assess any such gene–environment interactions. For 12 single nucleotide polymorphisms (SNPs) previously associated with the disease we compared relative risks for breast cancer across categories of ten established environmental risk factors[14] in a large prospective UK cohort.

Methods

Participants

During 1996–2001, 1·3 million middle-aged women (mean age 56 years [SD 5]) who had been invited for routine screening for breast cancer at 66 National Health Service (NHS) screening centres in the UK were recruited into the Million Women Study and completed a questionnaire. The study design and methods are described elsewhere[15] and are available at the Million Women Study website. Study participants have a unique NHS number, and are routinely followed up for cause-specific incident cancer and death via the NHS Central Registers. In 2005–08, women with breast cancer and randomly selected women without breast cancer were asked to participate in a genetic-susceptibility study. Ethics approval for the work was granted from the Oxford and Anglia Multi-Centre Research and Ethics Committee and the Eastern Multi-Centre Research and Ethics Committee.

Genotyping

Genotyping was done at the Centre National de Génotypage in Paris, France. Genotyping assays were designed and undertaken with use of the Taqman assay (Applied Biosystems, Foster City, CA, USA). Cases and controls were mixed for genotyping, and laboratory personnel were unaware of the case or control status of the samples. The genotyping success rate was at least 96% for each variant (range 96·3–98·8), and was 97% overall.

	Cases (n=7610)	Controls (n=10 196)	p value*
Age at menarche (years)	12·9 (1·6)	12·9 (1·5)	0·5
Parity (% parous)	6537 (86%)	9000 (88%)	<0·0001
Age at first birth in parous women (years)	24·9 (4·5)	24·6 (4·2)	0·0002
Breastfeeding (% of parous women who ever breastfed)	4036 (74%)	5626 (76%)	0·02
Menopausal status (% postmenopausal)	5906 (78%)	7789 (76%)	0·06
Age at natural menopause (years)†	49·9 (4·1)	49·5 (4·2)	<0·0001
Current use of hormone replacement therapy†	2948 (50%)	3256 (42%)	<0·0001
Body-mass index (kg/m²)	25·9 (4·4)	25·6 (4·2)	<0·0001
Height (cm)	163·1 (6·6)	162·7 (6·6)	<0·0001
Alcohol intake (g per day)	7·5 (8·2)	7·1 (7·7)	0·0005

Data are mean (SD) or number (%). Numbers do not always add up because of missing values. *For the effect of each environmental risk factor on breast cancer risk (with the factor subdivided into two groups, as described in the Methods section). †Postmenopausal women only.

Table: **Baseline characteristics of breast cancer cases and controls**

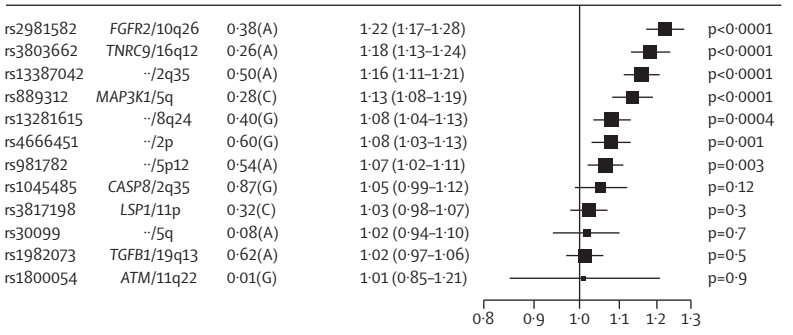


Figure 1: **Summary of findings for 12 SNPs in 7610 women with breast cancer and 10 196 randomly selected controls without breast cancer**

Statistical analysis

We examined the association between ten environmental (ie, reproductive, behavioural, or anthropometric) risk factors for breast cancer and 12 SNPs (*FGFR2*-rs2981582, *TNRC9*-rs3803662, 2q35-rs13387042, *MAP3K1*-rs889312, 8q24-rs13281615, 2p-rs4666451, 5p12-rs981782, *CASP8*-rs1045485, *LSP1*-rs3817198, 5q-rs30099, *TGFB1*-rs1982073, and *ATM*-rs1800054), and whether these environmental risk factors modified the relative risks of breast cancer associated with the SNPs. The ten environmental risk factors studied were: age at menarche, parity, age at first birth, breastfeeding, menopausal status, age at menopause, use of HRT, body-mass index, height, and alcohol consumption. Information about these factors was recorded prospectively at the time the women joined the study—ie, before the diagnosis of breast cancer in cases and at an equivalent time for controls. The effects of most of the ten factors on breast cancer incidence have already been shown to be highly significant for the entire Million Women Study cohort.[16–19]

To describe the main effect of each of the 12 SNPs on the relative risk for breast cancer, logistic regression models were applied, calculating the per-allele relative risks for the high-risk versus the low-risk allele, adjusted by age at recruitment (50–52, 53–55, 56–58, 59–61, and ≥62 years) and by ten regions in the UK (corresponding to UK cancer registration regions).

To study the associations between the genetic and environmental factors, we compared the means (for continuous variables) and proportions (for categorical variables) using ANOVA and conventional χ^2 tests of heterogeneity, respectively. Conventional p values are given but have been interpreted in view of the 120 (12 SNPs × ten other factors) comparisons made. With 120 tests, only an uncorrected p value of less than 0·0004 would be regarded as statistically significant.

For the main tests for gene–environment interaction, we compared the per-allele relative risks of breast cancer for each of the 12 SNPs across two levels of each of the ten other factors. For the continuous variables, women were divided into two groups of roughly equal size for the comparisons: age at menarche (<13 or ≥13 years); parity (nulliparous or parous); age at first birth (<25 or ≥25 years in parous women); age at menopause (<50 or ≥50 years, in postmenopausal never users of HRT); body-mass index (<25 or ≥25 kg/m²); height (<165 or ≥165 cm); and alcohol intake (<one or ≥one drink per day, as defined previously[18]). For the categorical environmental factors (menopausal status, use of HRT, and breastfeeding), women were categorised as: premenopausal/perimenopausal or postmenopausal never users of HRT; never/past or current users of HRT; and ever or never having breastfed in parous women). χ^2 tests were applied to assess heterogeneity in the per-allele relative risks of breast cancer by each environmental factor (assuming a linear relation of log risk with number of alleles). The per-allele relative risk estimates are presented together with conventional 95% CIs and p values, but are interpreted in view of the 120 tests for interaction.

For continuous variables we also tested for interaction between the value of the variable and the per-allele log risk, and we repeated the categorical analyses with three rather than two groups.

When results are presented as plots, the means, proportions, and relative risks are represented by squares (and the corresponding 95% CIs as lines), each with area inversely proportional to the variance of the logarithm of the values and thus providing an appropriate indication of the amount of statistical information included for that particular estimate. All statistical tests were two sided; analyses were done with Stata (version 10.0).

Role of the funding source

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. RCT, DB, GKR, JG, and VB had full access to all the data in the study and all authors had final responsibility for the decision to submit for publication.

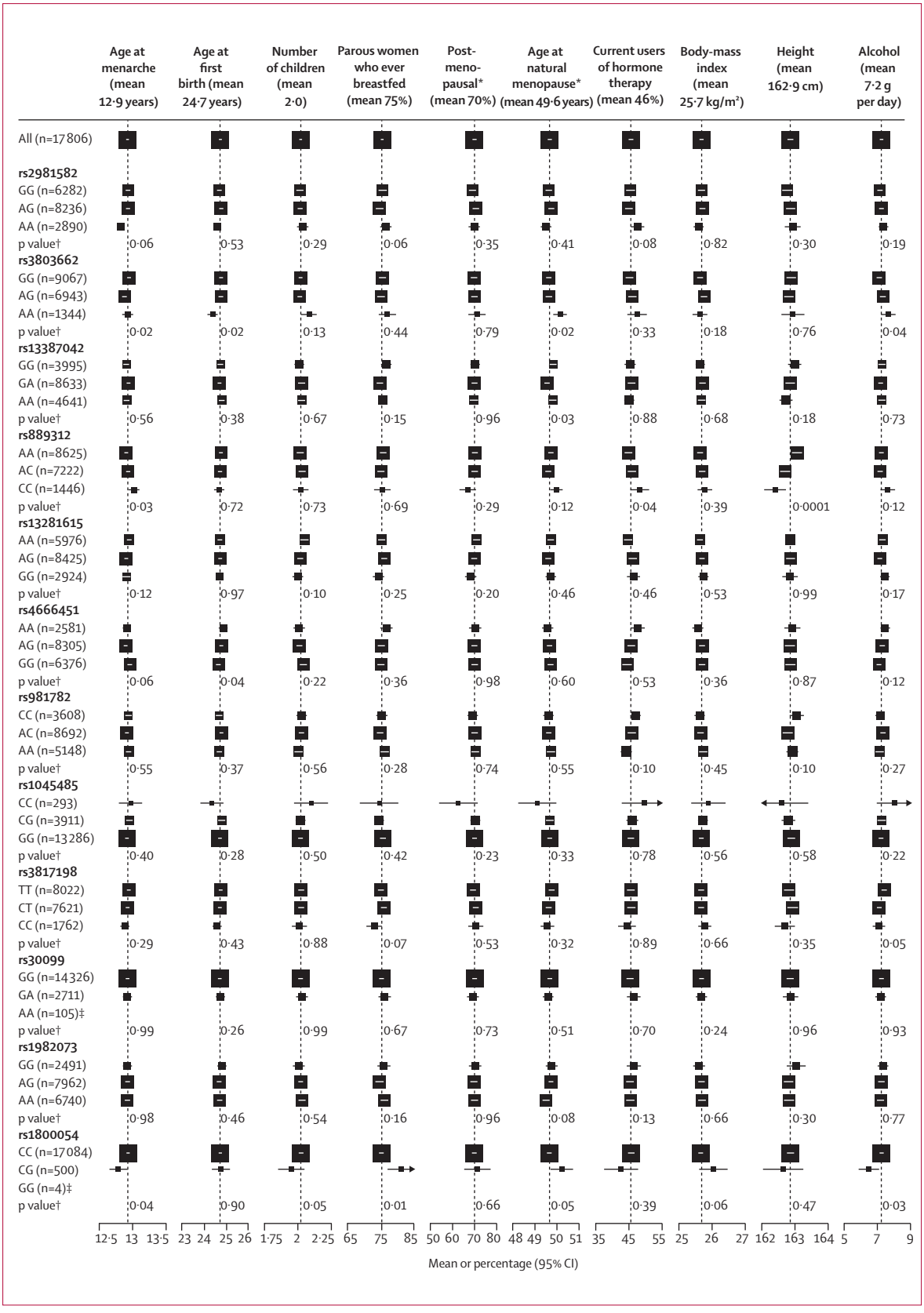
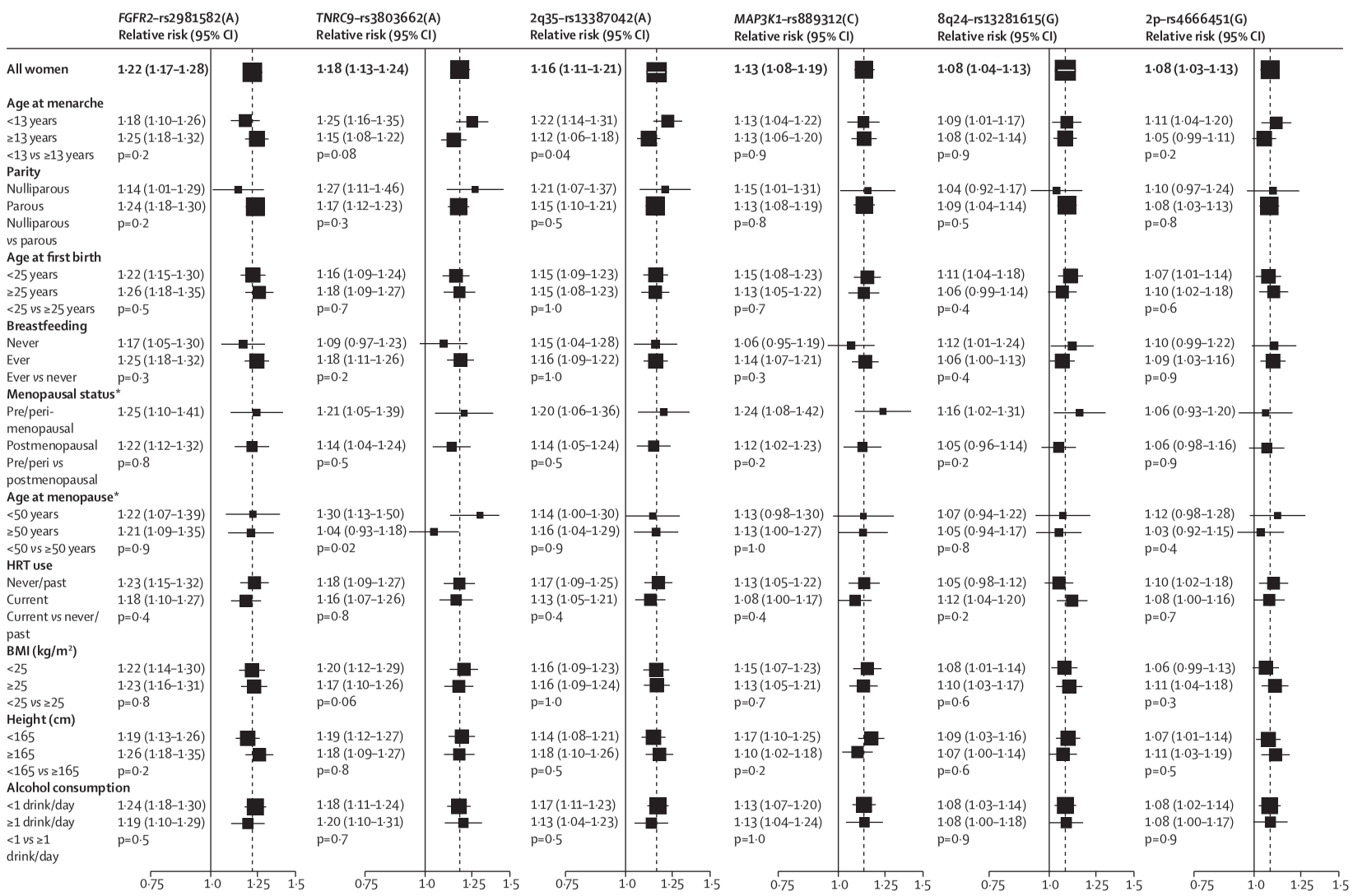


***Figure 2*: Distribution of environmental risk factors for breast cancer by genotype**

Figure shows means or proportions and 95% CIs. Numbers do not always add up because of missing values. *Never users of hormone replacement therapy. †Conventional p values are shown in the figure; all p values are non-significant after adjustment for multiple testing, apart from that for rs889132 by height, when adjusted p=0·01. ‡Insufficient data.


FGFR2–rs2981582(A) Relative risk (95% CI)
TNRC9–rs3803662(A) Relative risk (95% CI)
2q35–rs13387042(A) Relative risk (95% CI)
MAP3K1–rs889312(C) Relative risk (95% CI)
8q24–rs13281615(G) Relative risk (95% CI)
2p–rs4666451(G) Relative risk (95% CI)
All women 1·22 (1·17–1·28) 1·18 (1·13–1·24) 1·16 (1·11–1·21) 1·13 (1·08–1·19) 1·08 (1·04–1·13) 1·08 (1·03–1·13)
Age at menarche
<13 years 1·18 (1·10–1·26) 1·25 (1·16–1·35) 1·22 (1·14–1·31) 1·13 (1·04–1·22) 1·09 (1·01–1·17) 1·11 (1·04–1·20)
≥13 years 1·25 (1·18–1·32) 1·15 (1·08–1·22) 1·12 (1·06–1·18) 1·13 (1·06–1·20) 1·08 (1·02–1·14) 1·05 (0·99–1·11)
<13 vs ≥13 years p=0·2 p=0·08 p=0·04 p=0·9 p=0·9 p=0·2
Parity
Nulliparous 1·14 (1·01–1·29) 1·27 (1·11–1·46) 1·21 (1·07–1·37) 1·15 (1·01–1·31) 1·04 (0·92–1·17) 1·10 (0·97–1·24)
Parous 1·24 (1·18–1·30) 1·17 (1·12–1·23) 1·15 (1·10–1·21) 1·13 (1·08–1·19) 1·09 (1·04–1·14) 1·08 (1·03–1·13)
Nulliparous vs parous p=0·2 p=0·3 p=0·5 p=0·8 p=0·5 p=0·8
Age at first birth
<25 years 1·22 (1·15–1·30) 1·16 (1·09–1·24) 1·15 (1·09–1·23) 1·15 (1·08–1·23) 1·11 (1·04–1·18) 1·07 (1·01–1·14)
≥25 years 1·26 (1·18–1·35) 1·18 (1·09–1·27) 1·15 (1·08–1·23) 1·13 (1·05–1·22) 1·06 (0·99–1·14) 1·10 (1·02–1·18)
<25 vs ≥25 years p=0·5 p=0·7 p=1·0 p=0·7 p=0·4 p=0·6
Breastfeeding
Never 1·17 (1·05–1·30) 1·09 (0·97–1·23) 1·15 (1·04–1·28) 1·06 (0·95–1·19) 1·12 (1·01–1·24) 1·10 (0·99–1·22)
Ever 1·25 (1·18–1·32) 1·18 (1·11–1·26) 1·16 (1·09–1·22) 1·14 (1·07–1·21) 1·06 (1·00–1·13) 1·09 (1·03–1·16)
Ever vs never p=0·3 p=0·2 p=1·0 p=0·3 p=0·4 p=0·9
Menopausal status*
Pre/peri-menopausal 1·25 (1·10–1·41) 1·21 (1·05–1·39) 1·20 (1·06–1·36) 1·24 (1·08–1·42) 1·16 (1·02–1·31) 1·06 (0·93–1·20)
Postmenopausal 1·22 (1·12–1·32) 1·14 (1·04–1·24) 1·14 (1·05–1·24) 1·12 (1·02–1·23) 1·05 (0·96–1·14) 1·06 (0·98–1·16)
Pre/peri vs postmenopausal p=0·8 p=0·5 p=0·5 p=0·2 p=0·2 p=0·9
Age at menopause*
<50 years 1·22 (1·07–1·39) 1·30 (1·13–1·50) 1·14 (1·00–1·30) 1·13 (0·98–1·30) 1·07 (0·94–1·22) 1·12 (0·98–1·28)
≥50 years 1·21 (1·09–1·35) 1·04 (0·93–1·18) 1·16 (1·04–1·29) 1·13 (1·00–1·27) 1·05 (0·94–1·17) 1·03 (0·92–1·15)
<50 vs ≥50 years p=0·9 p=0·02 p=0·9 p=1·0 p=0·8 p=0·4
HRT use
Never/past 1·23 (1·15–1·32) 1·18 (1·09–1·27) 1·17 (1·09–1·25) 1·13 (1·05–1·22) 1·05 (0·98–1·12) 1·10 (1·02–1·18)
Current 1·18 (1·10–1·27) 1·16 (1·07–1·26) 1·13 (1·05–1·21) 1·08 (1·00–1·17) 1·12 (1·04–1·20) 1·08 (1·00–1·16)
Current vs never/past p=0·4 p=0·8 p=0·4 p=0·4 p=0·2 p=0·7
BMI (kg/m²)
<25 1·22 (1·14–1·30) 1·20 (1·12–1·29) 1·16 (1·09–1·23) 1·15 (1·07–1·23) 1·08 (1·01–1·14) 1·06 (0·99–1·13)
≥25 1·23 (1·16–1·31) 1·17 (1·10–1·26) 1·16 (1·09–1·24) 1·13 (1·05–1·21) 1·10 (1·03–1·17) 1·11 (1·04–1·18)
<25 vs ≥25 p=0·8 p=0·06 p=1·0 p=0·7 p=0·6 p=0·3
Height (cm)
<165 1·19 (1·13–1·26) 1·19 (1·12–1·27) 1·14 (1·08–1·21) 1·17 (1·10–1·25) 1·09 (1·03–1·16) 1·07 (1·01–1·14)
≥165 1·26 (1·18–1·35) 1·18 (1·09–1·27) 1·18 (1·10–1·26) 1·10 (1·02–1·18) 1·07 (1·00–1·14) 1·11 (1·03–1·19)
<165 vs ≥165 p=0·2 p=0·8 p=0·5 p=0·2 p=0·6 p=0·5
Alcohol consumption
<1 drink/day 1·24 (1·18–1·30) 1·18 (1·11–1·24) 1·17 (1·11–1·23) 1·13 (1·07–1·20) 1·08 (1·03–1·14) 1·08 (1·02–1·14)
≥1 drink/day 1·19 (1·10–1·29) 1·20 (1·10–1·31) 1·13 (1·04–1·23) 1·13 (1·04–1·24) 1·08 (1·00–1·18) 1·08 (1·00–1·17)
<1 vs ≥1 drink/day p=0·5 p=0·7 p=0·5 p=1·0 p=0·9 p=0·9
0·75 1·0 1·25 1·5

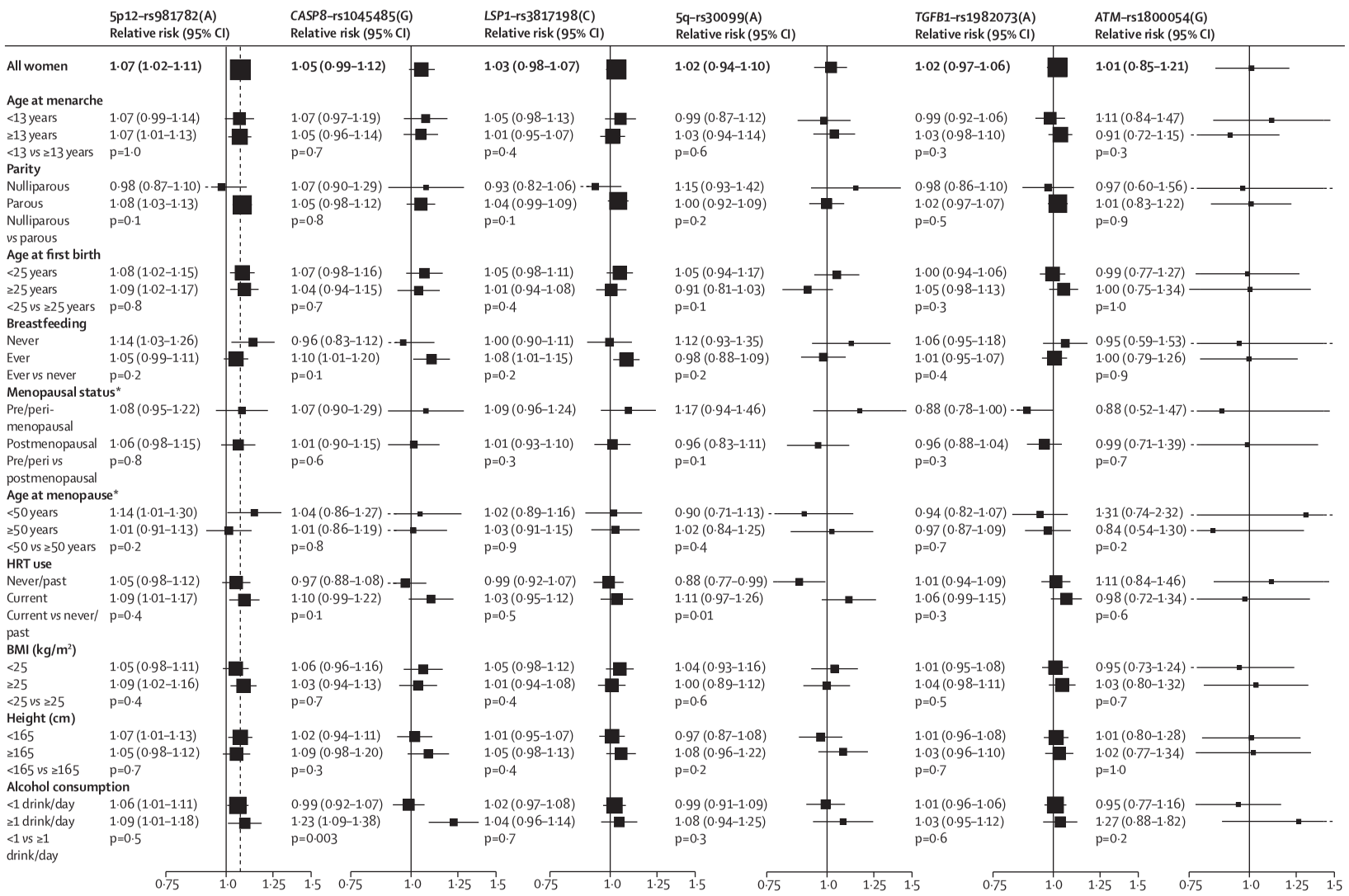


***Figure 3*: Per-allele relative risk (95% CI) of breast cancer by ten environmental risk factors for breast cancer**
Conventional p values are shown in the figure; all p values are non-significant after adjustment for multiple testing. HRT=hormone replacement therapy in postmenopausal women. *Never users of HRT.

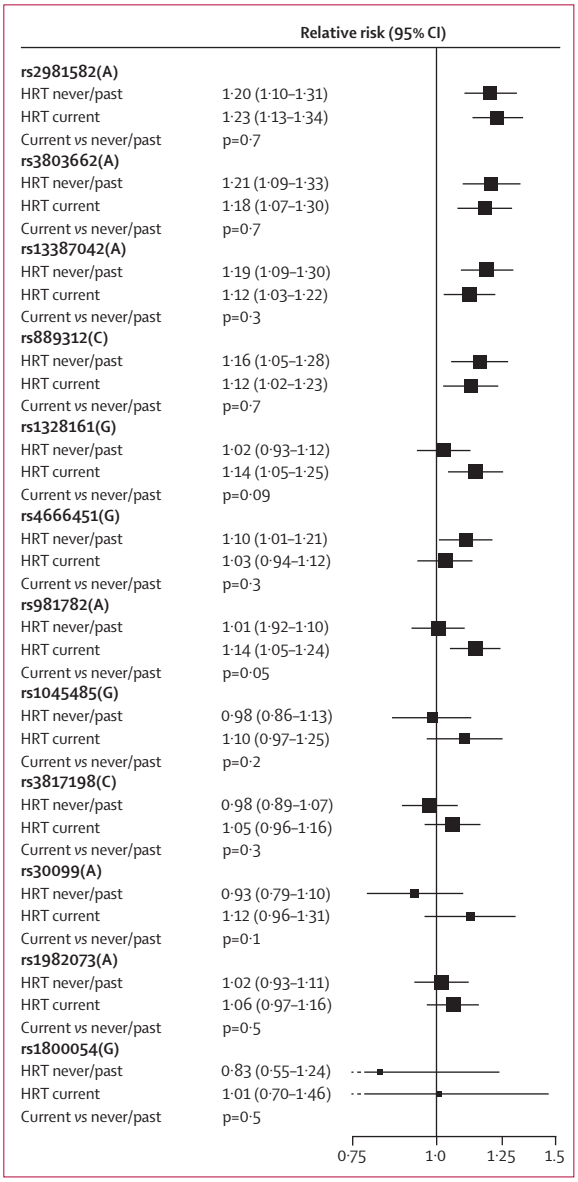


Figure 4: **Per-allele relative risk (95% CI) of oestrogen-receptor-positive breast cancer by use of hormone replacement therapy in postmenopausal women**

Conventional p values are shown in the figure; all p values are non-significant after adjustment for multiple testing.

Results

7610 women with incident breast cancer and 10 196 women without breast cancer were included in these analyses. The mean age of cases at cancer diagnosis was 60 years (SD 5·1). The table shows prospectively collected information about the characteristics of cases and controls, recorded an average of 4 years (SD 2·3) before cancer diagnosis for cases and at an equivalent time for controls. As expected from previously published results in this cohort,[16–19] cases were, on average, less likely to be parous, were older at first birth, had a higher body-mass index, consumed more alcohol, and were more likely to be current users of HRT than controls.

Figure 1 shows the high-risk allele frequency and the corresponding per-allele relative risks of breast cancer for each of the 12 SNPs. Results are arranged in order of magnitude of the per-allele relative risk, with the largest recorded for *FGFR2*-rs2981582, followed by *TNRC9*-rs3803662, 2q35-rs13387042, *MAP3K1*-rs889312, 8q24-rs13281615, 2p-rs4666451, 5p12-rs981782, and *CASP8*-rs1045485. The allele and genotype frequencies for each of the 12 loci in controls were similar to those reported for populations of European descent,[1,3,7] and with no substantial deviations from Hardy-Weinberg proportions in the controls (p>0·1 for 11 comparisons and p=0·04 for one comparison).

We examined whether any of the ten environmental risk factors for breast cancer were related to any of the 12 SNPs. Figure 2 shows the results, with SNPs ordered by magnitude of the per-allele relative risks. After allowance for multiple testing, there were no significant associations between 11 of the 12 polymorphisms and the ten environmental risk factors (figure 2). For the 12th polymorphism *MAP3K1*-rs889312, there was a significant association between the allele variants and height (p=0·0001, and after allowance for multiple tests with the Bonferroni correction, p=0·01). Homozygous carriers of the C allele, which is associated with an increased risk of breast cancer, were slightly shorter than were homozygous carriers of the AA allele (mean heights 162·4 cm [95% CI 162·1–162·7] and 163·1 cm [162·9–163·2], respectively) with heterozygous AC carriers being of intermediate height (162·7 cm [162·6–162·9]).

Figure 3 shows the per-allele relative risks of breast cancer for each of the 12 polymorphisms by the ten environmental factors. Of the 120 separate statistical tests for gene–environment interaction only four yielded a p value less than 0·05, and none remained significant after allowance for multiple comparisons with the Bonferroni correction (figure 3).

Of the multiple comparisons done, the strongest suggestion of a gene–environment interaction was between *CASP8*-rs1045485 and alcohol consumption; the per-allele relative risk of breast cancer for the common variant of *CASP8*-rs1045485 (ie, the high-risk G variant) was significantly greater in women who reported consuming one or more alcoholic drinks per day than in those consuming less alcohol (figure 3); but after allowance for multiple testing this result was not significant. For three other interactions, conventional p values were of borderline significance: *TNRC9* SNP rs3803662 for age at menopause ($p_{heterogeneity}$=0·02); 2q35-rs13387042 for age at menarche ($p_{heterogeneity}$=0·04); and 5q-rs30099 for use of HRT ($p_{heterogeneity}$=0·01). However, in view of the large numbers of tests done these could well be chance findings.

The findings did not differ materially when women were subdivided into three rather than two groups on the basis of the environmental risk factors (webappendix p 1), when the environmental variables were treated as continuous, or when results for homozygous and heterozygous carriers of each SNP were examined separately (webappendix p 2).

Current use of HRT (but not past use) is associated with an increased risk of breast cancer in this cohort,[16] and current use has a far greater effect on tumours that were oestrogen-receptor-positive (relative risk 2·44 [95% CI 2·27–2·63]) than those that were oestrogen-receptor-negative (1·32 [1·14–1·53]) ($p_{heterogeneity}$<0·0001). Furthermore, some of the SNPs are more strongly related to oestrogen-receptor-positive than to oestrogen-receptor-negative disease—eg, for *FGFR2*-rs2981582 the respective per-allele relative risks were 1·27 (1·21–1·34) and 1·01 (0·92–1·12) ($p_{heterogeneity}$<0·0001). Although there was no strong evidence for gene–environment interaction by use of HRT on all types of breast cancer (figure 3), we also studied oestrogen-receptor-positive disease separately. In a comparison of the per-allele relative risks for oestrogen-receptor-positive breast cancer there was again little evidence of heterogeneity by HRT use after allowance for multiple testing (figure 4). Nor were there any significant interactions for oestrogen-receptor-positive disease or for all breast cancer when the current users were further subdivided by the type of HRT used—eg, for *FGFR2* the per-allele relative risks associated with oestrogen-receptor-positive breast cancer were 1·15 (0·98–1·35) in current users of oestrogen-only HRT and 1·25 (1·12–1·40) in current users of oestrogen-progestin HRT (tests for interaction in current *vs* never/past users p=0·6 and p=0·6, respectively). The corresponding results by HRT type for all breast cancer were 1·09 (0·98–1·25) and 1·22 (1·11–1·34), respectively (tests for interaction, p=0·1 and p=0·9, respectively). Numbers are too small to study gene–environment interactions for oestrogen-receptor-negative breast cancer, since only about 20% of breast cancers are of this type.

A possible interaction between *FGFR2* and birthweight was reported in a study of 693 women with breast cancer,[20] but there was no evidence of this in our data (per-allele relative risks of 1·21 [1·13–1·30] for birthweight <3·5 kg and 1·24 [1·12–1·38] for ≥3·5 kg; p=0·9).

Results were similar when the small number of women reporting non-white ethnic origin were excluded (28 cases [0·4%] and 51 controls [0·5%]; data not shown).

Discussion

In this large-scale and systematic examination of 120 possible gene–environment interactions in 7610 women with breast cancer, only four were conventionally statistically significant (whereas six would be expected by chance alone) and none remained significant after appropriate allowance for the multiple tests that were done. Furthermore no strong relations between genotype and the environmental risk factors studied were detected, suggesting that the low-penetrance susceptibility loci investigated here do not generally affect breast cancer risk through mechanisms involving these environmental factors.

See Online for webappendix

The term gene–environment interactions is so widely used that we have retained it, even though several of the ten environmental risk factors for breast cancer (eg, height and body-mass index) have both environmental and genetic determinants. Selection of the ten environmental factors and of the 12 SNPs was based on reviews, meta-analyses, and results from genome-wide association studies published up to March, 2008 (when genotyping began).[1,3,7,14] Since then an additional five genetic susceptibility regions have been identified, although the associated relative risks for breast cancer are not large.[4–6] The only major environmental risk factor not studied here is postmenopausal hormone concentrations in blood,[14,21] because these were not measured in this cohort.

The magnitude of the genotypic relative risks for breast cancer (shown in figure 1) accord with other published results,[1,3,7,9] with the largest risks noted for *FGFR2*-rs2981582 and *TNRC9*-rs3803662 (relative risks of about 1·2 per allele) and 2q35-rs13387042 and *MAP3K1*-rs889312 (relative risks of about 1·15 per allele). The four previous studies examining gene–environment interactions have been comparatively small (including 1049, 456, 1749, and 685 women with breast cancer[10–13]) and have generally examined only two loci (*FGFR2*[10–13] and *MAP3K1*[13]). In our study there was more than 90% power, after allowing for multiple testing, to detect an interaction of at least 1·5 (on a multiplicative scale) between each of the risk factors and polymorphisms with a prevalence of at least 25% and a per-allele relative risk of at least 1·1. Thus, although there is sufficient power to assess reliably gene–environment interactions in the four or so SNPs most strongly associated with breast cancer, power to detect corresponding interactions with rarer, or low-risk, genetic variants is limited.

Most of the environmental factors are more strongly associated with breast cancer risk than are the genotypic factors that we studied. Current use of HRT is the strongest environmental risk factor in this cohort, agreement between the reported use of HRT and prescription data is excellent,[22] and the associated relative risks are much greater than for any of the SNPs studied here. There is therefore greater statistical power to assess the gene–environment interaction for HRT use than for the other exposures. Nevertheless, there was little evidence for interactions between genotype, HRT use, and breast cancer risk. HRT has a substantially greater effect on oestrogen-receptor-positive than on oestrogen-receptor-negative disease. However, even when analyses were restricted to oestrogen-receptor-

positive disease there was still little evidence for interactions (figure 4). The *FGFR2* gene is more strongly related to oestrogen-receptor-positive than to oestrogen-receptor-negative disease and is the locus most commonly investigated by others with respect to use of HRT.[11–13] The numbers of cases in previous studies were comparatively small, however, and a formal meta-analysis of published results is precluded by the incompatible ways in which they have been presented.

For factors other than use of HRT, we did not replicate any of the reported gene–environment interactions.[10,11,13] For *FGFR2*, there has been one previous report of interactions with parity[11] and with age at menarche,[11] but these were not reproduced by us or by another study.[13]

Although there was no convincing evidence for gene–environment interactions, we did find that homozygous carriers of the high-risk C allele *MAP3K1*-rs889312 were on average 0·7 cm shorter than were homozygous carriers of the lower-risk A allele. The encoded enzyme is a serine/threonine kinase, which is implicated in early development and growth,[23] but the association is in the opposite direction of that expected if this gene affected breast cancer risk via its effect on height. Taller women are at a slightly increased risk of breast cancer than are shorter women, but carriers of the high-risk allele are slightly shorter than are carriers of the low-risk allelles. Although a possible association between rs2981582 and rs3803662 genotypes and duration of breastfeeding has been reported previously,[9] we found little evidence for such effects.

Most of the women that we studied were postmenopausal, and we have no useful information about breast cancer risk in younger women. However, meta-analysis of the worldwide data for women whose breast cancer was diagnosed before age 50 years[24] showed little variation in familial risk by environmental risk factors.

Many statistical tests were done and, because there were few a-priori hypotheses, our findings are largely hypothesis generating. The function of most of the SNPs is unclear and future investigations might show them to be only markers of the true causal variants. Although our study is fairly large, with prospectively collected exposure data from one cohort, it still lacked power to assess moderate gene–environment interactions for all but the four or so SNPs most strongly related to breast cancer risk or to investigate interactions separately for causally distinct types of breast cancer, such as those with and without oestrogen receptors. Whereas thousands of cases were needed to characterise the main effects of environmental factors and of genetic factors on breast cancer risk, tens of thousands of cases will be needed to assess reliably a comprehensive range of biologically plausible gene–environment interactions.

Contributors

All authors contributed to the design and execution of this work, and to the preparation of the report. All authors had an opportunity to contribute to the interpretation of the results and to the redrafting of the report, and all authors approved the final report.

The Million Women Study Collaborators

Million Women Study Steering Committee Emily Banks, Valerie Beral, Ruth English, Jane Green, Julietta Patnick, Richard Peto, Gillian Reeves, Martin Vessey, Matthew Wallis. *NHS Breast Screening Centres collaborating in the Million Women Study (in alphabetical order)* Avon, Aylesbury, Barnsley, Basingstoke, Bedfordshire & Hertfordshire, Cambridge & Huntingdon, Chelmsford & Colchester, Chester, Cornwall, Crewe, Cumbria, Doncaster, Dorset, East Berkshire, East Cheshire, East Devon, East of Scotland, East Suffolk, East Sussex, Gateshead, Gloucestershire, Great Yarmouth, Hereford & Worcester, Kent (Canterbury, Rochester, Maidstone), Kings Lynn, Leicestershire, Liverpool, Manchester, Milton Keynes, Newcastle, North Birmingham, North East Scotland, North Lancashire, North Middlesex, North Nottingham, North of Scotland, North Tees, North Yorkshire, Nottingham, Oxford, Portsmouth, Rotherham, Sheffield, Shropshire, Somerset, South Birmingham, South East Scotland, South East Staffordshire, South Derbyshire, South Essex, South Lancashire, South West Scotland, Surrey, Warrington Halton St Helens & Knowsley, Warwickshire Solihull & Coventry, West Berkshire, West Devon, West London, West Suffolk, West Sussex, Wiltshire, Winchester, Wirral, and Wycombe. *Million Women Study Co-ordinating Centre staff* Simon Abbott, Naomi Allen, Miranda Armstrong, Krys Baker, Angela Balkwill, Vicky Benson, Valerie Beral, Judith Black, Anna Brown, Diana Bull, Benjamin Cairns, Andrew Chadwick, James Chivenga, Barbara Crossley, Francesca Crowe, Dave Ewart, Sarah Ewart, Lee Fletcher, Laura Gerrard, Adrian Goodill, Bryony Horner, Isobel Lingard, Jane Green, Winifred Gray, Joy Hooley, Sau Wan Kan, Carol Keene, Nicky Langston, Maria Jose Luque, Kath Moser, Lynn Pank, Kirstin Pirie, Gillian Reeves, Emma Sherman, Evelyn Sherry-Starmer, Moya Simmonds, Helena Strange, Sian Sweetland, Alison Timadjer, Sarah Tipper, Lyndsey Trickett, Ruth Travis, Joanna Watson, Steve Williams, Lucy Wright.

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

We thank all the women who participated in the study, the general practitioners who took blood samples, staff from the NHS Breast Screening Centres and the Cancer Epidemiology Unit, and Adrian Goodill for the preparation of the figures. The Million Women Study is supported by Cancer Research UK and the UK Medical Research Council.

References

1 Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007; **447:** 1087–93.

2 Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 2007; **39:** 870–74.

3 Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2007; **39:** 865–69.

4 Ahmed S, Thomas G, Ghoussaini M, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 2009; **41:** 585–90.

5 Thomas G, Jacobs KB, Kraft P, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* 2009; **41:** 579–84.

6 Zheng W, Long J, Gao YT, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 2009; **41:** 324–28.

7 Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* 2007; **39:** 352–58.

8 Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2008; **40:** 703–06.

9 Huijts PE, Vreeswijk MP, Kroeze-Jansema KH, et al. Clinical correlates of low-risk variants in FGFR2, TNRC9, MAP3K1, LSP1 and 8q24 in a Dutch cohort of incident breast cancer cases. *Breast Cancer Res* 2007; **9:** R78.

10 Liang J, Chen P, Hu Z, et al. Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. *Carcinogenesis* 2008; **29:** 2341–46.

11 Kawase T, Matsuo K, Suzuki T, et al. FGFR2 intronic polymorphisms interact with reproductive risk factors of breast cancer: results of a case control study in Japan. *Int J Cancer* 2009; **125:** 1946–52.

12 Prentice RL, Huang Y, Hinds DA, et al. Variation in the FGFR2 gene and the effects of postmenopausal hormone therapy on invasive breast cancer. *Cancer Epidemiol Biomarkers Prev* 2009; **18:** 3079–85.

13 Rebbeck TR, DeMichele A, Tran TV, et al. Hormone-dependent effects of FGFR2 and MAP3K1 in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. *Carcinogenesis* 2009; **30:** 269–74.

14 Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *Lancet Oncol* 2001; **2:** 133–40.

15 The Million Women Study Collaborative Group. The Million Women Study: design and characteristics of the study population. *Breast Cancer Res* 1999; **1:** 73–80.

16 The Million Women Study Collaborators. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* 2003; **362:** 419–27.

17 Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ* 2007; **335:** 1134.

18 Allen NE, Beral V, Casabonne D, et al. Moderate alcohol intake and cancer incidence in women. *J Natl Cancer Inst* 2009; **101:** 296–305.

19 Reeves GK, Pirie K, Green J, Bull D, Beral V. Reproductive factors and specific histological types of breast cancer: prospective study and meta-analysis. *Br J Cancer* 2009; **100:** 538–44.

20 Tamimi RM, Lagiou P, Czene K, et al. Birth weight, breast cancer susceptibility loci, and breast cancer risk. *Cancer Causes Control*. 2010; published online Jan 7. DOI:10.1007/S10552-009-9446-7.

21 Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst* 2002; **94:** 606–16.

22 Banks E, Beral V, Cameron R, et al. Agreement between general practice prescription data and self-reported use of hormone replacement therapy and treatment for various illnesses. *J Epidemiol Biostat* 2001; **6:** 357–63.

23 Craig EA, Stevens MV, Vaillancourt RR, Camenisch TD. MAP3Ks as central regulators of cell fate during development. *Dev Dyn* 2008; **237:** 3102–14.

24 The Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *Lancet* 2001; **358:** 1389–99.