

Value of Phenotypic and Single-Nucleotide Polymorphism Panel Markers in Predicting the Risk of Breast Cancer

Adam R Brentnall^{1*}, D Gareth Evans² and Jack Cuzick¹

¹Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK
²Genesis Breast Cancer Prevention Centre, University Hospital of South Manchester NHS Trust, Wythenshawe, Manchester M23 9LT, UK

Abstract

The risk of breast cancer from a number of SNPs (Single-Nucleotide Polymorphisms) has recently been estimated singly by COGS (Collaborative Oncological Gene-Environment Study). We assessed how the predicted risk from a panel of SNPs would compare with classical phenotypic factors including age, family history and parity, and how much it might add to risk assessment. The analysis was based on prospective data from ten thousand women of routine screening age enrolled into the UK Predicting Risk of Breast Cancer at Screening (PROCAS) study, and computer simulation SNP scores. We found that the current panel of 67 SNPs was less able to identify high-risk women than classical phenotypic factors, but if they can be treated independently, then in combination a substantially increased predictive effect might be seen. The proportion of women in the PROCAS cohort with a 10-year risk of more than 8% increased from 0.5% using age and the SNP67 score; to 1.1% using the phenotypic factors in the Tyrer-Cuzick model; to 3.3% when combined.

Keywords: Breast cancer; Risk; Single-nucleotide polymorphisms

Abbreviations: COGS: Collaborative Oncological Gene-Environment Study; NICE: National Institute for Health and Clinical Excellence; PROCAS: Predicting Risk of Breast Cancer at Screening Study; SNP: Single-Nucleotide Polymorphism; TC: Tyrer-Cuzick; UK: United Kingdom

Introduction

Breast cancer is the most common form of cancer affecting women. It is estimated that in the UK approximately one in eight women will develop the disease in their lifetime; in 2010 almost 50,000 women were diagnosed with invasive breast cancer and just over 11,500 died of it [1]. Thus there is a need to predict which women will develop the disease, and to apply measures to prevent it.

A wide body of research has focused on phenotypic breast cancer risk factors including age, family history, reproductive history and benign breast disease. The Tyrer-Cuzick (TC) risk evaluator uses family histories of breast and ovarian cancer in conjunction with personal factors such as parity, menopausal status and weight, to estimate 10-year risk through a single statistical model [2]. The performance of the model has been examined in different settings, and it is being used to assess the risk of all women recruited into the PROCAS study (predicting risk at breast cancer screening) in Manchester, UK [3-5].

A recent development has been the identification of SNPs associated with breast cancer risk, each with a small relative risk [6]. The objective of this article is compare how the risk attributable to a panel of these SNPs compares with that from classical phenotypic factors, when applied to a cohort of women from the UK screening program.

Material and Methods

The analysis was based on ten thousand women prospectively recruited into the PROCAS study (predicting risk at breast screening) in Manchester, UK. Each woman completed a questionnaire at entry to the study with information on all phenotypic factors used by the Tyrer-Cuzick risk evaluator (version 6.0). A full description of these women has been given elsewhere [5].

The primary outcome was the 10-year risk of developing breast

cancer. This was estimated for phenotypic factors through the TC model and for the SNP panel by multiplying the relative risk by the same age-specific rates used in the TC model.

A polygenic score was used to provide an overall relative risk from SNPs. For a single woman with known genotypes, each SNP i has an estimated odds ratio R_i for a risk allele with frequency M_i . There are three genotypes for each SNP with population frequencies assumed to be from Hardy-Weinberg equilibrium $M_{i1}=M_i^2$, $M_{i2}=(1-M_i)^2$ and $M_{i3}=1-M_{i1}-M_{i2}$. A normalised risk S_{ij} relative to the population for genotype $j=1, 2, 3$ was defined so that $\sum_{j=1,2,3} M_{ij} S_{ij}=1$. The polygenic risk score for a woman was the product of their genotype normalised risks.

To assess predicted risk distributions SNP genotypes were simulated independently. The odds ratios and population allele frequencies were taken from the recent COGS (Collaborative Oncological Gene-Environment Study) analysis and for comparison, earlier estimates of the first 18 SNPs [6,7]. 100 000 simulation replicates were used to assess SNP score distributions from all 67 SNPs and the most recent COGS data; and the first 18 SNPs with both the COGS and earlier estimates. Additionally, saliva samples were taken from 478 participants in the cohort, and the genotypes of SNPs in all 18 loci identified by [7] and given in Table 1 were tested as reported by [5]. The 10, 25, 50, 75 and 90% percentile points of phenotypic components of the TC model in the cohort were tabulated alongside risk conferred. The hypothesis that all SNP genotypes are independent was assessed by applying Fisher's method using p -values from pairwise Spearman rank correlation

***Corresponding author:** Adam R Brentnall, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK, Tel: +44 20 7882 3531; Fax: +44 20 7882 3890; E-mail: a.brentnall@qmul.ac.uk

Received October 26, 2013; **Accepted** November 27, 2013; **Published** December 06, 2013

Citation: Brentnall AR, Evans DG, Cuzick J (2013) Value of Phenotypic and Single-Nucleotide Polymorphism Panel Markers in Predicting the Risk of Breast Cancer. J Genet Syndr Gene Ther 4: 202. doi:10.4172/2157-7412.1000202

Copyright: © 2013 Brentnall AR, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

coefficients. Spearman correlation was calculated for 10-year TC risk and the PROCAS SNP score [5].

The SNP score was combined with the phenotypic factors by treating the TC model and SNP score as independent. The COGS and earlier risk estimates for the first 18 SNPs to be discovered were plotted against each other, and histograms were used to compare the predicted risk distributions.

Results

Table 2 shows the distribution of risk factors used by the TC model in the cohort, and their range of risk.

The hypothesis that the 18 SNP genotypes were uncorrelated was not rejected ($\chi^2_{306}=336.7$, $P=0.11$) in the 478 PROCAS women. A

Spearman correlation coefficient between the PROCAS SNP score and TC 10-year risk was -0.04 ($P=0.41$).

Figure 1a compares the spread of risk from the COGS and initial estimates from the first 18 SNPs. The log SNP score distribution is approximately normally distributed as expected from the central limit theorem; the estimated standard deviation of the log score was 0.43 for the SNP18 Turnbull score, 0.32 for the SNP18 COGS score and 0.44 for SNP67. The reason for the difference between old and new SNP18 risk distributions is shown by the estimated odds ratios for SNPs in Figure 1b, and is due to regression to the mean (see discussion).

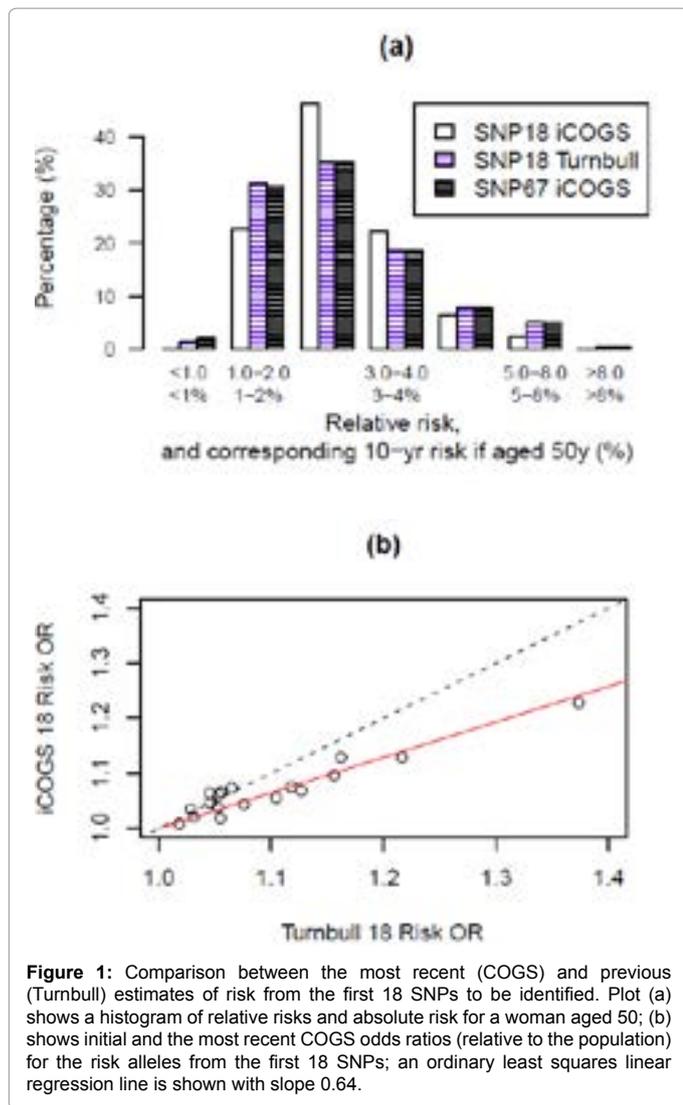
Figure 2 shows histograms of 10-year risk in the cohort. Age is an important risk factor and so it is included for comparison. The histograms show that SNP67 was less able to discriminate high-risk

SNP	Locus / Gene	Chm	Ma/Min Allele	MAF1 (%)	MAF2 (%)	RR1	RR2	Encodes / function
rs1562430	8q24	8	T/C	42%	-	0.85	-	-
rs909116	LSP1	11	T/C	47%	-	0.85	-	Intracellular F-actin binding protein
rs10995190	ZNF365	10	G/A	15%	16%	0.86	0.86	Zinc finger protein 365
rs1156287	COX11	17	A/G	29%	-	0.91	-	Catalyzes the electron transfer from reduced cytochrome c to oxygen
rs2380205*	10q	10	C/T	46%	-	0.94	-	On block with ANKRD16 (encoding ankyrin repeat domain 16) and FBXO18 (encoding the F-box protein, helicase 18) [7]
rs704010	ZMIZ1	10	G/A	39%	38%	1.07	1.08	Member of the PIAS (protein inhibitor of activated STAT) family
rs11249433	1p11.2	1	A/G	42%	40%	1.08	1.09	-
rs1011970	CDKN2A	9	G/T	17%	17%	1.09	1.06	Cyclin-dependent kinase inhibitors
rs9790879	5p12	5	T/C	40%	-	1.10	-	-
rs10931936	CASP8	2	C/T	26%	-	1.14	-	Member of the cysteine-aspartic acid protease (caspase) family
rs8009944	RAD51L1	14	A/C	25%	-	1.14	-	Member of the RAD51 family
rs614367	11q13	11	C/T	15%	15%	1.15	1.21	Plausible flanking genes [7]: MYEOV, CCND1, ORAOV1, FGF19, FGF4 and FGF3
rs4973768	SLC4A7	3	C/T	47%	47%	1.16	1.10	A sodium bicarbonate co-transporter
rs13387042	2q35	2	A/G	49%	51%	1.21	1.14	-
rs889312	MAP3K	5	A/C	28%	28%	1.22	1.12	A serine / threonine kinase
rs3757318	ESR1	6	G/A	7%	7%	1.30	1.16	An estrogen receptor
rs3803662	TOX3	16	G/A	26%	26%	1.30	1.24	Protein containing HMG-box
rs2981579	FGFR2	10	G/A	42%	40%	1.43	1.27	Member of the fibroblast growth factor receptor family

Table 1: Summary of the 18 SNPs genotyped in PROCAS. MAF1 and RR1 are the minor allele frequency and minor allele odds ratio from [7], MAF2 and RR2 are the same but from the most recent COGS estimates [6], when available; Chm is the Chromosome number. Gene function information was taken from the National Center for Biotechnology Information (NCBI) database called Entrez Gene. *rs713588 was used as a proxy for rs2380205.

	Number (% in 10000)	Percentile				
		10%	25%	50%	75%	90%
Enrolment (age)	10000 (100%)	50	52	57	64	68
10-year risk (%)		2.6%	2.7%	2.7%	2.8%	3.0%
Menarche (age)	9816 (98%)	11	12	13	14	15
Relative risk		1.08	1.08	1.01	0.99	0.94
Nulliparous (yes)	1333 (13%)			-		
Relative risk				1.28		
Parous (age first)	8619 (86%)	18	20	23	27	31
Relative risk		0.86	0.95	0.95	1.13	1.33
Post / peri-menopause (age)	5867 (59%)	42	46	50	52	55
Relative risk		0.80	0.93	1.06	1.06	1.20
Pre-menopause (age)	917 (9%)	47	48	50	51	56
Relative risk		1.08	1.08	1.08	1.09	1.20
Height (m)	9049 (90%)	1.55	1.57	1.63	1.65	1.70
Relative risk		0.91	0.91	1.01	1.05	1.24
BMI (kg/m²)	7441 (74%)	21.3	23.3	26.0	29.6	34.0
Relative risk (post menopause)		0.97	0.97	1.08	1.13	1.13
Affected mother and/or sister (min age)	378 (4%)	40	47	55	67	74
RR from mother when aged 50		1.8	1.8	1.7	1.7	1.7

Table 2: Summary of phenotypic risk factors in the cohort.



women than the TC model. However, the TC model is mainly based on uncommon high-risk phenotypes, and the SNP score was better at identifying lower-risk women because the relative risk distribution is symmetric on a log scale, and the baseline is low risk. A combination of the SNP67 score and the TC model might substantially improve the ability to identify high-risk women within this screening population (Table 3). In the high-risk group (>8% 10-year risk) the proportion from SNP67, TC and when combined was respectively 0.5%, 1.1%, 3.3%; the moderate-risk group (5-8% 10-year risk) was 5.7%, 8.2% and 9.5%.

Discussion

In this article we have examined the spread in risk from a panel of SNPs in comparison with classical phenotypic factors.

Table 2 showed the distribution of some phenotypic risk factors in the screening cohort. The distributions of age at menopause and current age for pre-menopausal women show that on average the pre-menopausal women in the cohort will undergo the menopause later than those who are postmenopausal. It is noticeable that the hormonal risk factors (age at menarche to BMI) altered the risk of a greater number of women than having an affected mother or sister did.

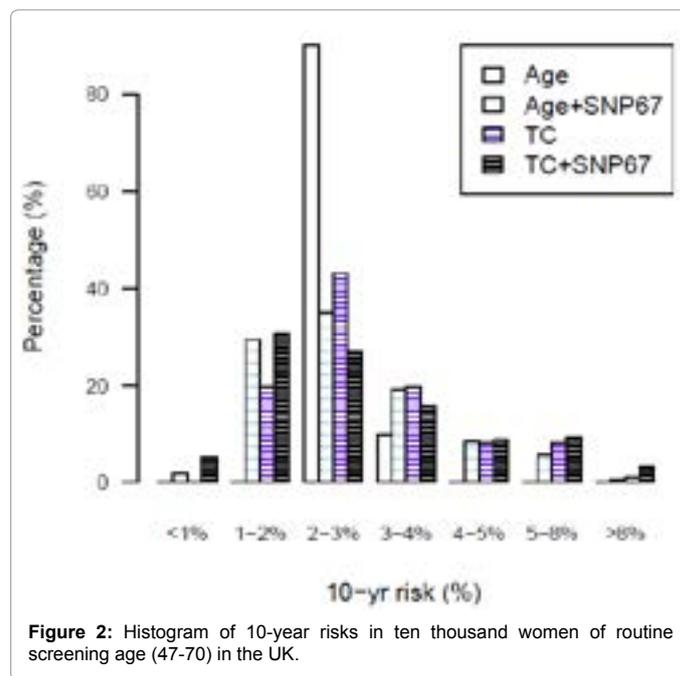
However, an affected first degree relative is still relevant and important because it confers a relatively large risk. SNPs in the first 18 loci to be discovered appeared to be uncorrelated with each other; the PROCAS SNP score also appeared uncorrelated with TC risk. These findings provide preliminary support to treating SNP scores and phenotypic risk from the TC model independently.

We found some optimism in the earlier risk estimates from the first 18 SNPs; Figure 1 showed that the 67 SNPs estimated by [6] had a similar spread to the first 18 SNPs from [7]. Although the COGS analysis used a very large data set, the true SNP67 risk distribution might also be less than was simulated. Thus, the analysis provides an indication of the maximum spread of risk that might be seen from a SNP score. More work would be helpful to assess the extent of optimism.

Figure 2 showed that 67 SNPs on their own might be less able to identify high-risk women than classical phenotypic factors. However, if they act independently then when combined with the TC model they would increase three-fold the number of women identified as being at high risk.

Mutations in BRCA 1 or 2 are known to confer much higher risks of breast cancer. However, they are very rare, being present in approximately 0.3% of the UK population [8]. Thus, testing the entire population for BRCA 1 or 2 would not change the risk distribution substantially. The distribution of risk from BRCA testing and age would look very similar to the age distribution in Figure 2, but approximately 0.3% would be moved into the high-risk group.

Breast density is a risk factor that is not presently incorporated



	Moderate risk (5-8%)	High risk (>8%)
Age+SNP67	5.7%	0.5%
TC	8.2%	1.1%
TC+SNP67	9.5%	3.3%

Table 3: Proportion of women from the cohort in moderate and high 10-year risk groups, if their risk would be assessed using age and the SNP67 score, the TC model, or the TC model in combination with the SNP67 score.

into the TC model, but is in some others [9]. It has roughly a four-fold difference in the relative risk from low to high groups. However, most women fall into the intermediate categories and so the overall spread of risk would be less than predicted for SNPs [10].

Finally, the assessment of breast cancer risk is important for prevention strategies. Most national screening programs only use age as a risk factor, where all women in an age range are invited to screening, but calibrated methods to assess risk for screening and other prevention strategies are being considered. In the UK the National Institute for Clinical Excellence (NICE) has published advice on the use of chemoprevention and risk-adapted screening for moderate and high-risk women [11]. Thus models that accurately identify larger numbers at high risk of breast cancer will have an impact on the health services, and on the health of women. In this context, SNPs might be useful in combination with classical phenotypic factors. However, validation work is needed to verify that the risk from all SNPs may be treated independently, and combines with other factors independently.

Acknowledgement

This work was supported by CRUK(C569/A10404).

Reference

1. <http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/incidence/CR-UK>
2. Tyrer J, Duffy SW, Cuzick J (2004) A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 23: 1111-1130.
3. Amir E, Evans DG, Shenton A, Lalloo F, Moran A, et al. (2003) Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. *J Med Genet* 40: 807-814.
4. Quante AS, Whittemore AS, Shriver T, Strauch K, Terry MB (2012) Breast cancer risk assessment across the risk continuum: genetic and nongenetic risk factors contributing to differential model performance. *Breast Cancer Res* 14: R144.
5. Evans DGR, Warwick J, Astley SM, Stavrinou P, Sahin S, et al. (2012) Assessing Individual Breast Cancer Risk within the UK National Health Service Breast Screening Program: A New Paradigm for Cancer Prevention. *Cancer Prev Res* 5: 943-951.
6. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, et al. (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 45: 353-361, 361e1-2.
7. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, et al. (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* 42: 504-507.
8. Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, et al. (2008) The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer* 98: 1457-1466.
9. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, et al. (2008) Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med* 148: 337-347.
10. McCormack VA, dos Santos Silva I (2006) Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 15: 1159-1169.
11. <http://guidance.nice.org.uk/CG164>

This article was originally published in a special issue, **Cancer Genetics** handled by Editor(s), Dr. Ahmed M Malki, Alexandria University, Egypt