

## The Role of Genome Sequencing in Personalized Breast Cancer Prevention

Weiva Sieh, Joseph H. Rothstein, Valerie McGuire, and Alice S. Whittemore

### Abstract

**Background:** There is uncertainty about the benefits of using genome-wide sequencing to implement personalized preventive strategies at the population level, with some projections suggesting little benefit. We used data for all currently known breast cancer susceptibility variants to assess the benefits and harms of targeting preventive efforts to a population subgroup at highest genomic risk of breast cancer.

**Methods:** We used the allele frequencies and effect sizes of 86 known breast cancer variants to estimate the population distribution of breast cancer risks and evaluate the strategy of targeting preventive efforts to those at highest risk. We compared the efficacy of this strategy with that of a "best-case" strategy based on a risk distribution estimated from breast cancer concordance in monozygous twins, and with strategies based on previously estimated risk distributions.

**Results:** Targeting those in the top 25% of the risk distribution would include approximately half of all future breast cancer cases, compared with 70% captured by the best-case strategy and 35% based on previously known variants. In addition, current evidence suggests that reducing exposure to modifiable nongenetic risk factors will have greatest benefit for those at highest genetic risk.

**Conclusions:** These estimates suggest that personalized breast cancer preventive strategies based on genome sequencing will bring greater gains in disease prevention than previously projected. Moreover, these gains will increase with increased understanding of the genetic etiology of breast cancer.

**Impact:** These results support the feasibility of using genome-wide sequencing to target the women who would benefit from mammography screening. *Cancer Epidemiol Biomarkers Prev*; 23(11); 2322–7. ©2014 AACR.

### Introduction

Describing an article in *Science Translational Medicine*, entitled "The predictive capacity of personal genome sequencing," by Roberts and colleagues (1), an editor wrote:

Imagine that everyone at birth could have their whole genome sequenced at negligible cost. Surely, this must be a worthwhile endeavor, given the list of luminaries who have already had this sequencing completed. But how well will such tests perform?

The authors address this question by using the incidence data of twins for a range of diseases to estimate the percentage of the population and the percentage of diseased cases whose genomes would classify them at high risk, and the risk among those whose genomes put them at

low risk relative to the population. They conclude that the genomes of most people would classify them at low risk for most of the diseases, and that the predictive value of such a negative test result would generally be small, because the disease risks among those who test negative would be similar to those in the general population. These conclusions have been criticized as deriving from assumptions that involve no information about the genetic factors relevant to the diseases studied (2). Here, we compare the breast cancer findings of Roberts and colleagues (1) with those obtained using information on the currently known breast cancer susceptibility loci.

To be specific, consider the strategy of sequencing the genomes of all young women, and using their genotypes at breast cancer loci to construct for each a genomic risk score, which specifies the rank of her inherited breast cancer risk relative to that of others in the population. These ranks can then be used to classify women into high- and low-risk categories, with high-risk women targeted for more intensive screening and preventive efforts. At present we do not fully know a person's breast cancer risk score, but genome sequencing coupled with current knowledge would allow us to assign her a partial score by combining her genotypes at all known breast cancer loci with the effect sizes of the risk alleles at these loci. Here, we shall evaluate and compare the efficacy of this strategy with that of (i) the best-case classification

Stanford University School of Medicine, Department of Health Research and Policy, Stanford, California.

**Corresponding Author:** Alice S. Whittemore, Department of Health Research and Policy, Stanford University School of Medicine, Redwood Building, Room T204, Stanford, CA 94305-5405. Phone: 650-723-5460; Fax: 650-725-6951; E-mail: alicesw@stanford.edu

doi: 10.1158/1055-9965.EPI-14-0559

©2014 American Association for Cancer Research.

theoretically possible if we knew the full risk scores and (ii) the estimates of Roberts and colleagues (1).

## Materials and Methods

### Population studied

As the frequencies and effect sizes of breast cancer susceptibility loci may differ by race/ethnicity, we restrict analyses to women of European ancestry. This study did not require approval from an ethical review board, as it involved only the use of published summary data.

### Statistical analysis

We modeled the lifetime probability of developing breast cancer (i.e., the absolute risk) for an individual with genomic risk score  $s$  as  $R = 1 - \exp(-ce^s)$ , where  $c$  is a positive constant. This model specifies a monotonically increasing relationship between risk  $R$  and risk score  $s$ . Therefore, the population percentile of the risk of an individual equals that of her risk score, and the efficacy of percentile-based stratification depends on the variances of the risk score distributions in the population and among future breast cancer cases. For the theoretically derived distribution of fully known risk scores, these variances can be estimated using the arguments of Pharoah and colleagues (3, 4) and Begg and Pike (5), as described in the Supplementary Materials and Methods.

To estimate the variance of the partially known risk scores, we modeled them as linear combinations of genotypes at a set of uncorrelated breast cancer loci from the literature, with coefficients given by their estimated effect sizes. Specifically, we listed all loci with validated breast cancer association in women of European ancestry, and then selected a subset of 86 uncorrelated loci. We chose 93 breast cancer susceptibility loci by reviewing the literature for established, replicated associations (6–15). We then selected a subset of 86 uncorrelated loci by computing all 4,278 pairwise correlation coefficients using the SNP Annotation and Proxy Search tool (<http://www.broad-institute.org/mpg/snap/ldsearch.php>) for the CEU population from the 1000 Genomes Project. We found seven pairs of loci with squared correlation coefficient exceeding 0.25 (16), and for each of these we selected the locus with the largest value of  $\beta p(1-p)$ , where  $\beta$  is the log relative risk associated with the variant and  $p$  is its allele frequency. The seven SNPs that we excluded were rs10069690, rs3215401, rs2943559, rs10759243, rs11199914, rs494406, and rs75915166.

The remaining 86 loci consist of (i) genes containing rare variants of high and moderate penetrance and (ii) SNPs identified in genome-wide association studies of breast cancer. We used the cumulative allele frequency and took the relative risk estimates for rare variants in breast cancer susceptibility genes to be the midpoints of the ranges spanned by the published studies. We also used the averages of the risk allele frequencies and relative risk (OR) estimates for SNPs that were associated with breast cancer in multiple genome-wide association studies.

Table 1 shows the 86 loci, and the frequencies and effect sizes of their risk alleles (6–15). We modeled the combined effects of these 86 loci by assuming that they act multiplicatively on a woman's cumulative hazard for breast cancer. As shown in the Supplementary Materials and Methods, this implies that her partially known risk score has the additive form  $s = \beta_1 g_1 + \dots + \beta_{86} g_{86}$ , where  $g_k = 0, 1, 2$  denotes her count of risk alleles at locus  $k$ , and  $\beta_k$  denotes the effect size of the risk allele at locus  $k$  as obtained from the literature,  $k = 1, \dots, 86$ . Determining the variance of the resulting partial scores is infeasible, as it would require summing over all  $3^{86} = 10^{41}$  possible multi-locus genotypes. However, it can be approximated by random genotype sampling as described in the Supplementary Materials and Methods.

### Performance of risk score-based classification

We estimated how well we can identify future breast cancer cases by classifying women into high-risk (targeted) and low-risk (untargeted) subgroups based on the percentiles of their fully and partially known risk scores (see Supplementary Materials and Methods for details). Specifically, we estimated the sensitivity (Sn), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), and risk in untargeted women relative to that of the population.

### Results

We estimated the population variance of the risk scores based on the 86 currently known breast cancer susceptibility variants to be 0.35. This variance, while lower than the estimate of 1.44 for the variance of the fully known risk scores determined using the arguments of Pharoah (3, 4), is nevertheless considerably higher than the value 0.07 obtained for the risk scores based on the seven loci known in 2008 (4).

Figure 1 shows the percentage of breast cancer cases included among women having the highest  $100(1-\alpha)$ th percent of risk scores, for  $0 < \alpha < 1$ . The curves correspond to the best-case classification with risk score variance equal to 1.44 (solid curve), the currently feasible classification based on partially known risk scores, with variance equal to 0.35 (dashed curve), and the classification based on the seven loci known in 2008 (4) with variance equal to 0.07 (dotted curve). As the efficacy of risk stratification for prevention depends on the population variance of the risk scores, these results indicate that (i) current genetic knowledge far exceeds that in 2008 (4); and (ii) despite these gains, considerably better stratification should still be possible in the future, as we better understand the etiology of this disease.

Table 2 shows additional measures of discrimination obtained by classifying women as high risk or low risk on the basis of their risk scores, where the high-risk group is defined as those whose scores exceed the  $100(1-\alpha)$ th percentile of the centered Gaussian risk score distribution. How do these results compare with the breast cancer predictions of Roberts and colleagues (1)? The latter were

**Table 1.** Risk-allele frequencies and relative risks for breast cancer susceptibility loci among women of European-American ancestry

|    | Locus             | Chromosome | Gene                       | Risk allele frequency (%) | Relative risk |
|----|-------------------|------------|----------------------------|---------------------------|---------------|
| 1  | rs11249433        | 1p         | <i>NOTCH2/FCGR1B</i>       | 39                        | 1.14          |
| 2  | Multiple variants | 1p         | <i>MUTYH</i>               | 0.5                       | 1.4–2.2       |
| 3  | rs616488          | 1p         | <i>PEX14</i>               | 67                        | 1.06          |
| 4  | rs11552449        | 1p         | <i>TPN2/BCL2L15</i>        | 17                        | 1.07          |
| 5  | rs4245739         | 1q         | <i>MDM4</i>                | 26                        | 1.14          |
| 6  | rs4849887         | 2p         | None                       | 90                        | 1.10          |
| 7  | Multiple variants | 2p         | <i>MSH6</i>                | 0.9                       | 4.9–4.9       |
| 8  | Multiple variants | 2p         | <i>MSH2</i>                | 0.01                      | 2.4–2.4       |
| 9  | rs12710696        | 2p         | None                       | 36                        | 1.11          |
| 10 | rs2016394         | 2q         | <i>METAP1D</i>             | 52                        | 1.05          |
| 11 | rs1550623         | 2q         | <i>CDCA7</i>               | 84                        | 1.06          |
| 12 | rs1045485         | 2q         | <i>CASP8</i>               | 0.85                      | 1.03          |
| 13 | rs13387042        | 2q         | <i>IGFBP2, IGBP5, TPN2</i> | 52                        | 1.12          |
| 14 | rs16857609        | 2q         | <i>DIRC3</i>               | 26                        | 1.08          |
| 15 | rs4973768         | 3p         | <i>SLC4A7/NEK10</i>        | 46                        | 1.11          |
| 16 | rs12493607        | 3p         | <i>TGFBR2</i>              | 35                        | 1.06          |
| 17 | rs6762644         | 3p         | <i>ITPR1/EGOT</i>          | 40                        | 1.07          |
| 18 | rs9790517         | 4q         | <i>TET2</i>                | 23                        | 1.05          |
| 19 | rs6828523         | 4q         | <i>ADAM29</i>              | 87                        | 1.11          |
| 20 | rs10941679        | 5p         | <i>MRPS30/HCN1</i>         | 26                        | 1.19          |
| 21 | rs7734992         | 5p         | <i>TERT</i>                | 43                        | 1.05          |
| 22 | rs889312          | 5q         | <i>MAP3K1/MEIR3</i>        | 28                        | 1.13          |
| 23 | rs10472076        | 5q         | <i>RAB3C</i>               | 38                        | 1.05          |
| 24 | rs1353747         | 5q         | <i>PDE4D</i>               | 90                        | 1.09          |
| 25 | rs1432679         | 5q         | <i>EBF1</i>                | 43                        | 1.07          |
| 26 | rs204247          | 6p         | <i>RANBP9</i>              | 43                        | 1.05          |
| 27 | rs17530068        | 6q         | None                       | 22                        | 1.09          |
| 28 | rs2046210         | 6q         | <i>ESR1</i>                | 34                        | 1.13          |
| 29 | rs3757318         | 6q         | <i>ESR1</i>                | 7                         | 1.21          |
| 30 | rs11242675        | 6q         | <i>FOXQ1</i>               | 61                        | 1.06          |
| 31 | rs720475          | 7q         | <i>ARHGEF5/NOBOX</i>       | 75                        | 1.06          |
| 32 | Multiple variants | 8q         | <i>NBN</i>                 | 0.9                       | 1.3–3.1       |
| 33 | rs9693444         | 8p         | None                       | 32                        | 1.07          |
| 34 | rs6472903         | 8q         | None                       | 82                        | 1.10          |
| 35 | rs2943559         | 8q         | <i>HNF4G</i>               | 7                         | 1.13          |
| 36 | rs13281615        | 8q         | <i>MYC</i>                 | 40                        | 1.08          |
| 37 | rs11780156        | 8q         | <i>MIR1208</i>             | 16                        | 1.07          |
| 38 | rs1011970         | 9p         | <i>CDKN2A/B</i>            | 17                        | 1.09          |
| 39 | rs865686          | 9q         | <i>KLF4/RAD23B</i>         | 61                        | 1.12          |
| 40 | rs10759243        | 9q         | None                       | 39                        | 1.06          |
| 41 | rs2380205         | 10p        | <i>ANKRD16</i>             | 56                        | 1.02          |
| 42 | rs7072776         | 10p        | <i>MLLT10/DNAJC1</i>       | 29                        | 1.07          |
| 43 | rs11814448        | 10p        | <i>DNAJC1</i>              | 2                         | 1.26          |
| 44 | rs10995190        | 10q        | <i>ZNF365</i>              | 85                        | 1.16          |
| 45 | rs704010          | 10q        | <i>ZMIZ1</i>               | 39                        | 1.07          |
| 46 | Multiple variants | 10q        | <i>PTEN</i>                | 0.01                      | 2.0–10.0      |
| 47 | rs7904519         | 10q        | <i>TCF7L2</i>              | 46                        | 1.06          |
| 48 | rs11199914        | 10q        | None                       | 68                        | 1.05          |
| 49 | rs2981579         | 10q        | <i>FGFR2</i>               | 38                        | 1.26          |

(Continued on the following page)

**Table 1.** Risk-allele frequencies and relative risks for breast cancer susceptibility loci among women of European-American ancestry (Cont'd)

|    | Locus             | Chromosome | Gene                    | Risk allele frequency (%) | Relative risk |
|----|-------------------|------------|-------------------------|---------------------------|---------------|
| 50 | rs3817198         | 11p        | <i>LSP1/H19</i>         | 30                        | 1.07          |
| 51 | rs3903072         | 11q        | <i>OVOL1</i>            | 53                        | 1.05          |
| 52 | rs614367          | 11q        | <i>CCND1/FGFs</i>       | 15                        | 1.15          |
| 53 | rs494406          | 11q        | <i>CCND1</i>            | 26                        | 1.07          |
| 54 | Multiple variants | 11q        | <i>ATM</i>              | 0.3                       | 2.0–3.0       |
| 55 | rs11820646        | 11q        | None                    | 59                        | 1.09          |
| 56 | rs10771399        | 12p        | <i>PTHLH</i>            | 88                        | 1.19          |
| 57 | rs12422552        | 12p        | None                    | 26                        | 1.05          |
| 58 | rs17356907        | 12q        | <i>NTN4</i>             | 70                        | 1.10          |
| 59 | rs1292011         | 12q        | <i>TBX3/MAPKAP5</i>     | 58                        | 1.10          |
| 60 | Multiple variants | 13q        | <i>BRCA2</i>            | 0.1                       | 9.0–21.0      |
| 61 | rs2236007         | 14q        | <i>PAX9/SLC25A21</i>    | 79                        | 1.08          |
| 62 | rs999737          | 14q        | <i>RAD51B</i>           | 78                        | 1.09          |
| 63 | rs2588809         | 14q        | <i>RAD51L1</i>          | 16                        | 1.08          |
| 64 | rs941764          | 14q        | <i>CCDC88C</i>          | 34                        | 1.06          |
| 65 | rs3803662         | 16q        | <i>TOX3/LOC643714</i>   | 25                        | 1.20          |
| 66 | rs17817449        | 16q        | <i>MIR1972-2-FTO</i>    | 60                        | 1.08          |
| 67 | rs11075995        | 16q        | <i>FTO</i>              | 24                        | 1.10          |
| 68 | Multiple variants | 16q        | <i>CDH1</i>             | 0.01                      | 2.0–10.0      |
| 69 | Multiple variants | 16p        | <i>PALB2</i>            | 0.01                      | 2.0–4.0       |
| 70 | rs13329835        | 16q        | <i>CDYL2</i>            | 22                        | 1.08          |
| 71 | Multiple variants | 17q        | <i>BRCA1</i>            | 0.06                      | 5.0–45.0      |
| 72 | Multiple variants | 17q        | <i>BRIP2</i>            | 0.1                       | 2.0–3.0       |
| 73 | Multiple variants | 17p        | <i>TP53</i>             | 0.01                      | 2.0–10.0      |
| 74 | rs6504950         | 17q        | <i>STXB4/COX11</i>      | 73                        | 1.05          |
| 75 | Multiple variants | 17q        | <i>RAD51C</i>           | 0.5                       | 3.2–3.5       |
| 76 | rs527616          | 18q        | None                    | 62                        | 1.05          |
| 77 | rs1436904         | 18q        | <i>CHST9</i>            | 60                        | 1.04          |
| 78 | Multiple variants | 19p        | <i>STK11</i>            | 0.01                      | 2.0–10.0      |
| 79 | rs8170            | 19p        | <i>MERIT40</i>          | 19                        | 1.25          |
| 80 | rs4808801         | 19p        | <i>SSBP4/ISYNA1/ELL</i> | 65                        | 1.06          |
| 81 | rs3760982         | 19q        | <i>KCNN4/ZNF283</i>     | 46                        | 1.06          |
| 82 | rs2284378         | 20q        | <i>RALY</i>             | 33                        | 1.16          |
| 83 | rs2823093         | 21q        | <i>NRIP1</i>            | 73                        | 1.09          |
| 84 | Multiple variants | 22q        | <i>CHEK2</i>            | 0.4                       | 2.0–3.0       |
| 85 | rs132390          | 22q        | <i>EMID1/RHBDD3</i>     | 3.6                       | 1.12          |
| 86 | rs6001930         | 22q        | <i>MKL1</i>             | 11                        | 1.12          |

based on a high-risk group defined as those whose risks exceed the 90th to 95th percentile of the population distribution. The authors estimated that this classification would target between 10% and 35% of all future breast cancer cases. In contrast, Table 2 shows that the percentage of cases targeted would be approximately 47% using the best-case classification and 32% using the currently feasible classification. In addition, the authors estimated that the ratio of risk among women classified at low risk relative to the population would be as high as 0.72 to 0.90, indicating poor specificity. Yet, Table 2 suggests that this relative risk is lower: 0.59 with the best-case classification and 0.75 with the currently feasible classification. Thus,

the present estimates provide more optimistic projections than those obtained using the theoretical model of Roberts and colleagues (1).

## Discussion

We have estimated the variance of breast cancer risk scores among women of European ancestry by using a multiplicative model for the joint effects of currently known breast cancer loci. We found that the distribution of these partially known risk scores has variance 0.35, which is similar in magnitude to the estimate of 0.28 obtained independently by Pashayan and colleagues

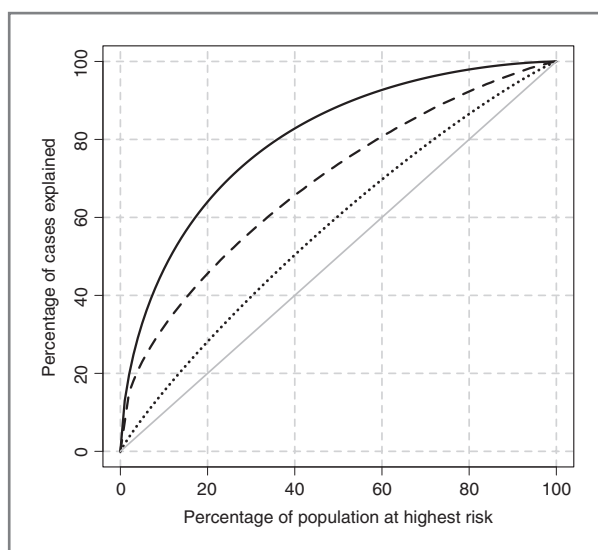


Figure 1. Percentage of all breast cancer cases explained by those at highest risk for the disease. Curves are shown for the best-case scenario when all breast cancer susceptibility alleles are known (solid curve), currently known susceptibility alleles (dashed curve), and seven alleles known in 2008 (ref. 4; dotted curve).

(17). We have compared the performance of targeted preventive measures based on the currently feasible partially known risk scores with those obtained using the theoretical distribution of fully known risk scores derived by Pharoah (3, 4). The results suggest that the predictive power of genome sequencing to determine breast cancer risk is considerably greater than that described by Roberts and colleagues (1), and the estimates contradict the authors' statement that "...our conclusions [...] represent an absolute upper bound that cannot be improved by improvements in technology or genetic knowledge."

To achieve the optimal predictive power represented by the "best-case" classification, we will need to identify the combined effects of all causal alleles for breast cancer. Moreover, better understanding of gene-environment interactions could further improve predictive power. For example, the performance measures described here underestimate the potential value of genomic-based risk classification. This is because a child's lifetime breast cancer risk is determined not only by her genome, but also by her future levels of nongenetic (lifestyle, environmental, and epigenetic) risk factors. Epidemiologic data support a multiplicative model for the joint effects of

**Table 2.** Impact<sup>a</sup> of targeting a subset of European-American women for breast cancer preventive strategies

| Percentage of population at high risk ( $\alpha$ )                              | Lifetime risk in low-risk (%) | Risk for high-risk relative to low-risk | Risk for low-risk relative to population | Percentage of cases who are high-risk ( $S_n$ ) | Percentage of noncases who are high-risk ( $1-S_p$ ) | Percentage of high-risk who are cases ( $PPV$ ) | Percentage of low-risk who are noncases ( $NPV$ ) | NNT <sup>b</sup> |
|---|-------------------------------|---|--|---|--|---|---|------------------|
| Current knowledge of genetic susceptibility alleles (risk score variance, 0.35) |                               |   |  |   |  |   |   |                  |
| 100   | —                             | —                                       | —  | 100   | 100  | 12.68 <sup>a</sup>                              | —   | 95               |
| 90  | 4.11                          | 3.32                                    | 0.32                                     | 96.76   | 89.02  | 13.63   | 95.89   | 88               |
| 75  | 5.20                          | 2.92                                    | 0.41                                     | 89.74   | 72.86  | 15.17   | 94.80   | 79               |
| 67  | 5.68                          | 2.84                                    | 0.45                                     | 85.22   | 64.35  | 16.13   | 94.32   | 74               |
| 50  | 6.64                          | 2.82                                    | 0.52                                     | 73.82   | 46.54  | 18.72   | 93.36   | 64               |
| 33  | 7.66                          | 2.99                                    | 0.60                                     | 59.54   | 29.15  | 22.88   | 92.34   | 52               |
| 25  | 8.20                          | 3.18                                    | 0.65                                     | 51.49   | 21.15  | 26.12   | 91.80   | 46               |
| 10  | 9.54                          | 4.30                                    | 0.75                                     | 32.31   | 6.76   | 40.97   | 90.46   | 29               |
| 0   | 12.68                         | —                                       | 1.00                                     | 0   | 0  | —   | 87.32   | —                |
| Best-case scenario (risk score variance, 1.44)                                  |                               |   |  |   |  |   |   |                  |
| 100   | —                             | —                                       | —  | 100   | 100  | 12.68 <sup>a</sup>                              | —   | 95               |
| 90  | 0.83                          | 16.85                                   | 0.07                                     | 99.34   | 88.64  | 14.00   | 99.17   | 86               |
| 75  | 1.55                          | 10.60                                   | 0.12                                     | 96.95   | 71.81  | 16.39   | 98.45   | 73               |
| 67  | 1.94                          | 9.24                                    | 0.15                                     | 94.94   | 62.94  | 17.97   | 98.06   | 67               |
| 50  | 2.92                          | 7.68                                    | 0.23                                     | 88.48   | 44.41  | 22.44   | 97.08   | 53               |
| 33  | 4.24                          | 7.04                                    | 0.33                                     | 77.61   | 26.52  | 29.82   | 95.76   | 40               |
| 25  | 5.07                          | 7.00                                    | 0.40                                     | 69.99   | 18.47  | 35.50   | 94.93   | 34               |
| 10  | 7.51                          | 7.88                                    | 0.59                                     | 46.67   | 4.68   | 59.18   | 92.49   | 20               |
| 0   | 12.68                         | —                                       | 1.00                                     | 0   | 0  | —   | 87.32   | —                |

<sup>a</sup>Assuming that the lifetime risk of developing breast cancer is 12.68% based upon estimates for the U.S. white female population (19).

<sup>b</sup>NNT, number of women who must be targeted to avoid one breast cancer death, assuming one sixth of all breast cancers are fatal and targeting prevents death from half of them.

genetic and nongenetic factors on breast cancer risk (18). Under this model, a modifiable nongenetic factor associated with an overall 50% increase in risk would add considerably more to the absolute risk of a female whose lifetime genetic risk is 36% (increasing it to 54%) than to that of a female whose lifetime genomic risk is 4% (increasing it only to 6%). Thus, high-risk women have considerably more to gain by appropriate choices of lifestyle factors than do low-risk women.

In conclusion, the data-based estimates presented here suggest that personalized breast cancer preventive strategies informed by genome sequencing may bring greater gains in cost-efficient disease prevention than previously projected. Moreover, these gains will increase as we gain increased understanding of the etiology of breast cancer.

### Disclosure of Potential Conflict of Interest

No potential conflicts of interest were disclosed.

### References

1. Roberts NJ, Vogelstein JT, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE. The predictive capacity of personal genome sequencing. *Sci Transl Med* 2012;4:133ra58.
2. Topol EJ. Comment on "the predictive capacity of personal genome sequencing". *Sci Transl Med* 2012;4:135le133.
3. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* 2002;31:33–6.
4. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* 2008;358:2796–803.
5. Begg CB, Pike MC. Comment on "the predictive capacity of personal genome sequencing". *Sci Transl Med* 2012;4:135le133.
6. Bogdanova N, Feshchenko S, Schurmann P, Waltes R, Wieland B, Hillemanns P, et al. Nijmegen breakage syndrome mutations and risk of breast cancer. *Int J Cancer* 2008;122:802–6.
7. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nat Genet* 2013;45:392–8.
8. Ghoussaini M, Pharoah PDP, Easton DF. Inherited genetic susceptibility to breast cancer. The beginning of the end or the end of the beginning? *Am J Pathol* 2013;183:1038–51.
9. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. Genetic susceptibility to breast cancer. *Mol Oncol* 2010;4:174–91.
10. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013;45:353–61.
11. Peng S, Lu B, Ruan W, Zhu Y, Sheng H, Lai M. Genetic polymorphisms and breast cancer risk: evidence from meta-analyses, pooled analyses, and genome-wide association studies. *Breast Cancer Res Treat* 2011;127:309–24.
12. Rennert G, Lejbkowitz F, Cohen I, Pinchev M, Rennert HS, Barnett-Griness O. MUTYH mutation carriers have increased breast cancer risk. *Cancer* 2012;118:1989–93.
13. Win AK, Lindor NM, Young JP, Macrae FA, Young GP, Williamson E, et al. Risks of primary extracolonic cancers following colorectal cancer in Lynch Syndrome. *J Natl Cancer Inst* 2012;104:1363–72.
14. Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet* 2010;42:410–414.
15. Thompson ER, Boyle SE, Johnson J, Ryland GL, Sawyer S, Choong DYH, et al. Analysis of RDA51C germline mutations in high risk breast and ovarian cancer families and ovarian cancer patients. *Hum Mutat* 2010;33:95–99.
16. International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–52.
17. Pashayan N, Hall A, Chowdhury S, Dent T, Pharoah PD, Burton H. Public health genomics and personalized prevention: lessons from the COGS project. *J Intern Med* 2013;274:451–6.
18. Li H, Beeghly-Fadiel A, Wen W, Lu W, Gao YT, Xiang YB, et al. Gene-environment interactions for breast cancer risk among Chinese women: a report from the Shanghai breast cancer genetics study. *Am J Epidemiol* 2013;177:161–70.
19. Howlander N, Noone A, Krapcho M, Garshell J, Neyman N, Altekruse S, et al. SEER Cancer Statistics Review, 1975–2010. National Cancer Institute, Bethesda, MD. Based on November 2012 SEER data submission, posted to the SEER website, April 2013. Available from: [http://seer.cancer.gov/csr/1975\\_2010/](http://seer.cancer.gov/csr/1975_2010/).

### Authors' Contributions

**Conception and design:** W. Sieh, A.S. Whittemore  
**Development of methodology:** W. Sieh, J.H. Rothstein, A.S. Whittemore  
**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** V. McGuire  
**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** W. Sieh, J.H. Rothstein, A.S. Whittemore  
**Writing, review, and/or revision of the manuscript:** W. Sieh, J.H. Rothstein, V. McGuire, A.S. Whittemore  
**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** V. McGuire  
**Study supervision:** W. Sieh, A.S. Whittemore

### Grant Support

This research is supported by grants K07CA143047 (to W. Sieh) and R01CA094069 (to A.S. Whittemore and J.H. Rothstein) from the U.S. National Cancer Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 22, 2014; revised August 1, 2014; accepted August 2, 2014; published OnlineFirst October 23, 2014.