

Accurate Prediction of *BRCA1* and *BRCA2* Heterozygous Genotype Using Expression Profiling after Induced DNA Damage

Zsofia Kote-Jarai,¹ Lucy Matthews,² Ana Osorio,^{1,4} Susan Shanley,^{1,5} Ian Giddings,² Francois Moreews,² Imogen Locke,^{1,5} D. Gareth Evans,⁶ Diana Eccles,⁷ The Carrier Clinic Collaborators,⁵ Richard D. Williams,³ Mark Girolami,⁸ Colin Campbell,⁹ and Ros Eeles^{1,5}

Abstract Purpose: In this study, the differential gene expression changes following radiation-induced DNA damage in healthy cells from *BRCA1/BRCA1* mutation carriers have been compared with controls using high-density microarray technology. We aimed to establish if *BRCA1/BRCA2* mutation carriers could be distinguished from noncarriers based on expression profiling of normal cells.

Experimental Design: Short-term primary fibroblast cultures were established from skin biopsies from 10 *BRCA1* and 10 *BRCA2* mutation carriers and 10 controls, all of whom had previously had breast cancer. The cells were subjected to 15 Gy ionizing irradiation to induce DNA damage. RNA was extracted from all cell cultures, preirradiation and at 1 hour postirradiation. For expression profiling, 15 K spotted cDNA microarrays manufactured by the Cancer Research UK DNA Microarray Facility were used. Statistical feature selection was used with a support vector machine (SVM) classifier to determine the best feature set for predicting *BRCA1* or *BRCA2* heterozygous genotype. To investigate prediction accuracy, a nonprobabilistic classifier (SVM) and a probabilistic Gaussian process classifier were used.

Results: In the task of distinguishing *BRCA1* and *BRCA2* mutation carriers from noncarriers and from each other following radiation-induced DNA damage, the SVM achieved 90%, and the Gaussian process classifier achieved 100% accuracy. This effect could not be achieved without irradiation. In addition, the SVM identified a set of *BRCA* genotype predictor genes.

Conclusions: We conclude that after irradiation-induced DNA damage, *BRCA1* and *BRCA2* mutation carrier cells have a distinctive expression phenotype, and this may have a future role in predicting genotypes, with application to clinical detection and classification of mutations.

It is estimated that 5% to 10% of breast cancer patients develop the disease due to the presence of a mutation in a breast cancer predisposition gene (1). A significant proportion of this population (about 50%) has a mutation in one of the known breast cancer predisposition genes (*BRCA1* or *BRCA2*). Besides the definite disease-causing deleterious mutations, small alterations, such as single base substitutions (missense mutations), are frequently found in these genes. Their functional effects are usually unknown, hence they are termed variants of uncertain

significance. Some of these variants of uncertain significance could also have a role in breast cancer predisposition, but it is not currently possible to establish their disease-causing effect. The available diagnostic tests for mutation analysis of *BRCA1/BRCA2* are time and labor intensive, expensive, and do not allow for the identification of all types of mutation. The aim of this study was to determine whether gene expression profiling could be used to distinguish between heterozygous *BRCA1* and *BRCA2* mutation carriers and control samples.

Authors' Affiliations: ¹Translational Cancer Genetics Team and Sections of ²Molecular Carcinogenesis and ³Paediatric Oncology, The Institute of Cancer Research, Sutton, Surrey, United Kingdom; ⁴Department of Human Genetics, Spanish National Cancer Centre, Madrid, Spain; ⁵Royal Marsden NHS Foundation Trust, London, United Kingdom; ⁶St. Mary's Hospital, Manchester, United Kingdom; ⁷Clinical Genetics, Princess Anne Hospital, Southampton, United Kingdom; ⁸Department of Computing Science, Bioinformatics Research Centre, University of Glasgow, Glasgow, United Kingdom; and ⁹Computational Intelligence Group, University of Bristol, Bristol, United Kingdom
Received 12/27/05; revised 2/7/06; accepted 3/23/06.

Grant support: Medical Research Council Discipline Hopping Award (M. Girolami), Cancer Research UK (L. Matthews, I. Giddings, F. Moreews, and the microarray production), NHMRC Australia (S. Shanley), Cancer Research UK grants C5047/A5463 (I. Locke) and C5047/A3354D (The Carrier Clinic and R. Eeles), Engineering and Physical Sciences Research Council grant GR/R96255/01 (C. Campbell), Breast Cancer Campaign project grant (Z. Kote-Jarai and the study), legacy of the late Marion Silcock (Z. Kote-Jarai

and the study), and Maxse/Knowles Research Fund (Z. Kote-Jarai and the study).

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.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

A. Osorio was a Haddow Fellow of The Institute of Cancer Research. D. Gareth Evans, D. Eccles, and R. Williams had no specific funding related to this work.

The Carrier Clinic Collaborators are Audrey Ardern-Jones, Elizabeth Bancroft, Kate Bishop, Elly Lynch, Rebecca Doherty, Sarah Thomas, Asher Salmon, Clare Turnbull, Sameer Jhavar.

Requests for reprints: Zsofia Kote-Jarai, Translational Cancer Research Team, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey, SM2 5NG United Kingdom. Phone: 44-208-661-3105; E-mail: zsofia.kote-jarai@icr.ac.uk.

© 2006 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-05-2805

The isolation of *BRCA1* (2) and *BRCA2* (3) stimulated intensive scientific interest. Although extensive data are now available on these genes, including nucleotide sequence, mutation spectrum, cellular localization, and protein structure, the exact molecular pathways in which *BRCA1* and *BRCA2* function and how their disruption promotes breast and ovarian tumorigenesis remain to be elucidated (4, 5). The products of both genes are large nuclear proteins: *BRCA1* has 1,863 amino acids, and *BRCA2* has 3,418 amino acids. Their amino acid sequences reveal little about their function, and although there are some similarities between their genetic structures, there is no sequence homology between them. However, a number of observations indicate that *BRCA1* and *BRCA2* function in similar pathways. Their tissue distribution and gene expression patterns are similar. Their expression levels are cell cycle regulated (6), and they both interact with *RAD51* (7, 8). *RAD51* plays a key role in homologous recombination and DNA double-strand break repair, suggesting a role for *BRCA1* and *BRCA2* in the DNA damage response (9). There is evidence that *BRCA1* actively participates in transcriptional regulation (10). *BRCA1* has a tandem BRCT domain at the COOH terminus that has transcriptional activation function, and it physically associates with the RNA polymerase II complex. *BRCA1* acts as a transcriptional coactivator of cyclic AMP-responsive element binding protein and E1A (11) and as a repressor of *MYC* (12). *BRCA2* also has a transcription activating function, localized to a highly conserved region at the NH₂ terminus, and interacts with *SMAD3* (13). It has also been shown that *BRCA2* is part of the Fanconi complementation group, and *BRCA2* and *FANCD1* are the same genes (14).

A number of studies have been published using cDNA microarrays to identify gene expression patterns in cancer cell lines and tumor samples. The aim of these studies has been to understand and classify tumors based on their global patterns of gene expression. For example, patient tumor samples can be divided into different histologic and prognostic groups based upon clustering algorithms using cDNA microarray data (15). It has been reported that *BRCA1* and *BRCA2* mutation status influences the somatic tumor gene expression profile (16, 17), but the profile in cells from healthy tissues in *BRCA1* mutation carriers after radiation-induced damage has only been reported by our group (18). We have shown that the *BRCA1* heterozygous genotype was predictable with high-accuracy fibroblast cultures from the breast. This result provided evidence that perhaps there is a heterozygous phenotype for *BRCA1*. The aim of the present study was to assess the potential of gene expression profiling in discriminating between *BRCA1* or *BRCA2* heterozygotes and controls using skin biopsies as a tissue source. Here, we show that it is possible to predict the *BRCA* genotype of these healthy tissue samples after irradiation-induced DNA damage using gene expression profiling.

Materials and Methods

Samples. Women were recruited from two centers: the Royal Marsden Hospital NHS Foundation Trust/The Institute of Cancer Research, Cancer Genetics Carrier Clinic and the Academic Unit of Medical Genetics at the St. Mary's Hospital Manchester, United Kingdom. Known *BRCA1* and *BRCA2* mutation carriers and sporadic cases were identified from the databases of these centers. All study participants had breast cancer with a median follow up of 6 to 7 years,

and at the time of the sample collection, they were healthy and not undergoing any treatment. Carriers and controls had a similar age of onset distribution (28-62 years, with means of 50, 47, and 45 years for *BRCA1*, *BRCA2*, and controls, respectively) and similar treatment regimes (chemotherapy and radiotherapy). Although mutation analysis was not undertaken in the control group, the likelihood of undetected *BRCA* mutations was minimized via the selection criteria (which excluded those with a family history of ovarian cancer or Ashkenazi origin and those of other than a family history of one case of postmenopausal breast cancer). Written informed consent was obtained from all patients before inclusion, and the study protocol was approved by the Royal Marsden Loco-regional Ethics Committee. The mutations in both genes spanned the whole length of the code; no specific mutations were overrepresented; and the ethnic origins in all three groups were very similar: all were White Caucasians.

The mutations in the *BRCA* samples are listed in Supplementary Table S1. Short-term primary fibroblast cultures were established from 3-mm skin punch biopsies obtained under local anesthesia from the buttock area. Biopsies were transported to the laboratory in DMEM with 10% FCS, cut into small fragments, and immediately explanted into 12.5-cm² culture flasks (Falcon) in 0.5 to 1 mL DMEM containing 10% FCS, 10 mmol/L HEPES with 50 units/mL penicillin, and 50 µg/mL streptomycin. One flask was set up per 3-mm punch biopsy; explanted biopsies were left undisturbed for 10 days, at which time the culture medium was removed along with any pieces of tissue unattached to the bottom of the flask. When fibroblasts reached confluence or became crowded in one area of the flask, they were passaged into a single 25-cm² flask (P1) and subsequently to a single 80-cm² flask (P2). Cells in four T80 flasks (P3) were allowed 10 days to reach confluence.

All fibroblast cultures used in this study were maintained in DMEM supplemented with 10% fetal bovine serum, penicillin (100 units/mL)/streptomycin (100 µg/mL), and 2 mmol/L glutamine. Confluent cells were irradiated with 15 Gy at a high dose rate (1.5 Gy/min) using a 250-kV X-ray machine.

Gene expression profiling. Total RNA was extracted from all cell cultures before irradiation and 1 hour after the radiation treatment using an RNeasy kit (Qiagen, Hilden, Germany). Universal Human Reference RNA (Stratagene, La Jolla, CA) was used as reference RNA.

Total RNA samples were fluorescently labeled using the CyScribe Post-labeling kit (GE Healthcare Ltd., Chalfont St. Giles, Buckinghamshire, United Kingdom) according to the manufacturer's instructions.

Equal amounts of 4 µg of labeled sample and reference cDNA were mixed and hybridized onto the microarrays. We used high-density cDNA microarrays manufactured by the CRUK DNA Microarray facilities at The Institute of Cancer Research, representing 14127 IMAGE cDNA clones. All supplementary data can be found at <http://www.icr.ac.uk/array/array.html>. Details of the clone set and hybridization conditions can also be found here and at the Gene Expression Omnibus, where all the data have been submitted in compliance with Minimum Information About a Microarray Experiment. The Gene Expression Omnibus accession number for this submission is GES3382. Image acquisition and analysis were done using a GenePix 4000B scanner and GenePix Pro 6 software (Axon Instruments, Inc., Sunnyvale, CA), respectively. Signal intensities for Cy3 and Cy5 channels were normalized using the Loess regression.

Data analysis. Of the 14,127 cDNA clones represented on the arrays, 8,080 clones were selected by a quality filter and used in subsequent analysis. The selection criterion was a signal intensity in at least one of the channels (red or green) of 2-fold greater than background in a minimum of 70% of the samples in each comparison. We analyzed our data using a support vector machine (SVM) for class comparison and class prediction followed by hierarchical clustering and principal component analysis. To evaluate class prediction on test data, we have used two types of classifiers: a SVM classifier (19) with a linear kernel and feature ranking using a statistical score (Fisher score, a *t* test, and a Mann-Whitney rank-based score) and second, a Gaussian process classifier (GPC; ref. 20). SVMs are known to give reliable prediction and

have frequently been used for classification tasks involving microarray data. However, they are nonprobabilistic; a class label is assigned to a new sample, but no information is given indicating the confidence in this assignment. Consequently, we have also investigated a probabilistic GPC, which assign probabilities for class membership.

For the SVM, we investigated the effects of feature selection: predictive accuracy was determined across a range starting with all features (data from 8,080 cDNA clones as mentioned above) followed by successive removal of the least discriminative feature (according to the statistical score used) through to the top two most discriminative features. With the GPC feature, selection was not evaluated given the slow training time. A sample set of 10 in each of the three classes has the statistical power to show significant differences. Given this size of the data set, the test error was evaluated using leave-one-out (LOO) cross-validation with the left-out data point excluded from the feature selection procedure, to provide an unbiased test statistic. To provide a baseline model (for null prediction), we can readily calculate the expected number of test errors and associated SD for random data. Thus, for a binary classification task with a balanced (10 + 10) split, the expected number of LOO test errors is 10 with a SD of 2.23 about this mean. In both the SVM and GPC, it was made sure that there was no contamination of the test point (e.g., the test point in LOO testing was not incorporated in the feature selection, etc.). The computation of the SVM classifier and GPC was done separately; thus, they validate each other. For subsequent visualization of our data, we applied hierarchical clustering and principal component analysis using the Genesis software package (21). The inputs for this were the expression data of the discriminatory features identified by feature selection with the SVM.

Results and Discussion

We have analyzed the expression profiles of normal skin fibroblast cultures from 10 *BRCA1* and 10 *BRCA2* mutation carriers and compared these with the profiles of 10 control skin fibroblast samples (with a very low probability of the presence of a *BRCA1* or *BRCA2* mutation; see Materials and Methods). All these samples were short-term primary cultures established from skin biopsies. Cell cultures were irradiated (15 Gy) to induce DNA damage, and the expression profiles of all 30 samples were analyzed both before and after irradiation. For post-irradiation RNA isolation, the 1-hour time point was chosen based on data from our previous study (18).

On the spotted cDNA microarray, 14,127 IMAGE clones were represented, covering approximately half of the human genome, of which 8,080 satisfied the quality filtering as described in Materials and Methods. We have used the expression data of these filtered clones in a class comparison analysis using a SVM classifier on all our pairs of classes: irradiated *BRCA1* mutation carriers and irradiated controls (*BRCA1.X* and *B0.X*, respectively), irradiated *BRCA2* mutation carriers and irradiated controls (*BRCA2.X* and *B0.X*, respectively), irradiated *BRCA1* mutation and irradiated *BRCA2* mutation carriers (*BRCA1.X* and *BRCA2.X*, respectively), *BRCA1* mutation carriers and controls without irradiation (*BRCA1* versus *B0*), *BRCA2* mutation carriers and controls without irradiation (*BRCA2* versus *B0*), and *BRCA1* mutation carriers and *BRCA2* mutation carriers without irradiation (*BRCA1* versus *BRCA2*). For each such (10 + 10) pairing, the test error of a SVM classifier was evaluated using LOO cross-validation. Although three types of feature selection were used (based on a *t* test, a Mann-Whitney score, and the Fisher score), there seemed to be little difference between these scores and the results quoted below are for a Fisher score.

The distinction between the irradiated *BRCA1.X* and *B0.X* was achieved with high predictive accuracy, one test error from 20 with LOO testing. The distinction between the irradiated *BRCA2* samples and controls (*BRCA2.X* versus *B0.X*) seemed to be less robust but was also significant; two to three LOO errors from 20, depending on the number of features after feature selection. Without irradiation, however, neither the *BRCA1* or the *BRCA2* carrier genotype could be predicted; LOO test errors were in the range of 6 to 8 and 5 to 8, respectively. *BRCA1.X* and *BRCA2.X* samples also showed a very different expression profile if compared with each other; the distinction was achieved with high predictive accuracy (95%): one LOO test error from 20 using an SVM. In all instances with irradiation, class distinction is achieved at a statistically significant level. We remarked earlier that for (10 + 10) binary classification with LOO testing, the expected number of test errors for a classifier trained on random data is 10 ± 2.23 : from this, we infer that observing three LOO test errors has a probability of occurrence of 8.2×10^{-4} ; two errors have a probability of 1.6×10^{-4} ; and one LOO error has a probability of occurrence of 2.6×10^{-5} . We also trained an SVM classifier to distinguish irradiated *BRCA1* and *BRCA2* samples taken as a single class *BRCA1/BRCA2* (20 samples) against the controls as second class (10 samples). Again, the prediction was achieved with high accuracy, two to three LOO test errors from 30 (the test error curves discussed above are provided as Supplementary Fig. S1A-C).

We have used an SVM for prediction given its wide application in classifying microarray data. However, although SVMs work well on binary classification tasks, they are less well suited to multiclass problems. In particular, the SVM assigns a class label to new instances, but does not assign a confidence to the labeling. Thus, we also used a probabilistic multiclass classification algorithm, a new GPC, trained using a variational Bayesian approach very suited to high-dimensional data sets (20). For the three-class task of distinguishing among irradiated *BRCA1*, *BRCA2*, and control samples, this algorithm gave zero LOO test error from 30 (if assigning the test sample to the class with associated highest probability). This result has a higher level of statistical significance than our probabilities reported for the SVM binary classifier: for the three-class (10 + 10 + 10) classification, the expected number of LOO test errors is 20.0 ± 2.58 , and the probability of observing zero test errors is upper bounded by 1.0×10^{-12} . Because GPC processes are slow to train, and because this result could not be improved, feature selection was not evaluated. The GPC has outperformed the SVM on the LOO test error and has the added advantage of assigning a confidence to the class label (Supplementary Fig. S2): for these reasons, we expect that these probabilistic classifiers offer the best approach to eventual clinical implementation.

The SVM feature selection by Fischer score provides us with a set of 200 discriminatory genes for each comparison. The list of top features for each classification is shown in the Supplementary Table S2A to C. Among these are oncogenes, cell cycle regulatory genes, and genes with function in transcription regulation and DNA damage repair. Interestingly in the *BRCA1* list, there are *STAT5*, *ATM*, *IL15R*, *CCNH*; in the *BRCA2* lists, there are *TGFA*, *SMURF2*, *SWI*-related *SMARCC4*, *E2F3*, and *CDKN1B* (*p27*). All these have been reported to be in a functional interaction with the *BRCA* genes. We have used these

top discriminative genes with the Genesis software package for subsequent analysis, such as hierarchical clustering and principal component analysis. The clustering diagram shows a clear separation for the BRCA1.X and B0.X classes and also a separation for BRCA2.X versus B0.X. and BRCA1.X versus BRCA2.X (Fig. 1A-C). Principal component analysis also separated the classes with the input data using the same top 80 features in each class comparison as above (Fig. 2A-C). The principal component analysis plot clearly shows that samples in the BRCA1.X class are very similar to each other and cluster together tightly. The BRCA2.X class represents samples with more diverse expression patterns, but clearly, all samples separate well from the control samples. This result confirms our previous finding that expression profiling can be used to predict the genotype of normal cells from BRCA1 mutation carriers, and we can now extend this statement to include cells from BRCA2 carriers. It is difficult to make comparisons between the predictor genes in this and in our previous study (18), as we

have used different clone sets on the cDNA microarrays, but few predictor genes for the BRCA1 genotype are common (*ATM*, *CDKN1B*, and *ADNP*). The previous study used a 6K selected clone set enriched for genes implicated in cancer development, apoptosis, DNA damage repair, and cell cycle. The present study has used a 15 K array, which covers approximately half of the human genome without selection for gene functions. In addition, in the present study, the tissue source was skin biopsy, whereas in our previous study, we used fibroblast cultures established from breast mastectomy specimens.

We can conclude that gene expression profiling after induced DNA damage can distinguish heterozygous BRCA1 and BRCA2 mutation carriers from controls with high accuracy. Two independent analysis (SVM and GPC) done with high statistical significance ($P = 10^{-4}$ to 10^{-12}), particularly the GPC with zero errors. Without induced DNA damage, however, expression profiling is not able to discriminate genotypes. This provides further evidence that both genes actively participate in DNA

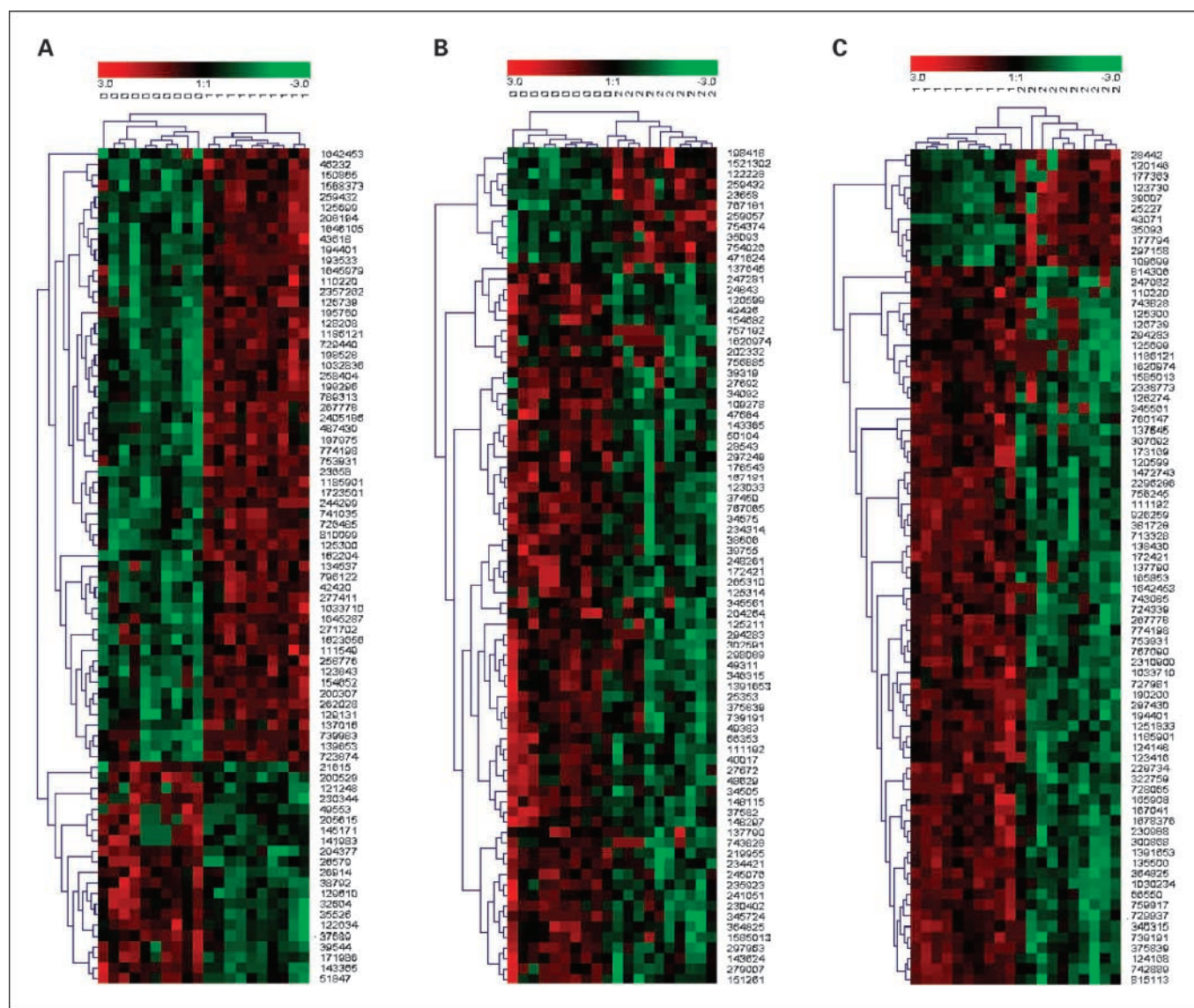


Fig. 1. Hierarchical clustering by Genesis software. The 80 most significant SVM features were used in this analysis for each irradiated class: (A) BRCA1 and controls, (B) BRCA2 and controls, (C) BRCA1 and BRCA2.

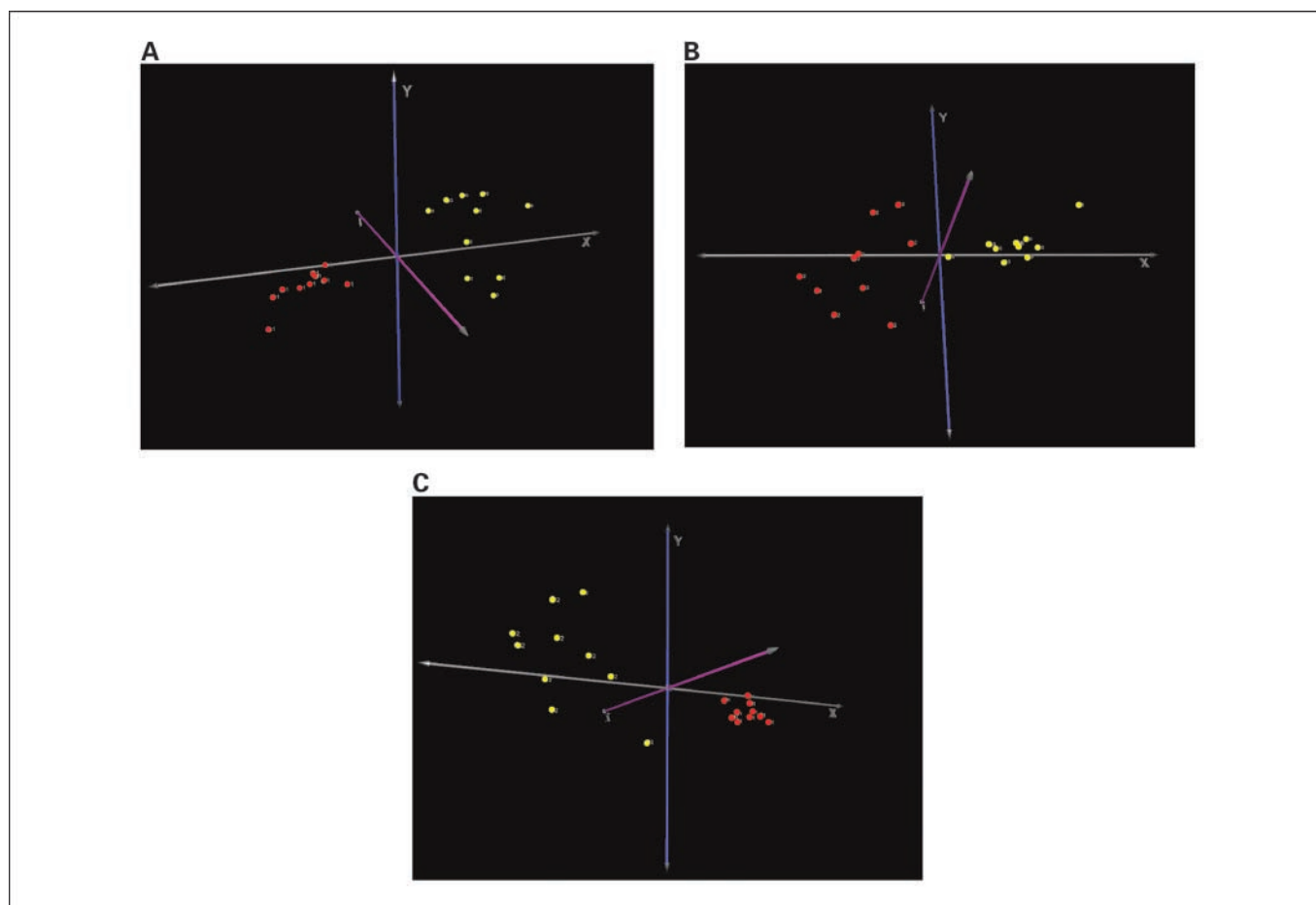


Fig. 2. Principal component analysis by Genesis. The 80 most significant SVM features were used in this analysis for each irradiated class: (A) *BRCA1* and controls, (B) *BRCA2* and controls, (C) *BRCA1* and *BRCA2*.

damage responses, and particularly the *BRCA1* gene, in gene expression regulation. Irradiation generally induces double-stranded DNA damage, and cells need to be in an active responding phase for expression profiling to be able to discriminate between mutation carriers and controls. Based on this study, we propose that gene expression profiling may be used for functional mutation detection in *BRCA1* and *BRCA2* heterozygous samples. It could also be a new and valuable tool for determining the significance of the variants of uncertain significance very often found in these genes. Such sequence variants (missense mutations and intronic variants), regularly reported as results of diagnostic mutation tests and which do not lead to the truncation of the protein products, are very difficult to classify as pathogenic. Counseling and clinical decision making for patients with such variants is a major challenge for clinicians. Much effort has been made towards the classification of these variants of uncertain significance mostly

based on algorithms and modeling the effect of the amino acid change in the proteins (22, 23). Here, we report a method that may lead to a functional test for these genes. Using state-of-the-art classification algorithms, we could correctly predict the genotype of all our 30 samples with confidence, and using the new GPC, we could assign a confidence measure to the class assignment, which could be very useful in future clinical practice. This study needs to be replicated before development as a clinical tool and then needs to be extended to studies of individuals with variants of uncertain significance to determine if these are functionally significant.

Acknowledgments

We thank all the patients and their clinicians for participating in this study, Judith Mills for providing her expertise of establishing skin fibroblast cultures, and Irene Granleese for supporting Caroline and Olive.

References

1. Easton D, Peto J. The contribution of inherited predisposition to cancer incidence. *Cancer Surv* 1990;9:395–416.
2. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* 1994;266:66–71.
3. Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* 1995;378:789–92.
4. Scully R, Livingston DM. In search of the tumour-suppressor functions of *BRCA1* and *BRCA2*. *Nature* 2000;408:429–32.
5. Venkitaraman AR. Cancer susceptibility and the functions of *BRCA1* and *BRCA2*. *Cell* 2002;108:171–82.
6. Marquis ST, Rajan JV, Wynshaw-Boris A, et al. The developmental pattern of *Brcal* expression implies a role in differentiation of the breast and other tissues. *Nat Genet* 1995;11:17–26.

7. Scully R, Chen J, Plug A, et al. Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell* 1997;88:265–75.
8. Sharan SK, Morimatsu M, Albrecht U, et al. Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. *Nature* 1997;386:804–10.
9. Kote-Jarai Z, Eeles RA. BRCA1, BRCA2 and their possible function in DNA damage response. *Br J Cancer* 1999;81:1099–102.
10. Monteiro ANA. BRCA1: exploring the links to transcription. *Trends Biochem Sci* 2000;10:469–74.
11. Pao GM, Janknecht R, Ruffner H, Hunter T, Verma IM. CBP/p300 interact with and function as transcriptional coactivators of BRCA1. *Proc Natl Acad Sci U S A* 2000;97:1020–5.
12. Wang Q, Zhang H, Kajino K, Greene MI. BRCA1 binds c-Myc and inhibits its transcriptional and transforming activity in cells. *Oncogene* 1998;17:1939–48.
13. Preobrazhenska O, Yakymovych M, Kanamoto T, et al. BRCA2 and Smad3 synergize in regulation of gene transcription. *Oncogene* 2002;21:5660–4.
14. Howlett NG, Taniguchi T, Olson S, et al. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* 2002;297:606–9.
15. Perou CM, Jeffrey SS, van de RM, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* 1999;96:9212–7.
16. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001;344:539–48.
17. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
18. Kote-Jarai Z, Williams R, Cattini N, et al. Gene expression profiling after radiation induced DNA damage is highly predictive of BRCA1 mutation carrier status. *Clin Cancer Res* 2004;10:958–63.
19. Campbell C. Kernel methods: a survey of current techniques. *Neurocomputing* 2002;48:63–84.
20. Girolami M, Rogers S. Variational Bayesian multinomial probit regression with Gaussian process priors neural computation (MIT Press) 2006 In press.
21. Sturn A, Quackenbush J, Trajanoski Z. Genesis: cluster analysis of microarray data. *Bioinformatics* 2002;18:207–8.
22. Phelan CM, Dapic V, Tice B, et al. Classification of BRCA1 missense variants of unknown clinical significance. *J Med Genet* 2005;42:138–46.
23. Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* 2006;43:295–305.