Prioritizing Candidate Genetic Modifiers of BRCA1 and BRCA2 Using a Combinatorial Analysis of Global Expression and Polymorphism Association Studies of Breast Cancer

Logan C. Walker and Amanda B. Spurdle

Abstract

Epidemiological evidence from different studies has shown that genes harboring sequence variations may modify breast cancer risk in BRCA1 and BRCA2 mutation carriers. Current attempts to identify genetic modifiers of BRCA1 and BRCA2 associated risk have focused on a candidate gene-based approach or the development of large genome-wide association studies. However, both methods have notable limitations. This chapter describes a novel approach for analyzing gene expression differences to prioritize candidate modifier genes for single nucleotide polymorphism association studies. The advantage that gives this strategy an edge over other candidate gene-based studies is its potential to identify candidate genes that interact with exogenous risk factors to cause or modify cancer, without detailed a priori knowledge of the molecular pathways involved.

Key words: Familial, Breast cancer, BRCA1, BRCA2, Genetic modifiers, Microarray, Irradiation, Genome wide association study, Lymphoblastoid cell lines

1. Introduction

Epidemiological evidence from family-based studies has indicated that genes harboring polymorphisms or mutations may modify breast cancer risk in BRCA1 and BRCA2 mutation carriers (1–3). As a result of this evidence, a number of studies have attempted to identify genetic modifiers of BRCA1- and BRCA2-associated risk by targeting candidate genes in breast cancer-related pathways, such as DNA repair and steroid hormone receptor signaling (4–10). To date, the identification of an SNP (rs1801320) in the 5’UTR of RAD51 has provided the best evidence for a genetic modifier in BRCA2 mutation carriers using the candidate gene approach (11).
Importantly, RAD51 is known to play an important role in the recombinational repair of radiation-induced double-strand DNA breaks (12), particularly through interaction with BRCA2 (13). Since ionizing radiation is known to be an environmental risk factor associated with breast cancer development in BRCA1 and BRCA2 mutation carriers (14), additional genes associated with irradiation response may also be candidates for potential risk modifiers. However, a major disadvantage of using the candidate gene (or candidate SNP) approach for identifying potential risk modifiers is the limited understanding of mechanisms and pathways that underlie breast cancer development in families carrying mutations in BRCA1 or BRCA2. An alternative and powerful approach that can overcome such issues is the use of genome-wide association (GWA) studies to identify candidate SNPs. The analysis of breast cancer risk-associated SNPs identified by a large population-based GWA study (15) has shown that several of these SNPs also appear to modify risk in BRCA1 and/or BRCA2 carriers (16). However, only some of the breast cancer-associated SNPs assessed were shown to modify risk in carriers, and some of those associated with risk showed effects in BRCA2 but not in BRCA1 (16), suggesting that additional approaches are required to more comprehensively identify modifiers of BRCA1 and BRCA2. While GWA studies specifically addressing risk among BRCA1 and/or BRCA2 carriers provide a more direct approach to identifying modifiers of these genes using an agnostic approach, GWA studies require large sample sizes to identify genetic modifiers with confidence. To address this issue, the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA) was established in 2005 to link clinical and epidemiological data from about 30 groups from around the world (17). Collaborative CIMBA projects now have access to more than 10,000 BRCA1 and 5,000 BRCA2 mutation carriers, and the potential power of >80% at a threshold of $P<0.0001$ to detect polymorphisms in BRCA1 carriers with minor allele frequencies >10% that confer risk ratios >1.2 (17). However, the GWA approach is still limited in that study designs involve predefined selection criteria for which SNPs identified from the initial whole genome scan are going to be analyzed in subsequent replication studies, a study design enforced by current genotyping costs. Due to insufficient power to detect low-risk SNPs, it is possible that cancer-associated SNPs that show no obvious correlation in the initial scan may be excluded from further study. Moreover, GWA studies are often limited in information about exogenous risk factors, such as environmental exposures, which confounds any effort to explore the effect of environmental factors in modifying gene–disease associations. Global gene expression analysis as a means to agnostically identify candidates has the potential to enhance the analysis of GWA studies.

We have previously explored the value of analyzing gene expression differences in irradiated lymphoblastoid cell lines from
25

Candidate Genetic Modifiers of BRCA1 Mutation Carriers

BRCA1 or BRCA2 mutation carriers and mutation-negative controls to prioritize candidate modifier genes for polymorphism association studies (18). Genes that discriminated between BRCA1 or BRCA2 mutation carriers and mutation-negative controls were identified and filtered based on cellular function. We theorized that genes which modify BRCA1 and BRCA2 are likely also to contribute to increased risk of developing breast cancer, and so aligned these refined gene lists against data derived through the National Cancer Institute’s Cancer Genetic Markers of Susceptibility (CGEMS) genome-wide association results for a large breast cancer case-control sample set (18). The intention was to identify genes whose expression is associated with BRCA1 and BRCA2 mutation status as a method for prioritizing candidate modifier genes for polymorphism association studies in large cohorts of BRCA1 and BRCA2 carriers. Assessing the effect of exogenous cellular stimulants by mRNA expression analysis was a way to target genes involved in biological pathways perturbed by such exogenous factors. We have used the example of irradiation to identify specific genetic risk factors that interact with this stimulant to cause or modify breast cancer in mutation carriers. However, similar array-based experiments could be designed that use exogenous stimulants to affect other biological pathways that appear to play a role in familial or population-based cancers, and so prioritize candidate genes for further study as potential genetic modifiers in the appropriate sample sets.

2. Materials

1. Epstein Barr virus-transformed lymphoblastoid cell lines derived from breast cancer-affected women in multicase families in which a BRCA1 or BRCA2 mutation had been identified. Samples used in the study were acquired as part of a research project approved by the familial breast cancer resource the Kathleen Cuningham Consortium for Breast Cancer (kConFab), Peter MacCallum Cancer Centre, Melbourne, Australia.

2. Epstein Barr virus-transformed lymphoblastoid cell lines derived from healthy female controls, recruited via the Australian Red Cross Blood Services for ethically approved genetic studies (Queensland Institute of Medical Research, Australia).

3. RPMI-1640 (Gibco Invitrogen) supplemented with 10% Serum Supreme (Lonza BioWhittaker) and 1% penicillin-streptomycin (Gibco Invitrogen).

4. Tissue culture flasks (25 cm²) (Techno Plastic Products).
The methodology describes the use of aligning gene expression data with genome-wide association data from a population-based breast cancer case-control study to identify candidate modifier genes of BRCA1 and BRCA2. This protocol utilizes the Human-6 version 1 BeadChips along with the Illumina Gene Expression System and Illumina BeadStation. Other array platforms (e.g. Affymetrix, Roche NimbleGen, Agilent, spotted cDNA arrays) may be adapted with appropriate probe annotation. An overview of the steps used in the methodology is presented in Fig. 1.

3. Methods

3.1. Cell Culture and Irradiation

1. Culture LCLs in a 25 cm² flask containing supplemented RPMI 1640 medium (20 mL) and passage when approaching confluence (see Note 1).
2. Twenty four hours prior to irradiation treatment, normalize cells to a concentration of 10 × 10⁶ cells per 20 mL of medium. This is carried out by calculating the total cell number using a hematocytometer and discarding excess cells. For example,
in a 20 mL cell culture that contained a total of $16 \times 10^6$ cells, 7.5 mL of cell suspension would be discarded to leave $10 \times 10^6$ cells in 12.5 mL of medium. Transfer these cells to a 50 mL tube and pellet by centrifugation at $300 \times g$ for 5 min. Remove the supernatant by aspiration and resuspend cells in 20 mL of fresh supplemented RPMI 1640 medium by gently pipetting. Transfer the cell suspension to a new 25 cm$^2$ flask and culture cells for a further 24 h.

3. Immediately prior to irradiation treatment, mix cells and transfer 10 mL of cell suspension to a 10 mL centrifuge tube. Centrifuge the cells at $300 \times g$ for 5 min and remove the supernatant by aspiration. Transfer the tubes containing the cell pellet into dry ice for 10–15 min to maintain the integrity of the cellular RNA (see Note 2). Cells can be stored at −80°C until required for RNA isolation. These cells are the untreated LCLs.
4. Irradiate the remaining $5 \times 10^6$ cells within the culture flask at 10 Gy with a GammaCell 40 irradiator that uses a calibrated Cesium-137 source (e.g. 1 Gy/min would deliver 10 Gy in 10 min). Remove the cells from the irradiator and incubate at 37°C for a further 30 min (see Note 3).

5. Pellet the treated LCLs in a 10 mL centrifuge tube and store at −80°C using methods identical to those used for the untreated LCLs in step 3.

### 3.2. RNA Isolation

1. To lyse cells, add 350 µL of RNeasy Buffer RLT and proceed immediately to the next step. Do not thaw frozen cells without first adding Buffer RLT.

2. Homogenize the lysate by passing 5–10 times through an 18-gauge needle using a sterile 1 mL syringe. Once homogenized, use the syringe to transfer the lysate to a 1.5 mL microcentrifuge tube.

3. Isolate and purify total RNA using the RNeasy Mini Kit. We made no modifications to the protocol for the purification of total RNA from animal cells using spin technology. Elute the RNA from the RNeasy spin column using 30 µL of RNase-free H$_2$O.

4. Assess the quality of RNA. We used the Agilent RNA 6000 Nano Chip kit and the Agilent 2100 Bioanalyzer. An RNA integrity number (RIN) greater than nine would be expected after purifying RNA from cell lines (see Note 4).

5. Quantify the RNA using a NanoDrop™ 1000 Spectrophotometer.

### 3.3. RNA Amplification

The Illumina TotalPrep RNA Amplification Kit is used to amplify and biotinylate the LCL derived RNA. The protocol for this procedure is well described in the instruction manual. Samples are most efficiently processed in batches but this approach can cause nonbiological variation or “batch effects” as evidenced by downstream microarray data analysis (see Note 5). Although statistical methods have been proposed that reduce “batch effects” (19), such effects are minimized if samples are processed in random order to prevent batches being overrepresented by samples belonging to a single biological class (e.g. \textit{BRCA1} or \textit{BRCA2} mutation status). It is also important that the incubation time used for the in vitro transcription step is the same for all samples.

### 3.4. Microarray Analysis

We have used Sentrix® Human-6 version 1.0 BeadChips that allow the simultaneous assessment of expression profiles of $>46,000$ transcripts (including $>23,000$ RefSeq fully annotated genes) for six RNA samples in parallel. Illumina have since superseded this version BeadChip (currently version 3.0) by redesigning
probe sequences derived from successive updates of NCBI RefSeq and UniGene databases. To reduce costs of microarray analysis, 8- or 12-sample Illumina BeadChips could be used in place of the 6-sample BeadChip.

3.4.1. cRNA Hybridization, Washing, and Signal Detection

Hybridize the cRNA samples to the BeadChip following the protocol that accompanies the Illumina Gene Expression System. Each step is adequately detailed from hybridization through to scanning and needs no modification (see Note 6). The protocol will differ slightly according to whether the 6-, 8-, or 12-sample BeadChip is being used.

3.4.2. Data Extraction and Preanalysis Processing

After the BeadChips are dried by centrifugation, scan each array by loading up to three BeadChips in the BeadArray Reader special tray (Illumina has since superseded this scanner with their new iScan System). Ensure that only the Cy3 channel is enabled and each scan uses the following settings: Gain = 2, PMT = 545 and Filter = 100%. Once all the slides have been scanned, import the data into BeadStudio (currently version 3.4) to form a single project containing all samples. Export the Sample Probe Profile to a *.txt file (see Note 7) ensuring the average signal intensity (AVG_Signal) and Detection_Score corresponding to each probe is included for each sample. Import the data into GeneSpring GX (Agilent Technologies; currently version 10.0) and normalize using the option of per chip normalization to the 50th percentile and per gene normalization to the median. Other normalization methods, such as quantile normalization, may also be considered at this stage. Filter the probes by excluding all that have an Illumina detection score of <0.99 in at least one sample. This will yield a subset of probes to use in downstream analyses.

3.5. Identify BRCA1- and BRCA2-Associated Genes

Identify genes that are differentially expressed between irradiated LCLs of mutation-negative controls and of the BRCA1 and BRCA2 classes. For this analysis, we import the normalized data into BRB-ArrayTools (http://linus.nci.nih.gov/BRB-ArrayTools.html) to utilize the multivariate permutation test that controls for false discoveries. The proportion of false discoveries allowed in the analysis can be selected by the user. We typically perform the test to provide 90% confidence that the number of false discoveries does not exceed 10%. To limit the analysis to the most differentially expressed genes, a stringent \( \alpha \)-level threshold (i.e. \( \leq 0.001 \)) is also applied to each two-sample \( T \)-test performed.

3.6. Identify Irradiation Response Genes

Irradiation response genes are defined as those genes whose expression levels change as a result of LCL exposure to irradiation. There are two methods which could be used to identify irradiation response genes:
Option 1: Compare microarray expression profiles of the treated (irradiated) and untreated normal LCL pairs using the paired T-test option. As in the previous section, restrict the number of false discoveries to 10% and set a $\alpha$-level threshold of 0.001. This gene list contains the irradiation responsive genes.

Option 2: In our study (18), we had expression data from irradiated LCLs derived from healthy controls, but we did not have expression data from the same LCLs prior to treatment. We were therefore unable to generate a list of irradiation responsive genes and so we utilized a gene list from a published study (20) (see Note 8). In this situation, where data is compared across two different arrays, it is necessary to exclude all genes that are not represented on both arrays.

3.7. Identify All Genes Tagged by SNPs in the CGEMS Database

1. Use gene symbols approved by the Human Genome Organization (HUGO) Gene Nomenclature Committee (www.genenames.org) to link the BRCA1-, BRCA2- and irradiation-associated gene lists with the CGEMS Breast Cancer GWA (currently Phase 1) dataset.

2. Where gene lists contain aliases for some gene symbols, convert each alias to an official symbol. Use the “Batch Lookup” tool in MatchMiner (21) or an equivalent annotation tool to translate the gene symbols and standardize the gene lists.

3. To identify all BRCA1-, BRCA2- and irradiation-associated genes tagged by SNPs in the CGEMS database, separately enter the HUGO gene symbol lists into the CGEMS Association Finding dataset. Select the Covariate Adjustment to return a $p$-value ($P \leq 0.05$) for an adjusted score test (see Note 9). The output table from CGEMS will contain a column with SNP-associated (tagged) gene symbols.

4. In many cases, the HUGO gene symbol will be listed alongside one or more aliases. Therefore, separate the terms so that the official symbol can be used as a search term. One way to do this is to import the output table as a *.txt file into Excel and separate each symbol using an “|” as a delimiter. This process will likely identify a high proportion of genes that are tagged with one or more SNPs. These SNP-tagged genes are to be retained for further analysis.

3.8. Alignment of GWA Study Association Results with BRCA1- and BRCA2-Associated Genes Categorized by Irradiation

1. Classify irradiation responsive genes in the BRCA1- and BRCA2-associated gene lists by aligning data against the irradiation gene list.

2. For each irradiation responsive and nonirradiation responsive gene list, identify those genes with $\geq 1$ SNP-tags (i.e. genes reported by CGEMS to be associated with breast cancer risk ($P < 0.05$)) and calculate the proportion of these genes in their respective gene lists.
3. Apply the chi-squared statistic to test the null hypothesis that irradiation responsive genes whose expression correlated with \textit{BRCA1} and/or \textit{BRCA2} mutation status are no more likely to be tagged by risk-associated SNPs in the CGEMS dataset.

4. To exclude potential bias as a result of gene size, obtain the transcriptional start and stop sites for each gene in their respective gene lists using the annotation tool MatchMiner and determine whether the average gene sizes differ significantly between the irradiation and nonirradiation response groups.

If genes defined by gene expression analysis to play an important role in irradiation response are potential breast cancer risk modifiers, we would predict that breast cancer associated SNPs would be more likely to be linked to these genes. These genes (and the breast cancer associated SNPs in them) would be suitable for prioritization as potential modifiers of \textit{BRCA1} and \textit{BRCA2} in large sample sets of \textit{BRCA1} and \textit{BRCA2} carriers.

4. Notes

1. We suggest that LCLs grown from frozen stocks be cultured over a period of at least 2 weeks to ensure that the cells no longer exhibit growth irregularities as a result of being cryogenically stored in DMSO.

2. Batching LCLs for treatment is an efficient way of carrying out the experiment; however, it is important to either freeze or homogenize the treated (or nontreated) cells in Buffer RLT immediately after centrifugation. Allowing cells to sit at room temperature increases the possibility of RNA degradation and may also allow gene expression levels to change so that treated LCLs may no longer reflect an irradiation induced profile.

3. A relatively early time-point of 30 min post irradiation was chosen to target early transcriptional response and to minimize possible downstream compensation effects. It has previously been shown that 10 Gy irradiation treatment of normal LCLs has an effect on the transcriptional response, with greatest change in mRNA levels for most genes within 1 h post-treatment (20). Interestingly, a study of mouse brain gene expression after whole-body low-dose irradiation also showed that expression changes in a large number of irradiation response genes can be measured at the 30 min time point (22).

4. RIN < 9 indicates partial RNA degradation. To avoid this, see Note 2.
5. We have observed expression profile differences between RNA amplification sample batches but not between array hybridization batches. These experimentally induced differences are likely to be enhanced when the biological difference between the samples with respect to gene expression magnitude is relatively small, which is the case with LCLs.

6. According to the protocol, the hybridization step can be between 16 and 20 h. We believe that to minimize batch bias, it is important that the incubation time used for the hybridization step is kept consistent for each BeadChip. We adhered strictly to an incubation time of 19 h and always found a high correlation ($r > 0.98$) for duplicate arrays.

7. Data can also be exported as a “Sample Gene Profile”. The “Sample Gene Profile” format takes the average value of all probes that target a specific gene regardless of whether the probe sequences differ. Therefore, probes that target specific mRNA isoforms for a particular gene may be combined with a probe that makes no such discrimination. For this reason, we have chosen to work with probe-based profiles.

8. A disadvantage with this method is that it might be difficult to find a study that has employed the similar treatment/post-treatment conditions (e.g. dose, posttreatment incubation period).

9. The rationale for choosing a nominal $P$-value cut-off of 0.05 was based on the following: (1) Comparing data from the breast cancer genome-wide association study of Easton et al. (15) with those from CGEMS revealed that two of the five top-ranking SNPs in named genes (TNRC9, LSP1) from Easton et al. had $P$-values greater than 0.05 in CGEMS, suggesting that a more stringent cut-off may exclude potential candidate risk associated SNPs; (2) Approximately 15,000–20,000 of the top-ranking SNPs from CGEMS Phase I will be selected for the Phase II replication study (http://cgems.cancer.gov/about/executive_summary.asp), and will therefore carry a $P$-value less than 0.04.

Acknowledgments

We thank Nic Waddell, Anette Ten Haaf, Georgia Chenevix-Trench, Denis Moss and kConFab for their contribution in developing the protocols described here. We also thank Nic Waddell for critical reading of the manuscript. This work was supported by the Susan G. Komen Breast Cancer Foundation and the NHMRC.
References


Cancer Susceptibility
Methods and Protocols
(Ed.) M. Webb
2010, XI, 295 p. 40 illus., 3 in color., Hardcover
ISBN: 978-1-60761-758-7
A Humana Press book