Mol Cell Toxicol (2017) 13:239-249
DOI 10.1007/s13273-017-0026-5

ORIGINAL PAPER

Meta-analysis of microarray and RNA-Seq gene expression datasets for carcinogenic risk: An assessment of Bisphenol A

Junghyun Jung[1], Changsoo Mok[1], Woosuk Lee[1] & Wonhee Jang[1]

Received: 26 May 2017 / Accepted: 30 May 2017

Abstract Bisphenol A (BPA) is an endocrine-disrupting chemical that is related to many diseases, including heart attacks and diabetes. Recently, several studies have reported the carcinogenic potential of BPA in rodents, yet carcinogenic effects of BPA in humans remains unclear. In this study, meta-analysis was applied to independent GEO datasets, based on 158 Affymetrix microarrays and 8 Illumina RNA-Seqs. Additionally, we performed functional enrichment analysis, disease similarity analysis based on Disease Ontology (DO) analysis, and network analysis. 1,993 (1,457 up-, 536 down-regulated) differentially expressed genes (DEGs) were identified from five GEO datasets by adjusting for batch effects. Using disease similarity analysis, we demonstrated that results of DO analysis of the top 20 diseases were highly related to breast cancer. Moreover, we showed that the DEGs were significantly enriched in gene expression datasets on human breast cancer tissue via gene set enrichment analysis. By performing network analysis, we finally identified 85 (68 up- and 17 down-regulated) DEGs, and some of their expression levels were validated by quantitative PCR. The identified DEGs were regarded as genetic markers for carcinogenic risks, indicating that BPA may be a potential carcinogenic chemical contributing to the cause of breast cancer in humans.

Keywords: Bisphenol A, Carcinogenic risks assessment, Toxicogenomics, Meta-analysis, Network analysis

Electronic supplementary material The online version of this article (doi: 10.1007/s13273-017-0026-5) contains supplementary material, which is available to authorized users.

[1]Department of Life Science, Dongguk University, Seoul 04620, Korea
Correspondence and requests for materials should be addressed to W. Jang (✉wany@dongguk.edu)

Introduction

For over 50 years, Bisphenol A (BPA) has been used as a raw material to make polycarbonate plastics for producing compact discs, auto parts, baby feeding bottles, plastic containers, coating materials for glass lenses, and shock absorbers[1,2]. It has also been used to synthesize epoxy resins that make linings of canned food, food packaging materials, and dental resins[3]. Because BPA released from polycarbonate plastics and epoxy resins can mimic estrogen hormones *in vivo*, safety issues have emerged[4,5]. *In vitro*, *in vivo*, and epidemiologic studies of BPA show that it is related to several diseases and adverse effects, including heart attack and coronary heart disease[6], cardiovascular disease[7], diabetes[8], arterial disease[9], and obesity[10]. Several researchers reported that prenatal exposure to BPA causes potential carcinogenic risks in rodent mammary tissues[11-13].

These studies provided valuable insight into the risks of BPA in human health; however, the carcinogenic risks of BPA in humans remains unclear[14,15]. Since 2007, diverse rat and mice *in vivo* studies have suggested that there is a significant increased cancer incidence rate in the female mammary gland and an increased number of prostatic intraepithelial neoplasms in male prostate organ[16,17]. In order to understand the mechanisms of BPA in carcinogenesis, a hybridization-based microarray technique can be a very powerful tool in the detection of the alteration of gene expression patterns[18,19]. The development of cDNA sequence-based approaches, such as RNA sequencing (RNA-seq), additionally allows the quantification of transcriptomes, as well as the detection of unannotated transcripts and isoforms[20,21]. Public microarray and RNA-Seq databases such as ArrayTrack, NCBI GEO, SRA, ArrayExpress at EBI, and the Stanford Microarray database are

available for study. The microarray- and RNA-Seq-based toxicogenomics data have also been continuously accumulating at public repositories and are currently available, providing useful information for scientific research and discovery[22]. The rapid growth of a variety of public databases and diverse bioinformatics tools make it possible to integrate heterogeneous datasets in different studies and platforms, as well as to identify underlying trends, which were previously very hard to determine from a single study.

In this study, a meta-analysis of the combined gene expression datasets containing data on the effects of BPA, obtained from the NCBI GEO, was performed. After pre-processing, we adjusted differences that can be considered as "batch effects" in each of the datasets. Batch effects are induced by time- and place-dependent, non-biological experimental variations, and are detected during combining multiple datasets[23,24]. In meta-analysis, adjusting for batch effects is an inevitable process in order to conduct an appropriate analysis free of non-biological variations, as heterogeneous multiple datasets are combined in order to increase statistical power during the process. The Rank Product method was performed, which offers the "rank" of genes by using geometric means in replicate microarray datasets[25]. The Rank product is a simple and powerful method to process heterogeneous datasets.

Herein, we conducted a meta-analysis on the carcinogenic risk assessment of BPA by combining microarray and RNA-seq gene expression datasets using the Rank Product method. The aims of this study were 1) to combine microarray and RNA-Seq data by adjusting for batch effects, 2) to identify robust, differentially expressed genes (DEGs) using the Rank Product method in the combined datasets, 3) to conduct enrichment analysis using Gene Ontology (GO), KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways, and disease similarity heatmaps based on Disease Ontology (DO), 4) to confirm correlations between our meta-analysis of DEGs and gene expression datasets obtained from cancer tissues using the Gene Set Enrichment Analysis (GSEA), and 5) to identify key DEGs through network analysis in order to demonstrate the carcinogenic risk of BPA.

Results

Dataset collection and preprocessing of BPA related toxicogenomic data

To identify consistently regulated genes upon BPA exposure across different studies, the GEO database was queried for expression studies of microarray and

Table 1. Toxicogenomics datasets for meta-analysis.

GEO series ID	GEO platform ID	Cell line	Dose of BPA	No. of arrays (Control : Treatment)	BPA-treatment time
GSE26884[50]	GPL 570 (Affymetrix Human Genome U133 Plus 2.0 Array)	MCF-10F (Normal-like human breast epithelial)	1 nM, 10 nM	6 (2 : 4)	Continuously two weeks, adding fresh media every day
GSE32158[50]		MCF-10F (Normal-like human breast epithelial)	1 nM, 10 nM	8 (2 : 6)	Continuously two weeks, adding fresh media every day
GSE17624[51]		Human Ishikawa (Endometrial adenocarcinoma)	1 pM, 100 pM, 10 nM, 1 μM	20 (4 : 16)	8 h
				20 (4 : 16)	24 h
				20 (4 : 16)	48 h
GSE50705[29]		MCF-7 (Mammary gland/ Breast cancer)	1.95 pM ~50 μM	84 (6 : 78)	48 h
GSE38234[52]	GPL11154 (Illumina HiSeq 2000)	T47D (Mammary gland/ Breast cancer)	100 nM	4 (2 : 2)	8 h
		ECC1 (Endometrium adenocarcinoma)	100 nM	4 (2 : 2)	8 h

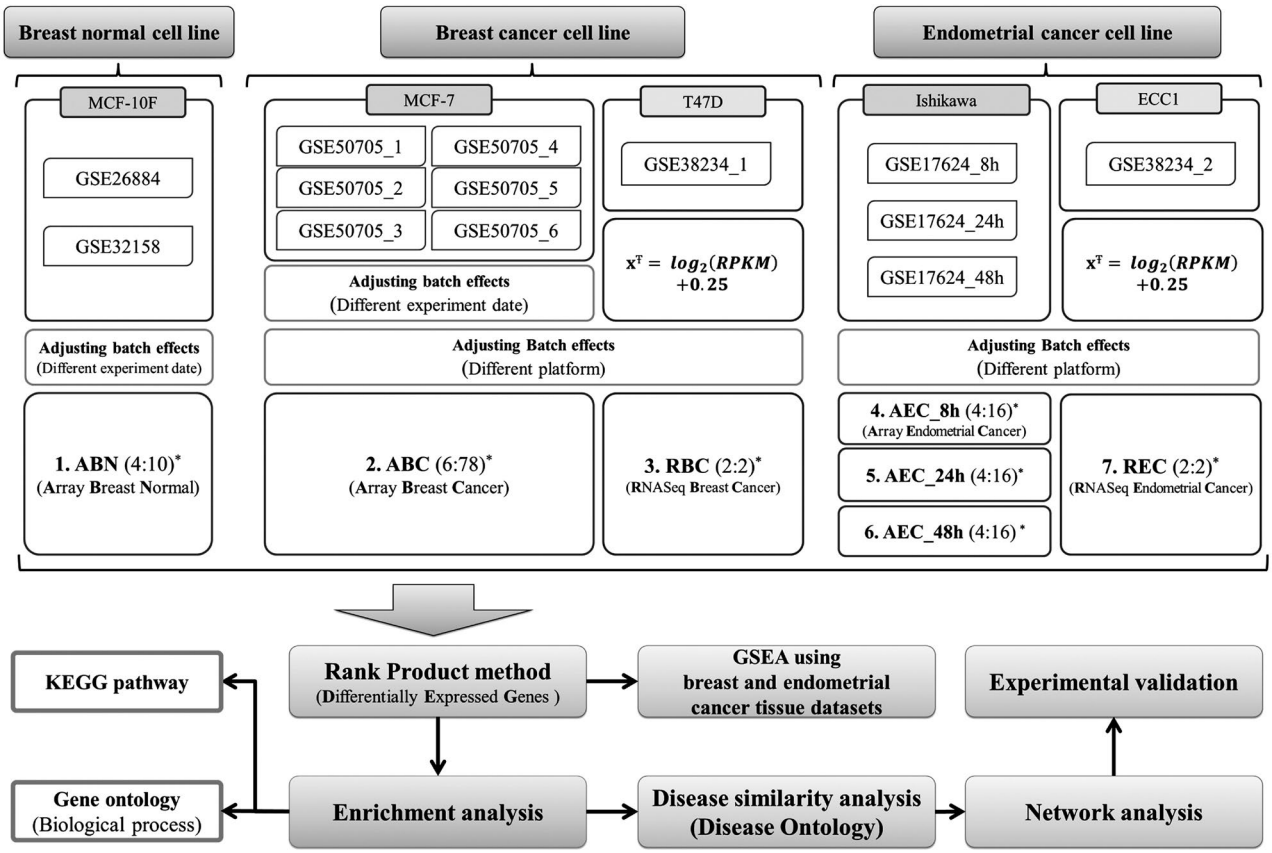

Figure 1. A flow chart of the meta-analysis process. The microarray and RNA-seq datasets were retrieved from NCBI GEO. All procedures were performed with R language using Bioconductor open-source packages and Cytoscape. *indicates the number of arrays (Control: Treatment). Ŧ indicates that 0.25 was added to the value to avoid the occurrence of $-\infty$ singularity at RPKM = 0.

RNA-Seq. The collection of datasets was exhibited in 5 GSEs (GEO Series), 2 GPL (GEO Platform) of 162 GSMs (GEO Samples) (Table 1). The 4 GSEs (GSE26884, GSE32158, GSE17624, and GSE50705) were from the Affymetrix Human Genome U133 Plus 2.0 Array (GPL 570) and 1 GSE (GSE38234) was from the Illumina HiSeq 2000 (GPL11154). The expression profiling studies contained 140 samples of BPA treatment and 26 control samples (Table 1). The multiple datasets were combined to perform a meta-analysis. To overcome platform differences of microarray and RNA-Seq data, a value of RNA-Seq RPKM was transformed using the Mooney *et al*.[26] method. The details of this method are described in the Materials and Methods section. Although the RNA-Seq data were transformed to integrate microarray data, we had a study-dependent variation called batch effects, which are regarded as non-biological deviations that occur when combining multiple datasets to conduct meta-analyses (Figure 1).

Meta-analysis for identification of consistently regulated genes

All of the adjusted gene expression datasets integrating 5 GSEs were analyzed to identify DEGs in replicate microarray experiments. Even after non-biological batch effects were adjusted for in 5 heterogeneous datasets, the biological differences still remained, such as differences in cell lines based on the origin of tissues and platform types. In order to identify genes that are only affected by BPA across the studies, we used the Rank Product algorithm, which is one of the most robust methods to conduct meta-analysis for identification of consistently regulated genes. Five different origins of tissues and two platform types were used to conduct the meta-analysis (Figure 1). The number of combined datasets of microarray and RNA-Seq data sorted by unique Entrez ID was counted to be 17,639 genes in total, and 1,993 (1,457 up- and 536 down-regulation) genes were identified (Figures 2B, S1B and Table S1) as relevant. Even with the higher power of

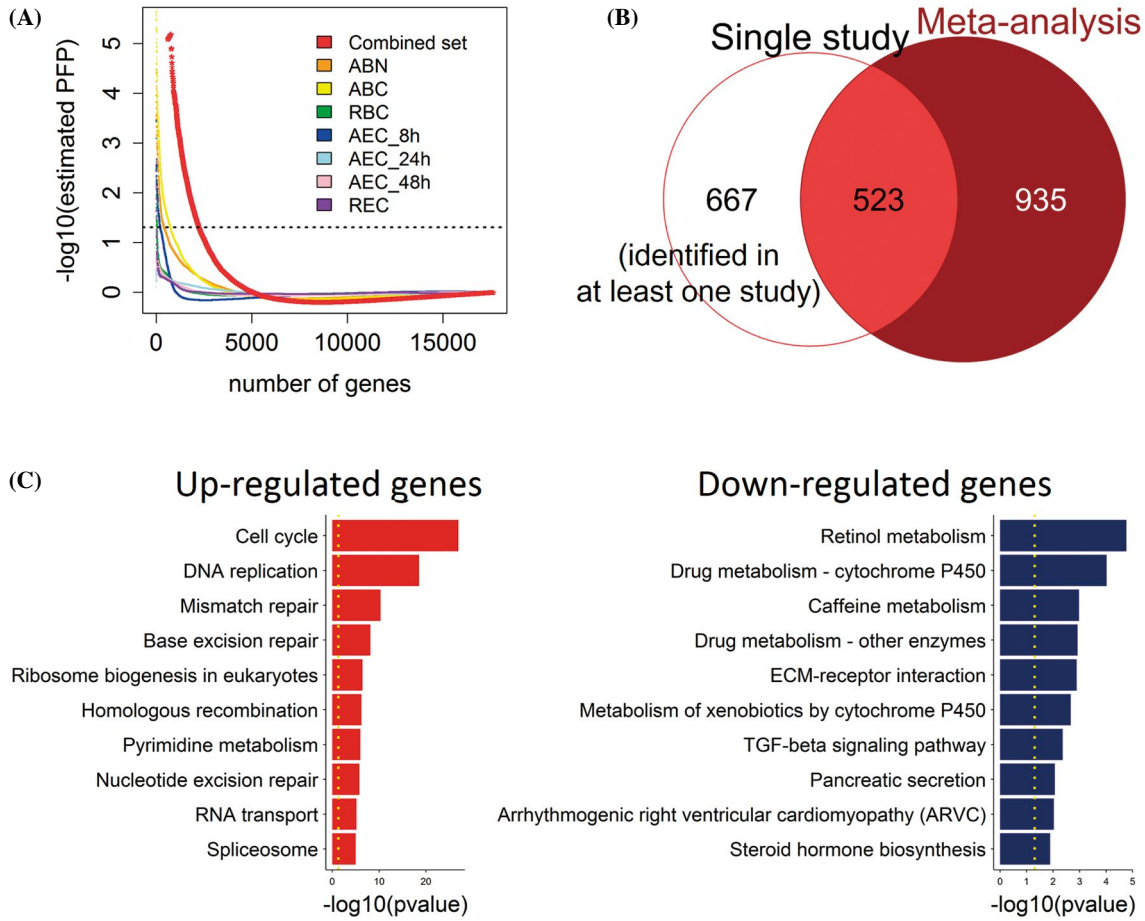

Figure 2. The results of meta-analysis and functional enrichment analysis. (A) A plot showing the statistical power of meta-analysis based on number of up-regulated genes and the corresponding pfp. Dotted lines indicate the threshold of DEGs. (B) A Venn diagram showing the overlap between up-regulated genes identified by at least one study and the meta-analysis (pfp <0.05). (C) Enrichment analysis of the KEGG pathway using up-regulated DEGs ($P<0.05$). Dotted lines indicate the threshold of DEGs. (D) Enrichment analysis of the KEGG pathway using down-regulated DEGs ($P<0.05$).

the meta-analysis, more genes were identified in combined datasets compared with each single study (Figures 2A and S1A). Notably, the Venn diagram showed that some genes were identified only in the meta-analysis (Figures 2B and S1B).

Functional enrichment analysis

To obtain insights into the biological interpretation of DEGs derived from meta-analysis, the selected up- and down-regulated DEGs were assigned to GO and KEGG pathway functional enrichment analysis. The GO and KEGG pathways provide a descriptive framework and functional annotation to the understanding of the biological roles of the DEGs[27]. The up-regulated genes were significantly over-represented in the cell cycle, nuclear division, and chromosome segregation, whereas the down-regulated genes were significantly enriched in multicellular organism processes, cell differentiation, and cell migration section ($P<0.01$) (Figure S1C). The over-represented KEGG pathway terms were associated with cell cycle, DNA replication, and several repair pathways in up-regulated genes, while retinol or drug metabolism pathways were found to be enriched in down-regulated genes ($P<0.05$) (Figure 2C).

Identifying connection of cancer and BPA

Generally, deregulation of the cell cycle is the basis of abnormal cell proliferation characterized by cancer[28]. The fact that the cell cycle, DNA replication, and several DNA repair pathways are significantly related to BPA treatment (Figures 2C and S1C) suggested that BPA exposure may be involved in tumorigenesis. To confirm whether our DEGs derived from meta-analy-

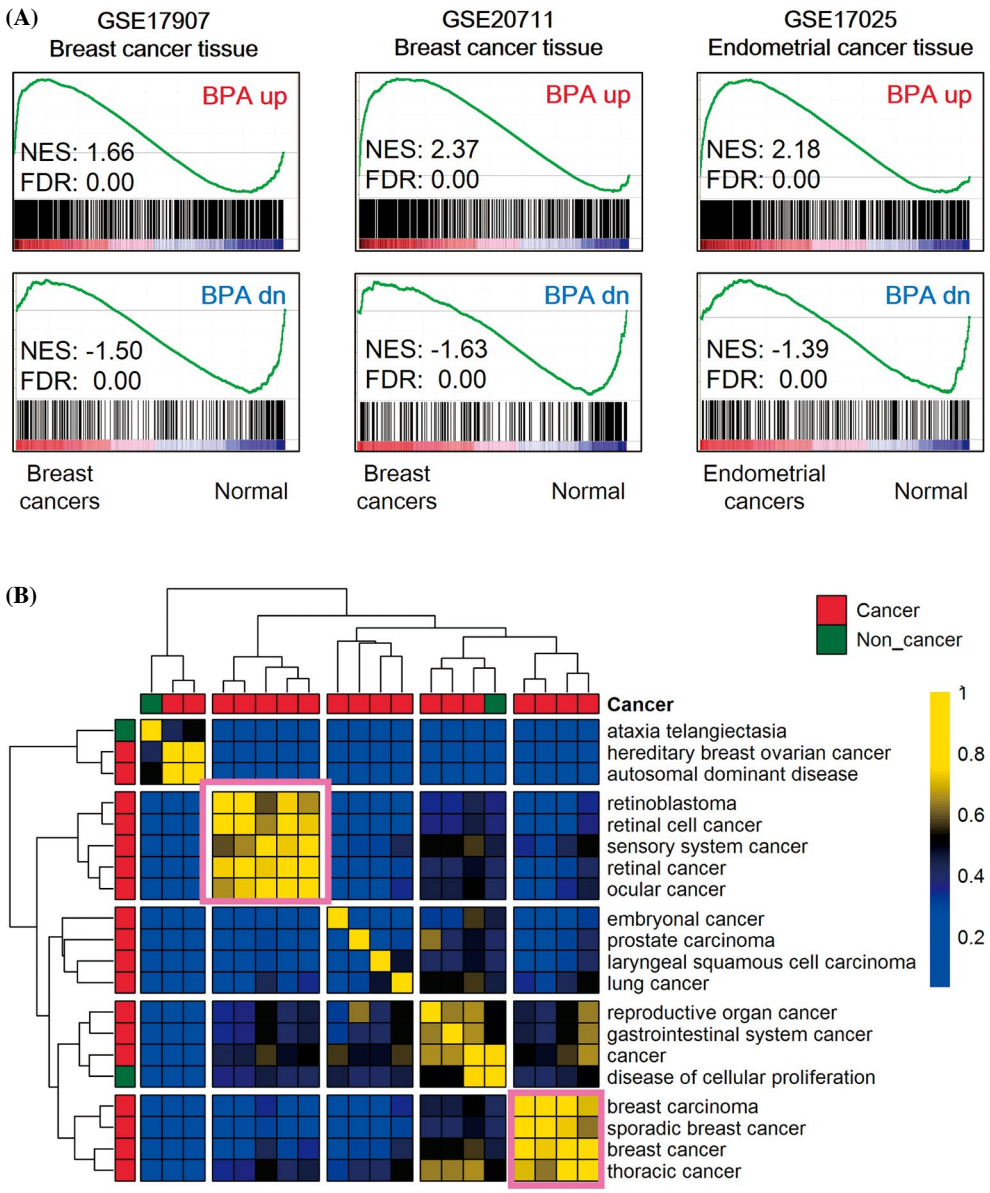

Figure 3. The correlation between BPA and cancer. (A) A GSEA plot showing that the DEGs from meta-analysis were significantly enriched in breast (GSE17907 and GSE17907) and endometrial (GSE17025) cancer datasets derived from human tissue. (B) The heatmap for disease similarity based on DO enrichment analysis of the BPA-related DEGs. The top 20 diseases of DO enrichment analysis were represented in each column and row of this heatmap. The gold color indicates greater similarities among the diseases. Column annotation of red and green colors also indicate cancer and non-cancer diseases, respectively. Using hierarchical clustering, the subgroups based on high similarity of DO terms are shown in the pink square.

ses were associated with various cancers, GSEA using cancer datasets was performed. Our gene sets of DEGs were applied to gene expression datasets from human breast cancer tissues. The GSEA results showed that up- and down-regulated DEGs were significantly enriched in two independent microarray datasets for human breast cancer tissues. Additionally, our DEGs were also enriched in gene expression datasets derived from human endometrial cancer tissues (Figure 3A).

Along with the GO and KEGG pathways, Disease ontology (DO) enrichment analysis was conducted to identify particular diseases related to DEGs under BPA exposure. Disease similarity analysis was also performed using the top 20 most significant diseases based on DO enrichment analysis. This provides a guideway to establish disease-disease relationships in order to fur-

ther elucidate the similarities among diseases, as well as to highlight variables of subgroups of diseases by identifying gene function in the perspective of a disease. The results of disease similarity analysis were shown as a similarity heatmap based on the top 20 diseases (Figure 3B). Among top 20 DO terms, 90% (18 of 20 terms) were associated with cancer. Notably, the terms of 5 eye-related and 4 breast-related cancers that contained 110 (80 up- and 30 down-regulated) DEGs were clustered by showing a high similarity via hierarchical clustering.

PPI network analysis and qPCR validation

Using the 110 DEGs derived from the disease similarity analysis, a PPI network was constructed to understand the interactions and molecular mechanisms of the DEGs. The network contained the 85 (68 up-, 17 down-regulated) proteins, including 12 transcription factors (TFs) which interact with one another (Figure 4A and Table S2). Using GO enrichment analysis, 3 main terms that overlapped with 5 TFs and 2 proteins were significantly related to cell cycle ($P < 10^{-23}$), response to chemical stimulus ($P < 10^{-20}$), and regulation of transcription ($P < 10^{-6}$). BRCA1 (BRCA1, DNA repair associated, Degree = 40), TP53 (tumor protein p53, Degree = 35), MYC (MYC proto-oncogene, Degree = 28), and SMAD3 (SMAD family member 3, Degree = 14) were identified in the core of the protein interaction network (Figure S2). In order to confirm that the genes are actually up- or down-regulated as predicted in the study, we performed a qPCR analysis on 5 genes (*BRCA1*, *BRCA2*, *TP53*, *TWIST1*, and *SMAD3*) using a normal human mammary epithelial cell line, MCF-10A. The validation results showed that the qPCR results were consistent with our results of meta-analysis and PPI network (Figure 4B).

Discussion

Many products made of BPA have been widely used for many decades. It has been widely reported that BPA can mimic estrogen in mammals and create various problem *in vivo*[4]. Although BPA does not exert strong estrogenic effects when compared to other endocrine-disrupting chemicals[29,30], it has been reported to be a potential carcinogen in mice[15,16]. Therefore, it has become a matter of interest to understand the underlying carcinogenic effects of BPA on humans. Here we demonstrate that BPA may indeed be carcinogenic in humans as shown in the enrichment analysis of the KEGG pathway, GO, and disease similarity heatmaps based on DO, using meta-analyses of heterogeneous microarray sets.

The meta-analysis for risk assessment combining several toxicogenomics datasets archived by public databases has been useful approaches in the identification of the underlying mechanisms of toxic chemicals[31]. The 5 different GSEs consisting of 1 normal cell line (MCF-10F) and 4 different cancer cell lines (Ishikawa, MCF-7, T47D, ECC1), divided into two different platforms, were used to analyze the potential carcinogenic risk of BPA in humans. The Illumina RNA-Seq and Affymetrix microarray platforms had a high correlation[32] and it has been reported that the Combat function for adjusting for batch effects performed better than other methods[33]. Our combined datasets were highly heterogeneous due to non-biological effects, so batch effects were adjusted for using the Combat function, and a robust RP algorithm was utilized to overcome these heterogeneities.

Our results indicated that up- and down-regulated DEGs were significantly enriched in human breast cancer tissues, suggesting that BPA may have carcinogenic effects on humans (Figure 3A). Via PPI network analysis, SMAD3, a putative key regulator, was identified (Figures 4 and S2). Previous research has shown that down-regulation of *SMAD3* using RNAi is responsible for tumorigenesis and metastasis in the human breast cancer cell line[34]. Additionally, PPARG is directly connected with SMAD3 (Figures 4 and S2A), and is a therapeutic target of breast cancer, as antagonist chemicals to suppress cell proliferation and motility can be used[35].

Our study has some limitations. First, BPA may not dose-dependently decrease or increase mRNA expression levels. Second, the expression of mRNA levels may have different expression patterns in different cell lines. Third, although qPCR was performed to validate the expression levels under BPA exposure, further research is necessary. Despite these limitations, our meta-analysis, which combined different studies adjusting for batch effects, was able to detect genes that were missing from the analysis of single study cases. Although a more thorough experimental study is necessary to confirm the carcinogenic effects of BPA, our results provide a robust set of genetic markers, including *BRCA1*, *BRCA2*, *TP53*, *TWIST1*, and *SMAD3* under BPA exposure.

In conclusion, we identified that the 1,993 robust genetic markers of BPA via the meta-analysis based on the gene expression profile of BPA exposure datasets. Although the DEGs derived from meta-analysis and their biological functions suggested that BPA-exposure is implicated in human breast carcinogenesis, further experimental studies are necessary to confirm whether our selected DEGs are indeed valid markers

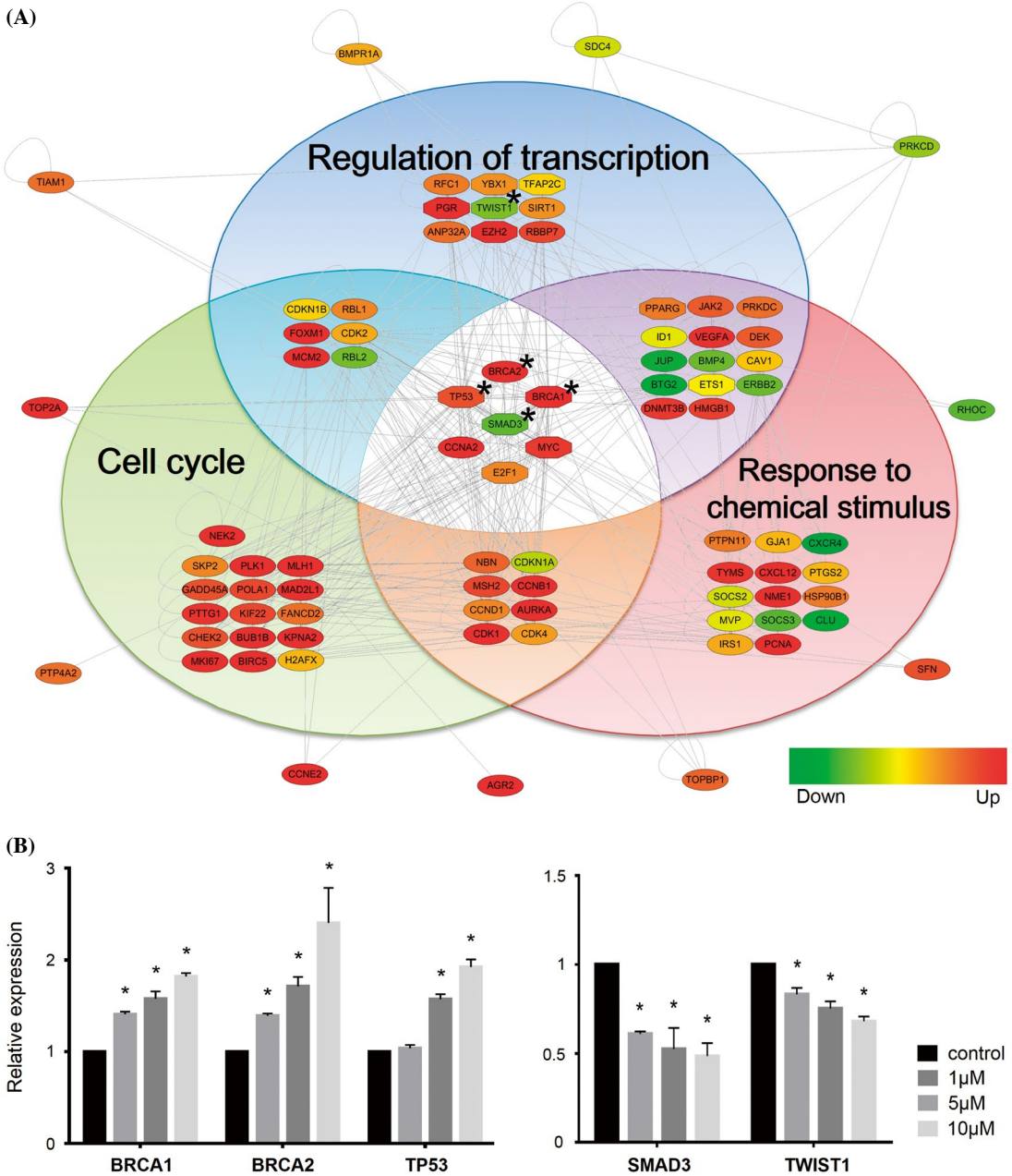

Figure 4. The construction of a protein-protein interaction network. (A) The protein network showing that nodes and edges represent proteins and their interactions, respectively. Proteins are indicated as ellipses and transcription factors as octagons. The node color represents the expression level of DEGs, where red color indicates up-regulation and green color indicates down-regulation. The full network obtained through similarity analysis of DO (Disease Ontology) reveals 85 nodes and 386 edges. The biological process terms of GO are represented by a Venn diagram, indicating significant over-representation of 3 BP terms. Black asterisks (*) represent genes that were validated by qPCR. (B) Bar graphs showing that qPCR validation. Data are presented as means ± SEM ($n \geq 3$). *indicates a significant difference compared with the control ($P < 0.05$).

for BPA-induced breast cancer. We believe that this research may provide additional insights into the risk assessment of BPA in humans.

Materials & Methods

Dataset collection of BPA related toxicogenomic data

The microarray gene expression datasets necessary for

meta-analysis to assess the carcinogenic risk of BPA were obtained from the NCBI GEO. The organism was *Homo sapiens* and the raw datasets for the research were limited to the GEO Series (GSEs) that are only related to BPA with expression profiling by Affymetrix array and Illumina high-throughput sequencing.

Data pre-processing

The raw datasets of microarray GSE were handled through the Bioconductor *affy* package[36] in R. The expression values were normalized by the RMA (Robust Multi-array Average) algorithm[37], and thus the processed datasets were given log2-transformed values. Each dataset was sorted by unique Entrez ID using *hgu133plus2.db*, which is the annotation package of GEO platform (GPL) 570. If duplicate Entrez IDs were detected, they were substituted by the average value of themselves. Before integrating the RNA-Seq and microarray data, normalized reads per kilobase of exon model per million mapped reads (RPKM) containing values of RNA-Seq data were transformed following Mooney *et al*.[26], using equation 1.

$$x = log_2(RPKM) + 0.25 \quad (1)$$

0.25 was added to avoid the occurrence of $-\infty$ singularity at RPKM = 0. The transformed RNA-seq data were combined with microarray datasets with unique Entrez ID.

Adjusting for batch effects

The surrogate variable analysis (*SVA*) R package[38] was used to adjust for batch effects of combined datasets from different GSEs to remove non-biological batch effects arising from different procedures, platforms, and to prevent the removal of meaningful biological effects. The Combat function in the SVA package, a parametric empirical Bayes method[24], was used for adjusting for known batch effects.

Differentially expressed genes using the Rank Product (RP)

The Rank Product analysis was used for calculating Rank Product (RP) values that were based on the fold change (FC) of the rank using the *PrankProd* R package[39]. Before the analysis, each batch was re-allocated to a respective control group and BPA-treatment group in order to calculate RP values to account for 7 separate origins, based on the difference in cell lines and platforms (Figure 1). After all of the genes in each batch were sorted by RP values, the geometric mean ranks were calculated across the origins. The up- and down-regulated differentially expressed genes (DEGs) were selected in the estimated percentage of false positive predictions (pfp) through 200 permutation tests. Because the Rank Product is a rank-based method and the FC values were heterogeneous, an FC value greater than 1 was identified as an up-regulated DEG and less than 1 identified as a down-regulated DEG.

Functional classification of GO, KEGG pathway and disease similarity heatmap based on disease ontology (DO)

GO terms[40] and the KEGG pathway[41] were used for functional enrichment analyses. The *GOstats* package[42] was used in conducting the Hypergeometric-based test and *KEGG.db* R package was used for the pathway analysis. In addition, the *clusterProfiler*[43] and Disease Ontology Semantic and Enrichment analysis (*DOSE*)[44] packages were used for enrichment analysis of DO and heatmaps for disease similarity, based on both Euclidean distance method and complete linkage method.

Gene set enrichment analysis (GSEA)

Using microarray datasets on breast cancer (GSE17907 and GSE20711)[45,46] and endometrial cancer (GSE17025)[47] tissues in humans, the microarrays of log FC values were regarded as a pre-ranked list for GSEA.

Construction of the BPA-related protein network analysis

The protein-protein interactions (PPIs) in *Homo sapiens* were obtained from the NCBI (ftp://ftp.ncbi.nlm.nih.gov/gene/GeneRIF/). The PPIs related to DEGs were imported into Cytoscape software[48]. A Cytoscape plug-in, BINGO[49], was used to identify over-representation of GO terms for biological processes.

Cell culture and BPA treatment

MCF-10A cells were a kind gift from Dr. Sun Jung Kim (Dongguk University, Korea). They were grown in MEBM (Lonza, Walkersville, MD, USA), supplemented with MEGM Single Quots (excluding CA-1000) (Lonza, Walkersville, MD, USA), 1% penicillin/streptomycin (Gibco-Life Technologies, Waltham, MA, USA), and 100 ng/mL of cholera toxin (List Biological Laboratories, Campbell, CA, USA). The cells were maintained in a humidified incubator at 37°C with 5% CO_2 and 95% air. Bisphenol A (BPA; Sigma, Saint Louis, MO, USA) was dissolved in dimethyl sulfoxide (DMSO; Sigma, Saint Louis, MO, USA). MCF-10A cells were seeded into 100 mm tissue culture dishes (SPL Life Science, Pocheon, Gyeonggi, Korea) at 5×10^5 cells for 24 hr, and treated with 1, 5, 10 μM BPA, and 0.01% DMSO as vehicle control for 6 days.

Real-time quantitative PCR

Total RNA was extracted by using TRIzol reagent (Invitrogen, Carlsbad, CA, USA), which was reversely transcribed using AMV reverse transcriptase (Promega, Madison, WI, USA) and oligo (dT) primers, according to the manufacturer's instructions. The resulting cDNA was used as a template for qPCR analysis using the SYBR® Premix Ex Taq™ (TAKARA BIO, Kusatsu, Shiga, Japan) with the CFX Connect™ Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA). The temperature profile of the reaction was 95°C for 2 min, followed by 40 cycles of denaturation at 95°C for 15 s, annealing at 61°C for 30 s, and extension at 72°C for 45 s. All reactions were performed in triplicates of three independent assays. The oligonucleotide primer sequences used in qPCR analyses were as follows: BRCA1 forward, 5′-CCTTCTACTGTCCTG GCTACTA-3′ and reverse, 5′-CAGATTTCCAAGGG AGACTTCA-3′; BRCA2 forward, 5′-CTCAGCCCAG ATGACTTCAAA-3′ and reverse, 5′-GGACTAACAG GTGGAGGTAAAG-3′; TP53 forward, 5′-GGGATGT TTGGGAGATGTAAG-3′ and reverse, 5′-CAGATAT GGGCCTTGAAGTTAG-3′; SMAD3 forward, 5′-ACAGGAGATGTAGGGAGAAGAA-3′ and reverse, 5′-CTCTAGCCAAGTCACACAGTAAG-3′; TWIST1 forward, 5′-CGGAGACCTAGATGTCATTGTTT-3′ and reverse, 5′-ACGCCCTGTTTCTTTGAATTTG-3′; GAPDH forward, 5′-GTGGTCTCCTCTGACTTCAAC-3′ and reverse, 5′-CCTGTTGCTGTAGCCAAATTC-3′. GAPDH was used as the internal control.

Statistical analysis

A One-way ANOVA followed by Tukey's multiple comparison tests was performed to determine the significant differences among groups. All values are expressed as means ± SEM. Statistical significance was considered as $P < 0.05$.

Acknowledgements This work was supported by the research program of Dongguk University 2013.

Conflict of Interest Junghyun Jung declares that he has no conflict of interest. Changsoo Mok declares that he has no conflict of interest. Woosuk Lee declares that he has no conflict of interest. Wonhee Jang declares that she has no conflict of interest.

Ethical Statement All data sources mentioned in the study are publically available summary level information that requires no ethical approval or consent.

Author Contributions JJ made substantial contributions to the conception and design of the acquisition, analysis, and interpretation of bioinformatics data and to the drafting of the manuscript; CM contributed the acquisition and analysis of molecular experiments; WL contributed the acquisition and analysis of bioinformatics data; WJ made contributions to the interpretation of data, critical revisions for intellectual content, and gave final approval of the version to be published.

References

1. Talsness, C. E., Andrade, A. J., Kuriyama, S. N., Taylor, J. A. & vom Saal, F. S. Components of plastic: experimental studies in animals and relevance for human health. *Philos Trans R Soc Lond B Biol Sci* **364**:2079-2096 (2009).
2. Park, H. G., Kim, J. Y. & Yeo, M. K. Paper-based Microfluidic Device for Bisphenol A Based Chemical Reaction and Image Analysis. *BioChip J* **10**:25-33 (2016).
3. Rubin, B. S. Bisphenol A: an endocrine disruptor with widespread exposure and multiple effects. *J Steroid Biochem Mol Biol* **127**:27-34 (2011).
4. Groff, T. Bisphenol A: invisible pollution. *Curr Opin Pediatr* **22**:524-529 (2010).
5. Vandenberg, L. N., Chahoud, I., Padmanabhan, V., Paumgartten, F. J. & Schoenfelder, G. Biomonitoring studies should be used by regulatory agencies to assess human exposure levels and safety of bisphenol A. *Environ Health Perspect* **118**:1051-1054 (2010).
6. Melzer, D., Rice, N. E., Lewis, C., Henley, W. E. & Galloway, T. S. Association of urinary bisphenol a concentration with heart disease: evidence from NHANES 2003/06. *PLoS One* **5**:e8673 (2010).
7. Magliano, D. J. & Lyons, J. G. Bisphenol A and diabetes, insulin resistance, cardiovascular disease and obesity: controversy in a (plastic) cup? *J Clin Endocrinol Metab* **98**:502-504 (2013).
8. Shankar, A. & Teppala, S. Relationship between urinary bisphenol A levels and diabetes mellitus. *J Clin Endocrinol Metab* **96**:3822-3826 (2011).
9. Melzer, D. *et al*. Urinary bisphenol A concentration and risk of future coronary artery disease in apparently healthy men and women. *Circulation* **125**:1482-1490 (2012).
10. Wang, T. *et al*. Urinary bisphenol A (BPA) concentration associates with obesity and insulin resistance. *J Clin Endocrinol Metab* **97**:E223-227 (2012).
11. Jenkins, S. *et al*. Oral Exposure to Bisphenol A Increases Dimethylbenzanthracene-Induced Mammary Cancer in Rats. *Environ Health Perspect* **117**:910-915 (2009).
12. Murray, T. J., Maffini, M. V., Ucci, A. A., Sonnenschein, C. & Soto, A. M. Induction of mammary gland ductal hyperplasias and carcinoma in situ following fetal bisphenol A exposure. *Reprod Toxicol* **23**:383-390 (2007).
13. Betancourt, A. M., Eltoum, I. A., Desmond, R. A., Russo, J. & Lamartiniere, C. A. In Utero Exposure to Bisphenol A Shifts the Window of Susceptibility for

Mammary Carcinogenesis in the Rat. *Environ Health Perspect* **118**:1614-1619 (2010).
14. Rochester, J. R. Bisphenol A and human health: A review of the literature. *Reprod Toxicol* **42**:132-155 (2013).
15. Seachrist, D. D. *et al*. A review of the carcinogenic potential of bisphenol A. *Reprod Toxicol* **59**:167-182 (2016).
16. Betancourt, A. M. *et al*. Altered carcinogenesis and proteome in mammary glands of rats after prepubertal exposures to the hormonally active chemicals bisphenol a and genistein. *J Nutr* **142**:1382S-1388S (2012).
17. Wong, R. L. Y. *et al*. Identification of Secretaglobin Scgb2a1 as a target for developmental reprogramming by BPA in the rat prostate. *Epigenetics-Us* **10**:127-134 (2015).
18. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-470 (1995).
19. Brown, P. O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nature Genetics* **21**:33-37 (1999).
20. Martin, J. A. & Wang, Z. Next-generation transcriptome assembly. *Nat Rev Genet* **12**:671-682 (2011).
21. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**: 57-63 (2009).
22. Kim, Y. J., Lee, N., Woo, S., Ryu, J. C. & Yum, S. Transcriptomic change as evidence for cadmium-induced endocrine disruption in marine fish model of medaka, Oryzias javanicus. *Mol Cell Toxicol* **12**:409-420 (2016).
23. Leek, J. T. *et al*. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* **11**:733-739 (2010).
24. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**:118-127 (2007).
25. Breitling, R., Armengaud, P., Amtmann, A. & Herzyk, P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Letters* **573**:83-92 (2004).
26. Mooney, M. *et al*. Comparative RNA-Seq and microarray analysis of gene expression changes in B-cell lymphomas of Canis familiaris. *PLoS One* **8**:e61088 (2013).
27. Jang, S. W., Oh, M. S., Yang, S. I. & Cho, E. M. Gene Expression Profiles of Human Neuroblastoma Cells Exposed to CuO Nanoparticles and Cu Ions. *BioChip J* **10**:140-149 (2016).
28. Williams, G. H. & Stoeber, K. The cell cycle and cancer. *J Pathol* **226**:352-364 (2012).
29. Shioda, T. *et al*. Expressomal approach for comprehensive analysis and visualization of ligand sensitivities of xenoestrogen responsive genes. *Proc Natl Acad Sci U S A* **110**:16508-16513 (2013).
30. Kang, S. Y., Kim, M. R., Kim, J. H. & Cho, H. H. Gene Expression Profiles of the Normal Myometrium After 48 and 96 Hours of Exposure to BPA. *BioChip J* **9**: 293-299 (2015).
31. Yim, W. C., Lee, B. M. & Kwon, Y. Cross-experimental analysis of microarray gene expression datasets for in silico risk assessment of TiO_2 nano-particles. *Mol Cell Toxicol* **8**:229-239 (2012).
32. Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* **9**:e78644 (2014).
33. Chen, C. *et al*. Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLoS One* **6**:e17238 (2011).
34. Petersen, M. *et al*. Smad2 and Smad3 have opposing roles in breast cancer bone metastasis by differentially affecting tumor angiogenesis. *Oncogene* **29**:1351-1361 (2010).
35. Zaytseva, Y. Y., Wallis, N. K., Southard, R. C. & Kilgore, M. W. The PPARgamma antagonist T0070907 suppresses breast cancer cell proliferation and motility via both PPARgamma-dependent and -independent mechanisms. *Anticancer Res* **31**:813-823 (2011).
36. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy - analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**:307-315 (2004).
37. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**:185-193 (2003).
38. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**:882-883 (2012).
39. Hong, F. *et al*. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* **22**:2825-2827 (2006).
40. Ashburner, M. *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **25**:25-29 (2000).
41. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**: D109-114 (2012).
42. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**:257-258 (2007).
43. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**:284-287 (2012).
44. Yu, G., Wang, L. G., Yan, G. R. & He, Q. Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**:608-609 (2015).
45. Sircoulomb, F. *et al*. Genome profiling of ERBB2-amplified breast cancers. *BMC Cancer* **10**:539 (2010).
46. Dedeurwaerder, S. *et al*. DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol Med* **3**:726-741 (2011).

47. Day, R. S. *et al*. Identifier mapping performance for integrating transcriptomics and proteomics experimental results. *BMC Bioinformatics* **12**:213 (2011).
48. Shannon, P. *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**:2498-2504 (2003).
49. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**:3448-3449 (2005).
50. Fernandez, S. V. *et al*. Expression and DNA methylation changes in human breast epithelial cells after bisphenol A exposure. *Int J Oncol* **41**:369-377 (2012).
51. Naciff, J. M. *et al*. The genomic response of Ishikawa cells to bisphenol A exposure is dose- and time-dependent. *Toxicology* **270**:137-149 (2010).
52. Gertz, J., Reddy, T. E., Varley, K. E., Garabedian, M. J. & Myers, R. M. Genistein and bisphenol A exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. *Genome Res* **22**:2153-2162 (2012).