# Network Analysis Reveals Novel Transcription Factors Associated With Biasphenol A Dose-Response

**Alexandra Maertens, Vy Tran, Andre Kleensang, Thomas Hartung**
Center for Alternatives to Animal Testing, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

## Background

Bisphenol A (BPA) is an industrial chemical used in the manufacture of plastic found in a number of consumer products such as thermal paper, canned foods and epoxy resin. Amongst the general population, exposure to BPA is widespread. BPA was presumed to have potentially estrogenic activity, as well as potential carcinogenicity, based on its structural similarity to DES (diethylstilbestrol) and other synthetic estrogens. Despite being subject to extensive study, a thorough understanding of molecular mechanism of BPA remains elusive.

## Materials and Methods

We constructed co-expression networks from a MCF7 RNA-seq data set that includes 1032 samples obtained from the ARCHS4 database, and a human breast cancer tissues data set with 1098 RNA-seq samples from The Human Genome Atlas.

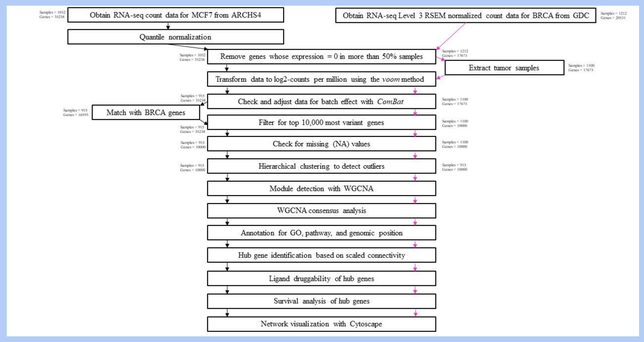



**Figure 1.** Analysis workflow. Black arrows indicate MCF7 workflow, and pink arrows indicate BRCA workflow.

## Module detection using WGCNA

Modules were defined using weighted gene co-expression network analysis, a scale-free network approach (Zhang & Horvath, 2005). The co-expression similarity matrix, calculated by the absolute value of Pearson correlation, $S = |cor(i,j)|$, was transformed into an adjacency matrix $A = [a_{ij}]$, in which $a_{ij} \in [0,1]$ denoting the connection strength between gene i and j. we picked a soft threshold power β of 6 based on the scale-free topology criterion.

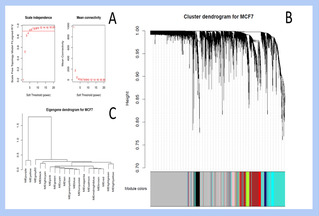



**Figure 2. Network analysis for MCF7.** (A) Scale free topology and mean connectivity as a function of soft threshold. (B) Hierarchical clustering dendrogram of MCF7 created with a soft threshold of 6. (C) Eigengene network.

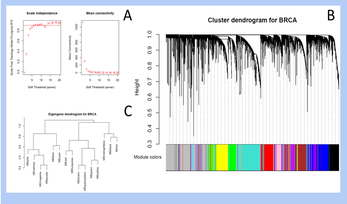



**Figure 3. Network analysis for BRCA.** (A) Scale free topology and mean connectivity as a function of soft threshold. (B) Hierarchical clustering dendrogram of BRCA created with a soft threshold of 6. (C) Eigengene network.

## Minimal conservation between MCF7 and BRCA gene networks

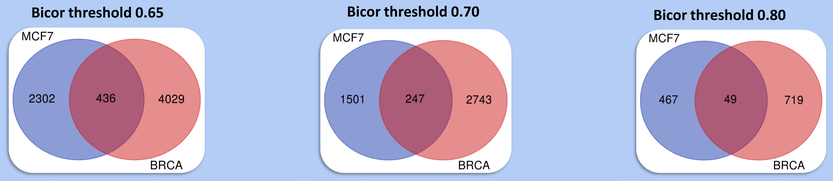



**Figure 3.** Overlapping genes between MCF7 and human breast cancer tissue networks at different gene expression correlation thresholds.

A WGCNA consensus network for MCF7 and BRCA together were generated using the 10,000 most variant genes for each dataset as determined by mean absolute deviation. Briefly, the network was derived based on a signed Spearman correlation using a β of 6 as a weight function. The topological overlap metric (TOM) was derived from the resulting adjacency matrix, and was used to cluster the modules using the blockwiseConsensusModules function, with min module size of 30, and the dynamic tree cut algorithm, with a height of 0.25 and a deep split level of 2.

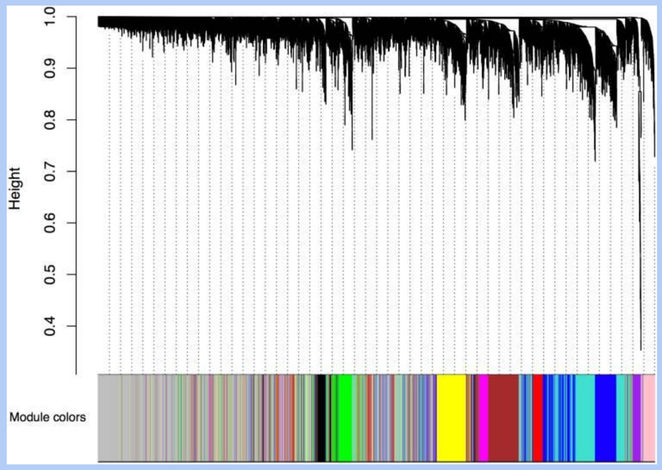



**Figure 4. Consensus analysis of MCF7 and BRCA.** (A) Consensus dendrogram. (B) Relating the consensus modules to MCF7-specific modules. (C) Relating the consensus modules to BRCA-specific modules.

## BRCA dataset points to genes with strong ligand-druggability scores not seen in MCF7

To examine the potential of top hub genes (scaled connectivity ≥ 0.4) in MCF7 and BRCA networks as drug targets, we uploaded the official gene symbols of top hub genes to canSAR Version 4, currently the largest cancer research and drug discovery knowledgebase in the world. CanSAR calculates ligand-based druggability by looking at the small molecule compounds that have been tested against the protein target or its homologues, and estimates how likely a target bind to drug-like compounds versus un-drug like compounds. The chemicals are weighted based on their bioactivity levels and the homology of their target to the initial protein query. Although there were druggable hub genes in both datasets, there was minimal overlap - highly draggable targets could therefore potential be missed if relying only on MCF7 data.

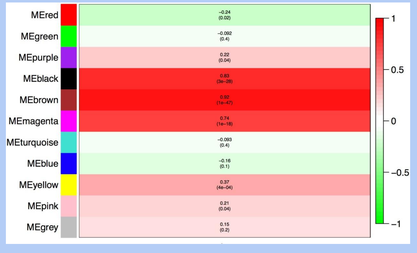



**Figure 5.** Hub genes with positive ligand druggability scores in MCF7 (A) and BRCA (B) networks. Venn diagram shows no overlapping in hub genes with positive ligand druggability scores between MCF7 and BRCA (C).

## Validation analysis with an estrogen-treated MCF7 dataset

To validate the above results, we used a small subset of MCF-7 cells (n = 88 samples) that were treated with estrogen for 48 hrs. The dataset was obtained from Gene Expression Omnibus (GSE50705).

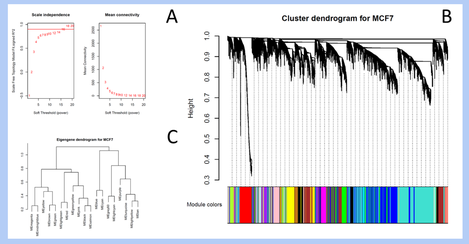



**Figure 7. Network analysis for estoregen-treated MCF7.** (A) Scale free topology and mean connectivity as a function of soft threshold. (B) Hierarchical clustering dendrogram of MCF7 created with a soft threshold of 6. (C) Eigengene network.

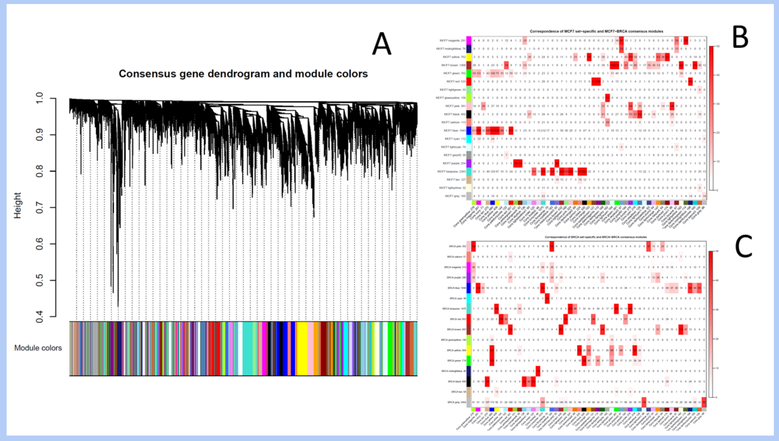



**Figure 8. Validation consensus analysis of estrogen-treated MCF7 and BRCA.** (A) Consensus dendrogram. (B) Relating the consensus modules to MCF7-specific modules. (C) Relating the consensus modules to BRCA-specific modules.

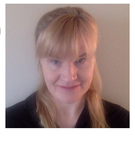
## Discussion

Our results indicate that results from MCF7 have to be used extremely cautiously as a proxy for human breast cancer physiology. To begin with, there is minimal overlap of gene expression levels even at the most basic metric of similarity (sorting genes by expression level). Moreover, when using a more sophisticated metric, such as weighted gene correlation network analysis, the hub genes are substantially different and more importantly, there are several genes that are "druggable" that are apparent in the BRCA dataset that would've been missed if relying solely on MCF7. Furthermore, our data cautions against the advisability of scaling-up transcriptomic data - in the case of MCF7 cells, owing to the genetic instability as well as artifacts intrinsic in tissue culture and transcriptomic data - a "big-data" approach may simply adding more noise rather than signal. Therefore, MCF7 cells -like all models - should be used in tandem with parallel approaches to provide added confidence in the data.

**Virtual poster!** Scan this QR code for an electronic version of this poster.

**About the authors**

**Vy Tran** is a 3rd year PhD student in bioinformatics with interest in systems biology and toxicology.
**Email:** vtran21@jhmi.edu



**Alexandra Maertens, PhD,** is a bioinformatician at the Center for Alternatives to Animal Testing, where she analyzes transcriptomic, metabolomic, and proteomic data.
**Email:** amaerte1@jhu.edu

**References**

1. Lachmann, A., D. Torre, A. B. Keenan, K. M. Jagodnik, H. J. Lee, L. Wang, M. C. Silverstein and A. Ma'ayan (2018). "Massive mining of publicly available RNA-seq data from human and mouse." Nat Commun 9(1): 1366.
2. Shioda T, Rosenthal NF, Coser KR, Suto M et al. Expressomal approach for comprehensive analysis and visualization of ligand sensitivities of xenoestrogen responsive genes. *Proc Natl Acad Sci U S A* 2013 Oct 8;110(41):16508-13.
3. Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol, 4*, Article17. doi:10.2202/1544-6115.1128