ESTROGEN-MEDIATED GENE REGULATION IN PATIENTS WITH CHRONIC MYELOID LEUKEMIA

Linda Y. Liu*[1,2], Chirag J. Patel[1,2], Atul J. Butte[2]

1. *Biomedical Informatics Training Program, Stanford University*
2. *Dept. of Pediatrics, Stanford School of Medicine and Center for Pediatric Bioinformatics, Lucile Packard Children's Hospital, Stanford, CA 94305, USA*

Email: lyl@stanford.edu (corresponding author), cjp@stanford.edu, abutte@stanford.edu

Understanding regulation of gene expression is crucial not only for developing knowledge of normal gene function, but also for understanding how dysregulation leads to disease. Some regulatory relationships have been well described, such as the role of the estrogen receptor on gene transcription in breast cancer progression. However, the role of estrogen signaling in other diseases, including other cancers, is less well understood. To address this question, we used estrogen-mediated gene expression data in conjunction with gene expression studies from breast cancer and chronic myeloid leukemia. Through this integrated approach, we show that estrogen regulation in a cell line is reproduced in a patient cohort with induced breast cancer phenotype, as well as in chronic myeloid leukemia. We show that the genes involved in leukemia that are also regulated by estrogen have a variety of enzymatic functions and are implicated in other cancer and disease pathways. This study will improve understanding of the regulatory role of the estrogen receptor in chronic myeloid leukemia, as well as in other cancers and diseases. These methods may also be applied to other gene regulatory systems.

Keywords: estrogen receptor, gene regulation, chronic myeloid leukemia

1. INTRODUCTION

The estrogen receptor regulatory system is a model of hormonal regulation of gene expression with important clinical implications in breast cancer, cardiovascular disease, and other diseases[1]. There have been many studies demonstrating the role of estrogen in breast cancer disease progression. Carroll et al used the estrogen-dependent breast cancer cell line MCF-7 to create a genome-wide map of estrogen receptor binding sites, enabling the discovery of genes that are up- or down-regulated immediately following estrogen exposure[2]. In this way, they discovered that the majority of estrogen receptor binding sites are actually not near gene promoters, in the conventional sense. They concluded that estrogen signaling may produce downstream indirect effects that require other genes acting together in a regulatory pathway.

Estrogen signaling is also potentially important in other diseases besides breast cancer. The estrogen receptor is present in two-thirds of ovarian tumors and the presence of estrogen has been shown to stimulate tumor growth[3]. There is also strong evidence that estrogen plays a pivotal role in progression of prostate cancer and endometrial cancer. On the contrary, estrogen can protect against colon cancer, cardiovascular and neurodegenerative diseases and osteoporosis[1]. Since estrogen levels in women change over time, this is an added layer of complexity in assessing susceptibility to disease among premenopausal and postmenopausal women.

To understand how estrogen affects disease processes, we must consider how it acts as a transcription factor. Transcriptional regulation is a complex process depending on the interaction of many regulatory protein complexes as well as the chromatin architecture of nuclear DNA[4]. In many cases cellular signaling cascades initiate transcription factor activity that affects several downstream

pathways[5]. The estrogen receptor is a gene transcription regulator that requires the steroid hormone estrogen to bind and activate the receptor. Once ligand binding occurs, the receptor-ligand complex moves into the nucleus, binds DNA, and can act as an activator or repressor of gene transcription. In this way, it can lead to coordinated changes in expression of many genes, affecting many pathways and biological processes.

Other studies have investigated the gene expression differences that distinguish breast tumor tissue from normal healthy tissue, independent of any regulatory hypotheses. Ince et al transformed normal breast tissue from premenopausal women into cells expressing breast adenocarcinomas via retroviral vectors expressing three genes[6]. They identified the genes that are significantly differentially expressed between normal and tumor phenotype, and further showed that the expression profile differed depending on the origin of the transformed cells. However, the authors did not investigate the effect of estrogen, if any, on the gene expression changes occurring in conjunction with cell transformation.

Despite these studies, to our knowledge, there has been relatively less published work quantifying the effect of estrogen-mediated regulation of gene expression in cancers of non-sex-specific tissues (i.e. not in breast, ovary, or prostate). Shen et al previously demonstrated that phytoestrogens, which are derived from plants and have similar structure and properties as estrogen, exhibit antiproliferative effects on leukemic cancer cells especially in combination with chemotherapeutic drugs[7]. However, there is some dispute about findings such as these, and phytoestrogen activity is not completely similar to that of estrogen. It is not clear what the relationship is between estrogen-mediated gene regulation and the disease mechanism of leukemia or cancers of other non-sex-specific tissues. Since breast and ovarian cancer are studied almost exclusively in women, the effect of estrogen on disease progression in men is poorly understood. Using an integrative approach that combines estrogen receptor binding information with gene expression data provides a unique opportunity to assess the effect of estrogen in a separate tissue. We wished to test the hypothesis that estrogen receptor regulates gene expression in leukemia, as suggested by Shen et al's result from phytoestrogens.

Chronic myeloid leukemia (CML) provides a good framework to address these questions. CML is a cancer of hematopoietic stem cells and affects both men and women of all age groups. Its incidence is 1-2 in 100,000 individuals, and accounts for 15-20% of adult leukemia cases in the Western Hemisphere[8]. CML is usually caused by a translocation between chromosomes 9 and 22 resulting in an aberrant fusion protein. This translocation is termed a Philadelphia chromosome and the protein fusion, BCR-ABL, encodes a constitutively active tyrosine kinase. The activation of oncogenic pathways and buildup of bone marrow progenitor cells can be treated with tyrosine kinase inhibitors[9]. A study by Diaz-Blanco et al used healthy controls and patients with chronic-phase myeloid leukemia and found a molecular signature of CD34+ stem cells[9].

Since the main mechanism of action in CML is caused by signal transduction due to abnormal phosphorylation activity by the kinase, many gene regulatory systems are potentially perturbed. However, it is unclear whether hormonal signaling affects or is affected by this mechanism. Therefore, we investigated the effect of estrogen signaling in a leukemia disease model incorporating both male and female subjects of various age groups. Using an informatics approach, we were able to integrate gene expression experiments obtained from different tissues and different patient populations. This made it possible to apply estrogen activity informa-

tion obtained from a sex-specific tissue to a different tissue. From this analysis, we found the novel association between estrogen regulation and chronic myeloid leukemia.

2. METHODS

2.1 Data extraction

We obtained estrogen receptor binding site data as well as a list of 294 genes found to be up- or down-regulated three hours following exposure to estrogen in the MCF-7 estrogen-dependent breast cancer cell line from Carroll et al (GSE11324; Affymetrix Human Genome U133 Plus 2.0 Array; GPL 570)[2]. The breast cancer and CML gene expression data were obtained from GEO by searching for cancer experiments containing both case and control samples.

Figure 1 shows a pipeline for the analysis described in Sections 2.2 and 2.3. The workflow was essentially the same for both the breast cancer and CML experiments.

2.2 Replication of Carroll et al results in a separate breast cancer set

The breast cancer samples (GSE6885; Affymetrix Human Genome U133 Plus 2.0 Array; GPL 570) were taken from breast primary epithelial cells (BPECs) in premenopausal women with normal healthy breast tissue. Half of the BPECs were induced to yield tumors closely resembling human breast adenocarcinomas[6]. There are five

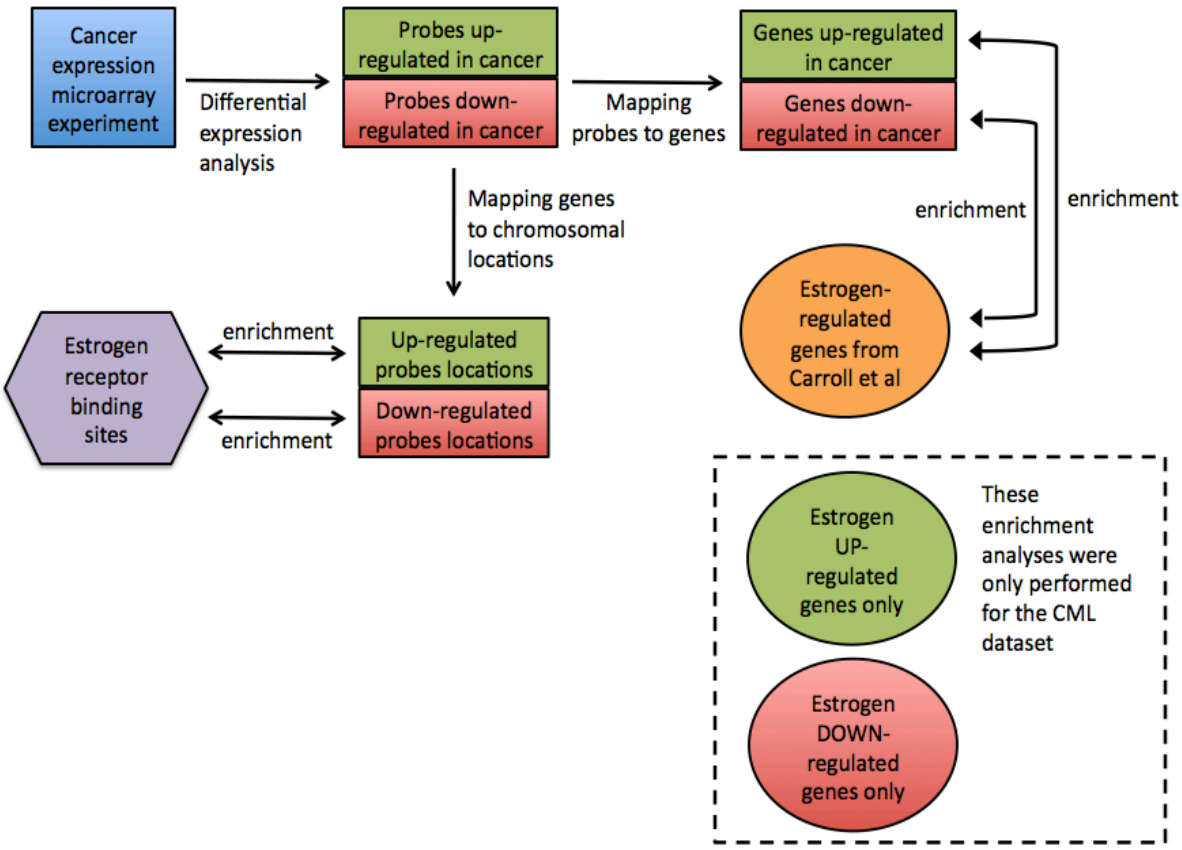

Figure 1. Overview of Methods. The starting cancer expression microarray dataset can refer to either the breast cancer or CML datasets. Differential expression analysis was performed using Significance Analysis of Microarrays (SAM), probe-to-gene mapping was performed using Array Information Library Universal Navigator (AILUN), and gene chromosomal locations were obtained from NCBI.

control samples and six tumor samples. We used Significance Analysis of Microarrays (SAM) to obtain significantly up- and down-regulated genes[10] and mapped probe identifiers to gene identifiers using a tool called Array Information Library Universal Navigator (AILUN)[11]. For the SAM analysis, we required significant genes to have a median false discovery rate of less than 10%. A small fraction of probes mapped to multiple gene identifiers, and in these cases we included all mapped genes. We then performed hypergeometric enrichment analysis of the significant up- and down-regulated breast cancer genes (separately) in the set of estrogen-regulated genes from Carroll et al. Additionally, we mapped the significant breast cancer genes to their chromosomal locations and performed enrichment analysis with the set of estrogen receptor binding sites from Carroll et al. A gene was defined to be near a binding site if the binding site is within 50kb of the transcription start site of the gene. This is consistent with the definition applied by Carroll et al[2]. In both cases the background set consists of all probes on the array, as both studies used the same array.

2.3 Integration of estrogen-regulated genes with CML data set

The CML samples (GSE5550; Affymetrix Human Genome Focus Array; GPL 201) were taken from CML patients and control individuals ages 24-69, of both sexes[9]. There are eight control samples and nine diseased samples. Similar to above, we used SAM to obtain significantly up- and down-regulated genes and AILUN to obtain gene identifiers. For the SAM analysis, we required significant genes to have a median false discovery rate less than 0.1% and a fold-change greater than 1.2 or less than 0.83 (the same thresholds applied by the original authors). We performed hypergeometric enrichment analysis of these genes with estrogen-regulated genes as well as estrogen receptor binding site locations. We additionally calculated enrichment separately for estrogen up-regulated genes and estrogen down-regulated genes. In this analysis, the background set consisted of all genes on the CML array. The array used by Carroll et al contained all these as well as additional genes. For the CML genes found to be regulated by estrogen, we used Onto Express to find enriched Gene Ontology categories, and Pathway Express to find enriched pathways in which the genes are implicated (Figure 2). Both tools are from the Onto Tools package[12]. We required a false discovery rate of less than 5% for the ontology and pathway enrichment results.

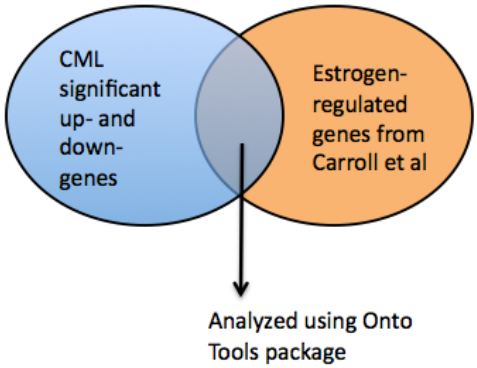

Figure 2. Venn diagram showing the intersection between CML significantly differentially expressed genes and estrogen-regulated genes.

3. RESULTS

We integrated estrogen receptor regulation and binding data with differentially expressed genes from breast cancer and chronic myeloid leukemia datasets. Our analysis incorporated hypergeometric tests of enrichment of

differentially expressed genes regulated by estrogen receptor.

3.1 Replication of Carroll et al results in a separate breast cancer set

At a false discovery rate < 10%, we found 56 up-regulated genes and 109 down-regulated genes from the breast cancer data set. After joining with the set of estrogen-regulated genes, we found that the genes up-regulated in breast cancer are enriched for estrogen

Table 1. Enrichment of breast cancer disease genes in estrogen-mediated gene set. BPLER refers to the cells in which a breast cancer phenotype was induced. ER stands for estrogen receptor. The background set consists of all genes on the microarray.

Gene Set	Hypergeometric p-value	Percent regulated by ER
All genes	N/A	1.4%
Up in BPLER	0.0068	5.4%
Down in BPLER	0.18	1.8%

Table 2. Enrichment of CML disease genes in estrogen-regulated gene set. ER stands for estrogen receptor. The background set consists of all genes on the CML microarray. A) Enrichment in the set of all estrogen-regulated genes. B) Enrichment in the set of genes up-regulated by estrogen only. C) Enrichment in the set of genes down-regulated by estrogen only.

A)

Gene Set	Hypergeometric p-value	Percent regulated by ER
All genes	N/A	1.8%
Up in CML	0.0078	2.6%
Down in CML	0.26	1.8%

B)

Gene Set	Hypergeometric p-value	Percent UP-regulated by ER
All genes	N/A	0.7%
Up in CML	1×10^{-6}	1.9%
Down in CML	0.34	0%

C)

Gene Set	Hypergeometric p-value	Percent DOWN-regulated by ER
All genes	N/A	1.0%
Up in CML	.86	0.7%
Down in CML	.11	1.8%

regulation, as expected (Table 1). The set of up-regulated genes in disease was observed to be nearly four times more likely to be regulated by estrogen than the background set, and three times more likely than down-regulated genes.

Additionally, we found that only 4% of the significant genes are located 50kb from an estrogen receptor binding site. This is in agreement with results from Carroll et al[2].

3.2 Integration of estrogen-regulated genes with CML data set

At a false discovery rate < 0.1% and fold-change > 1.2 or < 0.83, we found 1270 up-regulated genes and 57 down-regulated genes from the CML data set.

After joining these genes with the set of estrogen-regulated genes, we found that genes significantly up-regulated in CML are enriched for estrogen regulation (Table 2). Interestingly, the enrichment signal comes entirely from the genes that are up-regulated by estrogen, and not down-regulated by estrogen. The significant CML up-regulated genes are nearly three times as likely as the background set to be up-regulated by estrogen receptor.

Similar to the breast cancer analysis, we found that only 4% of the significant genes are located 50kb from an estrogen receptor binding site.

Observing an enrichment between differentially expressed CML disease genes and estrogen regulation, we sought to annotate these enriched genes. Table 3 shows the results of applying Onto Express to the list of CML up-regulated disease genes that are also regulated by estrogen receptor. There are 34 genes that fell into this category and we found that they are enriched for various enzymatic functions such as synthases and kinases, crucial to protein translation and cellular homeostasis. Table 4 shows that this set of 34 genes is also highly associated with many other cancer and important cellular signaling pathways. Many of these pathways would be highly impacted if these genes were mutated or dysregulated, as quantified by the impact factor associated with each pathway result. The impact factor includes a prob-

Table 3. Onto Express results for 34 estrogen-regulated CML disease genes. The categories shown are from the Gene Ontology graph expanded to seven levels. The p-value associated with each category is calculated from hypergeometric enrichment and corrected by false discovery rate. The fourth column shows the number of terms associated with the 33 genes that fell into a particular Gene Ontology category.

Gene Ontology Category	P-value	Corrected p-value	Terms	Percent of total
Pseudouridine synthase activity	1×10^{-5}	9×10^{-4}	2	5.9 %
Adenosylmethionine decarboxylase activity	0.004	0.034	1	2.9 %
NADPH activity	0.004	0.034	1	2.9 %
Phosophoribosylaminoimidazole carboxylase activity	0.004	0.034	1	2.9 %
Pseudouridylate synthase activity	0.004	0.034	1	2.9 %
Ribonuclease inhibitor activity	0.008	0.058	1	2.9 %
Adenylate kinase activity	0.02	0.089	1	2.9 %
Isomerase activity	0.028	0.102	2	5.9 %
Lyase activity	0.031	0.102	2	5.9 %

Table 4. Pathway Express results for 33 estrogen-mediated CML disease genes. Results are ranked by impact factor and FDR-corrected hypergeometric p-value. All pathways were obtained from KEGG.

Pathway name	Impact Factor	Total genes in pathway	Input genes in pathway	P-value
Basal transcription factors	5.16	37	2	0.0057
Acute myeloid leukemia	4.08	59	2	0.017
Colorectal cancer	3.36	84	2	0.035
Circadian rhythm	3.21	13	1	0.04
Wnt signaling pathway	2.48	152	2	0.084
Thyroid cancer	2.41	29	1	0.09

abilistic term that takes into account the proportion of differentially regulated genes in the pathway and the relative importance of each gene to the pathway's function[12].

4. DISCUSSION

In this study, we used informatics methods to integrate three different datasets using gene expression measurements. Through this analysis, we have shown that estrogen-mediated gene regulation results obtained from a breast cancer cell line can be replicated in a separate breast cancer data set with actual patients, and moreover can be extended to a completely different cancer and tissue. These techniques can be generalized and applied to other diseases and transcription factors to find more regulatory relationships implicated in disease progression.

The first experiment involving the breast cancer case-control study validated the approach of integrating the estrogen regulation data with an external data set. The results we obtained are completely consistent with the original Carroll et al study. The genes found to be significantly up-regulated in a breast cancer phenotype are also significantly enriched for estrogen regulation, while the down-regulated genes had no such association. This result was expected since estrogen presence is associated with more severe disease progression as well as higher risk for developing breast cancer. It is important to note that for this experiment we chose to use data from premenopausal women to ensure that estrogen levels would be high. It would be interesting to investigate if the same results hold in postmenopausal women.

For the SAM analysis of the breast cancer data, we chose a liberal false discovery rate cutoff of 10% since we were only trying to confirm the validity of integrating the two data sets. However, it was surprising that so few genes (165 genes) were differentially expressed at this threshold. A possible explanation is that these samples were not from diseased tissue, but rather healthy tissue that was induced to express a tumor-like phenotype using a few genes. There may be more gene expression changes that occur in the natural progression of tissue from healthy to cancerous.

Surprisingly, we found a strong estrogen signal in a chronic myeloid leukemia data set, which we did not expect since CML is caused by a chromosomal physical aberration that is not known to be affected by estrogen signaling. What is perhaps more surprising is that we saw this effect in a patient cohort that is very different from the breast cancer and estrogen receptor data sets. The CML patient samples came from both men and women, of

various ages, suggesting that the effect of estrogen receptor activity is strong even in men and postmenopausal women.

Here we have shown that CML disease-associated genes are enriched for being up-regulated by estrogen receptor, implying that estrogen exposure leads to expression of the disease. To our knowledge, there has been no previous documentation of the effect of estrogen receptor-mediated transcriptional regulation on CML disease progression. Our results contradict those from Shen et al claiming that estrogen-like compounds slow progression of leukemia. Here we show that estrogen up-regulates genes that are associated with disease phenotype.

The poor estrogen receptor binding site enrichment in both experiments confirms the indication in Carroll et al that estrogen has most of its effects in downstream pathways rather than direct, cis-regulatory effects. More work has yet to be done to identify the pathways that estrogen affects and how it affects them, the temporal nature of the signaling cascade, and what other regulators need to act in concert.

This work suggests that estrogen may have a role in regulating disease phenotype not only in breast cancer, but also in other diseases not known to be directly affected by estrogen receptor activity. One limitation to our and other informatics methods is the requirement for gene expression data measured on compatible platforms for integration. Additionally, our results are not a confirmation of estrogen activity in chronic myeloid leukemia, but rather present a hypothesis to initiate an experiment for clinical validation. Further investigation into the mechanism by which estrogen regulates gene transcription in disease can lead to new treatments with estrogen receptor as a therapeutic target.

References

1. Deroo B and K Korach. Estrogen receptors and human disease. *J Clin Invest* 2006 Mar;116(3):561-70.
2. Carroll JS, et al. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* 2006 Nov;38(11):1289-97.
3. Syed V, et al. Expression of gonadotropin receptor and growth responses to key reproductive hormones in normal and malignant human ovarian surface epithelial cells. *Cancer Res* 2001 Sep;61(18):6768-76.
4. Lanctot C, et al. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet* 2007 Feb;8:104-15.
5. Montminy M. Transcriptional regulation by cyclic AMP. *Ann Rev Biochemistry* 1997 July;66:807-22.
6. Ince TA, et al. Transformation of different human breast epithelial cell types leads to distinct tumor phenotypes. *Cancer Cell* 2007 Aug;12(2):160-70.
7. Shen J, et al. [Synergistic antileukemic effect of phytoestrogens and chemotherapeutic drugs on leukemic cell lines in vitro]. [Article in Chinese]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi* 2008 Apr;16(2):276-81.
8. Faderi S, et al. Chronic myelogenous leukemia: biology and therapy. *Annals Internal Med* 1999 Aug;131(3):207-19.
9. Diaz-Blanco E, et al. Molecular signature of CD34(+) hematopoietic stem and progenitor cells of patients with CML in chronic phase. *Leukemia* 2007 Mar;21(3):494-504.
10. Tusher VG, et al. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci* 2001 Apr;98(9):5116-21.
11. Chen R, et al. AILUN: reannotating gene expression data automatically. *Nat Methods* 2007 Nov;4(11):879.
12. Khatri P, et al. Onto-Tools: an ensemble of web-accessible, ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments. *Nucleic Acids Res* 2004 Jul;32.

Acknowledgments

The authors would like to acknowledge Joel Dudley, Rong Chen, Purvesh Khatri, Alex Skrenchuk, Boris Oskotsky, National Science Foundation Graduate Research Fellowship, Hewlett Packard Foundation, and Lucile Packard Foundation for Children's Health.