Gender-specific differences in expression in human lymphoblastoid cell lines

Wei Zhang, **Wasim K. Bleibel**, **Cheryl A. Roe**, **Nancy J. Cox**, and **M. Eileen Dolan**
Department of Medicine, University of Chicago, Chicago, Ilinois, USA

Abstract

Women and men have different risks for certain diseases and they often respond differently to treatment. These differences could be due to the sex-specific differences in the expression of genes related to primary disease susceptibility or pharmacodynamic targets. To evaluate the sex-specific pattern of gene expression, we compared gene expression levels using a publicly available microarray dataset of 233 (115 women and 118 men) lymphoblastoid cell lines. From the 4799 probes meeting a specified minimal level of expression, 10 genes ($P<0.005$, permutation adjusted false discovery rate less than 50%) located on autosomal chromosomes were identified using a permutation-based approach. These genes were found to be over-represented in certain gene ontology terms of biological process (cell adhesion, apoptosis, transcription and signal transduction), and molecular function (structural molecule activity, zinc ion binding, transcription factor activity and protein binding). A Kyoto Encyclopedia of Genes and Genomes pathway analysis indicated that two known pathways are over-represented: adherens junction and cytokine-cytokine receptor interaction.

Keywords

Centre d' Etude du Polymorphisme Humain; gene expression; lymphoblastoid cell lines; sex difference

Introduction

It is widely recognized that risks for many diseases differ between men and women. For example, women are more likely to get multiple sclerosis, rheumatoid arthritis and migraines; men are more susceptible to heart attacks and gout [1]. In human cancers, sex differences in prevalence are observed for not only hormone-dependent cancers like breast cancer and thyroid cancer (more common in women) but also other cancers including those of the lung, kidney, bladder and pancreas (more common in men) [1]. Environmental exposures can, of course, be different in men and women, but linkage studies on a variety of quantitative traits have shown that men and women may have a different genetic architecture for the same phenotype [2]. Moreover, even when men and women have the same disease, the disease may progress differently, and they may respond differently to treatment.

Sex is an important human variable that reflects the differences in reproductive systems, hormonal environment and gene expression. These differences, especially in gene expression,

Correspondence and requests for reprints to M. Eileen Dolan, PhD, Section of Hematology-Oncology, 5841 S. Maryland Avenue, Box MC2115, University of Chicago, Chicago, IL 60637, USA Tel: + 1 773 702 4441; fax: + 1 773 702 0963; e-mail: E-mail: edolan@medicine.bsd.uchicago.edu.

may contribute to observed sex differences in response to drugs. For example, studies have shown that women have more cytochrome CYP3A in the liver [3]. The overexpression of CYP3A, which metabolizes about 50% of drugs [4], modulates the effectiveness of drugs in women. Therefore, a genome-wide search for gender-specific differences in gene expression could provide more clues as to the underlying mechanisms for the discrepancy in pharmacodynamic effects of medications between men and women.

Microarray expression studies provide a high-throughput way to search the whole genome for sex differences in gene expression. Morley *et al.* [5] published a database of baseline expression levels of genes in lymphoblastoid cell lines derived from Centre d' Etude du Polymorphisme Humain individuals from the International HapMap Project using Affymetrix Human Genome Focus (HG-Focus) Array. The HG-Focus Array represents over 8500 verified human sequences from the NCBI RefSeq database. We compared the expression levels between men and women for genes with expression meeting a specified minimal threshold.

The objective of our work was to identify the genes with significant differential expression between men and women, as well as their functions and involvement in physiological pathways by searching GO (Gene Ontology) [6] and KEGG (Kyoto Encyclopedia of Genes and Genomes) [7] databases. Although the physiological functions are not clear for all the identified genes, our genome-wide search for genes with sex differences in expression provides clues to the underlying mechanisms for the discrepancy in pharmacodynamics between men and women.

Methods

Cell lines

Data from 355 expression microarrays (study GSE1485 [5], including replicates) were downloaded from the Gene Expression Omnibus Repository (http://www.ncbi.nlm.nih.gov/projects/geo/). Sample data came from 233 lymphoblastoid cell lines (including 94 grandparents) derived from individuals within 24 Centre d' Etude du Polymorphisme Humain families [5]. Among them, 115 individuals were women and 118 were men. The samples comprised family members representing three generations from 14 families and unrelated samples (grandparents) from 10 families (see Supplemental data for detailed family structure and age information). Samples were processed and hybridized using the HG-Focus Array containing 8793 probe sets [5].

Baseline expression levels measured by Affymetrix Human Genome Focus Array

The raw intensity data were normalized using a robust multiarray average [8] method to calculate probe signal levels. To be included in our studies, probe sets had to reach a threshold raw expression level of 50 in at least 5% of the arrays. Furthermore, X-linked or Y-linked probes were removed from the data set. After this filtering, 4799 probe sets on autosomal chromosomes remained in the analysis set. The raw signals were then log_2 transformed for further analyses.

Comparison of sex differences in expression

We used a permutation-based *t*-test to identify differentially expressed genes between men and women using all the 233 samples (115 women and 118 men) for more statistical power. The basic test used was the standard pooled variance *t*-statistic. To address the problem of multiple comparisons, the free step-down approach of Westfall-Young (W-Y) [9] was used to compute simultaneous *P* values that control the overall or family-wise error rate. The permutation-adjusted *P* values were then used as an empirical estimate of false discovery rate (FDR). The

permutation-adjusted *P* values, based on *B* = 10 000 permutations, were calculated using the software Permax, which was implemented in the *R* statistical package [10].

For genes of interest, we tested two additional models. The first model adjusted for family membership to account for the relatedness among family members, whereas the second adjusted for generation to take into account the possible effect of age on expression level. For the family-adjusted model, all unrelated samples (94 grandparents) were grouped into one cluster, whereas parents and children were grouped into a second cluster. For the generation-adjusted model, the samples were grouped into three clusters: 94 grandparents, 28 parents and 111 children, respectively. A general stratified statistic for comparing probe *i* between men and women is defined as

$$\sum_k W_k \left(\bar{X}_{1ik} - \bar{X}_{2ik} \right) = \sum_k W_k \frac{m_k}{n_{1k} n_{2k}} \sum_{j=1}^{n_{1k}} X_{ijk},$$

where $\bar{X}_{lik}$ is the average of the expression value of probe *i* in either men or women within cluster *k*, n_{1k} is the number of participants in group 1 (men) in cluster *k*, $n_{2k} = m_k - n_{1k}$ and w_k are cluster weights determined by cluster size. The one-sided individual *P* values for the family-adjusted and generation-adjusted models were calculated using software Permax in *R* [10].

To select probe sets to use in bioinformatics searches, we used a lenient cutoff value of unadjusted *P* < 0.005 and permutation-adjusted FDR < 50% based on the W-Y approach [9]. The differentially expressed probes were then matched to unique genes, which were subjected to further model testing and GO/KEGG pathway analyses.

Gene Ontology analysis

To more thoroughly characterize sets of functionally related genes differentially expressed between women and men, we used Onto-Express [11] to classify genes according to the following GO categories [6]: biological process, cellular component and molecular function. GO terms overrepresented in our set of genes relative to the filtered analysis set of 4799 probes were identified by a hypergeometric test controlled at a Benjamini-Hochberg (B-H) FDR [12] of 5%.

Kyoto Encyclopedia of Genes and Genomes pathway analysis

The differentially expressed genes were searched against the KEGG database [7] (release 37.0, January 2006) for known physiological pathways using Pathway-Express [11]. Pathways overrepresented in our set of genes relative to the filtered analysis set of 4799 probes were identified by a hypergeometric test controlled at a B-H FDR [12] of 5%.

Results and discussion

Comparison of expression and identification of differentially expressed genes

Using a permutation-based *t*-test following the W-Y approach, 10 probe sets out of the filtered data set of 4799 probes of the HG-Focus Array meeting a minimal expression level were found to be differentially expressed (*P* < 0.005, permutation-adjusted FDR < 50%). The main advantage of the W-Y approach is that it fully takes into consideration all dependencies between genes. This is extremely important for tightly correlated genes such as those involved in the same pathway. To avoid the complication of X-inactivation, we removed all X-linked or Y-linked probes before analysis. As expected, without filtering out X-linked or Y-linked probes, many of them would be selected for differential expression. A separate test including the X-

linked or Y-linked probes meeting a minimum expression level showed that 21 of them could be selected using our approach together with nine other probes (data not shown). Filtering X-linked or Y-linked probes improved our power of identifying differential genes. Therefore, all identified genes are located on autosomal chromosomes. The 10 differentially expressed probes correspond to 10 unique genes. Among these, four genes (*FOXO1A, JUP, KLF2, ZNF706*) had higher expression in women and six genes (*CTNND1, FEZ1, FFAR2, LGALS1, LMNA, TNFSF9*) had higher expression in men. As gene expression could be correlated among family members and because age and age-related hormonal levels could influence gene expression, we used general stratified models to adjust for family membership and generation. After separate adjustments for family membership and generation, respectively, these 10 genes still showed significant expression differences between men and women ($P < 0.01$). The statistics, *P* values and gene annotations for the 10 genes are shown in Table 1. In addition to the HG-Focus data, expression data were available from the Affymetrix HG-U133 Plus 2.0 Array on a subset of our samples (eight men and eight women, unpublished data). Even with such a small sample size, one of the 10 genes found above (*JUP*) also showed significantly higher expression in women (data not shown).

Gene Ontology analysis

The three organizing principles of GO are molecular function, biological process and cellular component [6]. A gene product has one or more molecular functions and is used in one or more biological processes; it might be associated with one or more cellular components. We searched GO annotations for our 10 genes of interest using Onto-Express [11]. GO terms with more than two hits and a corrected *P* value less than 0.05 on the basis of a reference data set of the filtered 4799 probes are presented in Table 2. Our 10 genes of interest were found to be overrepresented in four GO biological processes: cell adhesion, apoptosis, transcription and signal transduction, four GO molecular functions: structural molecule activity, zinc ion binding, transcription factor activity and protein binding and four GO cellular components: `integral to plasma membrane', cytoplasm, nucleus and cytoskeleton.

Kyoto Encyclopedia of Genes and Genomes analysis

The KEGG database [7] was searched for any known pathways involving our 10 genes of interest using Pathway-Express [11]. Two genes were found to be involved in two separate known pathways that are overrepresented (corrected $P < 0.05$) in our gene list (Table 2). Particularly, *CTNND1*, which is a member of the overrepresented GO categories of structural molecule activity and protein binding and whose expression is higher in men, was found to be involved in the adherens junction pathway. *TNFSF9*, which is also higher in men, was involved in another overrepresented pathway: cytokine-cytokine receptor interaction.

The fact that there are overrepresented GO terms and known pathways in the 10 differentially expressed genes between men and women suggests their roles in regulating expression and such physiological processes as drug response between sexes. The overall differences in expression and other physiological processes between men and women could be the comprehensive results of these differing regulations. The GO analysis suggests such evident biological processes as transcription activity (*KLF2*, *FOXO1A*) through molecular functions of zinc ion binding (*KLF2*) that could contribute to the gender differences in expression. Moreover, such GO biological processes as cell adhesion (*JUP*, *FEZ1*), apoptosis (*TNFSF9*, *LGALS1*) and signal transduction (*FFAR2*, *ZNF706*) that could possibly contribute to gender-specific physiological processes via different responses to the cellular microenvironment between men and women exist. A separate analysis using KEGG found that one of the final genes, *CTNND1*, is related to adherens junctions, which have been linked to *de novo* drug resistance in tumors [13]. Although the relationship between the overexpression of *CTNND1* in men and drug resistance is not clear, the E-cadherin-dependent intercellular adhesion may

enhance chemoresistance [14]. *TNFSF9*, a gene involved in another overrepresented pathway of cytokine-cytokine receptor interaction, has been found to be upregulated by treatment with certain drugs like irinotecan *in vitro* [15]. The fact that *TNFSF9* is involved in apoptosis signal transduction suggests its possible role in drug response. Although the precise mechanisms responsible for gender-specific differences in expression and physiological processes largely remain to be identified, our results provide some clues for further investigations.

One limitation of this study is the use of the HG-Focus Array, which represents only a fraction of the genes present in the human genome. Use of an expression chip with a more complete coverage of the genome such as the Affymetrix U133A Array or the Affymetrix Human Exon Array will enable a more comprehensive analysis regarding differences in expression owing to sex. In particular, data from the Exon Array will allow researchers to assess gender-specific effects at the exon level as well as at the gene level, and to extend investigation of gender influence to the area of alternative splicing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Bren L. Does sex make a difference? FDA Consumer Magazine 2005;39:10–15.

2. Weiss LA, Pan L, Abney M, Ober C. The sex-specific genetic architecture of quantitative traits in humans. Nat Genet 2006;38:218–222. [PubMed: 16429159]

3. Paine MF, Ludington SS, Chen ML, Stewart PW, Huang SM, Watkins PB. Do men and women differ in proximal small intestinal CYP3A or P-glucoprotein expression? Drug Metab Dispos 2005;33:426–433. [PubMed: 15608139]

4. Guengerich, FP. Human cytochrome P450 enzymes. In: Ortiz de Montellano, PR., editor. Cytochrome P450: structure, metabolism, and biochemistry. Plenum Press; New York, New York: 1995. p. 473-535.

5. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, et al. Genetic analysis of genome-wide variation in human gene expression. Nature 2004;430:743–747. [PubMed: 15269782]

6. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–29. [PubMed: 10802651]

7. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res 2004;32:D277–D280. [PubMed: 14681412]

8. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249–264. [PubMed: 12925520]

9. Westfall, PH.; Young, SS. Resampling-based multiple testing: examples and methods for P-value adjustment. Wiley Publishers; New York, New York: 1993.

10. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2005. http://www.R-project.org

11. Draghici S, Khatri P, Bhavsar P, Shah A, Krawetz S, Tainsky MA. Onto-Tools, the toolkit of the modern biologist: Onto-Express, Onto-Compare, Onto-Design and Onto-Translate. Nucleic Acids Res 2003;31:3775–3781. [PubMed: 12824416]

12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 1995;B57:289–300.

13. Hazlehurst LA, Landowski TH, Dalton WS. Role of the tumor microenvironment in mediating *de novo* resistance to drugs and physiological mediators of cell death. Oncogene 2003;22:7396–7402. [PubMed: 14576847]

14. Nakamura T, Kato Y, Fuji H, Horiuchi T, Chiba Y, Tanaka K. E-cadherin-dependent intercellular adhesion enhances chemoresistance. Int J Mol Med 2003;12:693–700. [PubMed: 14532995]

15. Minderman H, Conroy JM, O'Loughlin KL, McQuaid D, Quinn P, Li S, et al. In vitro and *in vivo* irinotecan-induced changes in expression profiles of cell cycle and apoptosis-associated genes in acute myeloid leukemia cells. Mol Cancer Ther 2005;4:885–900. [PubMed: 15956246]

Table 1

Genes that are differentially expressed between men and women

Probe set	Symbol	Title	t-score	Women (log₂)	Men (log₂)	Individual t-test P^a	Permutation-adjusted P^a	Family-adjusted P^a	Generation-adjusted P^a
203562_at	*FEZ1*	Fasciculation and elongation protein zeta 1 (zygin 1)	-3.82	6.69	7.26	$<1.0 \times 10^{-4}$	0.16	6.0×10^{-4}	3.0×10^{-4}
203411_s_at	*LMNA*	Lamin A/C	-3.66	8.36	8.59	3.0×10^{-4}	0.24	2.0×10^{-4}	$<1.0 \times 10^{-4}$
221345_at	*FFAR2*	Free fatty acid receptor 2	-3.66	4.86	5.13	$<1.0 \times 10^{-4}$	0.24	4.0×10^{-4}	1.0×10^{-4}
206907_at	*TNFSF9*	Tumor necrosis factor (ligand) super-family, member 9	-3.63	7.16	7.32	3.0×10^{-4}	0.26	5.0×10^{-4}	4.0×10^{-4}
201105_at	*LGALS1*	Lectin, galactoside-binding, soluble, 1 (galectin 1)	-3.57	10.75	10.95	2.0×10^{-4}	0.31	6.2×10^{-4}	3.0×10^{-4}
208862_s_at	*CTNND1*	Catenin (cadherin-as-sociated protein), delta 1	-3.42	6.40	6.55	3.0×10^{-4}	0.42	2.5×10^{-3}	4.0×10^{-4}
201015_s_at	*JUP*	Junction plakoglobin	3.36	5.69	5.39	3.0×10^{-4}	0.48	4.0×10^{-4}	6.0×10^{-4}
219371_s_at	*KLF2*	Kruppel-like factor 2 (lung)	3.36	6.82	6.55	3.0×10^{-4}	0.48	1.0×10^{-2}	5.0×10^{-4}
202724_s_at	*FOXO1A*	Forkhead box O1A (rhabdomyosarcoma)	3.41	5.81	5.65	4.0×10^{-4}	0.43	2.0×10^{-4}	2.0×10^{-4}
218059_at	*ZNF706*	Zinc finger protein 706	3.46	9.46	9.39	2.0×10^{-4}	0.39	5.7×10^{-3}	5.0×10^{-4}

[a] One-sided *P* value.

Table 2

Overrepresented GO terms and KEGG pathways for 10 differentially expressed genes and corrected *P* values

GO-biological process	Cell adhesion (3.5×10^{-3}) *JUP, FEZ1*	Transcription (1.8×10^{-2}) *KLF2, FOXO1A*	Signal transduction (1.9×10^{-2}) *FFAR2, TNFSF9*	Apoptosis (4.3×10^{-3}) *TNFSF9, LGALS1*
GO-molecular function	Transcription factor activity (3.0×10^{-2}) *KLF2, FOXO1A*	Structural molecule activity (6.8×10^{-5}) *CTNND1, LMNA, JUP*	Zinc ion binding (2.8×10^{-2}) *ZNF706, KLF2*	Protein binding (4.8×10^{-2}) *CTNND1, LMNA*
GO-cellular component	Cytoplasm (4.3×10^{-2}) *CTNND1, JUP*	Cytoskeleton (5.6×10^{-3}) *CTNND1, JUP*	Integral plasma membrane (2.7×10^{-2}) *FFAR2, TNFSF9*	Nucleus (2.3×10^{-2}) *LMNA, ZNF706, CTNND1, KLF2, FOXO1A*
KEGG pathway	Cytokine-cytokine receptor interaction (1.2×10^{-2}) *TNFSF9*	Adherens junction (3.5×10^{-3}) *CTNND1*		

GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.