Gene expression profiling of whole-blood samples from women exposed to hormone replacement therapy

Vanessa Dumeaux,[1,2,3] Jostein Johansen,[1] Anne-Lise Børresen-Dale,[3] and Eiliv Lund[1]

[1]Institute of Community Medicine, University of Tromsø, Tromsø, Norway; [2]Equipe E3N-EPIC, Institut National de la Sante et de la Recherche Medicale XR521, Institut Gustave-Roussy, Paris, France; and [3]Department of Genetics, Institute for Cancer Research, Faculty Division, The Norwegian Radium Hospital, University of Oslo, Oslo, Norway

Abstract

The American Women's Health Initiative study published in July 2002 caused considerable concern among hormone replacement therapy (HRT) users and prescribers in many countries. This study is an exploratory research comparing the genome-wide expression profile in whole-blood samples according to HRT use. Within the Norwegian Women and Cancer study, 100 postmenopausal women (50 HRT users and 50 non-HRT users) born between 1943 and 1949 with normal to high body mass index and no other medication use were selected. After total RNA extraction, amplification, and labeling, the samples were hybridized together with a common reference (Universal human reference RNA, Stratagen) to Agilent Human 1A oligoarrays (G4110b, Agilent Technologies) containing 20,173 unique genes. Differentially expressed genes were used to build a classifier using the nearest shrunken centroid method (PAM). Then, we tested the significant changes in single genes by different methods like t test, Significance Analysis of Microarrays, and Bayesian ANOVA analysis. Results did not reveal any distinct gene list which predicted accurately HRT exposure (error rate, 0.40). Classifier performance slightly improved (error rate, 0.26) including only women who were using continuous combined HRT treatment. According to the small amplitude of expression alterations observed in whole blood, more quantitative technique and larger sample sizes will be needed to be able to investigate whether significant single genes are differentially expressed in HRT versus non-HRT users. Taken cautiously, significant enrichments in biological process of genes with small changes after HRT use were observed (e.g., receptor and transporter activities, immune response, frizzled signaling pathway, actin filament organization, and glycogen metabolism). [Mol Cancer Ther 2006;5(4):868–76]

Introduction

The findings of the American Women's Health Initiative study, published in July 2002 (1), profoundly affected patient and clinician perceptions of the role of hormone therapy in the management of menopause. This placebo-controlled trial of an oral continuous combined conjugated equine estrogens plus medroxyprogesterone acetate regimen was prematurely discontinued because the overall health risks (breast cancer, heart attacks, stroke, and blood clots) exceeded the benefits (fewer fractures and colorectal cancers) and caused considerable concerns among hormone replacement therapy (HRT) users and prescribers in many countries. The results indicated that HRT (*a*) does not confer cardiovascular or cognitive protection; (*b*) increases breast cancer risk in women with a uterus; (*c*) increases stroke risk in women with hysterectomy; and (*d*) does not improve overall quality of life. HRT does, however, decrease fracture rates and vasomotor symptoms (1).

Quantitative data on global gene expression offer the opportunity not only to indicate or confirm exposure to specific chemical classes but also to identify biomarkers (single genes or patterns of gene expression) that are indicative of the initiation/presence of toxicant mechanisms or that can be linked to specific pathology (2, 3). Due to the pleiotropic activity of estrogens and progestins, more research is needed to better understand the effect of HRT. Furthermore, there is a great heterogeneity in the way individuals respond to hormonal treatment. We believe that large-scale gene expression study may present promising new insights on hormone effects through still unclear multiple pathways.

Peripheral blood is an easily accessible source of RNA for analysis of environmental exposure (3, 4). Indeed, blood constituents maintain homeostasis, effect immunity or inflammation, participate in stress signaling, and mediate cellular communication in vascular associated tissues, including those of the central nervous system. One of the challenges is to preserve valid gene expression profiles from sample collection to isolation and how to maximize RNA yields and purity for sensitive downstream analyses.

This study is then an exploratory research comparing the genome-wide expression profile of ~20,000 genes in whole-blood samples according to the use of HRT in a cross-sectional analysis within the prospective follow-up study "Norwegian Women and Cancer study (NOWAC)."

Table 1. Characteristics of postmenopausal women exposed to HRT

Type of HRT	Drug name	Description	n	Mean duration of use, y (SD)
Continuous combined preparation (28 d)	Activelle	1 mg 17β-estradiol + 0,5 mg norethisterone acetate	11	6.1 (3.8)
	Kliogest	2 mg 17β-estradiol + 1 mg norethisterone acetate	14	
Estrogen-only preparation (21 d)	Progynova	1 or 2 mg estradiol valerate	6	7.3 (3.1)
	Ovesterin	1 mg estriol	1	
Sequential combined treatment	Trisekvens	2 or 4 mg 17β-estradiol + 1 mg norethisterone acetate (12 + 10 + 6 d)	4	4.2 (3.0)
	Cyclabil	1 mg estradiol valeroate + 0.25 mg levonorgestrel (11 + 10 d)	1	
Tibolone	Livial	Continuous treatment (28 d) 2.5 mg tibolone	5	4.0 (2.7)

Materials and Methods

Study Population

The study population was drawn from women who are participants of the large prospective cohort NOWAC study (5).[4] In early fall 2003, a letter of invitation, together with a two-page questionnaire asking on various exposure factors including hormonal therapy, was sent to a random subsample of 500 women in NOWAC who had previously agreed to give blood samples. In late fall 2003, 393 women (response rate, 78.6%) had returned the questionnaire and a blood sample. We selected participants to obtain postmenopausal women born between 1943 and 1949 who did not use other medications and had a body mass index between 18.5 and 29.9 kg/m². In the sample size calculation, we defined $\alpha = 0.00005$ and setting $\beta = 0.01$ that provides 99% probability of detecting a 2-fold change in gene expression. The resulting total sample size was then 100 for $\sigma = 0.75$ (6). Thus, 50 women using HRT and 50 women who never used HRT corresponding to our inclusion criteria cited above were selected.

The study activities were approved by the National Data Inspection Boards and Regional Ethical Committee for Medical Research.

Blood Samples and Microarray Procedures

Blood was drawn in PAXgene tube, shipped to the institute at room temperature, and stored at −80°C. Total RNA was isolated from 2.5 mL of whole blood with the PAXgene Blood RNA System (PreAnalytiX, Hombrechtikon, Switzerland). Three hundred nanograms of total RNA were amplified and labeled with cyanine 5 for samples (and with cyanine 3 for the Universal human reference RNA, Stratagen, La Jolla, CA). Before and after the amplification/labeling procedure, RNA quantity (and labeling) was measured on the NanoDrop Spectrophotometer and RNA quality was checked on Agilent 2100 Bioanalyzer using RNA 6000 Nano Assay kit.

Microarray analyses were prepared essentially according to the instructions of the manufacturer (7). Briefly, labeled cRNAs were hybridized overnight to Agilent Human 1A (v2) oligoarrays (Agilent Technologies, Palo Alto, CA) containing 22,153 features representing 20,173 genes. After washing (use of stabilization and drying solution, Agilent Technologies), microarrays were scanned in Agilent model G2505A microarray scanner.

Images were processed using Agilent Feature Extraction software version A.7.5.1. The Feature Extraction Software provides normalized cyanine 3 and cyanine 5 channel intensity values for each spot on an array (in the gProcessedSignal and rProcessedSignal fields of the output files). The default settings were used for all options. Quality control algorithms in the software detected poor quality spots that were excluded from analysis and contained a nonzero value for any of the following fields: IsSaturated, IsFeatNonUnifOL, IsBGNonUnifOL, IsFeatPopnOL, IsBGPopnOL, and IsManualFlag. For a detailed description of the Agilent Feature Extraction software and the algorithms it uses, see the Agilent Feature Extraction version 7.5.1 User Manual. Briefly, Agilent Feature Extraction determines the raw spot value calculated as its foreground value minus its background value. A surrogate raw value is assigned when the foreground value does not exceed the background value by 2 SDs of the spot background pixel values. Intensity-based normalization between channels using a linear regression and a lowess curve-fit technique is then applied to remove any systematic dye incorporation biases. Microarray data with accession number GSE3492 are online at the GEO database.[5]

Data Analysis

Data analysis was done using R,[6] an open-source-interpreted computer language for statistical computation

---

[4] http://uit.no/kk/NOWAC/

[5] http://www.ncbi.nlm.nih.gov/geo/
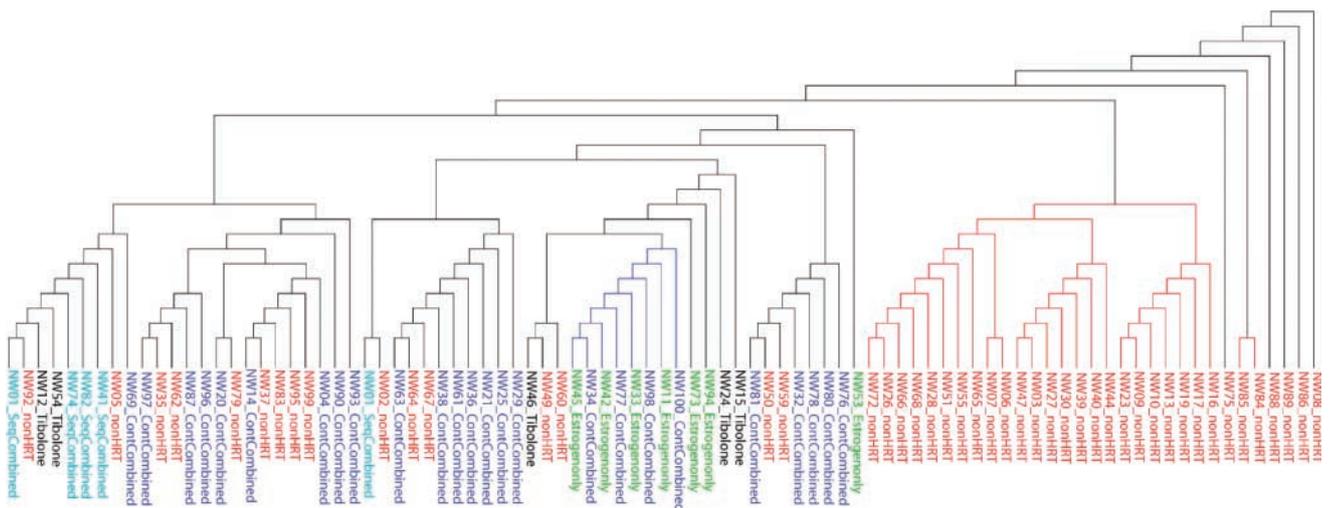[6] http://cran.r-project.org/

Figure 1. Samples tree after average linkage hierarchical clustering using Euclidian distance based on all 16,143 genes. *Red,* non-HRT (_nonHRT) users; *dark blue,* continuous combined treatment (_ContCombined) users; *green,* estrogen only treatment (_Estrogenonly) users; *light blue,* sequential combined treatment (_SeqCombined) users; *black,* Tibolone (_Tibolone) users.

and graphics, and tools from the Bioconductor project,[7] adapted to our needs. Genes with >20% of missing values were excluded (N = 4,030). Based on the results of Troyanskaya et al. (8), the missing values (4.5%) were imputed using the K-nearest neighbor–based method (K = 10). The resulting 16,143 genes were used to conduct the analyses.

We used a filter method based on the Student's t test to select the most important features. Comparing HRT with non-HRT users, we chose a nonrestrictive threshold for the noncorrected P value ($P < 0.03$) to allow genes to still have a positive effect and preserve information to build the classifier; 272 genes were in this way selected. Stratifying on specific HRT regimen, only continuous combined treatment (ethinyl estradiol-norethisterone acetate) were included in the HRT users group. A new Student test was conducted and 444 genes had $P < 0.03$. In this case, more genes were selected and filtering was finally based on $P < 0.02$, which gave approximately the same number of genes (N = 297) as in the analysis including all HRT users. After feature selection, genes were used to classify and predict the HRT exposure by the method of the nearest shrunken centroid method (9) using R functions for PAM software.[8]

Then, we searched our data for differentially expressed genes in blood RNA from HRT users versus non-HRT users. We used the usual t test and volcano plot for the visual identification of genes with altered expression (10). Accompanying this method, Significance Analysis of Microarrays (SAM), a software program developed at Stanford University (11), selects the most pertinent genes with control of the false discovery rate. In addition, we used BAMarray 2.0, a Java software package that imple-

ments the Bayesian ANOVA for microarray (BAM) algorithm (12). The BAM approach uses a special type of inferential regularization known as spike-and-slab shrinkage, which provides an optimal balance between total false detections (the total number of genes falsely identified as being differentially expressed) and total false nondetections (the total number of genes falsely identified as being nondifferentially expressed; ref. 12).

Finally, we assessed the biological functions enriched with top-ranking genes from the t test ($P < 0.03$) using the annotation defined by the Gene Ontology Consortium.[9] We chose a nonrestrictive threshold for the noncorrected P value to allow genes that might not have been selected on their own merits to still have a positive effect and preserve information contained in the Gene Ontology classes scores (13). The probability of observing a particular number of genes in one Gene Ontology biological process was tested using GOMiner (14). GOMiner performs analysis using genes associated with a node in Gene Ontology or genes associated with any children of that node, sometimes called ''inclusive analysis.'' A functional pathway was regarded when it had ≥2 genes contained in the top-ranking gene list and <500 genes defined in the Gene Ontology category. We reported the Gene Ontology categories for which the one-tailed Fisher exact P value based on underregulated, overregulated, or changed genes was <0.015.

Results

Study Population Characteristics

We obtained RNA from whole-blood samples without evidence of degradation on 47 non-HRT users and 47 HRT users. We excluded five women among the HRT users: one
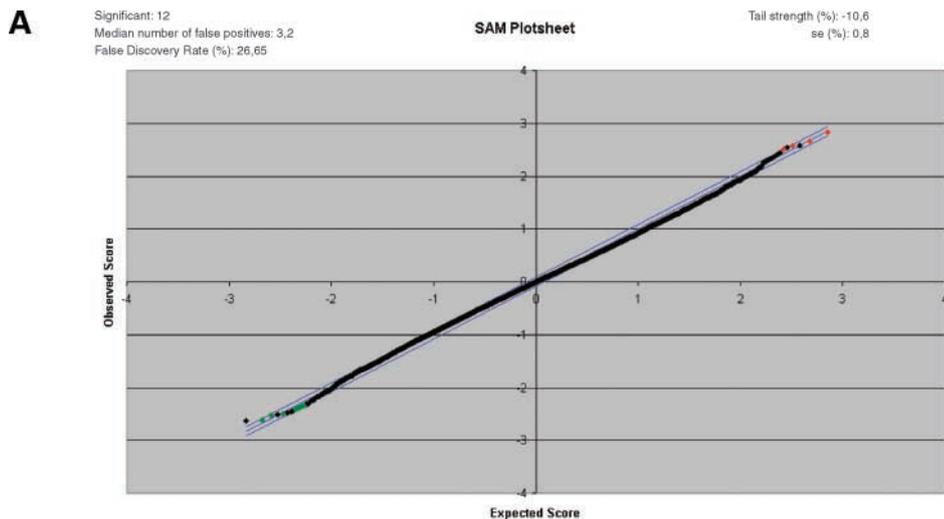
Figure 2. Partial least square analyses done on whole-blood gene expression from women who never used HRT (*blue*) and women who were using HRT (*green*). In **A**, women who were using all types of HRT regimens were included in HRT users group; in **B**, only women who were using continuous combined treatment were included in HRT users group.

was using a cream and four were using transdermal patch. Finally, 47 women non-HRT users and 42 HRT users were available for the analyses.

At the time of blood sampling, postmenopausal women who were using HRT were slightly younger than women who were not using HRT (56.6 and 57.3 among HRT users and non-HRT users, respectively). The mean body mass index was approximately the same among HRT users (24.2 ± 2.5 kg/m^2) and non-HRT users (24.3 ± 2.7 kg/m^2).

Among HRT users, the mean age at first HRT use was 49.1 years (SD 3.9), ranging from 35 to 55 years. HRT users had been on HRT for at least 1 year and for a maximum of 20 years (mean, 5.8; SD 3.5). We categorized HRT users by regimen: continuous combined preparation, estrogen-only preparation, sequential combined treatment, and tibolone (Table 1). A large majority of exposed women were using continuous combined treatment with either use of Activelle or Kliogest (Table 1).

Classification and Prediction of HRT Exposure

Unsupervised method like principal component analysis did not show any clear subgroup of samples with similar expression patterns that associated to HRT use even when only women using continuous combined treatment were

Figure 3. Volcano plot reveals a small proportion of genes that met our arbitrary criteria. *Horizontal line, P* = 0.01; *vertical lines,* 1.5-fold changes, both up- and down-regulated.

Figure 4. Significance Analysis of Microarrays. SAM plot (**A**) and list of significant genes (**B**). *Red* and *green,* up-regulated and down-regulated genes, respectively.

Probe Name	Gene Symbol	Gene Name	Score(d)*	Fold Change	q-value (%)†
A_23_P141275	HLF	hepatic leukemia factor	2.83	1.66	26.65
A_23_P53257	AVIL	advillin	2.65	1.57	26.65
A_23_P67871	SLC5A7	solute carrier family 5 (choline transporter). member 7	2.57	1.77	26.65
A_23_P216742	PAPPA	pregnancy-associated plasma protein A. pappalysin 1	2.50	1.67	26.65
A_23_P28181	PTHR2	parathyroid hormone receptor 2	-2.62	0.56	26.65
A_23_P53119	OR52J3	olfactory receptor. family 52. subfamily J. member 3	-2.53	0.66	26.65
A_23_P300090	DKFZP566M114	hypothetical protein DKFZp566M114	-2.50	0.64	26.65
A_23_P46879	PNLIP	pancreatic lipase	-2.42	0.64	26.65
A_23_P200670	FLJ23129	hypothetical protein FLJ23129	-2.40	0.52	26.65
A_23_P94288	DEFA4	defensin. alpha 4. corticostatin	-2.38	0.51	26.65
A_23_P80529	FLJ39873	hypothetical protein FLJ39873	-2.36	0.65	26.65
A_23_P208167	FPRL2	formyl peptide receptor-like 2	-2.35	0.51	26.65

* The *t*-statistic value

† The lowest False Discovery Rate at which the gene is called significant

included (data not shown). Average linkage hierarchical clustering analyses using Euclidean distance grouped 31 of 47 non-HRT users (Fig. 1). However, we did not observe distinct clusters according to type of HRT regimen (Fig. 1). We conducted supervised analysis using partial least square method (Fig. 2). First, all women using HRT were included in the HRT group and only a small difference in gene expression was observed in the two groups of women (Fig. 2A). Second, only women using continuous combined treatment (ethinyl estradiol-norethisterone acetate) were included in the HRT users group. A slightly greater difference was observed when only users of continuous combined treatment were included as HRT users (Fig. 2B).

Genes (N = 272) identified after feature selection using *t* test (see Materials and Methods) were used to build a classifier using the nearest shrunken centroid method. The PAM results did not reveal any distinct gene list which predicted accurately the HRT exposure (error rate, 0.40). When we considered in the HRT group only women who were using continuous combined treatment and after a new feature selection using *t* test (N = 297; see Materials and Methods), the performance of the classifier slightly improved (error rate, 0.26; see Supplement 1).[10] It was based on 103 genes (see Supplement 2)[10] and had a slightly higher specificity (78.7%) than sensitivity (64.0%).

---

[10] Supplementary material for this article is available at Molecular Cancer Therapeutics Online (http://mct.aacrjournals.org/).

Identifying Genes with Altered Expressions by *t* test, SAM, and BAM

Initially, we tried to find genes with large alterations of expressions and highly significant between HRT users and non-HRT users. Volcano plot (Fig. 3) shows that very few genes had a 1.5-fold greater mean expression levels between control and treatment groups.

When all the 16,143 genes were included in the Student test, we found 519 genes with altered gene expression in HRT users versus non-HRT users with an uncorrected $P < 0.05$. Weak signals of exposure-induced changes made it more difficult to conduct valid multiple testing adjustment. One of our approaches was to conduct the SAM statistics and to estimate the false discovery rate (Fig. 4). Few genes ($N = 12$) were found significantly differentially expressed among HRT users compared with non-HRT users with a global false discovery rate of 26.6% (Fig. 4A). Differentially expressed genes are listed in Fig. 4B. Another approach was to conduct a Bayesian ANOVA analysis (BAM) which controls false discovery rate but optimizes also the true positive rate

(Fig. 5A; Supplement 3).[10] Only five genes were found differentially expressed by both methods: three genes [hepatic leukemia factor (*HLF*), advillin (*AVIL*), and pregnancy-associated plasma protein A (*PAPPA*)] were up-regulated and two genes [hypothesitcal protein FLJ23129 (*FLJ23129*) and corticostatin (*DEFA4*)] were down-regulated among HRT users compared with non-HRT users.

BAM has been shown especially effective in multiclass analyses. We used the same method distinguishing the different groups of HRT regimen (Fig. 5B; Supplement 4).[10] Among 22 genes, 1 down-regulated and 6 up-regulated genes were more likely to be truly differentially expressed among users of continuous combined treatment compared with non-HRT users (Supplement 4).[10] It was too few users in other HRT regimen groups to conclude.

Functional Clustering Using Gene Ontology

Genes ($N = 272$) identified after feature selection were evaluated by hypergeometric distribution testing based on Gene Ontology annotations to define any bioprocesses enriched with the identified genes. We identified 14 functional pathways ($P < 0.015$) with significant enrichment of genes having altered expressions in response to all HRT (Table 2). We chose to classify pathways as down-regulated or up-regulated. Most of them were clearly down-regulated (9 of 14). We identified several genes involved in receptor activity: both thyroid receptor (*PAX8* and *TSHR*) and several rhodopsin-like receptor (*CNR1*, *DRD1*, *EDG8*, *FPRL2*, *GPR147*, *OR10H3*, *OR52J3*, *OPN1MW*, *OR4Q3*, *TBXA2R*, *TSHR*, *PAX8*, and *GP1BA*) processes were down-regulated. HRT use also decreased blood transporter activities: carboxylic acid and neutral amino-acid transporter activities. We observed also that HRT modulates immune functions (defense response to fungi). One of the significantly down-regulated pathway included genes that were related to structural constituent of eye lens (*BFSP2*, *CRYGS*, and *LIM2*). G protein–coupled receptor activity (unknown ligand), frizzled signaling pathway, response to hypoxia, glycogen metabolism, and actin filament organization were found up-regulated after HRT use. Including only combined treatment users and using the 262 genes with $P < 0.02$, five pathways were down-regulated and seven pathways were up-regulated among continuous combined treatment users (see Supplement 5).[10] Defense response to fungi, actin filament organization, and G protein receptor activity were found deregulated as we observed in analysis including all HRT users. Some transport activities (bile acid and protein) were down-regulated among continuous combined treatment users. Intracellular junction and metalloendopeptidase activity were also found down-regulated (see Supplement 5).[10] Finally, other pathways were found significantly up-regulated among users of continuous combined treatment compared with non-HRT users: vitamin metabolism, neutrophil chemotaxis, gluthatione transferase activity, and immune cell chemotaxis (see Supplement 5).[10]
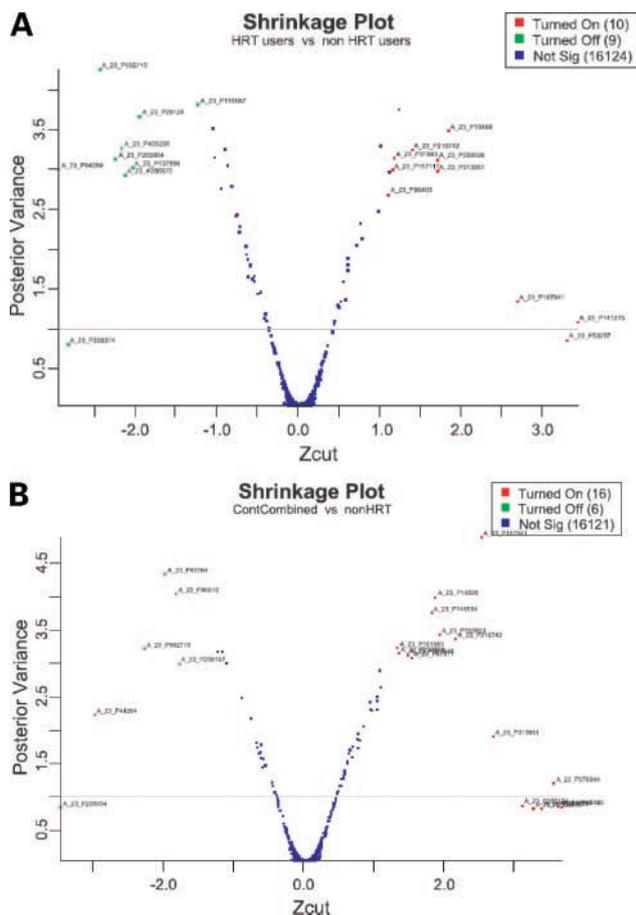

Figure 5. Genes differentially expressed among all HRT users (**A**) and in continuous combined treatment users only (**B**) compared with non-HRT users using Bayesian ANOVA method. Genes with a posteriori variance of ~1 are more likely to be truly differentially expressed.

Table 2. Significant bioprocesses based on gene ontology terms associated to the top ranking gene list (N = 272) selected by t test

Gene Ontology ID	Term	Total	Under	Gene symbol (under)*	Over	Gene symbol (over)*	P (under)	P (over)
Down-regulated pathways*								
4996	Thyroid-stimulating hormone receptor activity	2	2	*PAX8, TSHR*	0		0.0002	1
46943	Carboxylic acid transporter activity	73	6	*AKR1C4, SLC6A9, SLC1A7, SLC7A10, SLC38A3, ABCC3*	0		0.0003	1
15175	Neutral amino acid transporter activity	15	3	*SLC6A9, SLC7A10, SLC38A3*	0		0.0009	1
16307	Phosphatidylinositol phosphate kinase activity	8	2	*PIP5K2A, PIK3C2B*	0		0.0044	1
1584	Rhodopsin-like receptor activity	479	13	*CNR1, DRD1, EDG8, FPRL2, GPR147, OR10H3, OR52J3, OPN1MW, OR4Q3, TBXA2R, TSHR, GP1BA, PAX8*	3	*GPR84, GPR75, OR11H6*	0.009	0.573
5212	Structural constituent of eye lens	12	2	*CRYGS, LIM2*	0		0.0101	1
6805	Xenobiotic metabolism	35	3	*DEFA4, GNLY, GSTP1*	1	*DEFA6*	0.0102	0.195
122	Negative regulation of transcription from RNA polymerase II promoter	105	5	*HIPK2, IRF2, MTA2, RETN, ZFPM2*	1	*ZMYND11*	0.0115	0.480
3799	Defense response to fungi	13	2	*DEFA4, GNLY*	1	*DEFA6*	0.0118	0.077
Up-regulated pathways*								
16526	G protein coupled receptor activity, unknown ligand	28	0		3	*GPR116, GPR 75, GPR84*	1	0.0007
7222	Frizzled signaling pathway	15	0		2	*LEF1, TLE4*	1	0.0037
1666	Response to hypoxia	25	0		2	*EGLN1, HLF*	1	0.0103
6073	Glycogen metabolism	26	0		2	*EPM2A, PHKA1*	1	0.0111
7015	Actin filament organization	57	3	*PLS3, TMOD1, GLA*	2	*AVIL, PTPNS1, RND1*	0.037[†]	0.048[†]

NOTE: P values from Fisher's exact test are reported for down-regulated and up-regulated gene groups, respectively, among all HRT users compared with non-HRT users.

*Underregulated or overregulated among all HRT users compared with non-HRT users.

[†]P = 0.01, for changed (up- and down-regulated) genes.

Discussion

In our study, no significant difference in gene expression according to HRT exposure was observed from whole-blood RNA in normal women. Although no good classifier could be identified from our full data set, results were slightly better when HRT group was composed only by women who were using the same type of HRT regimen. Fold changes in gene expression had minor magnitude and made the detection of single genes in significant test almost impossible. Taken cautiously, significant enrichments in biological process of genes with small changes after HRT use were assessable.

Several reports have shown that total RNA isolated from whole blood drawn in PAXgene tube offer a good quality of RNA suitable for gene expression analysis by microarray technology (15, 16). Several cellular components are analyzed using the PAXgene technology (erythrocytes, reticulocytes, and granulocytes), and the gene expression profile obtained is then the combined gene expression profiles from each cell subpopulation, weighted according to its relative frequency in the cell mixture (4, 17). Recent studies show that the sensitivity for gene expression profiling from PAXgene tubes was reduced certainly due to the presence of reticulocyte-derived globin transcripts, which can represent up to 95% of the total cellular mRNA (18).

In this context, our exploratory study certainly lacks statistical power to identify single gene effect of minor amplitude. Other studies have found that gene expression from whole-blood RNA can be of little magnitude (2, 19). This can be, as explained above, due not only to the reduced sensitivity of gene expression profile from PAXgene but also to the gene activity per se of blood cells. It is possible that leukocyte gene expression data can be linked to clinical, biochemical, and pathologic changes in the human body to slightly adjust their response to physiologic disorders. However, one might expect that these adjustments are not of big amplitude to avoid any hyper- or

hypoactivity of the normal physiologic processes. An experiment conducted among ovariectomized rats (19) shows that there is a significant overlap of gene expression profiles in uterus tissue and blood cells after estrogen treatment. Furthermore, fold changes of gene expression had less amplitude in blood cells than in uterus tissue, which is in accordance with our results.

Our a priori sample size calculation was done to detect a 2-fold change in gene expression. The required sample size of an experiment depends on the variance component that we estimated from literature to be 0.75. From our data, we can estimate the distribution of the SDs (Fig. 6A) and then choose a value of variance as the upper 75th percentile of variance across all genes ($\sigma = 0.55$), which imply that the sample size calculations will ensure us of having the desired power to detect a chosen n-fold change for all but the 75% of most variable genes (20). Figure 6B shows the
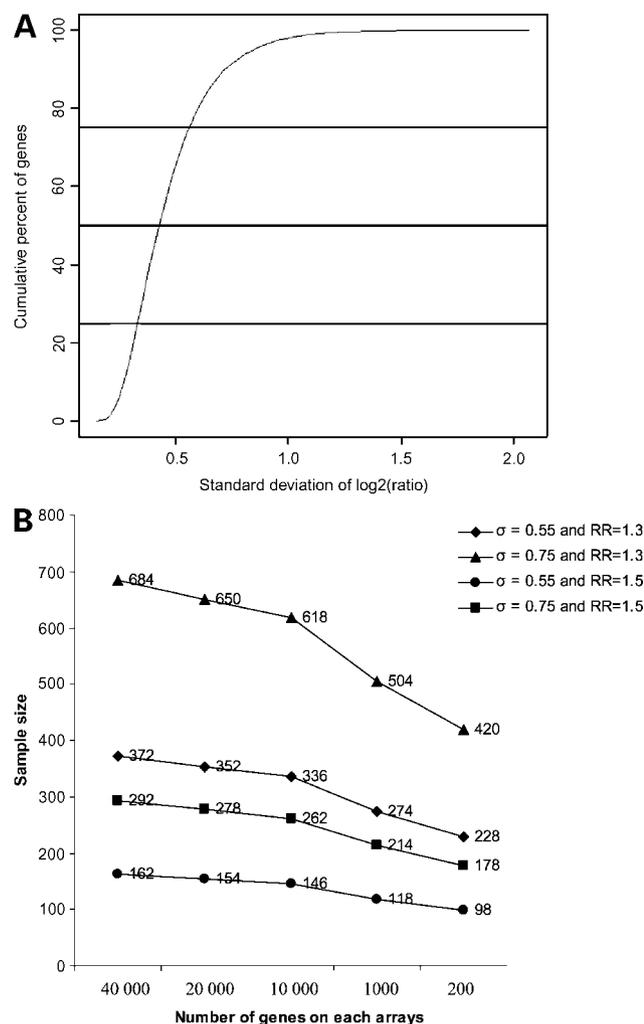
estimated sample size required to detect a 1.3- to 1.5-fold change in expression level for the 90th and 75th percentile genes for a false-positive rate equal to 1 and a power of 99%. According to our results, detection of 1.3-fold change in gene expression should give substantial insights about transcriptome changes in blood cells associated to HRT use. The required sample size would then be 354 and 154 for detection of 1.3- and 1.5-fold change, respectively. We may then have missed statistical power in our study to identify single genes differentially expressed after HRT exposure.

Furthermore, the lack of difference of gene expression profiles from HRT and non-HRT users blood cells may also be explained by exposure misclassification. Indeed, use of HRT may not correlate with physiologic hormone levels. Women who never take HRT may have higher endogenous hormone levels and so less need to use HRT. In parallel, HRT users may metabolize differently the administrated drug and, as a result, be exposed to different circulating hormone levels.

Although one gene expression change may be small and difficult to detect accurately in a significant test, significant enrichments in biological process of genes with small changes after HRT use were assessable. We identified some genes involved in receptor activity: thyroid and several rhodposin-like receptors which were down-regulated. Notably, in accordance with our results, data presented by Proenza et al. (21) showed that blood-cell amino acid pool absorbed by erythrocytes membrane decreased after estrogen treatment. Another finding is the HRT effect on structural constituents of eye lens for which three genes implicated in lens opacities are down-regulated in accordance with several studies (22–24). In parallel, there is extensive experimental evidence that sex steroids and insulin interact in their actions on tissues (25). Although this interpretation should be taken cautiously, our results also suggest that HRT exposure could regulate the early B-cell development through *TLE4* and *LEF1* (26–28). In the same manner, deregulated pathways identified after including only continuous combined treatment users could be discussed. Noteworthy, *GSTA3* has been found to specifically participate in the biosynthesis of testosterone and progesterone in steroidogenic tissues (29), and *GSTP1* has been implicated in breast cancer etiology (30). Although certainly of interest, these interpretations remain speculative given that, overall, no statistically significant classifier was identified and that statistical power was missing to identify single gene effects of low magnitude due to lack of technical sensitivity and small sample size.

Figure 6. SD and sample size estimations. **A,** histogram of SD. *X axis,* SD; *Y axis,* percentage of genes that has SD below the value of *X*. **B,** sample size calculations according to the number of genes tested, the estimated SD, and the fold-change detection.

References

1. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 2002;288:321–33.

2. Wang Z, Neuburg D, Li C, et al. Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. Environ Health Perspect 2005;113:233–41.

3. Lampe JW, Stepaniants SB, Mao M, et al. Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. Cancer Epidemiol Biomarkers Prev 2004;13:445–53.

4. Whitney AR, Diehn M, Popper SJ, et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci U S A 2003; 100:1896–901.

5. Lund E, Kumle M, Braaten T, et al. External validity in a population-based national prospective study-the Norwegian Women and Cancer Study (NOWAC). Cancer Causes Control 2003;14:1001–8.

6. Simon R, Radmacher MD, Dobbin K. Design of studies using DNA microarrays. Genet Epidemiol 2002;23:21–36.

7. Agilent Low RNA Input Fluorescent Linear Amplification Kit Protocol Rev.2.0. 5184-3523. Agilent Technologies 2003. Available from: http://www.agilent.com/chem/supplies.

8. Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays. Bioinformatics 2001;17:520–5.

9. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002;99:6567–72.

10. Yang YH, Speed T. Design and analysis of comparative microarray experiments. Chapman & Hall/CRC; 2003. p. 35–91.

11. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001; 98:5116–21.

12. Ishwaran H, Rao JP. Detecting differentially expressed genes in microarrays using Bayesian model selection. J Am Stat Assoc 2003;98:438–55.

13. Pavlidis P, Qin J, Arango V, Mann JJ, Sibille E. Using the gene ontology for microarray data mining: a comparison of methods and application to age effects in human prefrontal cortex. Neurochem Res 2004;29:1213–22.

14. Zeeberg BR, Feng W, Wang G, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol 2003;4:R28.

15. Thach DC, Lin B, Walter E, et al. Assessment of two methods for handling blood in collection tubes with RNA stabilizing agent for surveillance of gene expression profiles with high density microarrays. J Immunol Methods 2003;283:269–79.

16. Feezor RJ, Baker HV, Mindrinos M, et al. Whole blood and leukocyte RNA isolation for gene expression analyses. Physiol Genomics 2004;19: 247–54.

17. Szaniszlo P, Wang N, Sinha M, et al. Getting the right cells to the array: gene expression microarray analysis of cell mixtures and sorted cells. Cytometry A 2004;59:191–202.

18. Debey S, Schoenbeck U, Hellmich M, et al. Comparison of different isolation techniques prior gene expression profiling of blood derived cells: impact on physiological responses, on overall expression and the role of different cell types. Pharmacogenomics J 2004;4:193–207.

19. Rockett JC, Kavlock RJ, Lambright CR, et al. DNA arrays to monitor gene expression in rat blood and uterus following 17β-estradiol exposure: biomonitoring environmental effects using surrogate tissues. Toxicol Sci 2002;69:49–59.

20. Wei C, Li J, Bumgarner RE. Sample size for detecting differentially expressed genes in microarray experiments. BMC Genomics 2004;5:87.

21. Proenza AM, Oliver J, Palou A, Roca P. Breast and lung cancer are associated with a decrease in blood cell amino acid content. J Nutr Biochem 2003;14:133–8.

22. Younan C, Mitchell P, Cumming RG, Panchapakesan J, Rochtchina E, Hales AM. Hormone replacement therapy, reproductive factors, and the incidence of cataract and cataract surgery: the Blue Mountains Eye Study. Am J Epidemiol 2002;155:997–1006.

23. Cumming RG, Mitchell P. Hormone replacement therapy, reproductive factors, and cataract. The Blue Mountains Eye Study. Am J Epidemiol 1997;145:242–9.

24. Klein BE, Klein R, Ritter LL. Is there evidence of an estrogen effect on age-related lens opacities? The Beaver Dam Eye Study. Arch Ophthalmol 1994;112:85–91.

25. Livingstone C, Collison M. Sex steroids and insulin resistance. Clin Sci (Lond) 2002;102:151–66.

26. Reya T, O'Riordan M, Okamura R, et al. Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. Immunity 2000;13:15–24.

27. Jin ZX, Kishi H, Wei XC, Matsuda T, Saito S, Muraguchi A. Lymphoid enhancer-binding factor-1 binds and activates the recombination-activating gene-2 promoter together with c-Myb and Pax-5 in immature B cells. J Immunol 2002;169:3783–92.

28. Eberhard D, Jimenez G, Heavey B, Busslinger M. Transcriptional repression by Pax5 (BSAP) through interaction with corepressors of the Groucho family. EMBO J 2000;19:2292–303.

29. Johansson AS, Mannervik B. Human glutathione transferase A3–3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones. J Biol Chem 2001;276:33061–5.

30. Thompson PA, Ambrosone C. Molecular epidemiology of genetic polymorphisms in estrogen metabolizing enzymes in human breast cancer. J Natl Cancer Inst Monogr 2000;27:125–34.

Molecular Cancer Therapeutics

Gene expression profiling of whole-blood samples from women exposed to hormone replacement therapy

Vanessa Dumeaux, Jostein Johansen, Anne-Lise Børresen-Dale, et al.

Updated version	Access the most recent version of this article at: http://mct.aacrjournals.org/content/5/4/868
Supplementary Material	Access the most recent supplemental material at: http://mct.aacrjournals.org/content/suppl/2006/05/10/5.4.868.DC1

Cited articles	This article cites 28 articles, 7 of which you can access for free at: http://mct.aacrjournals.org/content/5/4/868.full#ref-list-1
Citing articles	This article has been cited by 1 HighWire-hosted articles. Access the articles at: http://mct.aacrjournals.org/content/5/4/868.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://mct.aacrjournals.org/content/5/4/868. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.