Gene-expression patterns predict phenotypes of immune-mediated thrombosis

Anil Potti, Andrea Bild, Holly K. Dressman, Deborah A. Lewis, Joseph R. Nevins, and Thomas L. Ortel

Antiphospholipid antibody syndrome (APS) is a complex autoimmune thrombotic disorder with defined clinical phenotypes. Although not all patients with elevated antiphospholipid antibody (aPLA) levels develop complications, the severity of these potential events mandates aggressive and extended lifelong antithrombotic therapy. One hundred twenty-nine patients (57 patients with APS and venous thromboembolism [VTE], 32 patients with VTE without aPLA, 32 patients with aPLA only, and 8 healthy patients) were enrolled. RNA from peripheral-blood collection was used for DNA microarray analysis. Patterns of gene expression that characterize APS as well as thrombosis in the presence of aPLA were identified by hierarchical clustering and binary regression methods. Gene-expression profiles identify and predict individuals with APS from patients with VTE without aPLA. Importantly, similar methods identified expression profiles that accurately predicted those patients with aPLA at high risk for thrombotic events. All profiles were validated in independent cohorts of patients. The ability to predict APS, but more importantly, those patients at risk for venous thrombosis, represents a paradigm for a genomic approach that can be applied to other populations of patients with venous thrombosis, providing for more effective clinical management of disease, while also reducing the possible underlying biologic processes. (Blood. 2006;107:1391-1396)

Introduction

More than 3 in 1000 individuals in the United States annually suffer from venous thromboembolism (VTE). Approximately 10% of these individuals have antiphospholipid antibodies (aPLAs), which, if persistent, define antiphospholipid syndrome (APS). APS identifies a subset of individuals at high risk for recurrent VTE and is defined as primary APS in the absence of an underlying rheumatologic disorder, such as systemic lupus erythematosus. Despite specific clinical and laboratory criteria (the Sapporo classification), diagnostic and prognostic tools in patients with primary APS are limited in their ability to predict adverse outcomes in patients with aPLA.

Multiple mechanisms have been offered to explain the etiology and pathogenesis of thrombosis in APS. Experimental evidence suggests that aPLAs play a causative role in vascular thrombosis and obstetric complications. However, in humans, a direct causal relationship between aPLAs and thrombosis has not been definitively proven. Further, a subset of individuals has elevated aPLA levels on serial laboratory testing, without clinical manifestations of the syndrome. Thus, determinants of a particular clinical phenotype (thrombotic vs asymptomatic) in individual patients with aPLAs remain unclear.

Numerous studies have demonstrated the power of gene-expression profiles to identify subtle distinctions that define important clinical phenotypes. The ability to classify and ultimately predict these phenotypes, including risk of a future event or response to a particular therapy, is central to our capacity to implement personalized medicine—health planning and treatment strategies customized to the individual rather than to the broad population of patients. The value and, simultaneously, the key challenge in genomics are its scale and complexity; the information content has the potential to identify unique characteristics of the disease state of an individual that will customize health care strategies. Although much of the recent success in using gene-expression patterns to classify and predict clinical phenotypes has been in cancer, where access to disease tissue affords a unique opportunity for study, the same principle holds for any circumstance where accessible biological material can be probed for expression patterns reflective of a relevant phenotype. We used peripheral blood to develop gene-expression information that might distinguish patients on the basis of their potential for recurrent thrombotic events. By doing so, we have developed expression signatures that accurately predict these important outcomes and can form the basis for clinical application to appropriately guide treatment decisions.

Patients, materials, and methods

Identification of patients and clinical data

The comparison groups of interest in our study were (1) APS patients with VTE versus patients with VTE who did not have aPLAs, and (2) APS patients with VTE versus asymptomatic subjects with aPLAs.

A.P., J.R.N., and T.L.O. participated in designing and performing the research; A.B. and H.K.D. analyzed the gene-expression data; D.A.L. performed the anticardiolipin testing; A.P., J.R.N., and T.L.O. wrote the paper, and all authors checked the final version of the manuscript.

Reprints: Anil Potti, Department of Medicine, Duke University Medical Center, Box 3841 Red Zone, Durham, NC 27710; e-mail: anil.potti@duke.edu.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 U.S.C. section 1734.

© 2006 by The American Society of Hematology
After institutional review board approval and appropriate informed consent, patients who fit the Sapporo criteria for primary APS² and were at least 8 weeks out from a VTE were eligible. In addition a cohort of individuals with persistent aPLAs without thrombosis (2 separate positive tests for aPLAs at least 6 weeks apart) were enrolled. To be eligible for the study, patients could not be pregnant or receiving immunosuppressant therapy.

Since the phenomenon of thrombosis or the use of warfarin may, by themselves, result in gene-expression profile differences, we identified a comparison group of patients with VTE but without aPLAs. These patients were age (± 3 years) and sex matched to the APS patients, were at least 8 weeks out from their VTE, and did not have an autoimmune disorder. Further, all non-APS patients with VTE had testing for aPLAs that was negative prior to enrollment, and the incidence of pulmonary embolism, a surrogate marker of extent of thrombosis, was similar between the non-APS VTE patients and the APS cohort. All subjects recruited in this study did not have active cancer, leukopenia, or leukocytosis at the time of enrollment. All patients were recruited from 2 academic institutions.

In the initial discovery mode of our study, a cohort of 33 subjects included 11 APS patients with VTE, 7 non-APS patients with VTE, 7 individuals with aPLAs without thrombosis, and 8 healthy controls. The discovery phase of the study was intended to address the fundamental issue of whether gene-expression data obtained from peripheral blood could possibly differentiate the phenotypes of interest. This was then expanded to describe training and validation cohorts that approximately represented a random two-third/one-third split, respectively. The training set (n = 75), including 36 APS patients with VTE, 21 non-APS patients with VTE, and 18 subjects with aPLAs only, was identified for the comparisons. The independent validation set (n = 46) included 21 APS patients, 11 with VTE without aPLAs, and 14 with aPLAs only. The training and validation cohorts were then systematically analyzed to further test our hypotheses that gene-expression patterns distinguish thrombotic phenotypes and predict thrombosis in patients with aPLAs.

Antiphospholipid testing

Lupus anticoagulants were detected using a dilute Russell viper venom (dRVV) clotting assay (American Diagnostics Inc, Stamford, CT). The lupus anticoagulant (LA) status was determined based on the International Society on Thrombosis and Haemostasis (ISTH) criteria (Table 1). IgG antibodies to cardiolipin were determined by ELISA as described by Kim et al. IgG antibodies to β2GPI were detected using the Asserachrom anti-β2GPI IgG kit from Diagnostica Stago, Inc (Parsippany, NJ).

RNA preparation

All patient samples were collected and stored at −80°C using the PAXgene blood RNA tubes, which enabled the extraction of 4 to 5 μg total RNA from 10 mL blood within 24 hours of the initial collection, using a previously published protocol.²⁴

RNA quality was assayed using Agilent bioanalyzer (Agilent Technologies, Silicon Genetics, Redwood City, CA). RNA was thereafter amplified and gene-expression data obtained. All steps involved in RNA processing, probe preparation, microarray hybridization, and data processing used MIAME (Minimal Information About a Microarray Experiment)—guidelines established by the Microarray Gene Expression Data Society.²⁵

DNA microarrays

Oligonucleotide arrays were printed at the Duke Microarray Facility using the Operon Human Genome Oligo Set Version 3.0 (Operon, Huntsville, AL) that includes 34 580 optimized 70mers, covering 24 650 genes.

Probe preparation and microarray hybridization

Total RNA (2 μg) from each sample and the reference (Universal Human Reference RNA, Stratagene) was used in probe preparation. Briefly, reverse transcription is driven by an oligo (dT) primer bearing a T7 promoter using ArrayScript. The cDNA then undergoes second-strand synthesis and clean-up to become a template for in vitro transcription with T7 RNA polymerase. To maximize RNA yield, Ambion’s proprietary MEGAscript in vitro transcription (IVT) technology is used to generate amplified RNA (aRNA). The antisense aRNA is then fluorescently labeled with Cy3 (reference) and Cy5 (sample). Sample and reference aRNAs were pooled, mixed with 1 × hybridization buffer (50% formamide, 5 × SSC, and 0.1% SDS), Cot-1 DNA, and poly-dA to limit nonspecific binding, and heated to 95°C for 2 minutes. This mixture was pipetted onto a microarray slide, a cover slip placed, and hybridized overnight at 42°C. The array was then washed at increasing stringencies and scanned on a GenePix 4000B microarray scanner (Axon Instruments, Foster City, CA). Detailed protocols are available on the Duke Microarray Facility web site.²⁶

Data processing

GENESPRING 6.1 was used to perform initial data analysis. Intensity-dependent (Lowess) normalization was performed on the entire dataset. Based on triplicates of each condition, a threshold of 2-fold change in expression relative to control and a 2-way ANOVA with a P value cutoff of 0.05, with an FDR of 25% was used to identify differentially expressed genes.

Table 1. Baseline clinical and laboratory parameters in the study cohorts

<table>
<thead>
<tr>
<th></th>
<th>APS with VTE</th>
<th>VTE without aPLAs</th>
<th>aPLAs only</th>
<th>Healthy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size, no. patients</td>
<td>57</td>
<td>32</td>
<td>32</td>
<td>8</td>
</tr>
<tr>
<td>Age range, y</td>
<td>34-71</td>
<td>31-65</td>
<td>41-69</td>
<td>29-58</td>
</tr>
<tr>
<td>Mean age, y</td>
<td>61</td>
<td>59</td>
<td>62</td>
<td>48</td>
</tr>
<tr>
<td>No. female/no. male</td>
<td>39/18</td>
<td>19/13</td>
<td>20/12</td>
<td>4/4</td>
</tr>
<tr>
<td>Mean age at onset of VTE, y</td>
<td>53.5</td>
<td>50.3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mean time of study after thrombosis, d (range)</td>
<td>88 (71-183)</td>
<td>78 (59-157)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Incidence of VTE, no./no. total (%)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>PE</td>
<td>7/57 (12.2)</td>
<td>3/32 (9.3)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>PE + DVT</td>
<td>14/57 (24.5)</td>
<td>8/32 (25)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>DVT</td>
<td>36/57 (63.2)</td>
<td>21/32 (65.7)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mean anticardiolipin, GPL (range)</td>
<td>34.5 (1-106)</td>
<td>6.3 (2-11)</td>
<td>32.7 (5-86)</td>
<td>8.5 (3-12)</td>
</tr>
<tr>
<td>Mean anti-β2GPI, glycoprotein, GAU/mL (range)</td>
<td>17.0 (1-72)</td>
<td>1.0 (1-1)</td>
<td>16.2 (2-53)</td>
<td>1.0 (1-1)</td>
</tr>
<tr>
<td>Lupus anticoagulant, no./no. total (%)*</td>
<td>45/81 (58.2)</td>
<td>0/28 (0)</td>
<td>27/29 (93.1)</td>
<td>0/8 (0)</td>
</tr>
<tr>
<td>Inherited thrombophilia, no./no. total (%)†</td>
<td>3/53 (0.05)</td>
<td>1/32 (0.03)</td>
<td>0/30 (0)</td>
<td>—</td>
</tr>
<tr>
<td>Smoking history, no./no. total (%)</td>
<td>11/57 (19.2)</td>
<td>6/32 (18.7)</td>
<td>7/32 (21.8)</td>
<td>2/8 (25)</td>
</tr>
</tbody>
</table>

APS indicates antiphospholipid syndrome; VTE, venous thromboembolism; PE, pulmonary embolism; DVT, deep vein thrombosis; aPLAs, antiphospholipid antibodies; and —, not applicable.

median value of expression or did not vary more than 2-fold across samples

step filtered out genes whose maximum expression did not exceed the

clinical phenotype. Applying a stringent statistical measure of

with aPLAs only, and healthy controls. The focus was to determine

included APS patients, non-APS patients with VTE, individuals

performed without a priori knowledge of the anticipated outcome,

We first used an unsupervised analysis, wherein clustering is

generate a single factor, called a metagene. A metagene represents a group

gene lists are provided online.27

Statistical analysis

Analysis was performed using metagene construction and binary prediction

analysis using MATLAB, as previously described, to analyze gene-

expression patterns predictive of breast cancer outcomes.28-30 The initial

step filtered out genes whose maximum expression did not exceed the

median value of expression or did not vary more than 2-fold across samples
to remove genes with extremely low expression levels or little variance.30

Next, we clustered the genes into groups based on their expression

patterns, with the notion that related genes share similar variances in

expression using K-means clustering. This algorithm randomly places
genomes into a predetermined number of groups. The genes are then shuffled

among the groups in an iterative fashion to maximize the distinction

between each group. The number of designated clusters was then varied

iteratively to further maximize differences between the clusters. The

resulting clusters contained approximately 50 genes and represented a

unique gene-expression pattern.

Singular value decomposition was performed on each cluster to
generate a single factor, called a metagene. A metagene represents a group

genes that together exhibit a consistent pattern of expression in relation to

an observable phenotype. The metagenes are then used in binary (0 = no

event, 1 = event) decision trees to partition the samples into subgroups.

In the trees, a metagene is used at a branch point to partition samples to 1 of 2
classifications based on similarity or dissimilarity of a sample’s gene-

expression pattern to the metagene. Each tree had several branches, and

hundreds of trees were generated to determine the metagenes that best

partitioned the samples. Within each metagene, we identified the genes that

contributed the most weight to the dominant expression pattern. The

resulting signature summarizes its constituent genes as a single expres-

sion profile.

To guard against overfitting, given the disproportionate number of

variables to samples, we also performed honest, out-of-sample cross-

validation analysis to test the stability and predictive capability of our

model. Each sample was left out of the data set one at a time. The model was

refitted (both the metagene factors and the partitioning used) using the

remaining samples, and the phenotype of the held-out case was then

predicted and the certainty of the classification was calculated. Given a

training set of expression vectors (of values across metagenes) representing

principal components, that represent the underlying structure

present in the data. Gene-expression profiles using a group of 50

genes were identified that can distinguish an APS patient from a

non-APS patient (Figure 2A, left).

This led to evaluations of predictive probabilities of each of the 2 states

(disease vs control) for each of the validation sets.

Results

A total of 129 patients were enrolled in the study (Table 1). Data

from the microarray studies was employed for 2 analyses: (1) identification of gene-expression patterns that could classify and predict APS in patients with VTE, and (2) identification of gene-expression patterns that classify and predict thrombosis in patients with aPLAs.

Gene-expression patterns distinguish thrombotic phenotypes

We first used an unsupervised analysis, wherein clustering is

performed without a priori knowledge of the anticipated outcome,

of the gene-expression data obtained in an initial cohort that

included APS patients, non-APS patients with VTE, individuals

with aPLAs only, and healthy controls. The focus was to determine

if there was an evident relationship between expression profiles and

clinical phenotype. Applying a stringent statistical measure of

significance with a criterion for a 2-fold difference (Bonferroni test of

significance), 106 genes were found to differ between APS patients and non-APS patients with thrombosis. Of note, both apoH (gene regulating B3 glycoprotein1) and MEKK1 (MAP kinase pathway gene), genes previously described in the molecular pathogenesis of APS,12,31 were included in that list of genes. In addition, other genes regulating inflammatory processes were identified, which may provide insight into the interaction between inflammatory and thrombotic pathways in the pathogenesis of thrombosis in APS. Clustering analysis revealed patterns of gene expression that separated APS patients from non-APS patients with VTE and healthy controls (Figure 1). Clearly, these distinct clinical phenotypes were represented by distinct gene-expression profiles.

Results from this initial analysis gave confidence in our

approach to evaluate gene-expression profiles in the comparison
groups described above.

Gene-expression patterns that predict APS in patients

with venous thromboembolism

Although the identification of structure in the gene-expression data

depicted by clinical phenotype was reassuring, this analysis does not

lend itself to the generation of predictive models that could be

used in a clinical setting. To address this question, we used

Bayesian binary regression analysis to develop models that could

classify and predict APS using a cohort of APS patients (n = 36)

and non-APS patients with thrombosis (n = 21). This uses methods

developed previously to classify and predict cancer phenotypes,

including the probability of disease recurrence.29,30 We initially

identified genes whose expression most highly correlated with the

2 phenotypic states. We then used this group of genes in a binary

regression analysis to elucidate patterns of gene expression, or

principal components, that represent the underlying structure

present in the data. Gene-expression profiles using a group of 50

genes were identified that can distinguish an APS patient from a

non-APS patient (Figure 2A, left).

The real test of whether a pattern reflects the phenotype of

interest is the ability to accurately predict the status of an unknown

sample. To assess this, we used leave-one-out cross-validation

analysis. One sample is removed, the remainder is then used for

generating the patterns for prediction, and then the removed sample

is used for prediction of whether it is an APS patient or not. As

shown in Figure 2B, the APS pattern accurately distinguished APS

patients from non-APS patients. The values on the horizontal axis

Figure 1. Patterns of gene expression that characterize clinical phenotypes.

Hierarchical clustering of the initial patient samples based on gene-expression

patterns. Each gene is represented by a single row, and each sample is represented

by a single column. The color heat map represents genes in a graded fashion along a

spectrum of activation, extending from strongly up-regulated genes in red to the
down-regulated genes in blue.
are estimates of the signature score from the regression, and the values on the vertical axis are estimated classification probabilities with the corresponding 95% probability intervals marked to indicate the uncertainty about these estimated values.

To demonstrate the use of the model developed from the initial training set, a separate validation cohort of 32 subjects (21 APS patients with VTE and 11 non-APS patients with VTE) was used to test performance of the predictive model. The 50 genes that formed the pattern developed in the training set distinguished APS and non-APS patients in the separate validation set (Figure 2A, right). More importantly, the predictive model also accurately identified the status of the validation samples in a true predictive test (Figure 2C).

Results from the binary regression model (50 most common discriminator genes or the metagene) were cross-referenced to the 106 genes found to be at least 2-fold different on the Bonferroni test of significance, in the discovery phase of the study, adding to the validity of this approach and supporting the results as clinically relevant.

**Gene-expression pattern that predicts thrombosis in patients with aPLA**

We next sought to generate expression profiles that would predict the potential for thrombosis in patients with aPLAs. The importance of this lies in the capacity to identify those patients with aPLA who are at highest risk for thrombosis and might benefit from preventive treatment strategies.

For this analysis, we used 32 patients with aPLA who did not have VTE compared to patterns obtained from 57 APS patients with VTE. The training set included 36 APS patients and 18 patients with aPLAs only. We used the same methods and strategy outlined in “Patients, materials, and methods,” initially identifying a group of 50 genes that discriminated the 2 comparison groups (Figure 3A, left).

We then used a leave-one-out cross-validation to assess the ability of the pattern to predict the potential for thrombosis in the relevant samples. As shown in Figure 3B, the profile accurately distinguished patients with thrombosis from individuals without thrombosis. This pattern of 50 genes selected in the training sample was then applied to a separate validation set (n = 35) to assess more rigorously the performance in a true predictive setting (Figure 3C).
3A, right). The predictive ability of the model developed in the training set also was robust in predicting the thrombotic state in the validation samples (Figure 3C). Based on these results, we conclude that it is possible to develop robust predictive models that can identify patients with aPLAs who are at risk for thrombosis.

An analysis of the 50 genes whose expression patterns provide the power to discriminate and predict thrombosis in patients with aPLAs revealed processes that we would infer from our current understanding of thrombosis, including APOE, factor X, and thrombomodulin. In addition, certain other genes were identified that have thus far not been directly linked to venous thrombosis, including those encoding for hypoxia inducible factor (HIF-1α), zinc finger proteins, matrix metalloproteinase19 (MMP19), interleukin 22 (IL22) receptor, and hematopoietic progenitor cell antigen (CD34) precursor.

Discussion

Autoimmune diseases affect 3% of the United States population. Although in many cases candidate autoantigens have been identified, the mechanism(s) whereby autoantibodies result in the clinical manifestations of the syndrome remain largely unclear. Analysis of gene-expression profiles offers the prospect of creating more precise determinations of disease phenotypes, pointing to the underlying mechanisms responsible for disease phenotypes. But, in addition to the opportunity to better understand the disease process, gene-expression profiles provide an opportunity for more effective diagnosis, identifying which patient is at risk for developing complications. Indeed, we have described gene-expression patterns from patient peripheral blood that can predict an individual’s predisposition to developing thrombosis, as well as the ability to identify patients with VTE who are at higher risk for recurrence.

The cross-validation and the confirmatory validation set analyses were able to correctly classify unknown samples of VTE with 85% accuracy for identification of patients with APS, and 100% accuracy for prediction of thrombosis in the blinded independent cohort of patients with aPLAs. Interestingly, in our first analysis (Figure 2B), one patient who was initially thought to have APS (from the clinical record) but was misclassified as non-APS with VTE (based on gene expression) was subsequently found to have been misdiagnosed as APS. Thus, even with the heterogeneous nature of the population studied, and even though we did not adjust for race, ethnicity, and number of thrombotic events, our approach as described identified candidate clusters of genes that clearly exhibit reliable discriminatory patterns between the APS and non-APS patient populations with venous thrombosis.

In the second analysis comparing expression profiles between APS patients with VTE versus patients with aPLAs but without thrombosis, it is possible that a major contributing factor to the different gene-expression patterns was the thrombotic event, and the differences in gene-expression profiles seen could have been due to this event. In a separate analysis, however, the set of discriminator genes identified in comparing patients with aPLAs without thrombosis and APS with thrombosis was different than the cluster of genes that differentiated non-APS patients with VTE from patients with aPLAs only (data shown online). A prospective study of asymptomatic patients identified with aPLAs who were followed over time would prove the value of gene-expression profiling as a predictor of venous thrombosis in such patients. It is also important to note that although it is likely that the cells contributing to the gene-expression profiles in our study cohorts were predominantly peripheral blood mononuclear cells (lymphocytes and monocytes), we did not eliminate platelet contamination, nonetheless, the presence of distinct profiles obtained from peripheral blood of patients makes the use of such a test very desirable in clinical practice.

While the development of expression profiles as tools to predict clinical outcomes was the driving impetus in these studies, knowledge of the gene-expression properties that define these phenotypes may identify genes that contribute to potential for disease. These genes may help identify critical pathways involved in the disease process, potentially opening the way for better clinical intervention. In addition, identification of genes whose expression relates to the disease process will provide candidates for polymorphism discovery that can be assessed for their role in defining germ-line susceptibility to these phenotypes. By identifying polymorphisms within high priority genes, we may identify combinations of single-nucleotide polymorphisms that lead to the development of new diagnostic tools to identify patients at risk, prior to the development of symptoms. Moreover, knowledge of germ-line variation(s) can represent another variable, coupled with the expression profile(s), in the development of predictive models.

Acknowledgments

The authors thank Laura Worfolk, PhD, from Diagnostica Stago for providing kits to run anticardiolipin antibody assays and Dr Geoffrey Ginsburg for critical review of the manuscript.

References


Gene-expression patterns predict phenotypes of immune-mediated thrombosis

Anil Potti, Andrea Bild, Holly K. Dressman, Deborah A. Lewis, Joseph R. Nevins and Thomas L. Ortel

Updated information and services can be found at:
http://www.bloodjournal.org/content/107/4/1391.full.html

Articles on similar topics can be found in the following Blood collections
- Clinical Trials and Observations (4539 articles)
- Gene Expression (1086 articles)
- Genomics (149 articles)
- Hemostasis, Thrombosis, and Vascular Biology (2485 articles)
- Immunobiology (5483 articles)

Information about reproducing this article in parts or in its entirety may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at:
http://www.bloodjournal.org/site/subscriptions/index.xhtml