Thrombosis Research 135 (2015) 659–665

Contents lists available at ScienceDirect

Thrombosis Research

journal homepage: www.elsevier.com/locate/thromres

Regular Article

Whole blood gene expression profiles distinguish clinical phenotypes of venous thromboembolism☆

Deborah A. Lewis [a], Sunil Suchindran [b], Michele G. Beckman [c], W. Craig Hooper [c], Althea M. Grant [c], John A. Heit [d], Marilyn Manco-Johnson [e], Stephan Moll [f], Claire S. Philipp [g], Kristy Kenney [c], Christine De Staercke [c], Meredith E. Pyle [c], Jen-Tsan Chi [h], Thomas L. Ortel [a,*]

[a] Thrombosis and Hemostasis Center, Division of Hematology, Duke University Medical Center, Durham, NC
[b] Center for Applied Genomics, Duke University School of Medicine, Durham NC
[c] Centers for Disease Control and Prevention, Atlanta, GA; Division of Cardiovascular Diseases
[d] Mayo Clinic, Rochester, MN
[e] University of Colorado and Children's Hospital, Aurora, CO
[f] University of North Carolina, Chapel Hill, NC
[g] Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ
[h] Department of Molecular Genetics and Microbiology and Center for Genomic and Computation Biology, Duke University Medical Center, Durham NC

ARTICLE INFO

Article history:
Received 4 November 2014
Received in revised form 15 January 2015
Accepted 2 February 2015
Available online 7 February 2015

Keywords:
venous thrombosis
deep-vein thrombosis
gene expression profiling

ABSTRACT

Introduction: Recurrent venous thromboembolism (VTE) occurs infrequently following a provoked event but occurs in up to 30% of individuals following an initial unprovoked event. There is limited understanding of the biological mechanisms that predispose patients to recurrent VTE.
Objectives: To identify whole blood gene expression profiles that distinguished patients with clinically distinct patterns of VTE.
Patients/Methods: We studied 107 patients with VTE separated into 3 groups: (1) 'low-risk' patients had one or more provoked VTE; (2) 'moderate-risk' patients had a single unprovoked VTE; (3) 'high-risk' patients had ≥2 unprovoked VTE. Each patient group was also compared to twenty-five individuals with no personal history of VTE. Total RNA from whole blood was isolated and hybridized to Illumina HT-12 V4 Beadchips to assay whole genome expression.
Results: Using class prediction analysis, we distinguished high-risk patients from low-risk patients and healthy controls with good receiver operating curve characteristics (AUC = 0.81 and 0.84, respectively). We also distinguished moderate-risk individuals and low-risk individuals from healthy controls with AUC's of 0.69 and 0.80, respectively. Using differential expression analysis, we identified several genes previously implicated in thrombotic disorders by genetic analyses, including *SELP*, *KLKB1*, *ANXA5*, and *CD46*. Protein levels for several of the identified genes were not significantly different between the different groups.
Conclusion: Gene expression profiles are capable of distinguishing patients with different clinical presentations of VTE, and genes relevant to VTE risk are frequently differentially expressed in these comparisons.

Abbreviations: *ANXA2*, annexin A2; *ANXA5*, annexin A5; AUC, area under the curve; *C1QB*, complement component 1, q subcomponent, B chain; *C5*, complement component 5; *CD46*, complement regulatory protein; CDC, Centers for Disease Control and Prevention; *CFH*, complement factor H; *CR1*, complement component receptor 1; *CR2*, complement component receptor 2; DVT, Deep vein thrombosis; *EDN1*, endothelin 1; *F2RL1*, coagulation factor II receptor-like 1; *F11*, coagulation factor XI; *HIF1A*, hypoxia inducible factor 1, alpha subunit; *ICAM-1*, intercellular adhesion molecule 1; *IGF1R*, insulin-like growth factor 1 receptor; *IL4*, interleukin 4; *ITGAL*, integrin alpha L chain; KEGG, Kyoto Encyclopedia of Genes and Genomes; *KLKB1*, kallikrein B; PE, pulmonary embolism; *PPARD*, peroxisome proliferator-activated receptor delta; *SELP*, selectin P; SERPING1, Serpin peptidase inhibitor Glade G (C1 inhibitor); SNPs, single nucleotide polymorphisms; *TF*, transferrin; VTE, venous thromboembolism

☆ Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

* Corresponding author at: Duke Thrombosis and Hemostasis Center, Box 3422 DUMC, Durham, NC 27710 USA. Tel.: +1 919 660 4363; fax: +1 919 681 1177.
E-mail address: thomas.ortel@duke.edu (T.L. Ortel).

http://dx.doi.org/10.1016/j.thromres.2015.02.003
0049-3848/

Introduction

Deep vein thrombosis (DVT) or pulmonary embolism (PE), referred to collectively as venous thromboembolism (VTE), affects approximately 350,000 to 600,000 individuals in the United States each year, and up to 100,000 will die from the thromboembolic event [1]. VTE may occur after transient exposures such as a surgical procedure, prolonged immobilization, or with the use of certain therapies, such as oral contraceptives and hormone replacement therapy, which is referred to as a provoked event [2]. VTE can also occur in the absence of any acquired risk factors, which is referred to as an unprovoked, or idiopathic, event [1,3,4]. Other factors that may increase an individual patient's risk for VTE include increased age, the presence of a thrombophilia [5], race/ethnicity, and a variety of medical conditions [6].

The current standard of care for patients with provoked VTE consists of therapeutic anticoagulation for three months [7]. In contrast, for patients with an unprovoked VTE, up to 30% will sustain a recurrent event within ten years of completing a standard course of therapy [3,8]. Consequently, it is recommended to consider an extended course of therapy for patients with an initial unprovoked event [7]. Continued anticoagulant therapy has been shown in several studies to significantly decrease the risk for recurrent VTE [6–8] but the risk of major bleeding in individuals after the first three months of therapy ranges from a baseline of 0.3% to ≥2.5% per year [7].

Determining which patients with VTE have a high risk for recurrent events, and balancing this risk with the potential for bleeding if anticoagulation is continued, is an important health concern. Multiple studies have investigated biomarkers to help predict which patients are at a higher risk for recurrent VTE [9]. Current evidence suggests that inherited thrombophilic disorders are not helpful to predict which patients with a first unprovoked VTE are at an increased risk for recurrent events [10]. In contrast, elevated D-dimer levels obtained after completing a standard course of anticoagulant therapy are associated with an increased risk for recurrent VTE [11]. Other biomarkers that have been associated with recurrent VTE include elevated levels of soluble p-selectin [12] and elevated thrombin generation [13].

Whole blood gene expression studies have been used in a variety of disorders including myocardial infarction and systemic lupus erythematosus [14,15]. We previously used whole blood gene expression profiles to distinguish patients with a single VTE from patients with recurrent VTE [16], but this study combined patients with provoked and unprovoked events. Here we extend this initial study by using clinically well-defined patient groups with the objectives of comparing individuals based on the type of VTE (provoked versus unprovoked) as well as by the number of events (single versus multiple). We used two distinct analytical approaches, class prediction analysis [17–20] and differential expression analysis [21] to identify means to distinguish among these patient groups. A group of healthy individuals was included to look for genes and pathways that are differentially expressed in healthy individuals compared to individuals with different types of VTE.

Material and Methods

Patient Population

Participants were enrolled in 2009 and 2010 at 4 sites participating in the Thrombosis and Hemostasis Centers Research and Prevention Network supported by the Centers for Disease Control and Prevention (CDC): Duke University Medical Center, Durham NC; Mayo Clinic, Rochester MN; University of North Carolina, Chapel Hill NC; and Rutgers Robert Wood Johnson Medical School, New Brunswick NJ. This Network consisted of Thrombosis and Hemostasis Centers that provided comprehensive specialty care to patients with thrombophilia and thrombotic disorders [22]. Study protocol and consent forms were approved by Institutional Review Boards at each site and at the CDC.

Patients with at least one VTE, defined as either PE or DVT of the leg or arm, with the first event occurring at age 18 years or older, and who were, at the time of enrollment, greater than 10 weeks from their most recent VTE, were approached for participation. The diagnosis of VTE was reviewed and objectively confirmed by the site investigator, based on clinical history and imaging data. Individuals with no prior history of VTE or known inherited clotting disorder and similar in age, gender, and race to the VTE case were identified at each site and approached to participate as controls. Patients with known antiphospholipid syndrome, active or prior malignancy (excluding skin cancer) at the time of VTE diagnosis, infection within the past two weeks of enrollment or currently pregnant were not included in this study.

Consenting VTE patients were allocated to 3 groups: (1) low-risk, defined as patients who had sustained 1 or more provoked VTE with no history of an unprovoked VTE; (2) moderate-risk, defined as patients who had sustained a single unprovoked VTE (with or without additional provoked VTE); and (3) high-risk, defined as patients who had sustained 2 or more unprovoked VTE (with or without additional provoked VTE). A provoked event was defined as a VTE occurring in a patient with a clear transient acquired risk factor for VTE, i.e. VTE occurring within 3 months after trauma, hospitalization, prolonged immobilization, or surgery and the post-operative setting; or in patients taking oral contraceptives or hormone replacement therapy; or during pregnancy or the post-partum period. Unprovoked events were defined as VTE occurring in the absence of any of these transient risk factors.

Patients with more than one VTE (provoked or unprovoked) had distinct clinical events that occurred at different points in time. Thromboembolic events affecting more than one vascular bed but occurring at the same time were considered to be a single event (*e.g.*, a patient presenting with PE and DVT).

Data and Sample Collection

Demographic and clinical information was collected from each participant through chart abstraction or in-person interview. Citrated plasma and serum samples were collected for each participant, processed, and stored at −80 °C at each site. Blood was simultaneously collected in PAXgene RNA tubes and stored according to the manufacturer's instructions. De-identified samples were shipped to the CDC Division of Blood Disorders' Molecular and Hemostasis Laboratories for analysis.

RNA Isolation and Microarray Hybridization

Total RNA was isolated from whole blood drawn into PAXgene tubes using the PAXgene Blood RNA kit (PreAnalytiX; Qiagen GmbH-USA). The quality and quantity of the RNA was confirmed using the Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). Samples with an A260/A280 ratio >2.19, or ≤70 ng of RNA, were excluded from the final analyses. RT PCR using probes for IL-1beta and CD141 was used to check RNA expression levels for several of the initial samples from each of the sites, to confirm comparable yields. RNA was amplified and the cRNA was biotinylated using the Ambion Illumina TotalPrep RNA Amplification Kit (Life Technologies, Carlsbad, CA). Following labeling, cRNA samples were hybridized to Illumina HT-12 V4 Beadchips to assay whole genome gene expression with over 47,000 probes against human transcripts.

Microarray Data Processing

A comprehensive quality control process was performed on all arrays using the lumi package in Bioconductor in the R environment for statistical computing [23,24]. Quality of the raw data was assessed using the percent of probes present, MA plots, boxplots of the expression distribution, and heatmaps to visualize the correlation between samples. Samples in which the percent of probes present was 15% or less were excluded, and all probes that were not detected in greater