EBioMedicine xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Whole Exome Sequencing Reveals Severe Thrombophilia in Acute Unprovoked Idiopathic Fatal Pulmonary Embolism

Matt Halvorsen [a], Ying Lin [c], Barbara A. Sampson [c], Dawei Wang [c], Bo Zhou [c], Lucy S. Eng [c], Sung Yon Um [c], Orrin Devinsky [b], David B. Goldstein [a], Yingying Tang [c,*]

[a] Institute for Genomic Medicine, Columbia University Medical Center, United States
[b] Department of Neurology, NYU Langone Medical Center, United States
[c] Molecular Genetics Laboratory, New York City Office of Chief Medical Examiner, United States

ARTICLE INFO

Article history:
Received 22 December 2016
Received in revised form 19 January 2017
Accepted 26 January 2017
Available online xxxx

Keywords:
Idiopathic fatal pulmonary embolism
Whole exome sequencing
Severe thrombophilia
Natural anticoagulant deficiency

ABSTRACT

Background: Acute unprovoked idiopathic fatal pulmonary embolism (IFPE) causes sudden death without an identifiable thrombogenic risk. We aimed to investigate the underlying genomic risks of IFPE through whole exome sequencing (WES).

Methods: We reviewed 14 years of consecutive out-of-hospital fatal pulmonary embolism records (n = 1478) from the ethnically diverse population of New York City. We selected 68 qualifying IFPE cases for WES. We compared the WES data of IFPE cases to those of 9332 controls to determine if there is an excess of rare damaging variants in the genome using ethnicity-matched controls in collapsing analyses.

Findings: We found nine of the 68 decedents (13·2%) who died of IFPE had at least one pathogenic or likely pathogenic variant in one of the three anti-coagulant genes: *SERPINC1* (Antithrombin III), *PROC*, and *PROS1*. The odds ratio of developing IFPE as a variant carrier for *SERPINC1* is 144·2 (95% CI, 26·3–779·4; $P = 1{\cdot}7 \times 10^{-7}$), for *PROC* is 85·6 (95% CI, 13·0–448·9; $P = 2.0 \times 10^{-5}$), and for *PROS1* is 56·4 (95% CI, 5·3–351·1; P = 0·001). The average age-at-death of anti-coagulant gene variant carriers is significantly younger than that of non-carriers (28·56 years versus 38·02 years; P = 0·01).

Interpretation: This study showed the important role of severe thrombophilia due to natural anti-coagulant deficiency in IFPE. Evaluating severe thrombophilia in out-of-hospital fatal PE beyond IFPE is warranted.

1. Introduction

Pulmonary embolism (PE), a complication of deep venous thrombosis (DVT), is a serious and growing public health problem, causing considerable morbidity and mortality worldwide. It is estimated that PE causes >500,000 deaths in the European Union (population 500 million) (Cohen et al., 2007) and a similar number of deaths in the United States (population 300 million) every year (Raskob GE, 2014). The annual incidence of PE and DVT increases with age (>50 years), is higher in people of African descent and lower in Asians, compared to white individuals, and is higher in men than in women when estrogen and pregnancy is not considered (Di Nisio et al., 2016).

Fatal PE events are investigated either in-hospital and emergency rooms, or out-of-hospital in the offices of Chief Medical Examiner. In most parts of the United States, Medical Examiners (ME) are responsible for determining the underlying cause of death due to acute fatal PE in the out-of-hospital setting, when unattended by a physician. After autopsy and review of clinical history, ME aims to identify one or more thrombogenic risk factors in Virchow's Triad (Wolberg et al., 2015): vascular damages (e.g. surgery or trauma), venous stasis (e.g. immobilization), and hypercoagulability (e.g. cancer). Frequently, ME finds no strong thrombogenic risk to explain a fatal PE in a young decedent, which prompts them to pursue a thrombophilia test. Heritable factors commonly assessed clinically confer mild thrombogenic risks: factor V Leiden (4–10 odds ratio (OR)) and prothrombin G20210A variants (2–4 OR). Mild thromobophilia plays little role in fatal PE as we reported previously (Tang et al., 2011). Severe thrombophilia due to deficiency of natural anticoagulants (antithrombin III, protein S, or protein C), are strong thrombogenic risk factors; however, the status of severe thrombophilia is largely unknown, particularly in the decedents who died of acute fatal PE events.

Acute unprovoked idiopathic fatal pulmonary embolism (IFPE) is pulmonary embolism leading to sudden death without identifiable thrombogenic risks and previous history. In this study, we aimed to identify the underlying genomic risk factors of IFPE through whole exome sequencing (WES). We reviewed 14 years of consecutive

* Corresponding author at: Molecular Genetics Laboratory, Office of Chief Medical Examiner, 421 East 26th Street, New York, NY 10016, United States.
E-mail address: ytang@ocme.nyc.gov (Y. Tang).

http://dx.doi.org/10.1016/j.ebiom.2017.01.037

out-of-hospital fatal PE records (n = 1478 cases) in the New York City Office of Chief Medical Examiner and identified IFPE decedents who were under the age of 50 years without a known thrombogenic risk and negative for the factor V Leiden and prothrombin G20210A variants. The WES testing results are reported here.

2. Methods

2.1. Case Selection

All out-of-hospital fatal PE deaths in the City of New York are investigated in the Office of Chief Medical Examiner (OCME). The detailed forensic investigation procedures include complete autopsy (gross and microscopic examination), toxicology, scene investigation, and review of clinical history (when possible). Postmortem specimens are routinely collected at autopsy and submitted by a ME for the molecular testing of two mild thrombophilia variants, factor V Leiden (FVL) and prothrombin (FII) G20210A, to the College of American Pathologists (CAP) accredited Molecular Genetics Laboratory within OCME, where the cases are accessioned and the specimens are tested. For this WES-based study, we reviewed 1478 consecutive cases with PE as immediate cause of death between 2001 and 2014 and selected 72 that met the selection criteria (see below). Four cases were removed due to insufficient DNA, and a total of 68 cases were selected for WES study at Institute for Genome Medicine (IGM) at Columbia University Medical Center (CUMC).

Inclusion criteria for the IFPE cases are: age under 50 years; BMI ≤ 30 kg/m^2; negative FVL and FII G20210A; no thrombogenic risks (i.e. various level of immobility, recent surgeries or trauma, arteriosclerotic and hypertensive cardiovascular disease, cancer et al.). A history of oral contraceptive use was *not* an exclusionary criterion for this study because oral (progestin-only) contraceptives and transdermal estradiol reportedly carry minimal or no thrombotic risk (Trenor et al., 2011), and the thrombotic risk is affected by estrogen dose, type of progestin, mechanism of delivery, and length of therapy, for which we often do not have the full information. Early pregnancy alone without the presence of other thrombogenic risk factors was also not an exclusionary criterion.

This study is *not* regulated by 45 CFR Part 46 because only cadaver specimens were used. OCME approved this study for diagnosis of the underlying cause of PE.

2.2. WES and Data Processing

Postmortem tissue samples (spleen, liver, or heart) preserved in *RNAlater*® (Qiagen, Valencia, CA) or dry bloodstain card samples were used for DNA extraction and subsequent WES. At least 1·5 μg of genomic DNA was used for WES. Sequencing was performed at IGM at CUMC on an Illumina Highseq 2500, using a Roche Nimblegen EZCap V3 capture kit.WES data were analyzed using the DNA sequence data alignment and variant calling pipeline utilized at IGM (see Supplemental Methods). The pipeline used in processing the cases matches that of controls, eliminating the possibility of case/control differences being confounded by differential processing.

2.3. Control Selection

We used samples that have been sequenced previously at the IGM as controls. Samples were required to be formally approved for control use, and to pass stringent bioinformatics QC at the IGM. Samples were only included if their self-declared ethnicity matched a group represented in the case cohort (African, Hispanic, White or South Asian ancestry). We additionally required controls to not be listed as having a broad phenotype characterized by cardiovascular or pulmonary disease. After controlling for sufficient coverage and a lack of cryptic relatedness (see details in Supplemental Methods), we were left with 68 cases and 9332 controls that passed QC and could be included in our case/control study. The control set included 1424 African Americans, 1917 Hispanics, 10 South Asians and 5981 Caucasians. A breakdown of the broad phenotypes represented across the control cohort was shown in Table S1, and the statistical summary of sequencing data for both controls and cases was shown in Table S2. We unfortunately lack easily accessible detailed clinical data for controls (including age); however, this should not affect our power to detect a strong genetic signal in cases relative to controls, if one is present.

2.4. Cases Versus Controls - Collapsing Analyses

For all collapsing analyses, we tested for the differential burden of rare nonsynonymous genetic variation within a protein coding *genetic unit* defined via particular criteria: for testing of known pathogenic burden, we defined the *genetic unit* as a set of 12 genes listed by OMIM as being associated with thrombophilia; for testing of general rare nonsynonymous variant burden, we defined the *genetic unit* as a single gene. All collapsing analyses were run under a dominant model, where samples were discretized into those with or without a single qualifying rare variant. For a single genetic unit we compared the portion of cases with a qualifying variant to the proportion of controls. For gene-level collapsing analyses done across the exome we require a P-value below the bonferroni-adjusted threshold to consider the result to be exome-wide significant (see details in Supplemental Methods). We additionally took all male cases and controls and performed gene-based collapsing analyses on X and Y chromosome genes only to identify evidence for any single gene where a haploid damaging nonsynonymous variant conferred PE risk.

For the variants that contributed to the collapsing analysis signal in the anticoagulant gene-set, we classified with regard to their clinical significance. We estimated the minor allele frequency (MAF) of a variant using population databases, e.g. ESP6500 data from NHLBI Grand Opportunity Exome Sequencing Project (ESP) (Exome Variant Server) and Exome Aggregation Consortium (ExAC) (Lek et al., 2016) and determined if they were previously reported by searching the variants in HGMD (Stenson et al., 2012) and ClinVar (Landrum et al., 2016a). If a variant is previously reported with strong evidence (family or function studies), the variant is classified as pathogenic variant. If a variant is not reported, but the same amino acid residue change has been previously reported with supporting evidence from family or function studies, the variant is classified as likely pathogenic variant. If a variant was not reported previously or there was insufficient information to classify it as benign or pathogenic, the variant was classified as a VUS (variant of uncertain significance).

2.5. Statistical Analysis

Calculations of odds ratios and P-values for collapsing analyses were done using a two-sided fisher's exact test (R statistical analysis software version 3.2.3). The significance of the age-at-death in carriers versus non-carriers of the anticoagulant gene variants was evaluated by Wilcoxon signed-rank test. The significant association of *SERPINC1* carriers and females was evaluated by binomial exact test (Graphpad prism 7).

3. Results

3.1. Characteristics of the Cases

We reviewed 14 years (between 2001 and 2014) of consecutive out-of-hospital fatal pulmonary embolism records (1478 cases) in the New York City. Consistent with our previous report (Tang et al., 2011), African Americans (non-Hispanic Blacks) represented a majority of deaths due to fatal PE (58%), followed by non-Hispanic Whites (25%), Hispanics (15%), and Asians (2%). This ethnic breakdown is in drastic contrast to the

Please cite this article as: Halvorsen, M., et al., Whole Exome Sequencing Reveals Severe Thrombophilia in Acute Unprovoked Idiopathic Fatal Pulmonary Embolism, EBioMedicine (2017), http://dx.doi.org/10.1016/j.ebiom.2017.01.037