Thrombosis Research 135 (2015) 404–409

Contents lists available at ScienceDirect

Thrombosis Research

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

Regular Article

Prediction of the pathogenicity of antithrombin sequence variations by in silico methods

Beate Luxembourg [a,b,*], Mathias D`Souza [a], Stephanie Körber [a], Erhard Seifried [a]

[a] Institute of Transfusion Medicine and Immunohaematology, DRK Blood Donor Service Baden-Württemberg – Hessen, University Hospital Frankfurt, Sandhofstr. 1, 60528 Frankfurt, Germany
[b] Department of Internal Medicine, University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany

ARTICLE INFO

Article history:
Received 3 July 2014
Received in revised form 30 October 2014
Accepted 30 November 2014
Available online 4 December 2014

Keywords:
Antithrombin
SERPINC1
Mutation
Sequence variation
In silico

ABSTRACT

Computational prediction tools have been developed to aid in the interpretation of novel sequence variations, but their utility within the diagnostic setting of antithrombin (AT) deficiency has not been evaluated to date. The aim of our study was to test the performance of different bioinformatic tools (Meta-SNP, MutPred, nsSNPAnalyzer, PANTHER, PhD-SNP, PMut, SIFT, SNAP, SNPs&Go, PolyPhen-2, PON-P2, and PredictSNP) in predicting the pathogenicity of AT sequence variations.

We analysed all naturally occurring *SERPINC1* missense mutations that have been previously characterised to be damaging with regard to the secretion or function of the AT molecule. Additionally, we analysed all reported non-synonymous exonic polymorphisms within *SERPINC1* with a population allele frequency >1.0%.

The in silico tools had accuracies of 62-96%, sensitivities of 59-98%, and specificities of 33-100% for the prediction of the pathogenicity of AT sequence variations; receiver operating characteristic analysis had area under the curves between 0.54-0.97. When mutations were grouped according to their effect on the phenotype of AT deficiency [type I or type II with a thrombin (IIRS) or heparin (IIHBS) binding defect or pleiotropic effects (IIPE)], we observed the lowest performance characteristics of the tools for mutations causing AT deficiency type IIHBS. Only three tools (MutPred, PhD-SNP, PolyPhen-2) detected mutants causing type IIHBS AT deficiency with high sensitivity (93%), the sensitivities of the other tools ranged between 36% and 79%.

This study demonstrates that bioinformatic tools are useful for pathogenicity prediction for AT sequence variations, but they have substantially different performance characteristics, particularly for type IIHBS AT deficiency.

© 2014 Elsevier Ltd. All rights reserved.

Introduction

Antithrombin (AT) is a major blood coagulation protease inhibitor, and mutations in the gene encoding AT, *SERPINC1*, cause AT deficiency that predisposes affected individuals towards venous and arterial thromboembolism [1]. *SERPINC1* spans 13.5 kb of genomic DNA and consists of 7 exons [2]. More than 300 naturally occurring *SERPINC1* mutations have been previously reported, and these are mainly missense mutations [3–5]. Missense mutations can have a major impact on the secretion or function of AT and can result in severe AT deficiency [3].

Although single amino acid changes of a protein are often pathogenic, they do not necessarily lead to structural or functional consequences associated with disease. Laboratory characterisation of individual variants is necessary to demonstrate that a variant can directly impair the structure or function of a protein. Most of the previously reported *SERPINC1* mutations are private mutations [3,4]. Wet-lab experiments to prove the pathogenicity of mutants from a single patient or family are limited, therefore, only a fraction of naturally occurring *SERPINC1* mutations have been experimentally characterised.

In silico methods have been developed to aid in the prediction of functionally deleterious and disease-associated variants [6]. Sequence-based in silico tools base their predictions on the physicochemical properties of amino acids [7]. Sequence homology-based methods calculate the effect of amino acid substitutions based on multiple sequence alignments and rely on the observation that disease-associated mutations are overrepresented at highly conserved positions [8]. Structure-based methods exploit structural protein information such as amino acid side chain conformation and residue-residue contacts. Moreover, several tools use machine-learning techniques, e.g., random forests [9,10], support vector machines [11], or neural networks [12]. The performance of in silico tools has not yet been investigated in relation to pathogenicity prediction for AT sequence variations.

Abbreviations: AT, antithrombin; AUC, area under the curve; HGMD, Human Gene Mutation Database; HGVS, Human Genome Variation Society; NCBI, National Center for Biotechnology Information; NPV, negative predictive value; PPV, positive predictive value; Q2, accuracy; ROC, receiver operating characteristic curve; SNP, single nucleotide polymorphism; subPSEC score, substitution position-specific evolutionary conservation score.

* Corresponding author at: Institute of Transfusion Medicine and Immunohaematology, DRK Blood Donor Service Baden-Württemberg - Hessen, Frankfurt a. M., Germany, Sandhofstr. 1, 60528 Frankfurt a.M., Germany. Tel.: +49 69 6782 353; fax: +49 69 6782 346.
E-mail address: beate.luxembourg@kgu.de (B. Luxembourg).

http://dx.doi.org/10.1016/j.thromres.2014.11.022
0049-3848/© 2014 Elsevier Ltd. All rights reserved.

AT belongs to the superfamily of serpins, which share a highly conserved three-dimensional fold comprised of three ß-sheets and typically nine α-helices connected by loop segments in which the loop at the "top" of the molecule presents the reactive centre [13]. After cleavage of the 32-amino acid signal peptide, mature AT with a molecular mass of 58 kDa encompassing 432 amino acid residues circulates in plasma in a low activity state. AT becomes an efficient inhibitor only upon binding glycosaminoglycans, such as heparin, that enhances the anticoagulant activity of AT towards its main targets, thrombin, factor IXa, and factor Xa, by the induction of allosteric activation changes and a bridging mechanism [14]. By acting as a pseudo-substrate, AT traps the target protease in an irreversible 1:1 complex [14].

The particular features of AT and its functional requirements render this molecule particularly sensitive to missense mutations. Coding sequence variations may induce functional effects or affect the RNA stability, folding or secretion of the molecule. Moreover, the conformational and functional sensitivity of AT might explain why missense mutations can have functional consequences by different mechanisms.

According to the phenotype, AT deficiency has been classified into 1) type I, characterised by a parallel reduction of the AT antigen and activity and 2) type II, defined by a reduction of the AT activity and normal or almost normal AT antigen levels. According to the localisation of the functional defect, type II AT deficiency has been further subdivided into a) type II with a heparin binding defect (type IIHBS), b) type II with a reactive site defect (type IIRS), and c) type II with pleiotropic defects (type IIPE).

By analysing AT mutants in a previous study [5], we observed prediction discrepancies between SIFT and PolyPhen, two in silico tools recommended in guidelines for the interpretation and reporting of unclassified variants in clinical molecular genetics [15]. Therefore, we aimed to systematically analyse the performance of different in silico methods in pathogenicity prediction for AT sequence variants in a dataset of previously reported *SERPINC1* mutations.

Subjects, Materials, and Methods

Identification of Mutations and Single Nucleotide Polymorphisms (SNPs) in SERPINC1

To identify missense mutations with experimentally proven pathogenic effects on the AT molecule, a literature search in the database PubMed was performed for all mutations listed in the AT Mutation Database [3] and the Human Gene Mutation Database (HGMD; 4). For the PubMed search, we used the search terms "antithrombin" and the mutated amino acid position according to the AT Mutation Database (e.g., "antithrombin and Arg25Cys", "antithrombin and R25C", "antithrombin and arg and 25 and cys") and also according to the Human Genome Variation Society (HGVS).

To identify *SERPINC1* missense mutations not listed in the databases, another PubMed search up to January 2014 was performed with the following search terms: "antithrombin and missense mutation", "antithrombin and mutation and heparin binding", "antithrombin and heparin binding and crossed immunoelectrophoresis", "antithrombin and mutation and thrombin binding", "antithrombin and type I and secretion", "antithrombin and type I and expression", "antithrombin and type I and RNA", "antithrombin and transfected cells", "antithrombin and thermolabile", "antithrombin and polymerisation or antithrombin and polymerization", and "antithrombin and conformational".

We only considered articles published in English.

The retrieved literature was examined for studies that investigated the effect of the mutation. The pathogenicity of a *SERPINC1* sequence variation was considered established if it led to AT deficiency in affected individuals and to an:

- impaired secretion of mutant AT proven by in vitro gene expression studies,
- intracellular retention of mutant AT proven by immunohistochemical or immunofluorescence analysis,
- impaired heparin activation or impaired heparin binding to mutant AT detected by enzyme kinetic studies with recombinant or purified AT mutants, direct binding studies, crossed immunoelectrophoresis, or heparin sepharose chromatography,
- impaired inactivation of thrombin or factor Xa by AT detected by in vitro binding studies or enzyme kinetic studies with recombinant or purified AT mutants, or
- conformationally unstable AT with loss of protease or heparin binding capacity, detected by heat stability assays and/or electrophoretic analysis.

The mutations were classified according to their functional and antigen level and effect on the AT molecule into mutations causing type I, IIHBS, IIRS, or IIPE AT deficiency.

Exonic non-synonymous single nucleotide polymorphisms (SNPs) with a population allele frequency of >1.0% were identified from the SNP database of the National Center for Biotechnology Information (NCBI; 16).

In Silico Methods

We evaluated 12 in silico tools: I) Meta-SNP, II) MutPred version 1.2., III) nsSNPAnalyzer (nonsynonymous SNP Analyzer), IV) PANTHER version 8.1 (Protein Analysis Through Evolutionary Relationships), V) PhD-SNP (Predictor of human Deleterious SNP), VI) PMut (Prediction of pathological mutations), VII) SIFT (Sorting Intolerant From Tolerant), VIII) SNAP (Screening for non-acceptable polymorphisms), IX) SNPs&Go (classifying human SNPs by including Gene Ontology), X) Poly-Phen-2 version 2.2.2 (Polymorphism Phenotyping-2), XI) PON-P2 (Pathogenic-or-not-Pipeline 2), and XII) PredictSNP 1.0.

The AT amino acid sequence of the NCBI protein accession number NP_000479.1 was used as the query sequence in all programs except SNPs&Go that used the Swiss-Prot code ANT3_HUMAN only. All tools were employed using their default settings if not otherwise stated.

SIFT (available at http://sift.jcvi.org/www/SIFT_seq_submit2.html) and PANTHER (http://www.pantherdb.org/tools/csnpScoreForm.jsp) are sequence homology-based tools [17,18]. A SIFT score <0.05 was considered as damaging, a score ≥0.05 as benign. SIFT default settings were: database UniRef90 2011 Apr; median conservation of sequence 3.0; allowance to remove sequences more than 90 percent identical to query. From the alignment to hidden Markov models, PANTHER calculates the substitution position-specific evolutionary conservation (subPSEC) score that ranges from 0 (neutral) to −10 (most likely to be damaging) [17]. A subPSEC score of −3 was used as cut-off for the dichotomisation of the effect of the mutants.

MutPred (http://mutpred.mutdb.org) is a sequence-homology based tool that also analyses structural and functional protein properties [19]. The probability that an amino acid substitution is deleterious/disease-related is expressed as the general score (g). A g score of >0.5-1 was considered disease-related.

PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/index.shtml) is a supervised machine-learning Na ve Bayes classifier that uses multiple sequence alignment, information from the three-dimensional protein structure, and sequence annotations from the UniProtKB/Swiss-Prot database to predict whether an amino acid substitution will affect protein stability or function [20]. Two prediction models of PolyPhen-2 trained on different datasets are available, HumDiv and HumVar [21]. HumDiv is recommended for the prediction of mutants with even mildly deleterious effects whereas HumVar is prone to identify those with more drastic effects [21]. The PolyPhen-2 outputs for each model were dichotomised into damaging (categories probably damaging and possibly damaging) and benign.

Download English Version:

https://daneshyari.com/en/article/6000745

Download Persian Version:

https://daneshyari.com/article/6000745

Daneshyari.com