

**Real-world clinical applicability of pathogenicity predictors assessed on *SERPINA1* mutations
in alpha-1-antitrypsin deficiency**

Edoardo Giacomuzzi^{1°}, Mattia Laffranchi^{2°}, Romina Berardelli², Viola Ravasio¹, Ilaria Ferrarotti³,
Bibek Gooptu⁴, Giuseppe Borsani¹, Annamaria Fra^{2*}

¹Division of Biology and Genetics, Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

²Experimental Oncology and Immunology, Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

³Centre for Diagnosis of Inherited Alpha-1 Antitrypsin Deficiency, Department of Internal Medicine and Therapeutics, University of Pavia, Pavia, Italy

⁴Leicester Institute of Structural and Chemical Biology / NIHR Leicester BRC – Respiratory, University of Leicester, Leicester, UK

[°]These authors contributed equally to this work

*Corresponding author:

Annamaria Fra

Dept. of Molecular and Translational Medicine,

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/humu.23562](https://doi.org/10.1002/humu.23562).

This article is protected by copyright. All rights reserved.

University of Brescia

25123 Brescia, Italy

E-mail: annamaria.fra@unibs.it

Grants from: Fondazione Cariplo (2013-0967), Regione Lombardia-Fondazione Cariplo (2015-1080), Italian Association Alfa1-AT and Medical Research Council UK.

ABSTRACT

The growth of publicly available data informing upon genetic variations, mechanisms of disease and disease sub-phenotypes offers great potential for personalised medicine. Computational approaches are likely required to assess large numbers of novel genetic variants. However, the integration of genetic, structural and pathophysiological data still represents a challenge for computational predictions and their clinical use. We addressed these issues for alpha-1-antitrypsin deficiency, a disease mediated by mutations in the *SERPINA1* gene encoding alpha-1-antitrypsin. We compiled a comprehensive database of *SERPINA1* coding mutations and assigned them apparent pathological relevance based upon available data. ‘Benign’ and ‘Pathogenic’ mutations were used to assess performance of 31 pathogenicity predictors. Well-performing algorithms clustered the subset of variants known to be severely pathogenic with high scores. Eight new mutations identified in the ExAC database and achieving high scores were selected for characterisation in cell models and showed secretory deficiency and polymer formation, supporting the predictive power of our computational approach. The behaviour of the pathogenic new variants and consistent outliers were rationalised by considering the protein structural context and residue conservation. These findings highlight the potential of computational methods to provide meaningful predictions of the pathogenic significance of novel mutations and identify areas for further investigation.

KEYWORDS: Alpha-1-Antitrypsin Deficiency; Serpins; Serpinopathies; Alpha-1-antitrypsin polymers; Pathogenicity prediction, ExAC database.

ABBREVIATIONS: α 1AT, Alpha1-Antitrypsin; ER, Endoplasmic Reticulum; ExAC, Exome Aggregation Consortium; HNE, Human Neutrophil Elastase; COPD, Chronic Obstructive Pulmonary Disease.

INTRODUCTION

Increasing access to genome sequencing technology supports increased personalisation of care in which disease risk can be stratified by genotype. The most straightforward applications of this will predictably occur for monogenic disorders, where many genetic variations are now found during wide screening of individuals with complex disease phenotypes, or in whom no disease is clinically apparent. Each gene is likely to host many such variants, though each one will tend to be rare. *In vitro*, cellular, and/or *in vivo* characterisation to define pathogenicity for all identified mutations is unlikely to be practical. Deleteriousness prediction scores have been developed to address this challenge and so distinguish benign and pathogenic alleles computationally (Niroula and Vihinen, 2016).

We chose to test how well general prediction algorithms predicted pathogenicity for variants of the *SERPINA1* gene (MIM# 107400) that encodes the major circulating anti-protease, alpha-1-antitrypsin (α 1AT). This system represents an interesting test. Pathogenic mutations cause the monogenic disorder alpha-1-antitrypsin deficiency (MIM# 613490), a relatively common ‘rare disease’ in individuals of North European descent, whose molecular mechanisms are amongst the best characterised for any disease (Lomas et al., 2016). It is characterized by deficiency of circulating

α 1AT. Common clinical manifestations are emphysema and/or hepatic cirrhosis. Almost all cases involve missense point mutations that cause the α 1AT variants to misfold within the endoplasmic reticulum (ER) of hepatocytes, which normally secrete the protein into the circulation. The synthesised mutant polypeptides are therefore prone to ER-associated degradation (ERAD) and/or polymerisation within the ER to varying degrees depending on the nature of the mutation. Emphysema and the associated pulmonary syndrome, chronic obstructive pulmonary disease (COPD), arise from loss-of-function due to reduction of anti-elastase activity that renders lung tissue vulnerable to dysregulated proteolysis. Hepatic disease ensues from gain-of-toxic-function effects of misfolding and polymerisation within hepatocytes. The function and dysfunction of α 1AT is intimately related to the three-dimensional structure of the protein that, due to its metastable native conformation, is prone to dramatic conformational changes (Lomas et al., 2016).

Coding mutations of the *SERPINA1* gene are most clearly defined by the amino-acid alterations involved. For α 1AT the latter are conventionally numbered according to the residue number in the mature secreted protein, lacking the 24 N-terminal residues of the signal peptide. This conventional system differs from that recommended by the Human Genome Variation Society (HGVS), which includes the initial 24 amino acids and is adopted hereafter as preferred annotation. In addition, classical definition of α 1AT variants includes iso-electric focusing (IEF) profiling (Ferrarotti et al., 2007). Wild-type α 1AT is designated PI*M (where PI refers to α ₁-Proteinase Inhibitor, an alternative name for α 1AT), mutants that migrate more cathodally are assigned letters occurring later in the alphabet whilst more anodal species are assigned earlier letters. Variants may also be named after a geographical location related to their discovery, typically the birthplace of the index case. The PI*M phenotype represents five normal polymorphisms named M1A, M1V, M2, M3 and M4, which have very similar IEF mobility (Luisetti and Seersholm, 2004; Ferrarotti et al., 2007). The coincidence of similar IEF in these five variants with physiological equivalence as ‘normal variants’ is serendipitous. M1A is believed to be the ancestral human allele, while M1V (p.A237V) contains the A213V substitution and it is the most frequent *SERPINA1* allele. M3 and M4 are characterized by p.E400D

and p.R125H variations respectively, in M1V background, whereas M2 (p.R125H; E400D) contains both substitutions.

The most common deficiency allele associated with clinically severe α 1AT deficiency is the Z variant (p.E366K), while the common α 1AT S variant (p.E288V) is associated with milder α 1AT deficiency (de Serres and Blanco, 2012). Although plasma levels of α 1AT in SS homozygotes are reduced to approximately 60% compared to healthy subjects, these appear sufficient to provide protection from damage to lung tissue (Hazari et al., 2017). Relatedly, the S variant is less polymerogenic compared to Z α 1AT, with more efficient clearance of the misfolded protein by degradation pathways (Curiel et al., 1989). On the other hand, compound heterozygotes with both S and Z alleles of *SERPINA1* (SZ) are at increased risk of α 1AT deficiency-related lung and liver diseases (Turino et al., 1996; American Thoracic Society and European Respiratory Society, 2003; Ferrarotti et al., 2005; Laffranchi et al., 2018). All the other observed α 1AT alleles have frequencies <0.005 and in many cases have been described in a single case or family. However, cumulatively, ‘rare’ mutations are likely to affect a substantial number of people. The reported frequency of rare α 1AT alleles, generally in compound heterozygosity with Z, is higher in Southern Europe (Ferrarotti et al., 2005; Piras et al., 2013). Among rare *SERPINA1* variants, the so-called null alleles are characterised by complete absence of protein in the bloodstream and are conventionally designed by Q0, followed by a given name. In most cases these are mutations generated by premature stop codons, splicing site alterations or large deletions. They are associated with high risk of lung disease, but not with liver disease (Luisetti and Seersholm, 2004; Ferrarotti et al., 2014). Unsurprisingly, the information about the clinical presentation of most rare pathological point mutants of α 1AT is limited. Those studied in more detail are polymerogenic mutants, such as M(malton) (p.F75del) (Graham et al., 1989), S(iiyama) (p.S77F) (Lomas et al., 1993) and King’s (p.H358D) (Miranda et al., 2010). Others such as the I (p.R63C) (Graham et al., 1989; Ronzoni et al., 2016), Queen’s (p.K178N) (Nyon et al., 2012), Baghdad (p.A360P) (Haq et al., 2016), Trento (p.E99V) (Miranda et al., 2017) and P(brescia) (p.G249R) (Medicina et al., 2009) are associated with milder polymerogenicity and minor plasma deficiency. The circulating deficiency of

α 1AT may be exacerbated by supervening reduced activity due to decreased binding affinity and inhibitory capacity against elastase, as reported for Z, F (p.R247C), Queen's and Baghdad mutations (Okayama et al., 1991; Cook et al., 1996; Nyon et al., 2012; Haq et al., 2016). A point mutation at the reactive site of α 1AT in the Pittsburgh variant (p.M382R) switches specificity from inhibition of elastase to inhibition of the clotting factors thrombin and Factor XIa, resulting in fatal bleeding events (Owen et al., 1983). Besides the anti-protease activity, amino acid variations may also potentially affect other anti-inflammatory functions of AAT (Jonigk et al., 2013; Janciauskiene et al., 2018).

Most computational tools used to predict pathogenicity use machine learning algorithms trained on known pathogenic variants to compute a damage score for every possible missense substitution. They integrate various combinations of features including conservation estimates, physico-chemical properties of the amino-acids, secondary structure, domain information and substitution matrices. Examples of these approaches include CADD (Kircher et al., 2014) and PolyPhen-2 (Adzhubei et al., 2010). In addition, 'meta-predictors' such as REVEL (Ioannidis et al., 2016) integrate results from previous prediction tools to improve classification performance. Successful gene-specific pathogenicity predictors have been developed for genes with large panels of mutations identified such as *CFTR* (Masica et al., 2012) and *BRCA1* (Starita et al., 2015).

Here, we report an updated catalogue of missense mutations described in the *SERPINA1* gene, together with a detailed characterization of its genetic variability across the human populations represented within large genomic databases. We then compare the performance of 31 prediction tools for their ability to discriminate previously characterized α 1AT variants and finally characterize in cellular models a panel of predicted pathogenic variants found in the Exome Aggregation Consortium (ExAC) database.

MATERIALS AND METHODS

Genetic variability of *SERPINA1* gene in human population

To compile the catalogue of α 1AT variants in Supp. Table S1, we integrated data from public databases and literature reports, last accessed in July 2017. We consulted public databases of clinically relevant variants, namely ClinVar (Landrum et al., 2014), HGMD (Stenson et al., 2014) and LOVD (Fokkema et al., 2011), queried for “*SERPINA1*”. Literature was searched in PubMed using as keywords “*SERPINA1*”, “alpha α 1-AT”, “ α 1-AT”, “ α 1 proteinase inhibitor”, combined with “variant”, “allele”, “mutant” and “rare”.

Genetic variability of *SERPINA1* gene in different human populations was assessed by combining data repositories from ExAC v.0.3.1 (~63,000 subjects) (<http://exac.broadinstitute.org>), UK10K (~3,500 subjects) (<http://www.uk10k.org>) and Greater Middle East (GME) Variome Project (~2,500 subjects) (<http://igm.ucsd.edu/gme>) (Supp. Table S2). The frequency of carriers of Z and S/T mutations, singularly and combined, and for all other pathogenic variants (P/ P* rare variants in Table 1 and loss-of-function variants) was calculated from allele counts in ExAC and GME populations. Estimations were not performed on UK10K data since this dataset includes data from twins, which could inflate observed allele frequencies for pathogenic variants.

Evaluation of pathogenicity/conservation predictors

We evaluate 24 deleteriousness and 7 conservation predictors for their ability to correctly classify α 1AT variations (Supp. Table S3). We included predictors reported in dbNSFP 3.2 (Liu et al., 2011, 2016) and 4 recently developed predictors: PON-P2 (6), REVEL (Ioannidis et al., 2016), iFISH (Wang and Wei, 2016) and M-CAP1.0 (Jagadeesh et al., 2016). Values of each score for all possible

missense variants in *SERPINA1* gene were retrieved from dbNSFP 3.2 or from predictor websites (Supp. Table S4). The M-CAP1.0 score was developed only for variants with MAF <0.01, so values for S/T and Z variants were not reported and set to missing. We used SERPINA1 protein sequence (NP_000286.3) as query in pBLAST search against database of primates sequences and selected 8 orthologs with $\geq 95\%$ sequence identity. By this method we identified 40 amino acid substitutions that were used as benign variants in subsequent analysis, as previously described (Riera et al., 2016) (Supp. Table S5 and Supp. Figure S1). Our final set of benign variants included also the 3 M background alleles for a total of 43 B variants. Pathogenic variants were P and P* variants reported in Table 1. Performance of each score in classifying benign and pathogenic variants was evaluated using the Overall Performance Metric (OPM), as defined in (Niroula et al., 2015), and AUC of ROC curves, calculated using ROCR package (Sing et al., 2005). For each score we calculated an optimal threshold, defined as the score value resulting in maximum OPM value, and then used this value to assess sensibility, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and Matthews correlation coefficient (MCC).

Distribution of REVEL scores and clusters definition for *SERPINA1* variants

Since REVEL emerged as the best predictor based on OPM value, we analysed how this score is distributed across all possible missense variants in *SERPINA1* gene. First, we evaluated the optimal number of variants subgroups that could be defined based on REVEL score. Analysis of within sum of squares across clusters suggested 3 groups as the best solution (Supp. Figure S2). These 3 groups are defined as: cluster 3, high scoring variants with REVEL value >0.618; cluster 2, with mid-range values between 0.354 and 0.618; cluster 1, with low scoring variants <0.354. For P/P* and tested variants we also calculated the classification concordance among the top 5 and top 10 performing predictors, using the optimal thresholds defined in Supp Table S3.

Expression vectors and cell transfection

Vectors encoding for the α_1 AT variants were obtained by site-directed mutagenesis of M1V (Medicina et al., 2009) using the QuikChange II Mutagenesis Kit (Agilent) with primers listed in Supp. Table S6. HEK293T/17 (ATCC#CRL-11268) or Hepa1-6 cells (ATCC#CRL-1830) were maintained in DMEM-10% FBS (Sigma) and transfected by PEI 'Max' (Polysciences Inc) or Lipofectamine2000 (Thermo Fisher) as previously described (Ronzoni et al., 2016; Miranda et al., 2017). 24h after transfection we collected the cell media and lysed the cells in 1% NP40/20 mM Tris-HCl pH 7.4/150 mM NaCl/10 mM N-ethylmaleimide/protease inhibitors, then discarding nuclei by 30' centrifugation at 800g.

SDS-PAGE, Native-PAGE and immunoblots

Lysates and media of transfected cells were analysed either by 7.5% SDS-PAGE or 8% Native-PAGE and immunoblots revealed by anti- α_1 AT (DAKO) followed by HRP-anti-rabbit antibodies (Thermo Fisher) and ECL (Euroclone) as described previously (Fra et al., 2012; Miranda et al., 2017).

Sandwich ELISA

Quantification of α_1 -AT in culture media was performed by sandwich ELISA as described (Miranda et al., 2017). Briefly, 96-well plates (Costar 3590) were coated with rabbit polyclonal anti- α_1 AT (DAKO) (2 μ g/ml) and saturated for 1 h at 37°C with a blocking buffer (PBS, 0.25% BSA). Serial dilutions (1:1.5) in PBS/0.1% BSA of purified α_1 AT (Merck) and cell media were added to the plates and incubated at 37°C for 1 h. After washing in PBS/0.05% Tween-20, wells were incubated for 1h at 37°C with Sheep anti- α_1 AT-HRP (Abcam) in PBS/0.1% BSA, further washed and revealed with the

TMB substrate (Sigma). The reaction was blocked by adding 3M HCl and the absorbance at 450 nm measured by an ELISA plate reader (EnSight, PerkinElmer).

Anti-elastase assay

The culture media of cells expressing α 1AT variants were diluted in PBS and incubated at 37°C for 30 min with equimolar porcine pancreatic elastase (PPE) (Sigma). Samples were then separated by 7.5% SDS-PAGE and detected by immunoblot by a polyclonal antibody anti- α 1AT (DAKO).

RESULTS

Updated catalogue of missense variants causing α 1AT deficiency

We integrated information from the publicly accessible databases ClinVar, HGMD and LOVD with published literature to compile an updated list of known missense variants in *SERPINA1* (Supp. Table S1). The M1V allele (RefSeq: NM_000295.4) is used as the reference sequence. Missense variations are annotated with nucleotide and amino acid substitutions reported with both HGVS and conventional α 1AT nomenclature. Whenever available, we reported the background allele and the name given to the α 1AT variant, as well as relevant references. Seven cases of double (2 aminoacids) substitution alleles were identified. In some cases, mutations occur on different haplotype backgrounds, resulting in different allele names, as in the case of S/T or Z/Z(augsburg). Four missense alleles associated with undetectable α 1AT plasma levels are annotated as Q0. However, these variants differ from the majority of Q0 type mutations, since they carry missense variations and not premature stop codons, splicing site alterations or large deletions (Ferrarotti et al., 2014).

Clinical significance of known variants was assessed after critical revision of published studies. Variants were classified as pathogenic (P) (Table 1) if clearly associated with disease based on clinical and experimental evidence according to the following criteria: (i) they were reported in at least two unrelated subjects with α 1AT deficiency-associated diseases; (ii) they were reported in single cases with manifestations of disease and have been further characterized by biochemical experiments and/or in cellular models. When a robust classification was not possible, the variants were catalogued as uncertain (U). In most cases, our classification correlates with that of ClinVar, where the variants are annotated in six main categories (P, B, LP, LB, U or O for pathological, benign, likely pathological, likely benign, uncertain and other, respectively) (Supp. Table S1). Among the pathogenic variants in Table 1, we further sub-classified 7 mutations as severe (P*), based on very low plasma levels *in vivo* and high tendency to form intracellular polymers when assessed *in vitro*. The distribution of pathogenic variants appears clustered in the tertiary α 1AT structure (PDB ID: 3NE4) (Figure 1).

***SERPINA1* variations in human population databases**

Different public databases have been established to collect genome/exome sequencing data worldwide. Querying for non-synonymous *SERPINA1* variants we retrieved 184 variants from ExAC v.0.3.1 (including 7 population groups), 38 from UK10K database (including UK individuals) and 20 from GME repository (including subjects from 6 regions in Greater Middle East) (Supp. Table S2). Most missense variants reported in ExAC populations are ultra-rare (Allelic Frequencies, AF < 0.0001), with 77 seen only once in the overall combined dataset (Supp. Table S7). Population specific variants are observed in UK10K (n=15) and GME (n=1) that are not reported in ExAC (Supp. Figure S3).

Several of the published variants are found in at least one of the population databases (Supp. Table S2). The polymorphisms defining the background M alleles (p.V237A, p.E400D, p.R125H) have global frequency of 0.215, 0.276 and 0.169, respectively, with differential distribution in different sub-populations (Figure 2A). The common pathogenic S (p.E288V) and Z (p.E366K) alleles show the larger AF values in the global population: 0.0201 and 0.0117, respectively (Figure 2B). Frequency of carriers of Z, S in the different populations are reported in Supp. Table S8. Notably, 11 Z homozygotes and 48 S homozygotes are reported in the ExAC database and at least 1 homozygous individual is reported for variants F, M(malton) and M(procida). A large proportion of carriers of AAT variants identified in Asian (GME, EAS and SAS) and non-Finnish European (NFE) groups carry a rare pathogenic allele, while Z and S variants prevail in other populations (Figure 2C). The observed AF and distribution are comparable to previous estimation in general populations of 97 countries worldwide (de Serres and Blanco, 2012).

Assessment of mutations by pathogenicity predictors and conservation scores

Several bioinformatics tools have been developed to predict the deleteriousness of protein variations, each considering a different set of features to evaluate amino acid and/or nucleotide substitutions. We considered 31 deleteriousness/conservation predictors for their ability to predict pathogenicity of the *SERPINA1* variants listed in Table 1 (P and P*). We considered as benign (B) the three common M polymorphisms, together with a panel of substitutions found by alignments with primate $\alpha 1$ AT orthologs (Supp. Table S5 and Supp. Figure S1), as performed previously (Riera et al., 2016). Overall, we used 32 P/P* and 43 B substitutions to calibrate the predictive algorithms outputs and ranked them according to their Overall Performance Metric (OPM) at optimal thresholds (Supp. Table S3). The REVEL algorithm had the highest OPM (0.93, Figure 3A), resulting in well distinguished distribution of scores comparing P/P* and B variants (Figure 3B). Overall, the 10 best performing predictors reach a consistent classification for all P mutations, except E(taurisano), N(hartford city), Pittsburgh and

p.Q129P (Supp. Table S9). Notably, P* mutations consistently achieved extreme pathogenicity scores. Missense variants reported in ExAC and all other possible missense substitutions (Other) predominantly scored below the optimal REVEL threshold, consistent with the expectation that most possible missense mutations will have little or no functional impact. Score distributions for the 10 best-performing tools are reported in Supp. Figure S10.

Clustering analysis of α 1AT missense variants using REVEL

We applied REVEL to all possible missense variants in *SERPINA1* gene (Supp. Table S10) to evaluate their pathogenic/neutral significance. Based on k-means clustering, we defined 3 groups that optimise classification: Cluster 1 (0-0.354), Cluster 2 (0.355-0.618), and Cluster 3 (0.618-1). Known P mutations all belong to cluster 2 or 3, above the 0.477 optimal threshold of REVEL, all P* mutations belong to Cluster 3 (Supp. Figure S5B), while U variants are distributed equally across the 3 clusters (Supp. Figure S5B). Looking at the overall distribution of all possible missense variants and that of ExAC and GME, the greater proportion falls in Clusters 1 and 2 (Figure 4 A, B). We mapped where variants belonging to Cluster 3 occur in the structure of native α 1AT (Figure 4C). This distribution of Cluster 3 variants spans in a network of residues whose structural dynamics were shown to be most affected by known mutations (Fra et al., 2012; Nyon et al., 2012).

Characterization of new α 1AT variants in cellular models

We correlated high REVEL scores against biological readouts in cellular models for a range of variants (highlighted in Figure 4B). Seven Cluster 3 variants assessed were previously uncharacterised whilst p.P279T, previously reported in the M2(pont-eveque) and in the double mutant M(frankfurt) as benign due to normal AAT serum levels and lack of hepatic or pulmonary symptoms (Joly et al., 2014; Faber et al., 1994), achieved the highest REVEL score. Two variants from Cluster 2 (p.A55P,

p.G282R) and two from Cluster 1 (p.S71R, p.M409V) were evaluated for comparison (Supp. Table S11).

The variants were first expressed transiently in HEK293T cells. Wild-type M1V and Z α 1AT were also expressed as reference. Formation of polymers both in cell extracts (Figure 5A) and culture media (Figure 5B) was analysed by native PAGE. Intracellular and secreted Z α 1AT was overwhelmingly polymeric, consistently with previous *in vitro* and *in vivo* studies (Fra et al., 2012, 2016; Tan et al., 2014; Miranda et al., 2017). The expression of p.P313S, p.P279T and p.G282R variants showed polymer/monomer proportions similar to Z α 1AT (Figure 5A). Conversely p.M409V resembles M α 1AT with monomer predominance and the other variants showed intermediate aggregation profiles. The α 1AT variants secreted by HEK293T transfected cells were functional, as they were able to react with protease porcine pancreatic elastase (PPE) (Figure 5C). Z, p.P279T and p.P313S were less functional due to a lower intrinsic activity, as published for the Z mutant (Ogushi et al., 1987), and/or to the presence of inactive polymers in the media. Similar results were obtained by expression in Hepa 1-6 cells in which we also analysed the p.A55P and p.S71R variants, respectively classified in cluster 2 and cluster 1 (Figure 5 D-E).

Secretion was evaluated by quantifying α 1AT in the culture media of both HEK293 and Hepa transfected cells by sandwich ELISA (Figure 5F). Consistent with previous studies (Fra et al., 2012; Ronzoni et al., 2016; Miranda et al., 2017; Laffranchi et al., 2018), these cell lines reproduced the secretion deficiency phenotype observed *in vivo* for Z variant homozygotes (~15% relative to M homozygotes). All Cluster 3 variants showed statistically significant reductions in α 1AT secretion relative to M α 1AT. Notably, the levels of p.P279T in HEK293T and Hepa cells (6.1 \pm 1.8% and 15.0 \pm 5.3% respectively) and p.P313S (3.4 \pm 2.0% and 1.5 \pm 1.0% respectively) were severely reduced in the media, similar to levels of Z α 1AT. Cluster 3 variants p.L90R, p.G216C and p.G331R showed more severe secretion deficiency in the Hepa cell line compared to HEK293T cells. A substantial reduction was also observed for the Cluster 2 variants p.A55P and p.G282R, while Cluster 1 variants

(p.M409V and p.S71R) were secreted at similar levels as M α 1AT. The observed secretion levels of the variants tested in the two cell models were then plotted against their REVEL scores (Figure 5G). Linear regression analysis suggested positive correlations between the two parameters in both cell models, although the datasets are relatively small.

Taken together, these biological findings support the potential of such algorithms to prospectively predict clinical significance of α 1AT variants with reasonable accuracy.

DISCUSSION

Several *in silico* tools have been developed to predict the deleterious effect of missense variants. These have become even more important in view of the increasing number of variants identified by high-throughput screenings of healthy individuals as well as of large cohorts of patients. Here we have compared the performance of multiple algorithms for a single gene, *SERPINA1*, where several point mutations are associated with the α 1AT deficiency through loss- and gain-of-function mechanisms.

A comprehensive database of mutations and their various classifications is a vital interface to evaluate the actual performance of computational tools, and at present it seems that this has to be accomplished manually by experts in the biological field. We have compiled such a database for *SERPINA1* and α 1AT deficiency and have identified important issues affecting predictions, which are likely of general relevance to other genes and disorders. First, many mutations identified from literature review were not cross-referenced on publicly accessible databases. Mutations may be described in the literature in terms of nucleic acid changes or simply at the amino-acid level. Conventional residue numbering in proteins generated from a precursor polypeptide (e.g. all secreted eukaryotic proteins) can diverge from the numbering of residues by HGSV. Existing assignments of the pathogenicity of variants may be based on varying degrees of evidence. Initiatives within a field can, as here, produce

a high-quality database of variants to cross-check with general databases and may be optimised for biomedical relevance. This can be of great use in the short term for those active in the field. However, without dedicated ongoing resource to maintain them, these inevitably become obsolete or lost.

We used well-supported pathogenic (P) and known or inferred benign variants (B) to calibrate and comparatively evaluate the performance of a large number of different algorithms designed for prediction of pathogenicity. The performance varied considerably among different algorithms suggesting that different predictors may be better suited for specific genes and hence diseases. A subset of the pathogenic variants for which the existing evidence strongly supported severe pathogenicity (P*) was used as an internal validator of the predictors' performance. The distribution of scores for P* mutants tended to cluster for the best performing predictors at the end of the scale indicating greatest pathogenic potential. None of the assessed algorithms identified as pathogenic the Pittsburgh variant whose mutation at the active site switches the anti-protease specificity from elastase to thrombin and Factor XIa. This is presumably because the mutated amino acid affects reactivity with the target protease rather than the protein structure.

As REVEL performed best in discriminating variants of known pathogenicity from those classified as benign, this tool was used to assess the pathogenic potential of all possible *SERPINA1* variants and those observed in the ExAC population database. The majority of these were rare variants of unknown pathogenic potential, since they lack an experimental or clinical characterisation. Most of possible missense variants were predicted to be benign by the REVEL algorithm. However, several novel variants identified within the ExAC cohort achieved scores above the optimal REVEL threshold. Eight of these variants, along with two new ones predicted as benign and with the wild type M and mutant Z as controls, were characterised *in vitro* using two distinct cellular models for α 1AT deficiency. We determined the presence of extracellular and intracellular polymers of α 1AT, the secreted levels and anti-protease function. The approach performed well in positively predicting mutations associated with reduced secretion mainly due to intracellular polymerisation, with varying

degrees of similarity to the Z variant. Although the number of variants characterised in cellular models is limited, our observations show that the secretion defect correlates with distribution among REVEL clusters and supports the hypothesis that the secretion levels might correlate with REVEL scores in a continuous manner. The secretion levels observed in our two cell models, though broadly similar, differ for some tested variants, likely reflecting differences among cell types in the proteostatic mechanisms (Fra et al., 2017; Sala et al., 2017).

The striking effects of p.P313S, p.P279T, and p.G282R mutations can be rationalised in terms of the structure of native α 1AT (Supp. Figure S6). Cluster 3 variations p.P313S and p.P279T both replace highly conserved, hydrophobic proline residues that terminate β -strands with small polar residues that are associated with less rigid and kinked backbone behaviour. Proline 313 is one of the most highly conserved residues across the serpin superfamily (~96% conservation) (Irving et al., 2000). It terminates strand 2 of β -sheet C (s2C) and interacts most directly with other similarly highly conserved residues (Supp. Figure S6B). Structurally, ‘latch’ interactions between these residues constitute a hydrophobic interaction network stabilising the flank region of the β -barrel formed by β -sheets B and C (Fra et al., 2012), which may nucleate folding of the metastable serpin native state (Tsutsui et al., 2012). The lateral part of this acts as a steric regulator of serpin conformational change and so is known as the ‘gate’ region. In addition to such latch mutation effects, the p.P313S variation likely has additional effects upon the s2C terminus that may also affect folding and/or conformational stability. The p.P279T mutation affects the C-terminal residue of strand 3 of β -sheet B (s3B). It also arises at a sharp boundary between conserved hydrophobic and polar regions and so may make this less biochemically distinct. The p.G282R variation replaces a small, flexible uncharged residue with a bulky, polar residue within the turn between strand s3B and helix G. This amino acid substitution changes a DEGK sequence to DERK where two acidic residues are immediately followed by two basic residues. Its effects may therefore be mediated by abnormal interactions between these large charged residues and close structural elements (s3B and helix G). The DEGK amino acid stretch has

low conservation among α 1AT orthologs, likely explaining why the p.G282R mutation seemed an outlier with greater observed severity relative to its low REVEL score when compared with the correlations observed for other variants.

The REVEL algorithm and the clustering analysis therefore seem promising for identifying novel variants of interest for study *in vitro* and has potential for clinical utility. Caveats remain. Reduced secretion observed in a cell model cannot be directly translated into clinical significance and disease risk. Similarly, defining thresholds of polymer accumulation in cell models to stratify clinically significant risk of liver disease may be challenging. Interestingly, studies on Z variant homozygotes have indicated that genetic modifiers affect disease penetrance (Joly et al., 2017). Though the nature of these modifiers remains to be fully elucidated, it is possible they might modify the intracellular handling of different *SERPINA1* variants to different degrees. Thus, it remains challenging to translate predictions from algorithms like those assessed in this study into a stratified prediction of the risk of lung and/or liver disease.

Overall, the work presented here provides a detailed picture of genetic variability in *SERPINA1* and represents a real-world example of the potential and challenges of using genomic data and computational approaches to evaluate the functional and clinical impact of new variants identified in genetic disorders.

CONFLICT OF INTEREST

The authors have nothing to declare.

ACKNOWLEDGEMENTS

The authors thank Marco Lancini, Giovanni Sgrò and Leonardo Lanfranchi (University of Brescia) for technical assistance.

This article is protected by copyright. All rights reserved.

REFERENCES

- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7:248–249.
- Cook L, Burdon JG, Brenton S, Knight KR, Janus ED. 1996. Kinetic characterisation of alpha-1-antitrypsin F as an inhibitor of human neutrophil elastase. *Pathology* 28:242–247.
- Curiel DT, Chytil A, Courtney M, Crystal RG. 1989. Serum alpha 1-antitrypsin deficiency associated with the common S-type (Glu264Val) mutation results from intracellular degradation of alpha 1-antitrypsin prior to secretion. *J Biol Chem* 264:10477–10486.
- Faber JP, Poller W, Weidinger S, Kirchgesser M, Schwaab R, Bidlingmaier F, Olek K. 1994. Identification and DNA sequence analysis of 15 new alpha 1-antitrypsin variants, including two PI*Q0 alleles and one deficient PI*M allele. *Am J Hum Genet* 55:1113–1121.
- Ferrarotti I, Baccheschi J, Zorzetto M, Tinelli C, Corda L, Balbi B, Campo I, Pozzi E, Faa G, Coni P, Massi G, Stella G, et al. 2005. Prevalence and phenotype of subjects carrying rare variants in the Italian registry for alpha1-antitrypsin deficiency. *J Med Genet* 42:282–287.
- Ferrarotti I, Carroll TP, Ottaviani S, Fra AM, O'Brien G, Molloy K, Corda L, Medicina D, Curran DR, McElvaney NG, Luisetti M. 2014. Identification and characterisation of eight novel SERPINA1 Null mutations. *Orphanet J Rare Dis* 9:172.
- Ferrarotti I, Scabini R, Campo I, Ottaviani S, Zorzetto M, Gorrini M, Luisetti M. 2007. Laboratory diagnosis of alpha1-antitrypsin deficiency. *Transl Res* 150:267–274.
- Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, Dunnen JT den. 2011. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* 32:557–563.

Fra A, Cosmi F, Ordoñez A, Berardelli R, Perez J, Guadagno NA, Corda L, Marciniak SJ, Lomas DA, Miranda E. 2016. Polymers of Z α 1-antitrypsin are secreted in cell models of disease. *Eur Respir J* 47:1005–1009.

Fra A, Yoboue ED, Sitia R. 2017. Cysteines as Redox Molecular Switches and Targets of Disease. *Front Mol Neurosci* 10:167.

Fra AM, Gooptu B, Ferrarotti I, Miranda E, Scabini R, Ronzoni R, Benini F, Corda L, Medicina D, Luisetti M, Schiaffonati L. 2012. Three new alpha1-antitrypsin deficiency variants help to define a C-terminal region regulating conformational change and polymerization. *PLoS ONE* 7:e38405.

Graham A, Kalsheker NA, Newton CR, Bamforth FJ, Powell SJ, Markham AF. 1989. Molecular characterisation of three alpha-1-antitrypsin deficiency variants: proteinase inhibitor (Pi) nullcardiff (Asp256Val); PiMmalton (Phe51deletion) and PiI (Arg39Cys). *Hum Genet* 84:55–58.

Haq I, Irving JA, Saleh AD, Dron L, Regan-Mochrie GL, Motamedi-Shad N, Hurst JR, Gooptu B, Lomas DA. 2016. Deficiency Mutations of Alpha-1 Antitrypsin. Effects on Folding, Function, and Polymerization. *Am J Respir Cell Mol Biol* 54:71–80.

Hazari YM, Bashir A, Habib M, Bashir S, Habib H, Qasim MA, Shah NN, Haq E, Teckman J, Fazili KM. 2017. Alpha-1-antitrypsin deficiency: Genetic variations, clinical manifestations and therapeutic interventions. *Mutat Res* 773:14–25.

Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, et al. 2016. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 99:877–885.

Irving JA, Pike RN, Lesk AM, Whisstock JC. 2000. Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Res* 10:1845–1864.

Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, Bernstein JA, Bejerano G. 2016. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet* 48:1581–1586.

Janciauskiene S, Wrenger S, Immenschuh S, Olejnicka B, Greulich T, Welte T, Chorostowska-Wynimko J. 2018. The Multifaceted Effects of Alpha1-Antitrypsin on Neutrophil Functions. *Front Pharmacol* 9:341.

Joly P, Lacan P, Chapuis-Cellier C, Garcia C, Bererd M, Francina A. 2014. Molecular characterization of 7 new alpha-1 anti-trypsin (A1AT) variants including two with an associated deficient phenotype. *Clin Chim Acta* 427:21–22.

Joly P, Vignaud H, Di Martino J, Ruiz M, Garin R, Restier L, Belmalih A, Marchal C, Cullin C, Arveiler B, Fergelot P, Gitler AD, et al. 2017. ERAD defects and the HFE-H63D variant are associated with increased risk of liver damages in Alpha 1-Antitrypsin Deficiency. *PLoS ONE* 12:e0179369.

Jonigk D, Al-Omari M, Maegel L, Müller M, Izykowski N, Hong J, Hong K, Kim S-H, Dorsch M, Mahadeva R, Laenger F, Kreipe H, et al. 2013. Anti-inflammatory and immunomodulatory properties of α 1-antitrypsin without inhibition of elastase. *Proc Natl Acad Sci USA* 110:15007–15012.

Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46:310–315.

Laffranchi M, Berardelli R, Ronzoni R, Lomas DA, Fra A. 2018. Heteropolymerization of α -1-antitrypsin mutants in cell models mimicking heterozygosity. *Hum Mol Genet* 27:1785–1793.

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42:D980-985.

Liu X, Jian X, Boerwinkle E. 2011. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* 32:894–899.

Liu X, Wu C, Li C, Boerwinkle E. 2016. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat* 37:235–241.

Lomas DA, Finch JT, Seyama K, Nukiwa T, Carrell RW. 1993. Alpha 1-antitrypsin Siiyama (Ser53-->Phe). Further evidence for intracellular loop-sheet polymerization. *J Biol Chem* 268:15333–15335.

Lomas DA, Hurst JR, Gooptu B. 2016. Update on alpha-1 antitrypsin deficiency: New therapies. *J Hepatol* 65:413–424.

Luisetti M, Seersholm N. 2004. Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. *Thorax* 59:164–169.

Masica DL, Sosnay PR, Cutting GR, Karchin R. 2012. Phenotype-optimized sequence ensembles substantially improve prediction of disease-causing mutation in cystic fibrosis. *Hum Mutat* 33:1267–1274.

Medicina D, Montani N, Fra AM, Tiberio L, Corda L, Miranda E, Pezzini A, Bonetti F, Ingrassia R, Scabini R, Facchetti F, Schiaffonati L. 2009. Molecular characterization of the new defective P(brescia) alpha1-antitrypsin allele. *Hum Mutat* 30: E771-781.

Miranda E, Ferrarotti I, Berardelli R, Laffranchi M, Cerea M, Gangemi F, Haq I, Ottaviani S, Lomas DA, Irving JA, Fra A. 2017. The pathological Trento variant of alpha-1-antitrypsin (E75V) shows nonclassical behaviour during polymerization. *FEBS J* 284:2110–2126.

Miranda E, Pérez J, Ekeowa UI, Hadzic N, Kalsheker N, Gooptu B, Portmann B, Belorgey D, Hill M, Chambers S, Teckman J, Alexander GJ, et al. 2010. A novel monoclonal antibody to characterize

pathogenic polymers in liver disease associated with alpha1-antitrypsin deficiency. *Hepatology* 52:1078–1088.

Niroula A, Urolagin S, Vihinen M. 2015. PON-P2: prediction method for fast and reliable identification of harmful variants. *PLoS ONE* 10:e0117380.

Niroula A, Vihinen M. 2016. Variation Interpretation Predictors: Principles, Types, Performance, and Choice. *Hum Mutat* 37:579–597.

Nyon MP, Segu L, Cabrita LD, Lévy GR, Kirkpatrick J, Roussel BD, Patschull AOM, Barrett TE, Ekeowa UI, Kerr R, Waudby CA, Kalsheker N, et al. 2012. Structural dynamics associated with intermediate formation in an archetypal conformational disease. *Structure* 20:504–512.

Ogushi F, Fells GA, Hubbard RC, Straus SD, Crystal RG. 1987. Z-type alpha 1-antitrypsin is less competent than M1-type alpha 1-antitrypsin as an inhibitor of neutrophil elastase. *J Clin Invest* 80:1366–1374.

Okayama H, Brantly M, Holmes M, Crystal RG. 1991. Characterization of the molecular basis of the alpha 1-antitrypsin F allele. *Am J Hum Genet* 48:1154–1158.

Owen MC, Brennan SO, Lewis JH, Carrell RW. 1983. Mutation of antitrypsin to antithrombin. alpha 1-antitrypsin Pittsburgh (358 Met leads to Arg), a fatal bleeding disorder. *N Engl J Med* 309:694–698.

Piras B, Ferrarotti I, Lara B, Martinez MT, Bustamante A, Ottaviani S, Pirina P, Luisetti M, Miravittles M. 2013. Clinical phenotypes of Italian and Spanish patients with α 1-antitrypsin deficiency. *Eur Respir J* 42:54–64.

Riera C, Padilla N, Cruz X de la. 2016. The Complementarity Between Protein-Specific and General Pathogenicity Predictors for Amino Acid Substitutions. *Hum Mutat* 37:1013–1024.

Ronzoni R, Berardelli R, Medicina D, Sitia R, Gooptu B, Fra AM. 2016. Aberrant disulphide bonding contributes to the ER retention of α 1-antitrypsin deficiency variants. *Hum Mol Genet* 25:642–650.

Sala AJ, Bott LC, Morimoto RI. 2017. Shaping proteostasis at the cellular, tissue, and organismal level. *J Cell Biol* 216:1231–1241.

Serres FJ de, Blanco I. 2012. Prevalence of α 1-antitrypsin deficiency alleles PI*S and PI*Z worldwide and effective screening for each of the five phenotypic classes PI*MS, PI*MZ, PI*SS, PI*SZ, and PI*ZZ: a comprehensive review. *Ther Adv Respir Dis* 6:277–295.

Sing T, Sander O, Beerenwinkel N, Lengauer T. 2005. ROCR: visualizing classifier performance in R. *Bioinformatics* 21:3940–3941.

Starita LM, Young DL, Islam M, Kitzman JO, Gullingsrud J, Hause RJ, Fowler DM, Parvin JD, Shendure J, Fields S. 2015. Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. *Genetics* 200:413–422.

Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. 2014. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 133:1–9.

Tan L, Dickens JA, Demeo DL, Miranda E, Perez J, Rashid ST, Day J, Ordoñez A, Marciniak SJ, Haq I, Barker AF, Campbell EJ, et al. 2014. Circulating polymers in α 1-antitrypsin deficiency. *Eur Respir J* 43:1501–1504.

Tsutsui Y, Cruz RD, Wintrobe PL. 2012. Folding mechanism of the metastable serpin α 1-antitrypsin. *PNAS* 109:4467–4472.

Turino GM, Barker AF, Brantly ML, Cohen AB, Connelly RP, Crystal RG, Eden E, Schluchter MD, Stoller JK. 1996. Clinical features of individuals with PI*SZ phenotype of alpha 1-antitrypsin deficiency. alpha 1-Antitrypsin Deficiency Registry Study Group. Am J Respir Crit Care Med 154:1718–1725.

Wang M, Wei L. 2016. iFish: predicting the pathogenicity of human nonsynonymous variants using gene-specific/family-specific attributes and classifiers. Sci Rep 6:31321.

FIGURE LEGENDS

Figure 1. Distribution of pathogenic variants in $\alpha 1$ AT. (A) Localization of pathological missense variants reported in Table 1 within the secondary structure of $\alpha 1$ AT. (B) Distribution of known pathological missense variants in the tertiary structure of $\alpha 1$ AT (PDB: 3NE4) (PyMOL; <https://pymol.org>). P and P* variants are respectively represented in orange and red.



Figure 2. Frequencies of known *SERPINA1* alleles in population databases. The graphs represent the allelic frequencies in ExAC and GME populations for (A) M haplotypes and (B) the common Z and S/T alleles. (C) Percentage of carriers of rare pathogenic missense and LoF mutations compared to carriers of Z and S alleles. The numbers of carriers in each population are reported in the bars (N) and detailed in Supp. Table S8. The total number of subjects in the considered ExAC populations are: ALL, (total population) 60,706 subjects; AFR (African), 5,203 subjects; AMR (American), 5,789 subjects; EAS (East Asian), 4,327 subjects; FIN (Finnish European), 3,307 subjects; NFE (Non-Finnish European), 33,370 subjects; OTH (Other populations) in ExAC, 454 subjects; SAS = South Asian, 8,256 subjects. The Great Middle East population (GME) includes 993 subjects.

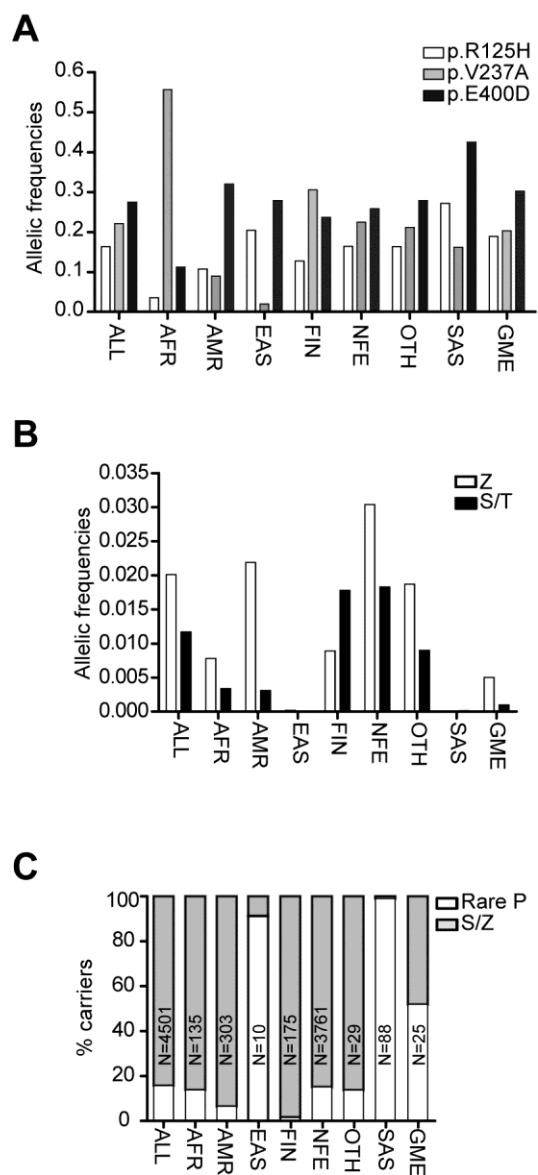


Figure 3. Performance of pathogenicity predictors. (A) Ability of 31 predictors in distinguishing pathogenic variants was ranked using OPM. (B) Violin plot of REVEL showing score distributions of P, P* and B α 1AT variants, as well as for amino acid variations reported in ExAC and all other possible variations based on single nucleotide change (Other). Dashed line indicates optimal threshold.

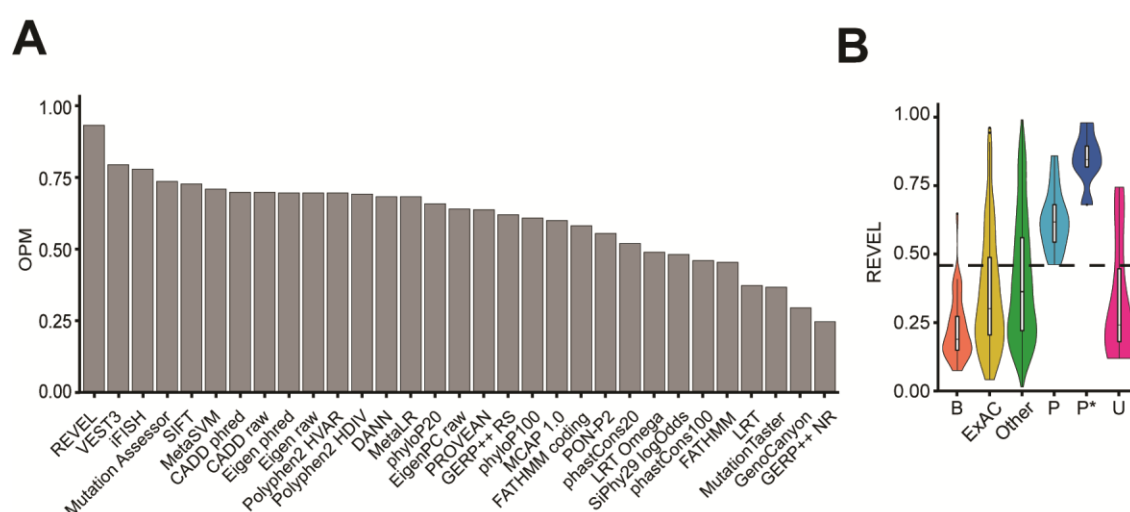


Figure 4. Clustering analysis of REVEL scores for α 1AT variants. (A) Stacked bars represent the distribution of all *SERPINA1* missense variants among the 3 clusters based on REVEL scores (Red: cluster 3; Green: cluster 2; Blue: cluster 1). (B) The scatter plot represents distribution of ExAC and GME missense variants and indicates those selected for our cellular studies. (C) Representation of amino acid whose substitution falls in cluster 3 (in red) in the α 1AT 3D structure (PDB: 3NE4). Spheres represent amino acid variants tested in this study with the cluster colour code of panel A.

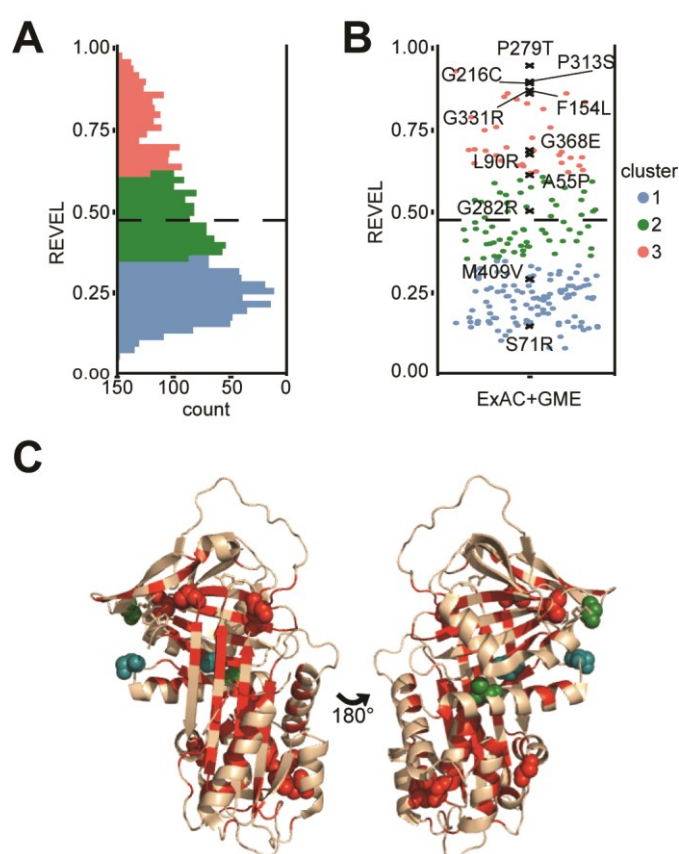


Figure 5. Characterization of new α 1AT variants in HEK293T and Hepa cells. (A, B) Immunoblots with anti- α 1AT pAb of cell extracts (A) and media (B) from HEK293T cells expressing the indicated variants, resolved by 7.5% SDS-PAGE or 8% Native-PAGE. Cell media were analysed by native PAGE with either equal volumes/lane (upper panel) or equal α 1AT amount/lane (lower panel). (C) Anti-elastase activity of α 1AT variants expressed in HEK293T. Cell media of transfected cells were incubated with (+) or without (-) an equimolar concentration of Porcine Pancreatic Elastase (PPE) for 30 min at 37°C and the complexes (MW 68kDa) were resolved from unreacted α 1AT monomers (52 kDa) by 7.5% SDS-PAGE and immunoblot with anti- α 1AT pAb. (D) Immunoblots with anti- α 1AT pAb of media from Hepa 1-6 cells expressing the indicated variants, resolved by 8% Native-PAGE. (E) Anti-elastase activity of α 1AT variants expressed in Hepa 1-6 cells, assayed as in panel C. (F) α 1AT levels in cell media from transfected HEK293T or Hepa cells were quantified by sandwich ELISA and represented as percentages of the wild-type M levels (mean \pm SEM, n=2) (1-way ANOVA, $p<0.0001$; two-tailed unpaired t-test between each variant and M α 1AT, * $p<0.05$, ** $p<0.001$, *** $p<0.0001$). (G) Linear Correlation ($p<0.001$) between REVEL scores and the secretion percentages of the α 1AT variants reported in panel F.

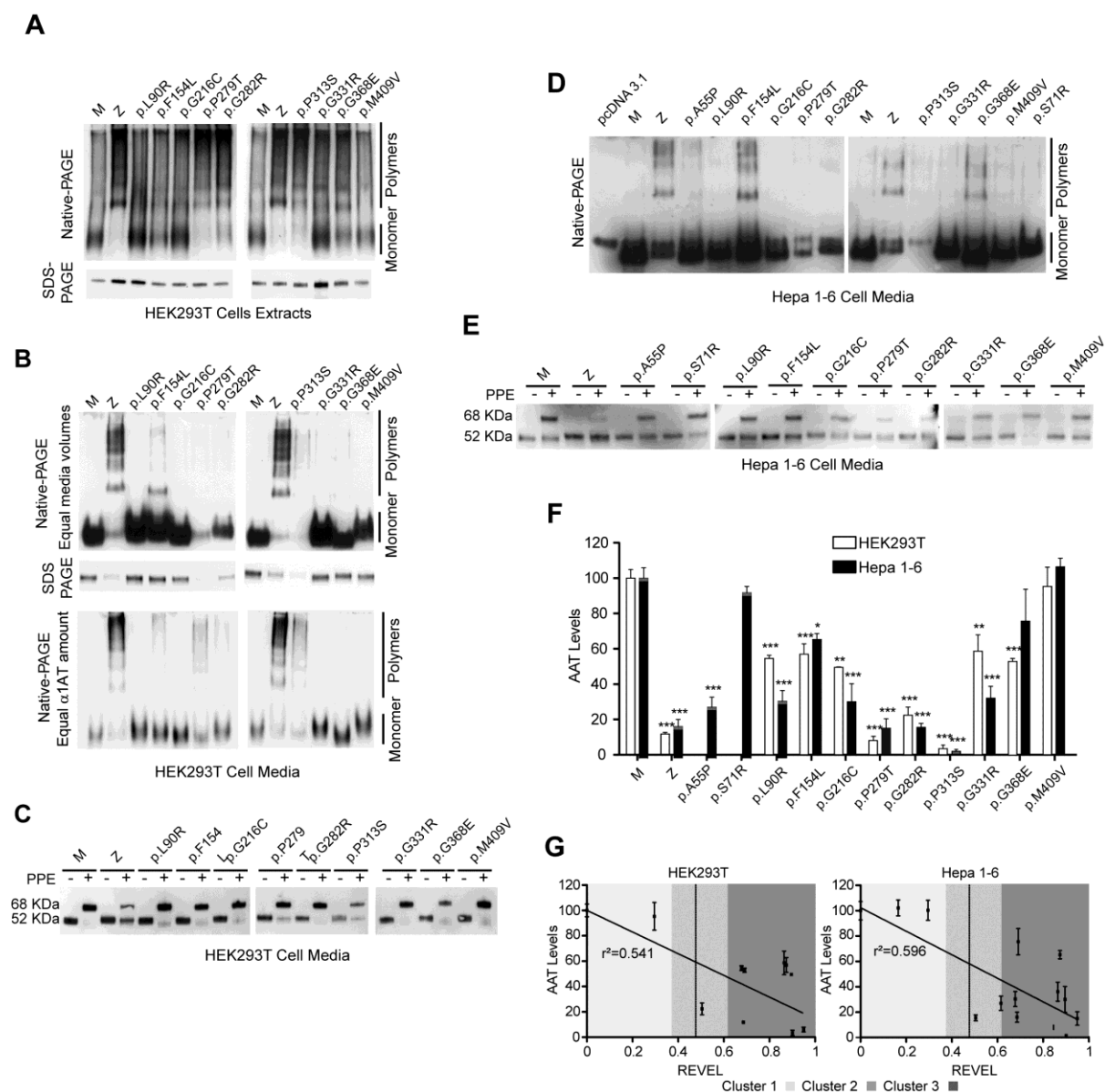


Table 1. Dataset of α 1AT variants used for evaluation of predictors' performance (P, pathogenic; P*, severe).

AA variation HGVS/ conventional	Allele name	P/P*	Clinical Phenotype
p.F59C/ F35C	Brixia	P	Def
p.R63C/ R39C	I	P	Def, Lung
p.L65P/ L41P	M(procida)	P	Def, Lung
p.F75del/ F51del	M(palermo) M(malton)/ M(cagliari)	P*	Def, Lung, Liver
p.S77F/ S53F	S(iiyama)	P*	Def, Lung, Liver
p.G91E/ G67E	M(mineral springs)	P	Def, Lung
p.T92I/ T68I	Q0(lisbon)	P	Def, Lung
p.E99V/ E75V	Trento	P	Def, Lung
p.T109M/ T85M	Z(bristol)	P	Lung, Liver
p.II16D/ I92D	Q0(ludwigshafen)	P	Def, Lung
p.Q129P/ Q105P	NA	P	Def
p.K178N/ K154N	Queen's	P	Def
p.E186G/ E162G	P(gaia)	P	Lung
p.V234E/ V210E	M1(pierre-bénite)	P	Def, Lung
p.R247C/ R223C	F	P	Def, Lung, Dysfunction
p.G249R/ G225R	P(brescia)	P	Def, Lung, Liver
p.D280V/ D256V	P(duarte) Q0(cardiff)/ (Plowell)	P	Def, Lung
p.K283I/ K259I	M(pisa)	P	Def, Lung
p.L287P/ L263P	Q0(gaia)	P	Def, Lung
p.E288V/ E264V	S/T	P	Def, Lung

p.T292I/ T268I	N(hartford city)	P	Def
p.G344R/ G320R	P(salt lake)/ P(lyon)	P	Lung
p.V357M/ V333M	NA	P	Def, Liver
p.H358D/ H334D	King's	P*	Def, Lung, Liver
p.A360P/ A336P	Baghdad	P	Def
p.A360T/ A336T	W(bethesda)	P	Def, Lung, Liver
p.E366K/ E342K	Z Z(augsburg)/ Z(tun)	P*	Def, Lung, Liver
p.M382R/ M358R	Pittsburgh	P	Bleeding
p.K392E/ K368E	E(tauriano)	P	Def, Lung
p.P393L/ P369L	M(heerlen)	P*	Def, Lung
p.P393S/ P369S	M(wurzburg)/ M(val d'hebron)	P*	Def, Lung, Liver
p.M409T/ M385T	NA	P	Def, Lung
p.P415H/ P391K	Y(orzinuovi)	P*	Def, Liver