Amyloid
The Journal of Protein Folding Disorders

http://informahealthcare.com/amy
ISSN: 1350-6129 (print), 1744-2818 (electronic)

Amyloid, Early Online: 1–9
 DOI: 10.3109/13506129.2014.900487

informa
healthcare

ORIGINAL ARTICLE

In silico analysis of *TTR* gene (coding and non-coding regions, and interactive network) and its implications in transthyretin-related amyloidosis

Renato Polimanti[1], Marco Di Girolamo[2], Dario Manfellotto[2], and Maria Fuciarelli[1]

[1]*Department of Biology, University of Rome "Tor Vergata", Rome, Italy and* [2]*Department of Medicine, Clinical Pathophysiology Center, AFaR – "San Giovanni Calibita" Fatebenefratelli Hospital, Isola Tiberina, Rome, Italy*

Abstract

Introduction: Transthyretin (TTR)-related amyloidosis is a life-threatening disease. Currently, several questions about the pathogenic mechanisms of TTR-related amyloidosis remain unanswered.
Methods: We have investigated various TTR-related issues using different *in silico* approaches.
Results: Using an amino acid similarity-based analysis, we have indicated the most relevant TTR secondary structures in determining mutation impact. Our amyloidogenic propensity analysis of TTR missense substitutions has highlighted a similar pattern for wild-type and mutated TTR amino β acid sequences. However, some mutations present differences with respect to the general distribution. We have identified non-coding variants in *cis*-regulatory elements of the *TTR* gene, and our analysis on V122I-related haplotypes has indicated differences in non-coding regulatory variants, suggesting differences among V122I carriers. The analysis of methylation status indicated CpG sites that may affect *TTR* expression. Finally, our interactive network analysis revealed functional partners of TTR that may play a modifier role in the pathogenesis of TTR-related amyloidosis.
Discussion and conclusion: Our data provided new insights into the pathogenesis of TTR-related amyloidosis that, if they were to be confirmed through experimental investigations, could significantly improve our understanding of the disease.

Abbreviations: APOA1: apolipoprotein A-1; AβPP: amyloid beta precursor protein; cP_{del}: probability of missense variants to be deleterious; FGA: fibrinogen alpha chain; GERP: genomic evolutionary rate profiling; HGMD: human genome mutation database; HSPG2: heparan sulfate proteoglycan 2; is-rSNP: in silico-regulatory SNP detection; PASTA: prediction of amyloid structure aggregation; RBP4: retinol binding protein 4; STRING: search tool for the retrieval of interacting genes/proteins; TFBSs: transcription factor binding sites; TTR: transthyretin

Keywords

Amyloidogenic propensity, bioinformatics, *cis*-regulatory elements, interactive network, V122I mutation

History

Received 27 November 2013
Revised 29 January 2014
Accepted 24 February 2014
Published online 29 April 2014

Introduction

The recent explosion of high-throughput technologies has yielded vast quantity of data pertinent to the human genome both in terms of variability and regulation mechanisms [1]. Furthermore, international consortia are currently focusing their attention on important questions (e.g. the 1000 Genomes Project consortium on genetic variability and the ENCODE consortium on the functional elements of human genome). These consortia released some of their data, in order to provide useful tools to the scientific community [2,3]. These data can be used to perform computational studies that generate new hypotheses for directing experimental research [4–6]. Much information could be gleaned about monogenic disorders by performing computational studies on the subject [7]. Indeed, although the cause of the disease is often recognized, several questions remain about its pathogenic mechanisms and therapies.

Hereditary transthyretin (TTR)-related amyloidosis is a monogenic disease with high phenotypic heterogeneity among affected individuals [8,9]. Mutations in the *TTR* gene are known to cause the disease. However, incomplete penetrance, variance in age-of-onset, diverse disease progression, and variability in symptoms and affected organs are widely recognized in carriers of some *TTR* mutations, such as the V30M mutation [10]. Information about human genome organization strongly suggests that several factors jointly contribute to gene functionality. Therefore, the phenotypic presentation is likely modulated by other modifier factors.

Address for correspondence: Renato Polimanti, Department of Biology, University of Rome ''Tor Vergata'', Via della Ricerca Scientifica 1, 00133 Rome, Italy. Tel: +390672594310. Fax: +39062023500. E-mail: renatopolimanti@alice.it

Different authors have found evidences of modifier factors in the pathogenesis of TTR-related amyloidosis, but these factors have not been observed in all studies conducted on the subject. Some investigations have focused their attention on modifier genes. *TTR* mutations are required for the disease, but other genes may govern the clinical phenotype [11–14]. Mitochondrial DNA has also been investigated in relation to the phenotypic variation of TTR-related amyloidosis with interesting outcomes [15–17]. Finally, *cis*-regulatory elements in coding regions of the *TTR* gene have also been analyzed in order to understand the phenotypic variation of the disease [18–21].

In addition to the heterogeneity in its genotype–phenotype correlations, other important questions relating to the pathogenesis of TTR-related amyloidosis must be explored as well. Because of its invariably fatal outcome, the most relevant issue concerns the therapy. TTR amyloidogenesis is caused by tetramer dissociation, and, consequently, monomer misfolding and aggregation [22]. Accordingly, current studies on therapies have focused on small stabilizing molecules that inhibit TTR amyloid fibril formation and limit tetramer dissociation [23]. Therefore, new insights into amyloidogenic processes of TTR could be useful to improve therapy research.

Computational approaches based on genome-scale databases and softwares can be used to collect findings that may lead research in new directions. The aim of this study is to analyze database information about the *TTR* gene and provide new insight into the pathogenesis of TTR-related amyloidosis. Specifically, we investigated the following subjects: non-coding regions (to verify the presence of *cis*-regulatory elements that may modulate TTR function); coding regions (to identify encoded amino acid residue that may change the amyloidogenic propensity of monomers); the phylogenetic relationships of V122I-related haplotypes (to understand whether functional differences may be present among haplotypes carriers of the same amyloidogenic mutation); and, finally, the TTR-related interaction network (to identify the genes/proteins potentially involved, as modifier factors, in disease pathogenesis).

Methods

Information about amyloidogenic mutations was extracted from the ''Mutations in Hereditary Amyloidosis'' database (available at http://amyloidosismutations.com). Genotypic data and haplotype information were extracted from the phase 1 of 1000 Genomes Project (available at http://browser.1000genomes.org/index.html). Specifically, we downloaded a VCF (Variant Call Format) file of a 20 kb region, in which the *TTR* gene (7257 bp), 6731 bp upstream and 6015 bp downstream (Chromosome 18: 29 165 000–29 185 000) are included.

To identify regions that impact gene function, genomic conservation, based on a comparative analysis, was analyzed using these tools: VISTA [24], Genomic Evolutionary Rate Profiling [25] and Sequence Information-Based Decision Model [26]. The thresholds, proposed by the authors, were used in the analyses of this study.

We estimated the cP_{del}, combining the results of four tools that predict the functional impact of missense variants on the basis of amino acid sequence similarity. Specifically, we utilized polyphen-2 [27], Sorting Intolerant From Tolerant (SIFT) [28], Panther [29], and Predictor of human Deleterious Single Nucleotide Polymorphisms (PhD-SNP) [30]. Details are available in a previous study [31].

We investigated the amyloidogenic propensity of non-synonymous variants using the TANGO, Prediction of Amyloid STructure Aggregation (PASTA) and AGGRESCAN tools [32–34]. TANGO evaluates the beta-aggregation percentage [33]. PASTA permits to estimate the single-residue amyloidogenic propensity [$h(K)$] [35]. AGGRESCAN analyzes amino acid aggregation-propensity value [36]. We performed amyloidogenic propensity analyses for all amino acid sequences with *TTR* missense variants, including the wild-type sequence.

We used the is-rSNP (*in silico*-regulatory SNP detection) algorithm to predict the transcription factor binding sites (TFBSs) in the *TTR* sequence [37]. We set the significance level for the is-rSNP analysis at 0.05 and JASPAR database was used as reference. In order to increase the sensibility of our analysis, we also considered ENCODE data about TFBS presence in *TTR* gene and its surroundings (available at http://encodeproject.org/ENCODE/).

We reconstructed the phylogenetic relationships of the *TTR* haplotypes using a Median Joining Network [38]. Variation sites were weighted in accordance to their relative nucleotide conservation status.

To identify genes/proteins potentially associated with disease pathogenesis, we analyzed the TTR-related interaction network using STRING v.9.05 (Search Tool for the Retrieval of Interacting Genes/Proteins) [39]. During our STRING analysis, we excluded textmining from the active prediction methods and only considered outcomes with the highest confidence (STRING score >0.900). Therefore, our string analysis is based only on co-expression, experimental and database evidences, and it is not biased by textmining effects.

Results

First, we analyzed the genetic variation of the *TTR* coding regions. In particular, we focused our attention on missense substitutions. While the database of 1000 Genomes Project and the ''Mutations in Hereditary Amyloidosis'' databases into consideration, we collected information on 130 missense substitutions (Supplemental Table 1). For each variant, we reported the amino acid position in the TTR precursor and TTR monomer, the source database, the amyloidogenic status (based on information found in the ''Mutations in Hereditary Amyloidosis'' database and The Human Gene Mutation Database (HGMD; available at http://www.hgmd.cf.ac.uk/), the location in TTR secondary structures (UniProt data were considered; available at http://www.uniprot.org), and the estimated probability of missense variants to be deleterious (cP_{del}). To estimate the cPdel, we combined the outcomes of four tools that predict the functional impact of missense variants on the basis of amino acid sequence similarity (see Methods section). Considering the amyloidogenic status, we observed that amyloidogenic variants showed a higher cP_{del} than non-amyloidogenic variants (0.740 versus 0.534, respectively). We determined that the missense substitutions

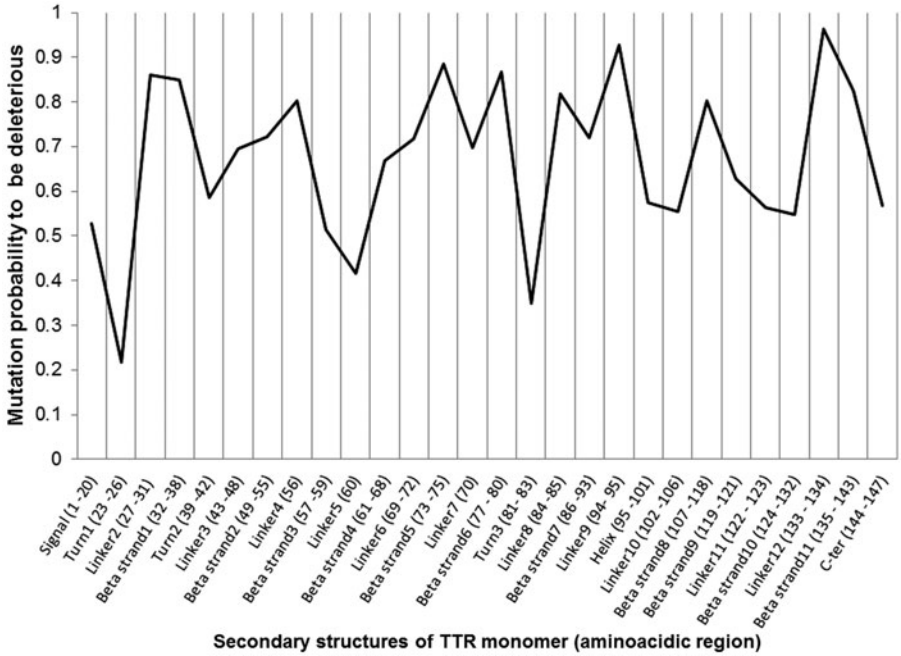

Figure 1. cP_{del} values in the secondary structures of the TTR monomer. The precursor cleaved signal is also reported in this analysis. The reported aminoacidic regions are referred to the aminoacidic sequence of TTR precursor.

with unknown amyloidogenic status have cP_{del} (0.516) comparable to the non-amyloidogenic variants (0.534). Subsequently, we analyzed the distribution of the cP_{del} average in the secondary structure of the TTR monomer (Figure 1). We also reported the precursor cleaved signal in our analysis. The top-3 cP_{del} values were observed in Linker12 ($cP_{del} = 0.963$; amino acid region from residue 113 to residue 114), Linker9 ($cP_{del} = 0.927$; amino acid region from residue 74 to residue 75) and Beta strand5 ($cP_{del} = 0.885$, amino acid region from residue 53 to residue 55). The three lowest cP_{del} values were observed in Turn1 ($cP_{del} = 0.217$; amino acid region from residue 3 to residue 6), Turn3 ($cP_{del} = 0.350$, amino acid region from residue 61 to residue 63) and Linker5 ($cP_{del} = 0.416$, amino acid residue 40).

To understand the effects of missense substitution on the amyloidogenic propensity of the TTR monomer, we analyzed all amino acid sequences derived from missense substitutions using the TANGO, PASTA and AGGRESCAN algorithms. In this analysis, we also considered the wild-type sequence and excluded the variants located in cleaved signal. In Figure 2, we report the average, maximum and minimum values estimated in each amino acid residue observed in the 129 tested amino acid sequences. All softwares indicated that a similar pattern of amyloidogenic propensity was present in all tested sequences, including the wild-type one. In TANGO analysis, we observed six regions with high amyloidogenic propensity: from residues (1) 12 to 20, (2) 25 to 33, (3) 68 to 73, (4) 93 to 98, (5) 105 to 111 and (6) 114 to 123 (Figure 2A and Supplemental Figure 1A). In PASTA analysis, we observed five amino acid regions with high amyloidogenic propensity: (1) a large region from residues 12 to 37 with a strong core between residue 26 and 33, (2) from residue 63 to 73, (3) from residues 90 to 98, (4) from residues 105 to 111 and (5) from residue 116 to 124 (Figure 2B and Supplemental Figure 1B). In AGGRESCAN, we observed six amino acid regions with high amyloidogenic propensity: from residue (1) 12 to 19, (2) 25 to 34, (3) 65 to 72, (4) 79 to 85, (5) 92 to 96 and (6) 105 to 123 (Figure 2C and Supplemental Figure 1C). Although there is a similar amyloidogenic pattern among TTR amino acid sequences, certain missense variants seem to show diverse amyloidogenic propensity with respect to the general pattern. In TANGO analysis, some sequences showed very high value of amyloidogenic propensity with respect to the general distribution: T96I (from residues 93 to 97), A109V (107 to 111) and P113L (107 to 122) (Supplemental Figure 1A). In PASTA analysis, sequences with missense variants within the amino acid region between residues 28 and 32 showed altered amyloidogenic propensity with respect to the other TTR amino acid sequences (Supplemental Figure 1B). AGGRESCAN indicated that missense variants altered the amyloidogenic propensity of the region, in which the genetic variation is located (Supplemental Figure 1C).

After analyzing the coding regions, we focused our attention on non-coding variants in order to collect information on the presence of *cis*-regulatory elements in the *TTR* gene and its surrounding regions. Specifically, we analyzed the genotypic data and haplotypic information of Phase 1 of the 1000 Genomes project. In Supplemental Table 2, we report the 271 considered genetic variants. To evaluate the impact of these variants on *TTR* regulation, we performed different *in silico* analyses (see Methods section). We also considered the data collected through the ENCODE project on methylation status and the presence of TFBSs (Supplemental Figure 2). Regarding the TFBS analyses, the data of ENCODE project and is-rSNP algorithm gave different results. This is likely due to the approaches used to perform the TFBS analyses. Since both the approaches are validated, we considered ENCODE and is-rSNP outcomes, in order to increase the sensibility of our analysis. Regarding methylation status, ENCODE project indicated the presence of different CpG sites (Supplemental Figure 2). In Supplemental Table 3, we report the positive outcomes of

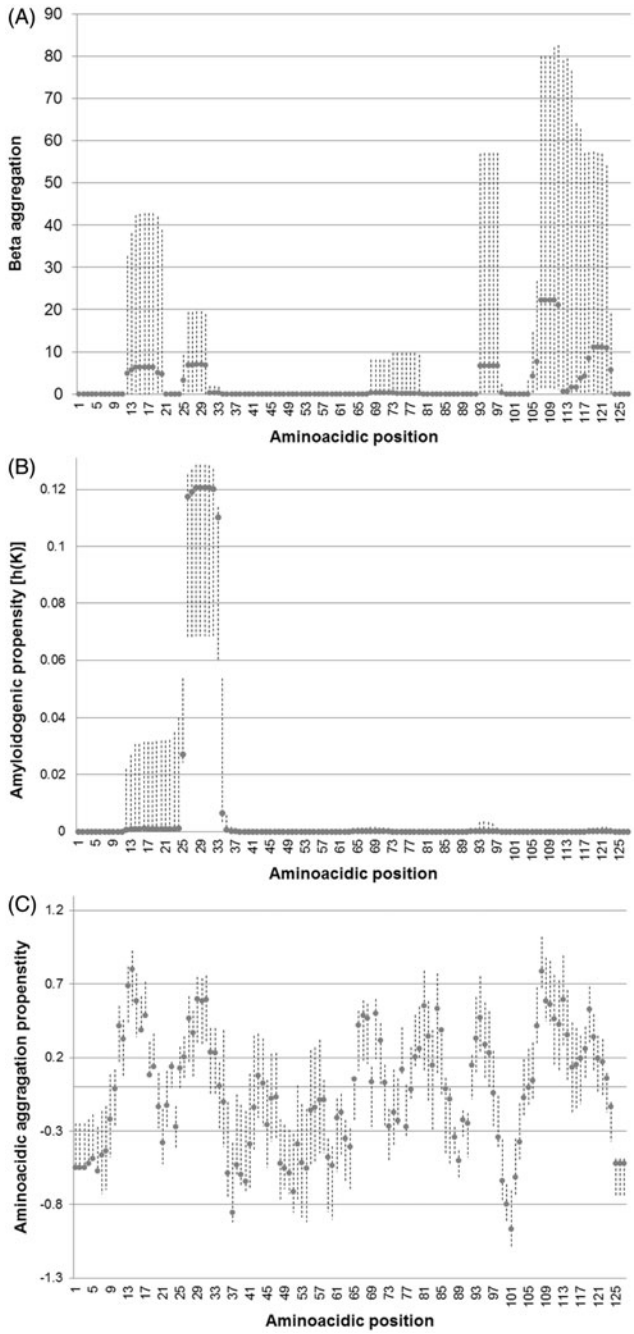

Figure 2. Distribution of estimated amyloidogenic propensity (average: grey dot; maximum-minimum range: dotted line) in each amino acid position performed by TANGO (A), PASTA (B) and AGGRESCAN (C). The reported aminoacidic positions are referred to the aminoacidic sequence of TTR monomer.

these analyses: 63 of 271 genetic variants may have a functional impact on the *TTR* gene, and 59 of these 63 variants are located in non-coding regions. While analyzing the coding variants with genotypic information in Phase 1 of the 1000 Genomes Project, we observed three non-amyloidogenic mutations (i.e. G6S, R104H and T119M), two synonymous substitutions (i.e. S110S, and T119T) and the amyloidogenic mutation (V122I). To analyze the relationship among V122I-related haplotypes, we performed a phylogenetic analysis using the Median Joining Network. According to data gleaned from the 1000 Genomes Project, there are 308 *TTR* haplotypes, and five of these haplotypes are carriers of the V122I amyloidogenic mutation: h65, h214, h216, h234 and h236. Phylogenetic analysis revealed that h65 may have originated independently from the other V122I-related haplotypes (Figure 3). In accordance with this phylogenetic outcome, we observed that the h65 haplotype showed a different variation pattern in loci potentially associated with gene regulation (Figure 4).

Finally, our analysis of TTR-related interactive network indicated the presence of five predicted functional partners: RBP4 (Retinol Binding Protein 4; STRING score = 0.994), APOA1 (Apolipoprotein A-1; STRING score = 0.982), HSPG2 (Heparan Sulfate Proteoglycan 2; STRING score = 0.959), AβPP (Amyloid Beta Precursor Protein; STRING score = 0.959) and FGA (Fibrinogen Alpha Chain; STRING score = 0.917) (Supplemental Figure 3).

Discussion

This study has provided an *in silico* analysis of database information pertinent to coding and non-coding regions, as well as an interactive network of *TTR*. It is our hope that this analysis will enable us to obtain new testable hypotheses for further experimental investigations.

Our data about coding variants are mainly focused on missense substitutions. As we expected, amyloidogenic mutations were more likely to be deleterious than non-amyloidogenic variants. Since a number of studies investigated TTR protein using crystallographic experiments, we decided to use an independent approach based on amino acid sequence similarity. This analysis permitted us to estimate the probability of a *TTR* mutation to be deleterious for human health, considering both the impact on protein function and protein structure. The variants with an unknown amyloidogenic status showed a cP_{del} similar to the non-amyloidogenic ones. This is likely due to the fact that the participants of the 1000 Genomes Project are apparently healthy individuals. However, some *TTR* variants with an unknown status reached high cP_{del}: S85P (0.781), T96I (0.805) and P113L (0.979). Since the participants of 1000 Genomes Project are mostly younger than 60, these variants may be associated with late-onset phenotype, such as other *TTR* mutations (e.g. F64L) [40]. Considering the secondary structures of the TTR monomer, we observed the top-3 cP_{del} averages in Linker12, Linker9 and Beta strand5. Among the variants located in these protein regions, we observed the presence of 12 amyloidogenic mutations (86%), 1 non-amyloidogenic variant (7%) and 1 locus with an unknown status (7%). Conversely, in the lowest 3 cPdel averages (i.e. Turn1, Turn3 and Linker5), we identified three non-amyloidogenic mutations (50%) and three variants with an unknown status (50%). To date, about 200 TTR-related X-ray crystal structures have been solved, providing relevant information about the relationship between structure and function [41]. Our data about TTR secondary structures were extracted from an algorithm based on amino acid sequence similarity. In other words, this information was obtained independently from the X-ray analyses. Therefore, we can observe some non-amyloidogenic variants with cPdel higher than amyloidogenic mutation average (i.e. D74H, G101S and T119M). This is likely due to nearness to other relevant amino acid positions.

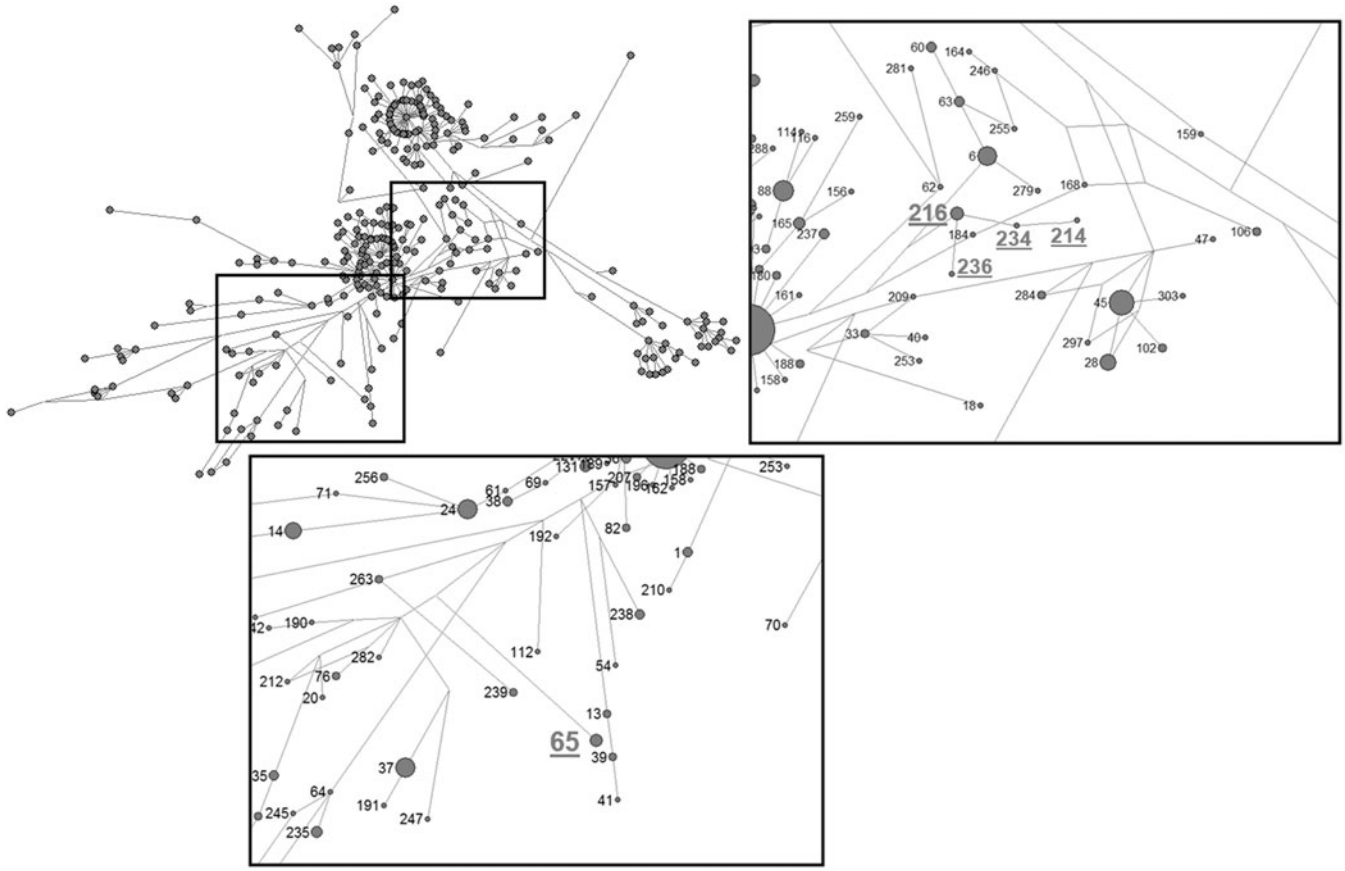

Figure 3. Median joining network of *TTR* haplotypes observed in the participants of the 1000 Genomes Project. The regions with V122I-related haplotypes (bold and underlined) are zoomed. Each circle represents a different haplotype, and size is proportional to frequency in the zoomed regions.

Pos_Chr18	29167736	29167905	29169115	29169213	29169215	29169825	29169933	29170483	29170634	29170730	29172476	29173680	29173795	29174331	29175237	29176460	29176873	29176971	29178379	**29178618**	29179040	29182352	29182361
ID	rs181315042	rs1667244	rs80322202	rs67292379	18:29169215	rs3764479	rs13381522	rs3764478	18:29170634	rs58616646	rs723744	rs1080093	rs1080094	rs141035204	rs75517067	rs3764476	rs7235277	rs3794884	rs1667251	**rs76992529**	rs1791228	rs1791229	rs1667252
Location	Upstream	Upstream	Upstream	Upstream	Upstream	NearGene-5	NearGene-5	NearGene-5	NearGene-5	NearGene-5	intron	intron	intron	intron	intron	intron	intron	intron	intron	**V122I**	NearGene-3	Downstream	Downstream
Function analysis	TFBS (ENCODE)	TFBS (is-rSNP)	none	none	none	none	none	none	none	none	Conservate (VISTA)	none	none	none	Conservate (GERP)	TFBS (is-rSNP)	none	Conservate (GERP)	none	**ATTR mutation**	Conservate (GERP, SiNBaD)	TFBS (is-rSNP)	TFBS (is-rSNP)
Ancestral Allele	C	A	C	G	A	A	C	G	CAT	C	G	C	A	A	G	C	G	T	T	**G**	C	T	T
h65	A	G	T	GA	AG	A	T	G	CAT	T	G	G	G	T	A	C	G	T	G	**A**	T	G	C
h214	C	G	C	GA	A	G	C	G	C	C	T	G	G	A	G	A	C	G	G	**A**	T	G	T
h216	C	G	C	GA	AG	G	C	G	CAT	C	T	G	G	A	G	A	C	G	G	**A**	T	G	T
h234	C	G	C	GA	AG	G	C	G	C	C	T	G	G	A	G	A	C	G	G	**A**	T	G	T
h236	C	G	C	GA	AG	G	C	T	CAT	C	T	G	G	A	G	A	C	G	G	**A**	T	G	T

Figure 4. Nucleotide changes among V122I-related haplotypes. V122I mutation is shown in black cells. Variants potentially associated with functional impact are shown in dark grey cells. Variants with none functional involvements are shown in light grey cells. TFBS: Transcription factor binding sites.

Indeed, D74H is located in one of the top-3 cPdel secondary structures, Linker9. Our approach based on an amino acid sequence similarity is independent from the X-ray analysis and it may provide another way to analyze *TTR* coding regions.

To analyze amyloidogenic propensity of TTR monomer, we used AGGRESCAN, PASTA and TANGO algorithms. To the best of our knowledge, no other studies have performed this systematic investigation. The three softwares showed overlapping results: similar amino acid regions with amyloidogenic propensity have been identified. Furthermore, this analysis indicated that wild-type sequence and the sequences carriers of missense substitutions (both amyloidogenic and non-amyloidogenic) have a similar amyloidogenic propensity. This outcome agrees with the current literature data: the disease-causing mutations do not increase the amyloidogenicity of TTR monomer, but their harmful effect is to decrease the stability of the TTR tetramer prompting an increase in monomer misfolding [22,42]. However, slight differences in some of the disease-causing mutations are present, suggesting that the amyloidogenic processes may be different in certain *TTR* mutations. Further studies may focus their attention on this topic, in order to determine whether these amyloidogenic differences may alter the efficiency of stabilizing molecules that are currently used in therapies of TTR-related amyloidosis.

We investigated coding regions of *TTR* gene with two different points of view: cPdel estimated the presence of amino acid regions that may have an high probability to be deleterious for human health; aggregation predictors evaluated the effects of the missense substitution on amyloidogenic propensity in order to understand if *TTR* mutations not only affect TTR tetramer instability but also amyloidogenic processes. Considering both the findings of these analyses, we can affirm that research about anti-amyloidogenic strategy should focus on the TTR tetrameric structure. All TTR monomers, including wild-type form, tend to form amyloid deposits (i.e. slight differences in amyloidogenic propensity), whereas amyloidogenic mutations alter tetrameric structure in different qualitative and quantitative ways (i.e. great variability in cPdel outcomes).

Our analysis of non-coding regions of the *TTR* gene suggests the presence of 59 non-coding variants potentially associated with gene function. This supports the hypothesis that the *cis*-regulatory elements of the *TTR* gene modulate the effect of mutations on the disease phenotype. To the best of our knowledge, only four previous studies have investigated the role of non-coding variants in disease pathogenesis. The works of Olsson et al. [19] and Norgren et al. [20] explored the effect of an SNP in *TTR* 3′ UTR in Swedish patients carriers of V30M mutation, indicating that this genetic variant did not affect *TTR* mRNA expression. A previous study on V30M carriers analyzed the coding and non-coding regions of *TTR* gene and its surroundings, highlighting that *TTR* downstream region may modulate the onset of the symptoms [18]. Our previous study on human variation of *TTR* gene suggests that non-coding variants have a role in determining phenotypic presentation in African patients [21]. This analysis in which we highlighted the presence of multiple functional non-coding variants furnishes the basis to develop further studies on this topic. Moreover, the identification of V122I amyloidogenic mutation in the participant of 1000 Genomes Project permits to explore the presence of functional non-coding variants in the haplotypes carriers of V122I mutation. Our analysis seems to agree with this hypothesis. Two haplotype clusters are present: one cluster composed only of the h65 haplotype, and another cluster composed of the h214, h216, h234 and h236 haplotypes. Our phylogenetic analysis is in accordance with the geographical distribution of these haplotype clusters. All haplotypes are present in populations with African ancestry, in agreement with the known distribution of the V122I mutation, which is the most recurrent disease-causing mutation in African individuals [10]. Haplotype h65 is present in Kenya and African–American populations, whereas the haplotypes of the second cluster are present in Nigeria and African–American populations. The presence of both clusters in African–American populations is likely due to the geographical diversity of their common ancestors [43], whereas the distribution in African populations indicated a separation between eastern (h214, h216, h234 and h236) and western (h65) Africa. Our *in silico* analyses indicated that functional differences may be present among V122I-related haplotypes. Specifically, we observed three h65-specific functional variants (i.e. rs181315042, rs75517067 and rs1667252), while the other V122I-related haplotypes showed three other specific variants (i.e. rs723744, rs3764476 and rs3794884). Among these functional variants, three were located in TFBSs (i.e. rs181315042, rs1667252 and rs3764476), and the others showed significant nucleotide conservation (i.e. rs75517067, rs723744 and rs3794884). Considering variants in TFBSs, ENCODE data indicated that rs181315042 was located in binding sites of different transcription factors (i.e. TCF4, SP1, FOXA1, HDAC2, HNF4A, FOXA2 and p300). The is-rSNP algorithm highlighted that the consensus motifs of TFBSs are located in the regions of rs1667252 (i.e. Dbx2, Pou2f1, Alx1_2, Vax2, Hoxd1, Phox2a, Hoxc6, Lbx2, Pax6, Vax1, Pou3f2, Nkx6-3, Lhx5, Nkx6-1_1, Lhx3, en, Shox2, En1, Gsx2, Hoxb7, Emx2, SMP1, Hoxc4, Hoxa3, Pou3f1, Dlx1, Msx1, Barx2, H2.0, Awh, CG32105, Hoxa5, Pou3f4, Hoxc8, Nkx1-1, Hoxa1, Hoxb3, Lim1, YOX1, Uncx, Pou1f1, Hoxb4, Lhx1, En2, Prop1, and E5) and rs3764476 (LM184, Ascl2_1, Tcfap2c_2, and GCANCTGNY).

Patients with the V122I mutation are mainly characterized by cardiomyopathy, but symptoms consistent with neuropathy can also be present [10,44,45]. Our data about non-coding variants in the V122I-related haplotype suggest that the differences among V122I patients may be explained by *cis*-regulatory elements in the *TTR* gene. Furthermore, our hypothesis may also explain the concordance between the epidemiological differences in V30M patients of Swedish and Portuguese descent and the diverse origins of the V30M mutation in Sweden and Portugal [21,46]. Indeed, two V30M-related haplotypes with different *cis*-regulatory variants are likely present in Swedish and Portuguese patients, respectively. Therefore, the replication of our *in silico* study about non-coding regions on V30M carriers, also coupled with laboratory analysis, may have relevant impact in the understanding of the genotype–phenotype correlation present in V30M carriers.

Besides genetic variation in coding and non-coding regions, we used ENCODE project also to obtain information about methylation in *TTR* gene and its surroundings. Different CpG sites are present, but no genetic variants alter these sites. However, methylation status likely affects *TTR* expression. In particular, two CpG sites upstream transcription start site (i.e. cg18038361 and cg04008843) may play a relevant role in gene expression. Indeed, a recent study about the relationship between epigenomics, genomics and transcriptomics indicated the role of upstream CpG sites in determining significant expression to methylation associations [47]. To the best of our knowledge, no studies have explored the role of epigenetics in the TTR-related amyloidosis. The data about CpG sites support further epigenetic studies.

Finally, our analysis of the interactive network indicated five predicted functional partners of TTR. These proteins interact with TTR in physiological condition. The disease-causing mutations trigger the onset of the disease, and alterations in expression or activity of these predicted partners may participate in determining the phenotype of TTR-related amyloidosis (i.e. penetrance, age-of-onset, symptoms). Although some authors have provided some insight about this hypothesis, the role of modifier genes/protein remains still uncovered, Our *in silico* investigation about TTR interactions revealed five potential interactions: RBP4, APOA1, HSPG2, AβPP and FGA. Regarding RBP4, a previous study hypothesized that *RBP4* genetic variants modulate the disease phenotype in V30M patients [11]. Other studies confirmed the strong TTR–RBP4 interaction [48,49]. Moreover, Afolabi et al. [50] observed that low protein intake causes a significant reduction of HDL-APOA1 coupled with declined synthesis rate of TTR and RBP. This finding may underline diet-related interactions in hepatic proteins that may have an impact on disease phenotype. Previous *in vitro* study indicated that HSPG2 can bind the TTR protein [51]. This confirms our prediction and, moreover, increased levels of HSPG2 may be related with stabilization enhancement of TTR tetramer. A recent study indicated increase that *TTR* is an AβPP-dependent gene, suggesting that AβPP-mediated signaling pathway may play an important role in maintaining TTR levels [52]. This may have a relevant impact in disease presentation: alteration of AβPP activity may modulate the amyloidogenic process caused by *TTR* mutations. As for APOA1, the TTR-FGA interaction seems to be due by dietary alterations: animal experiments with MALDI/TOF mass spectrometry indicated that diabetic rats showed altered proteomic profile, in which FGA and TTR are also included [53]. Our bioinformatic analysis and the data available in literature suggest that these proteins may be involved in the pathogenesis of TTR-related amyloidosis, along with the pathways indicated. Furthermore, previous studies have indicated the presence of relevant pathways that are involved in the onset and progression of TTR-related amyloidosis, such as the endoplasmatic reticulum-associated degradation and the endoplasmatic reticulum-assisted folding [54]. Alteration of expression/activity of modifier genes/proteins/pathways may drive the clinical presentation of the disease in the affected patients. To date, few efforts were done to explore this hypothesis, but our data and literature strongly suggest its potential relevance. In addition, the data about APOA1 and FGA advise that diet may also play a significant role in TTR interactoma.

Conclusion

Our *in silico* study shed further light on topics pertaining to TTR-related amyloidosis, such as the link between TTR secondary structures and mutation impact, the relationship between mutation and amyloidogenic propensity, the presence of *cis*-regulatory elements in the *TTR* gene and its surrounding regions and the identification of modifier genes/proteins. However, some of the presented data (i.e. information about non-coding variants and epigenetic modifications) are only bioinformatics predictions, not yet confirmed by experimental analysis. This limitation is due to the explorative bioinformatic approach used to test multiple hypotheses in order to identify the most reliable ones to be confirmed with laboratory investigation. Therefore, further molecular analyses are necessary to deepen our computational insights, verifying their impact on disease pathogenesis.

Acknowledgements

The authors are grateful to the research groups of 1000 Genomes Project for their essential contribution.

Declaration of interest

The authors report no conflicts of interest. This study was supported by AFaR (Associazione Fatebenefratelli per la Ricerca).

References

1. Berger B, Peng J, Singh M. Computational solutions for omics data. Nat Rev Genet 2013;14:333–6.
2. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.
3. Encode Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 2004;306:636–40.
4. Polimanti R, Piacentini S, Fuciarelli M. HapMap-based study of human soluble glutathione S-transferase enzymes: the role of natural selection in shaping the single nucleotide polymorphism diversity of xenobiotic-metabolizing genes. Pharmacogenet Genomics 2011;21:665–72.
5. Polimanti R, Piacentini S, Manfellotto D, Fuciarelli M. Human genetic variation of CYP450 superfamily: analysis of functional diversity in worldwide populations. Pharmacogenomics 2012;13:1951–60.
6. Polimanti R, Fuciarelli M, Destro-Bisol G, Battaggia C. Functional diversity of glutathione peroxidase gene family among human populations: implications for genetic predisposition to disease and drug response. Pharmacogenomics 2013;14:1037–45.
7. Zhao Y, Clark WT, Mort M, Cooper DN, Radivojac P, Mooney SD. Prediction of functional regulatory SNPs in monogenic and complex disease. Hum Mutat 2011;32:1183–90.
8. Benson MD, Kincaid JC. The molecular biology and clinical features of amyloid neuropathy. Muscle Nerve 2007;36:411–23.
9. Hornsten R, Pennlert J, Wiklund U, Lindqvist P, Jensen SM, Suhr OB. Heart complications in familial transthyretin amyloidosis: impact of age and gender. Amyloid 2010;17:63–8.
10. Coelho T, Maurer MS, Suhr OB. THAOS – The Transthyretin Amyloidosis Outcomes Survey: initial report on clinical manifestations in patients with hereditary and wild-type transthyretin amyloidosis. Curr Med Res Opin 2013;29:63–76.

11. Soares ML, Coelho T, Sousa A, Batalov S, Conceicao I, Sales-Luis ML, Ritchie MD, et al. Susceptibility and modifier genes in Portuguese transthyretin V30M amyloid polyneuropathy: complexity in a single-gene disease. Hum Mol Genet 2005;14: 543–53.
12. Saporta MA, Plante-Bordeneuve V, Misrahi M, Cruz MW. Discordant expression of familial amyloid polyneuropathy in monozygotic Brazilian twins. Amyloid 2009;16:38–41.
13. Dardiotis E, Koutsou P, Zamba-Papanicolaou E, Vonta I, Hadjivassiliou M, Hadjigeorgiou G, Cariolou M, et al. Complement C1Q polymorphisms modulate onset in familial amyloidotic polyneuropathy TTR Val30Met. J Neurol Sci 2009; 284:158–62.
14. Gomes RA, Franco C, Da Costa G, Planchon S, Renaut J, Ribeiro RM, Pinto F, et al. The proteome response to amyloid protein expression *in vivo*. PLoS One 2012;7:e50123.
15. Olsson M, Hellman U, Plante-Bordeneuve V, Jonasson J, Lang K, Suhr OB. Mitochondrial haplogroup is associated with the phenotype of familial amyloidosis with polyneuropathy in Swedish and French patients. Clin Genet 2009;75:163–8.
16. Bonaiti B, Olsson M, Hellman U, Suhr O, Bonaiti-Pellie C, Plante-Bordeneuve V. TTR familial amyloid polyneuropathy: does a mitochondrial polymorphism entirely explain the parent-of-origin difference in penetrance? Eur J Hum Genet 2010;18: 948–52.
17. Liu JY, Jiang XM, Zhang M, Guo YJ. Analysis of mitochondrial haplogroups associated with TTR Val30Ala familial amyloidotic polyneuropathy in Chinese patients. Int J Neurosci 2012;122:716–8.
18. Soares ML, Coelho T, Sousa A, Holmgren G, Saraiva MJ, Kastner DL, Buxbaum JN. Haplotypes and DNA sequence variation within and surrounding the transthyretin gene: genotype–phenotype correlations in familial amyloid polyneuropathy (V30M) in Portugal and Sweden. Eur J Hum Genet 2004;12:225–37.
19. Olsson M, Norgren N, Obayashi K, Plante-Bordeneuve V, Suhr OB, Cederquist K, Jonasson J. A possible role for miRNA silencing in disease phenotype variation in Swedish transthyretin V30M carriers. BMC Med Genet 2010;11:130.
20. Norgren N, Hellman U, Ericzon BG, Olsson M, Suhr OB. Allele specific expression of the transthyretin gene in Swedish patients with hereditary transthyretin amyloidosis (ATTR V30M) is similar between the two alleles. PLoS One 2012;7:e49981.
21. Polimanti R, Di Girolamo M, Manfellotto D, Fuciarelli M. Functional variation of the TTR gene among human populations and its correlation with amyloidosis phenotypes. Amyloid 2013: 20: 256–62.
22. Hammarstrom P, Jiang X, Hurshman AR, Powers ET, Kelly JW. Sequence-dependent denaturation energetics: a major determinant in amyloid disease diversity. Proc Natl Acad Sci USA 2002;99 (Suppl 4):16427–32.
23. Bulawa CE, Connelly S, Devit M, Wang L, Weigel C, Fleming JA, Packman J, et al. Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. Proc Natl Acad Sci U S A 2012;109:9629–34.
24. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res 2004;32:W273–9.
25. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 2010;6:e1001025.
26. Lehmann KV, Chen T. Exploring functional variant discovery in non-coding regions with SInBaD. Nucleic Acids Res 2013; 41:e7.
27. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7: 248–9.
28. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073–81.
29. Thomas PD, Kejariwal A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. Proc Natl Acad Sci USA 2004;101:15398–403.
30. Brunham LR, Singaraja RR, Pape TD, Kejariwal A, Thomas PD, Hayden MR. Accurate prediction of the functional significance of single nucleotide polymorphisms and mutations in the ABCA1 gene. PLoS Genet 2005;1:e83.
31. Squitti R, Siotto M, Bucossi S, Polimanti R. In silico investigation of the ATP7B gene: insights from functional prediction of non-synonymous substitution to protein structure. Biometals 2014;27: 53–64.
32. Trovato A, Seno F, Tosatto SC. The PASTA server for protein aggregation prediction. Protein Eng Des Sel 2007;20:521–3.
33. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 2004;22: 1302–6.
34. Conchillo-Sole O, de Groot NS, Aviles FX, Vendrell J, Daura X, Ventura S. AGGRESCAN: a server for the prediction and evaluation of ''hot spots'' of aggregation in polypeptides. BMC Bioinform 2007;8:65.
35. Trovato A, Chiti F, Maritan A, Seno F. Insight into the structure of amyloid fibrils from the analysis of globular proteins. PLoS Comput Biol 2006;2:e170.
36. Sanchez de Groot N, Pallares I, Aviles FX, Vendrell J, Ventura S. Prediction of ''hot spots'' of aggregation in disease-linked polypeptides. BMC Struct Biol 2005;5:18.
37. Macintyre G, Bailey J, Haviv I, Kowalczyk A. is-rSNP: a novel technique for in silico regulatory SNP detection. Bioinformatics 2010;26:i524–30.
38. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 1999;16: 37–48.
39. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, et al. STRING v9.1: protein–protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013;41:D808–15.
40. Russo M, Mazzeo A, Stancanelli C, Di Leo R, Gentile L, Di Bella G, Minutoli F, et al. Transthyretin-related familial amyloidotic polyneuropathy: description of a cohort of patients with Leu64 mutation and late onset. J Peripher Nerv Syst 2012;17: 385–90.
41. Palaninathan SK. Nearly 200 X-ray crystal structures of transthyretin: what do they tell us about this protein and the design of drugs for TTR amyloidoses? Curr Med Chem 2012;19:2324–42.
42. Cendron L, Trovato A, Seno F, Folli C, Alfieri B, Zanotti G, Berni R. Amyloidogenic potential of transthyretin variants: insights from structural and computational analyses. J Biol Chem 2009;284: 25832–41.
43. Ely B, Wilson JL, Jackson F, Jackson BA. African–American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. BMC Biol 2006;4:34.
44. Buxbaum J, Alexander A, Koziol J, Tagoe C, Fox E, Kitzman D. Significance of the amyloidogenic transthyretin Val 122 Ile allele in African Americans in the Arteriosclerosis Risk in Communities (ARIC) and Cardiovascular Health (CHS) Studies. Am Heart J 2010;159:864–70.
45. Ruberg FL, Maurer MS, Judge DP, Zeldenrust S, Skinner M, Kim AY, Falk RH, et al. Prospective evaluation of the morbidity and mortality of wild-type and V122I mutant transthyretin amyloid cardiomyopathy: the Transthyretin Amyloidosis Cardiac Study (TRACS). Am Heart J 2012;164:222-8 e221.
46. Zaros C, Genin E, Hellman U, Saporta MA, Languille L, Wadington-Cruz M, Suhr O, et al. On the origin of the transthyretin Val30Met familial amyloid polyneuropathy. Ann Hum Genet 2008; 72:478–84.
47. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, Bryois J, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife 2013;2:e00523.
48. Bobbert T, Raila J, Schwarz F, Mai K, Henze A, Pfeiffer AF, Schweigert FJ, et al. Relation between retinol, retinol-binding protein 4, transthyretin and carotid intima media thickness. Atherosclerosis 2010;213:549–51.
49. Ackermann EJ, Guo S, Booten S, Alvarado L, Benson M, Hughes S, Monia BP. Clinical development of an antisense therapy for the treatment of transthyretin-associated polyneuropathy. Amyloid 2012;19 Suppl 1:43–4.

50. Afolabi PR, Jahoor F, Gibson NR, Jackson AA. Response of hepatic proteins to the lowering of habitual dietary protein to the recommended safe level of intake. Am J Physiol Endocrinol Metab 2004;287:E327–30.
51. Smeland S, Kolset SO, Lyon M, Norum KR, Blomhoff R. Binding of perlecan to transthyretin *in vitro*. Biochem J 1997;326:829–36.
52. Li H, Wang B, Wang Z, Guo Q, Tabuchi K, Hammer RE, Südhof TC, et al. Soluble amyloid precursor protein (APP) regulates transthyretin and Klotho gene expression without rescuing the essential function of APP. Proc Natl Acad Sci USA 2010;107:17362–7.
53. Choi SA, Yun JW, Park HS, Choi JW. Hypoglycemic dipeptide cyclo (His-Pro) significantly altered plasma proteome in streptozocin-induced diabetic rats and genetically-diabetic (ob/ob) mice. Mol Biol Rep 2013;40:1753–65.
54. Sekijima Y, Wiseman RL, Matteson J, Hammarstrom P, Miller SR, Sawkar AR, Balch WE, et al. The biological and chemical basis for tissue-selective amyloid disease. Cell 2005;121:73–85.