Unravelling modulators of age-at-onset variability in FAP ATTRV30M

Andreia Dias
Mestrado em Bioquímica
Departamento de Química e Bioquímica
2016

Orientadora
Dra Carolina Lemos, PhD, Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto

Coorientadora
Diana Santos, MSc, Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.
O Presidente do Júri,

Porto, _____/_____/_____

ACKNOWLEDGMENTS

A realização desta dissertação de mestrado só foi possível devido à contribuição e colaboração de várias pessoas e instituições. Assim, não posso deixar de revelar os meus sinceros agradecimentos à Doutora Carolina Lemos, minha orientadora, que me permitiu fazer parte deste projeto tão interessante e promissor. Agradecer-lhe por todo o apoio sempre que necessitei e me permitiu evoluir ao transmitir-me todo o conhecimento inerente à concretização deste projeto. Muito obrigada por todo o tempo que me dedicou, e levo comigo a certeza que não poderia ter sido melhor orientada.

Não posso também deixar de agradecer à minha co-orientadora, Mestre Diana Santos, que esteve sempre do meu lado e foi sempre incansável ao tentar solucionar qualquer adversidade que surgisse. Foi sem dúvida um dos meus pilares na realização desta dissertação, sem a qual este trabalho não teria sido possível.

Agradecer também ao Miguel Alves-Ferreira por toda a ajuda e sugestões prestadas que permitiram enriquecer esta dissertação.

Agradeço também ao Professor Jorge Sequeiros que permitiu a realização da minha dissertação na UnIGENe, à Doutora Alda Sousa que me auxiliou com o seu conhecimento científico e à Doutora Isabel Alonso por todo o apoio prestado.

À Doutora Teresa Coelho e à restante equipa da Unidade Corino de Andrade do Centro Hospitalar do Porto, onde são seguidos todos os doentes que participaram neste estudo.

Quero também expressar os meus agradecimentos a todos os membros do grupo UnIGENe, nomeadamente à Conceição, Mariana, Sara, Marlene, e à Patrícia e Cheila do Centro Genética Preditiva e Preventiva que me proporcionaram dia após dia momentos de descontração e convívio, dos quais levo uma boa amizade.

Agradeço ao Grupo Regeneração Nervosa e Modelação em Doenças Neurodegenerativas, em particular à Doutora Márcia Liz, por todo o *feedback* prestado em relação a trabalhos anteriores em *Drosophila*, que serviram de impulso para a análise do gene *RAC1* em doentes com PAF ATTRV30M.

Aos meus amigos, Sofia, Alexandra, Ricardo Abreu, Carla e Rita que estiveram sempre comigo quando mais precisei, apesar da distância por vezes ter sido uma entrave e me deram sempre um ombro amigo nos momentos de mais stress.

Às minhas companheiras e amigas de curso, Francisca e Inês, pelos momentos que passamos ao longo destes dois anos e que sem dúvida tornaram este mestrado ainda mais enriquecedor.

Ao José Gouveia que esteve sempre em todos os momentos e me deu sempre uma palavra de conforto quando algumas adversidades surgiram, e pela compreensão e paciência sempre demonstrada.

Por fim, agradeço aos meus pais e à minha irmã que são os meus alicerces e me permitiram sempre seguir os meus sonhos. Pela compreensão sempre demonstrada, pelo amor e preocupação, e por me apoiarem incondicionalmente. A eles estou eternamente grata.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

ACV	Abnormal conjunctival vessels
AD	Alzheimer's disease
ALS	Amyotrophic lateral sclerosis
AO	Age-at-onset
APCS	Amyloid P component, serum
ApoE	Apolipoprotein E
AR	Androgen receptor
ATTR	Amyloidogenic *TTR*
BGN	Biglycan
C	Cytosine
C1QA	Complement C1q A chain
C1QC	Complement C1q C chain
CBP	CREB-binding protein
CGPP	Centro de Genética Preditiva e Preventiva
CHP	Centro Hospitalar do Porto
CLR	Collagen-like region
CNPD	National Commission for Data Protection
CNS	Central nervous system
CSF	Cerebrospinal fluid
CV	Consensus value
CVC	Cross-validation consistency
CVS	Chorionic villus sampling
DCs	Dendritic cells
DMSO	Dimethyl sulfoxide
DN	Dominant-negative
DNA	Deoxyribonucleic acid
DRG	Dorsal root ganglia neurons
ER	Endoplasmatic reticulum
ESE	Exonic splicing enhancer
ESRD	End-stage renal disease
ESS	Exonic splicing silencer
Evi-1	Ecotropic viral integration site
FAP	Familial amyloid polyneuropathy
G	Guanine
GAP	GTPase activating proteins
GDP	Guanosine diphosphate
GEEs	Generalized estimating equations

GEF	Guanine nucleotide exchange factor
GTP	Guanosine triphosphate
HAT	Histone acetyltransferase
HD	Huntington's disease
HDAC	Histone deacetylase
HSD17β1	Hydroxysteroid (17-β) dehydrogenase 1
HSF	Human splicing finder
HSP27	Heat shock protein 27
HSPC300	Haematopoietic stem/progenitor cell protein 300
IRF	Interferon regulatory factor
KCS	Keratoconjunctivitis sicca
LD	Linkage disequilibrium
MDR	Multifactor dimensionality reduction
MEK1	MAP kinase/ERK kinase 1
MEK2	MAP kinase/ERK kinase 2
miRNA	microRNA
mtDNA	mitochondrial DNA
NERF	New-ETS-related factor
NGAL	Lipocalin 2
NMD	Nonsense-mediated decay
PAK	P21-activated kinase
PCR	Polymerase chain reaction
PD	Parkinson disease
PGD	Pre-implantation genetic diagnosis
PND	Prenatal diagnosis
PST	Presymptomatic testing
PTCs	Premature termination codons
RAC1	Ras-related C3 botulinum toxin substrate 1
RAGE	Receptor for advanced glycation endproducts
RBP	Retinol binding protein
RhoGDIs	Rho Guanine nucleotide dissociation inhibitors
RNA	Ribonucleic acid
SAA	Protein serum AA
SNP	Single nucleotide polymorphism
SPSS	Statistical package for the social sciences
SR protein	Serine-arginine rich protein
TBA	Testing balance accuracy

TF	Transcription factor
TFBS	Transcription factor binding site
TTR	Transthyretin
UCA	Unidade Corino de Andrade
UTR	Untranslated region
VGCC	Voltage-gated calcium channels
WAVE	Velprolin homologue protein
WT	Wild-Type
YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta

ABSTRACT

Familial Amyloid Polyneuropathy (FAP ATTRV30M) is an autosomal dominant systemic amyloidosis caused by a point mutation in *TTR* gene (chr18q12.1) that is characterized by deposition of amyloid substance. Until now more than 100 mutations in *TTR* gene have been described, and the most frequent disease causing variant in Portugal is the V30M. FAP ATTRV30M shows a wide variation in age-at-onset (AO) [19-82 years], in the Portuguese population. Within some families, offspring often exhibit anticipation in AO when compared to their parents.

However, V30M alone does not explain the AO variability of the disease observed in the same family. Previous studies in *Drosophila* expressing TTRV30M demonstrated that *RAC1* may be associated with phenotypic variability and therefore may be a good candidate gene for the study of AO variation in FAP ATTRV30M patients. In addition, previous findings in a Cypriot population for *C1QA* and *C1QC* showed an association of these genes with AO. Our aim was to unravel if these candidate genes associated with *TTR* pathways may act as genetic modifiers of AO in FAP ATTRV30M, studying for the first time the *RAC1* gene as a genetic modifier and to assess the involvement of C1Q complement genes in order to confirm if the same happens in our population.

We analyzed 267 patients from 117 families. To search for variants in the three genes, all exons and flanking regions were genotyped by automatic bidirectional sequencing. To account for non-independency of AO between members of the same family, we used generalized estimating equations (GEEs) with IBM SPSS software. Afterwards, an intensive *in silico* analysis was performed, in order to predict differences in regulation of gene expression and to analyze a putative gene-gene interaction.

Our results showed that some variants in *RAC1*, *C1QA* and *C1QC* genes were associated with late-onset and they can have a protective role in FAP ATTRV30M patients. On the other hand, most variants were found to be associated with an early-onset acting as risk factors for FAP ATTRV30M carriers in *RAC1* and in *C1QC* gene).
Importantly, we found some interesting results in the *in silico* analysis, since we found some unreported modifications in transcription factors binding and miRNAs binding (*RAC1* and *C1QC* genes) and in the mechanism of splicing (*C1QA* gene).

Noteworthy, a strong interaction between *C1QA* and *C1QC* genes was found showing that these two genes may act together as genetic modifiers. With this study, we confirmed the role of *C1Q* complement genes in FAP Portuguese patients and we revealed for the first time the role of *RAC1* gene as a genetic modifier of AO variability. Importantly, we found very interesting and unreported results in the *in silico* analysis since alterations in these mechanisms can lead to dysregulation of gene expression, which can have an impact in AO and phenotypic variability.

Keywords: Familial Amyloid Polyneuropathy (FAP), V30M, *RAC1, C1QA*, *C1QC*, age-at-onset (AO), variability.

RESUMO

A Polineuropatia Amiloidótica Familiar (PAF ATTRV30M) é uma amiloidose sistémica autossómica dominante causada por uma mutação pontual no gene da *TTR* (chr18q12.1), que se caracteriza pela deposição de substância amilóide. Atualmente existem mais de 100 mutações descritas no gene da *TTR*, sendo que a mais frequente em Portugal é a substituição V30M. Na população portuguesa, a PAF ATTRV30M apresenta uma grande variação na idade de início [19-82 anos] e, em algumas famílias, a descendência apresenta uma antecipação na idade de início comparativamente aos progenitores. Contudo, a mutação V30M por si só não explica a variabilidade da idade de início da doença, observada dentro da mesma família.

Estudos anteriores em *Drosophila*, expressando a TTRV30M, demonstraram que o gene *RAC1* pode estar associado com a variabilidade fenotípica. Assim, este seria um bom gene candidato para o estudo da variabilidade da idade de início nos doentes PAF ATTRV30M. Para além disso, resultados anteriores numa população cipriota evidenciaram uma associação dos genes *C1QA* e *C1QC* com a idade de início.

O nosso objetivo foi determinar se estes genes candidatos associados com os mecanismos da *TTR* poderiam atuar como modificadores da idade de início em PAF ATTRV30M, analisando pela primeira vez o gene *RAC1* e confirmando o envolvimento dos genes do complemento C1Q na nossa população.

Analisámos 267 doentes de 117 famílias e para a pesquisa de polimorfismos nos três genes, todos os exões e regiões flanqueantes foram genotipadas por sequenciação automática bidireccional. De modo a ter em conta a não-independência da idade de início entre membros da mesma família, utilizamos o método *generalized estimating equations* (GEEs) recorrendo ao IBM SPSS software. Posteriormente realizou-se uma análise *in silico* de modo a prever diferenças na regulação da expressão génica e para analisar potenciais interações entre os genes.

Os nossos resultados demonstraram que algumas variantes nos genes *RAC1*, *C1QA* e *C1QC* estavam associadas com uma idade de início tardia, tendo um papel protetor nos doentes PAF ATTRV30M. Contudo, a maioria das variantes encontradas nos genes *RAC1* e *C1QC* estavam associadas com uma idade de início precoce, atuando como fatores de risco.

Na análise *in silico* encontramos alguns resultados interessantes tendo identificado algumas alterações não descritas na ligação de fatores de transcrição e ligação de miRNAs nos genes *RAC1* e *C1QC* e no mecanismo de *splicing* no gene *C1QA*. É de realçar a forte interação encontrada entre os genes *C1QA* e *C1QC* demonstrando que estes dois genes podem atuar em conjunto como modificadores genéticos.

Com este estudo, confirmamos o papel dos genes do complemento *C1Q* nos doentes PAF portugueses e revelamos pela primeira vez o papel do gene *RAC1* como um modificador genético da variabilidade da idade de início. Encontramos resultados importantes e até agora não descritos na análise *in silico* já que alterações nestes mecanismos podem levar à desregulação da expressão génica, que pode ter impacto na idade de início e na variabilidade fenotípica.

Palavras-Chave: Polineuropatia Amiloidótica Familiar (PAF), V30M, *RAC1, C1QA, C1QC*, idade de início, variabilidade.

1. INTRODUCTION

1.1. Amyloid definition and classification

The Nomenclature Committee of the International Society of Amyloidosis has defined amyloid as “extracellular depositions of protein fibrils with characteristic appearance in electron microscope, typical X-ray diffraction pattern, and affinity for Congo red with concomitant green birefringence” [1].

Therefore, amyloidosis are a particular group of diseases characterized by deposition of insoluble proteins that form amyloid fibrils with 10-12 nm. In addition, the amyloid deposits also present secondary components such as serum amyloid P component [2].

The current nomenclature is based on the nature of amyloid fibril protein precursor which is designated protein A and then adding the suffix that is an abbreviated form of the precursor protein name. We can find many types of amyloidosis according to the precursor protein of amyloid: transthyretin (TTR), apolipoprotein A1, gelsolin [2], fibrinogen Aα-chain, lysozyme, cystatin C [3] and amyloid β [4].

Thus, *TTR*-related amyloidosis are designated ATTR. In addition, there are also some rules for the classification of the variants of fibrillary protein according to the different gene mutations. In these cases, the classification was based on the location of the amino acid residue preceded by the abbreviation of normal amino acid and followed by the abbreviation of mutant amino acid, for example, ATTR V30M [5].

The different types of diseases can be identified through the protein name, namely systemic AA-amyloidosis and systemic AL-amyloidosis. The first one was formerly named secondary or reactive amyloidosis and is a consequence of long-lasting high expression of the acute-phase amyloid protein serum AA (SAA); the second form of amyloidosis referred contains fibrils which are resulting from monoclonal immunoglobulin light chains, produced by a plasma cell clone in the bone marrow. Rarely, AL-amyloidosis may appear as a localized amyloid type [6].

1.2. History of Familial Amyloid Polyneuropathy (FAP)

In the History of FAP, there can be distinguished essentially four time periods along its investigation – an “incubation period” that runs from 1939 to 1952; one second period until 1980; a third one, which is defined by the explosion of scientific knowledge of the disease; and finally a fourth period that begins with liver transplantation in 1991 and remains until nowadays [7].

In 1939, Corino de Andrade observed in Hospital Santo António, a woman with 37 years-old who had been living for several years in Póvoa de Varzim. In this case, Corino de Andrade found that this women had a peripheral neuropathy, gastrointestinal and sphincter perturbations that was known as “foot disease” [8]. In order to clarify this study, Corino de Andrade observed twelve more patients and he concluded that they were dealing with a new clinical entity [8]. In 1952, Andrade led an extensive anatomo-clinical study composed by 74 cases and this revealed a high frequency of the disease in the population [9]. In 1964, Becker described FAP as an autosomal dominant systemic amyloidosis, and revealed for the first time the presence of anticipation in some Portuguese families [10].

In 1978 there was an evolution in the FAP characterization, when Pedro Costa found that the transthyretin (TTR) was the responsible protein for amyloid deposits [11]. And a few years later, Tawara and Saraiva described a point mutation in the *TTR* (V30M), which is characterized by the substitution of a valine for methionine at position 30 in Portuguese, Japanese and Swedish families with FAP [12], [13].

1.3. Transthyretin amyloidosis

The *TTR* amyloidosis can be presented in two forms: hereditary (caused by mutations in the *TTR* gene) and *wild-type.* The hereditary form of FAP is inherited as an autosomal dominant disease [14] which means that only one mutant allele is required to develop pathology [15].

FAP is characterized by systemic extracellular deposition of *TTR* amyloid fibrils, especially in peripheral nerves, autonomic nervous system tissues, heart, kidneys, eyes and gastrointestinal tract [16]. These deposits, in the peripheral nerves, occur in the epi-, peri- or endoneurium [4] and they are formed either within the nerve bundles or vessels. In most cases it takes the form of rounded masses with more or less regular outlines but it may appear in the form of striate or irregular deposits [8].

TTR is a soluble plasma protein that binds to thyroxine (T4) and retinol binding protein (RBP) in the blood and cerebrospinal fluid [3]. Some studies demonstrated that soluble TTR self-assembles into amyloid fibrils as a consequence of point mutations in *TTR* gene originating the amyloidogenic *TTR* (ATTR) [16].

The native form of TTR is a tetramer composed of four equal subunits each of which contain 127 residues which form a homotetrameric protein with 56kDa [17]. Each monomer contains eight β-strands (A-H) and the global tetramer has two thyroxine-binding sites and four RBP-binding sites [15]. The molecular structure has determined

by X-ray analysis as shown in **Figure 1**. Two TTR monomers link edge-to-edge to form a dimer which is stabilized through antiparallel hydrogen-binding between H-H and F-F strands. Subsequently occurs the connection of two dimers through hydrophobic interactions between the A-B Loop of one monomer and the H strand of the opposite dimer [18].

TTR has an intermediate affinity for binding to T4 and the ability to carry approximately 15% of the total circulating T4, so it is one of the T4 transport proteins in plasma and in the cerebrospinal fluid (CSF) [19].

The *TTR* gene is localized on the long arm of chromosome 18 and this gene spans ≈ 7 kb containing four exons [15]. It is mostly synthesized by the liver (approx. 95%) and the remaining 5% is produced by the choroid plexus [20], as *TTR* does not cross the blood-brain barrier [18]. Furthermore, even in small quantities, it can also be synthesized in other organs such as in retina, pancreas and intestine [19].

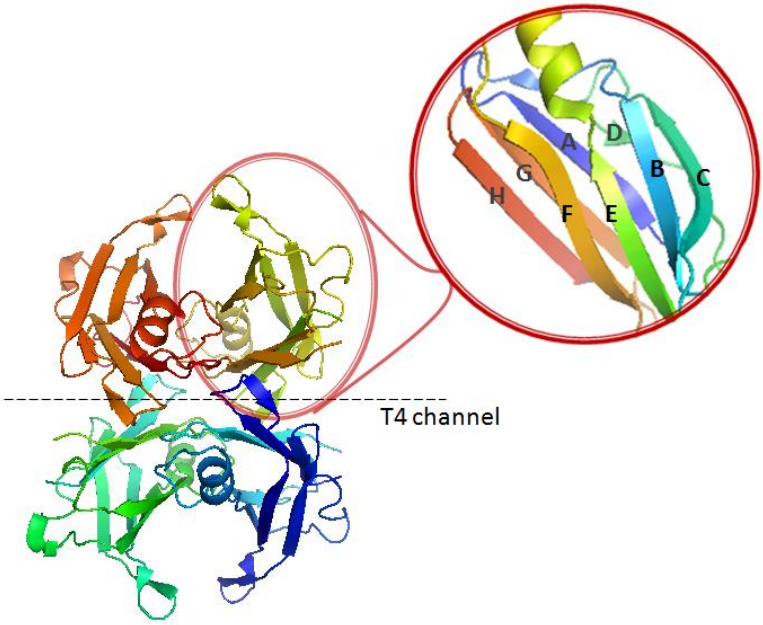

Figure 1: X-ray structure of transthyretin. In native state, *TTR* contains eight β-strands (A-H) and a small α-helix presented in red. The dimers bind together originating two hydrophobic pockets where T4 binds (T4 channel). PDB code: 1DVQ(Adapted from [21]) .

1.3.1. Mutations in *TTR*

TTR mutations are the most frequent cause of hereditary amyloidosis and more than 100 mutations worldwide have been described which report to different phenotypes varying from cardiac to neurologic disease [22]. The most common pathogenic substitution in FAP is a replacement of valine by methionine at position 30,

V30M (V50M, following HGVS) [2]. This change results in a substitution from a Guanine to an Adenine in the second exon of the *TTR* gene [19].

Some studies with heterozygous for V30M mutation demonstrated that the ratio of wild-type *TTR* and *TTR* V30M is 2: 1 and 1: 2 in plasma and amyloid fibrils, respectively [23]. The pathogenic mutations might decrease the stability of TTR tetramers leading to dissociation into monomers [2].

As previously mentioned, there are mutations that result in more cardiac component of the disease such as L111M, V122I (in the African-American population)[3], I68L and T60A while others are more associated with neuropathy such as V30M as illustrated in **Figure 2** [24].

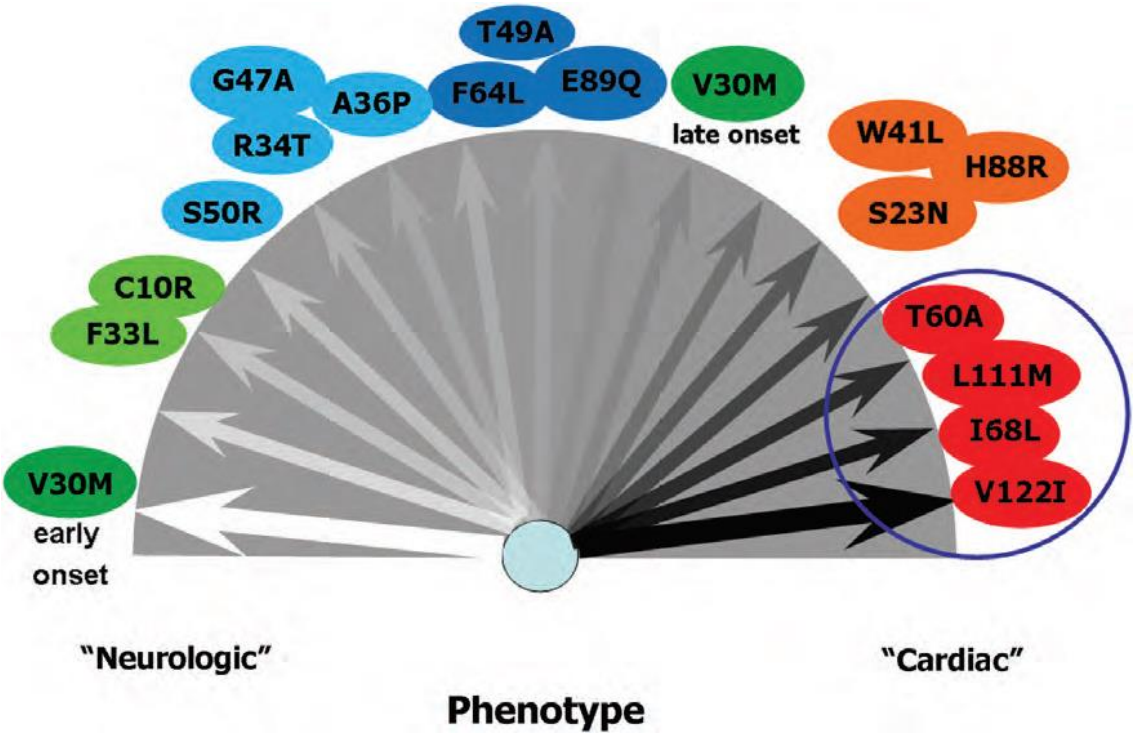

Figure 2: Spectrum of correlations between genotype-phenotype in transthyretin-related amyloidosis [25].

Apart from the already cited mutations, there are other known mutations such as the case of *TTR* L55P and *TTR* T119M. The first one is a variant with high amyloidogenic potential and consequently with a very severe phenotype, with patients presenting early-onset (≤ 40 years). On the other hand, the *TTR* T119M is a non-pathogenic variant [19], and it has been described in compound heterozygous carriers of both *TTR* V30M and *TTR* T119M, providing a protective function over the effect of *TTR* V30M since these patients present more stable TTR tetramers leading to a delayed disease course [26].

1.3.2. Disease pathogenesis

There are many factors involved in the pathogenesis of the disease, including fibril precursor protein and the metabolic processing of TTR.

TTR is a very important plasma protein (20–40 mg/dl) and it has a plasma persistence time of only 1–2 days, which must represent a substantial burden to the catabolic mechanisms of plasma protein turnover. Furthermore, another factor of TTR pathogenesis is the selective involvement of certain systems. The major location of amyloid deposition in FAP *A*TTRV30M is the peripheral nervous system, although the gastrointestinal system is also often involved. However, the most frequent cause of death is due to abnormalities in the cardiorespiratory system [15].

The partially formation of unfolded monomers and amyloid aggregates is closely related with thermodynamic stability of *TTR* variants since ATTR presents lower thermodynamic stability [18].

In the case of V30M mutation, the crystal structure has revealed that this substitution results in a small conformational change that is transmitted through the protein core to Cys10. Additionally, it is known that this substitution entails the separation of the two β-sheets of each monomer, leading to a change in the conformation of thyroxine-binding site [18].

Several hypotheses have been developed to explain the mechanisms of TTR fibril formation, but two of them were emphasized by Gustavsson A. et al.: the first one comprises proteolytic processing of the TTR molecule which produce fibrils with C-terminal fragments; and the second involves full-length ATTR [19], [27].

As several studies demonstrated that asymptomatic carriers with *TTR* V30M presented nonfibrillar TTR aggregates and the existence of nonfibrillar and fibrillar TTR aggregates appears in patients with late-onset (≥50 years), it can be concluded that TTR aggregates in the nonfibrillar form were unable to cause toxicity whereas in the fibrillary form cause cytotoxicity [19].

Some studies carried out by Miroy and Coworkers had shown that when T4 binds to TTR, it is able to efficiently inhibit TTR fibrillogenesis *in vitro* as a result of stabilizing the tetramer. Therefore, it was possible to demonstrate that TTR binds to a variety of small compounds with molecular structural similarities with T4, which function as competitors of T4. These compounds have been the subject of several studies as therapeutic alternative in FAP ATTRV30M, albeit with limitations, once many stabilizers do not produce any effect when tested *ex vivo* [19].

The mechanism of TTR aggregation is shown in **Figure 3,** which mutations in TTR lead to dissociation of tetramers in monomers, which posteriorly can aggregate. The low-molecular-mass diffusible species binds to the lipid membrane disrupting the same, which allows the calcium influx through voltage-gated calcium channels (VGCC) from extracellular medium. On the other hand, TTR can bind to a receptor for advanced glycation endproducts (RAGE) and disturb the kinase cascade, which induce ER stress with consequently release of calcium from intracellular stores. Therefore these mechanisms cause cellular damage, death and apoptosis [18].

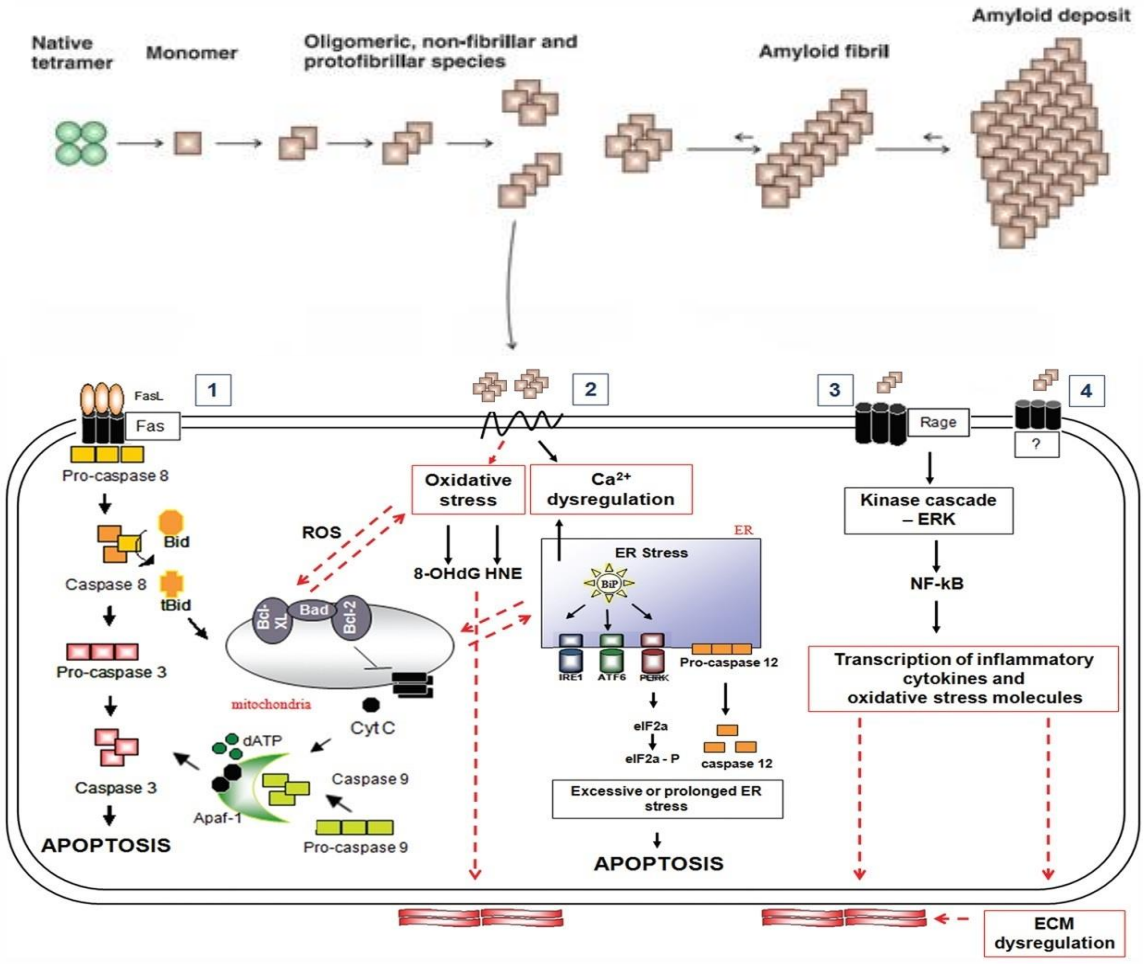

Figure 3: Mechanism of TTR aggregation in FAP: 1. Small TTR aggregates stimulate the FAS receptor which in turn leads to the activation of the caspase cascade and promote apoptosis. 2. TTR aggregates cause disruption of the cellular membrane and consequently, leads to calcium influx from the extracellular medium and calcium efflux from the ER. This allows the activation of ER stress and posteriorly the apoptosis, through ER stress – associate caspase-12 activation. Oxidative stress is also activated by TTR 3. TTR aggregates interact with RAGE stimulating a kinase cascade that activates NFkB, which in turn activates the transcription of inflammatory cytokines and oxidative stress molecules and consequently increase levels of ROS. 4. Cellular toxicity can also be explained by the existence of still unknown receptors for TTR aggregates which are induced by extracellular small non-fibrillar TTR aggregates (Adapted from [18];[19]).

1.4. Clinical Manifestations

FAP ATTRV30M has a higher neurological component since it is evident the sensorimotor and autonomic polyneuropathy, due to the involvement of non-myelinated and small myelinated fibers [28]. Therefore, this disease is commonly characterized by the presence of peripheral nerve dysfunction and autonomic dysfunction, which is evidenced in the diagram of the **Figure 4.**

1.4.1. Peripheral Nerve dysfunction

In Portugal, patients showed the first symptoms in their 30's, and these symptoms consists initially in discomfort in the feet, including numbness and spontaneous pains [2]. This signs begins in the small nerve fibers altering thermal sensitivity, with decreased pinprick sensation. However, light touch and proprioception are relatively preserved. Posteriorly, the neurologic deficit progresses with sensory loss towards the proximal lower limbs which implies difficulty in walking. At this stage there is also a weight loss occurrence and muscle wasting. In addition, Carpal tunnel syndrome is an early but nonspecific manifestation of *TTR*-FAP [24]. Concerning to autonomic nervous system it was possible to observe the early loss of unmyelinated nervous fibers which plays an important function in the distal component of the autonomic nervous system. The progressive denervation of numerous organs can contribute to the most frequent clinical manifestations, namely gastrointestinal, glandular, vesical, sexual and cardiovascular [29].

1.4.2. Autonomic dysfunction

The autonomic neuropathy occurs in individuals with early-onset FAP ATTRV30M. The circulatory, gastrointestinal and genitourinary systems are affected [2].

Regarding cardiovascular disturbances, there are many causes of its occurrence, namely, autonomic disturbances of peripheral blood flow, blood pressure and cardiac arrhythmia, because of amyloid accumulation in the heart's conductive system[30]. The first symptom observed is the modification of diastolic function, and only in the later stages of the disease, it is possible to observe the cardiac ventricular hypertrophy and systolic malfunction [28].

Concerning to gastrointestinal tract, it is regulated by central, intrinsic and endocrine-paracrine system. Autonomic nervous system includes alternating periods of constipation and diarrhea, nausea, and vomiting and neurogenic bladder [24]. It is thought that gastrointestinal disturbances may be caused by amyloid infiltration and destruction of the celiac ganglion, the vagus nerve and the sympathetic chain ganglia and nerves [30].

Another system commonly affected is the ocular system, with various symptoms such as vitreous opacity, dry eye, glaucoma, pupillary disorders [24], abnormal conjunctival vessels (ACV) and keratoconjunctivitis sicca (KCS) [30].

Concerning to renal manifestations, in the FAP ATTRV30M patients different levels of proteinuria, renal insufficiency, and progression to end-stage renal disease (ESRD) have been found [31].

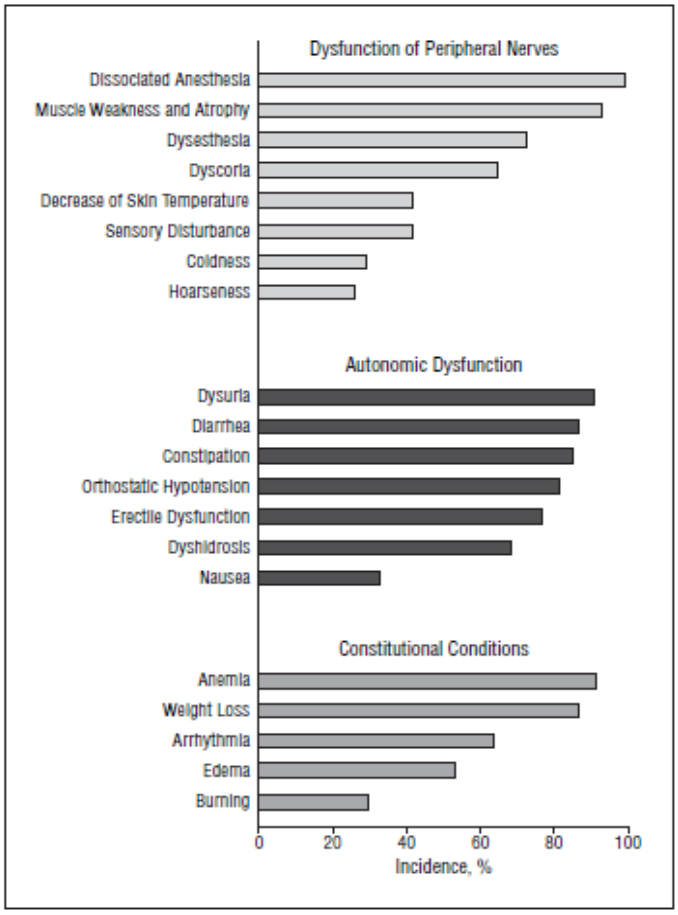

Figure 4: Clinical manifestations in patients with FAP ATTRV30M (Adapted from [32]).

1.5. Epidemiology

V30M is the major variant found in Portugal, but it is described in many other countries such as Brazil [33], Sweden [34], Japan [35], France[36], Spain [37],

USA[38], China [39], Italy and Greece [40]. The geographical distribution of FAP ATTRV30M is shown in **Figure 5.**

In northern Sweden, V30M mutation is associated with late-onset cases [15] and it has a low penetrance (11% by 50 years)[24]. On the other hand, in northern Portugal the same mutation is associated with an earlier onset [15] and has a higher penetrance (80% by 50 years) [24]. Since FAP ATTRV30M penetrance is variable from country to country, some authors propose that this difference may be explained through a combination of genetic and nutrition factors [20]. In addition, penetrance is higher when the disease is maternally inherited, which suggest that differences in penetrance may be explained by mitochondrial polymorphism [3] or due to an imprinting phenomenon [41].

Epidemiological studies in northern Portugal (Póvoa de Varzim and Vila do Conde) shows a prevalence rate of 1/1000 [42] and the incidence is estimated to be one in 538 individuals [24]. Therefore, Portugal has a high prevalence of the disease as Sweden and Japan [20].

Figure 5: Geographic distribution of FAP ATTRV30M (Adapted from [2]).

1.6. Age-at-onset (AO)

In Portuguese patients, the range is from 19 to 82 years and the mean AO of the disease is around 33.5 years [43] and the patients have a rapidly progressive sensory-motor and autonomic neuropathy which causes death past 10-20 years [42].

In Portugal, more and more late-onset cases (≥50 years) have been ascertained as well as aged-asymptomatic carriers. So, in previous studies, our group,

found that late-onset cases and aged-asymptomatic carriers aggregate in families, and that, just like in Sweden, isolated cases with late AO often descend from aged-asymptomatic carriers [44], [45].

In some families the mutation is transmitted through several generations without expressing any symptoms. However, this protective effect disappears in the following generations leading to a process called genetic anticipation. Anticipation is characterized by the existence of a considerably earlier AO (≤ 40 years) in younger generations with subsequent increase of the disease severity. In addition, in FAP ATTRV30M it has been observed that early and late-onset cases often coexist in the same family with progeny showing anticipation **(Figure 6)** [44].

Some studies carried out by our group demonstrate that this genetic anticipation is significantly higher when the mutation is transmitted by the mother than the father. So, the anticipation is more pronounced in mother-son pairs [44]. In spite of the mechanism underlying this maternal effect is not yet elucidated, there was some hypothesis that may explain this effect such as due an imprinting phenomenon or due to existence of a mitochondrial DNA polymorphism that regulate gene expression [41, 46].

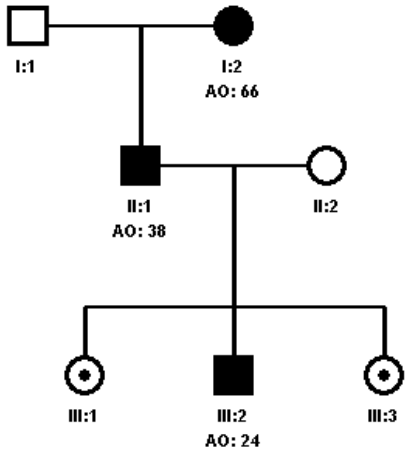

Figure 6: Portuguese family showing anticipation throughout three generations.

However, due to low penetrance and late-onset of *TTR* mutations in some regions, FAP ATTRV30M can exist as sporadic cases [2].

Portuguese and Japanese foci of patients with FAP ATTRV30M have been described as early AO (33 years), while Swedish patients with FAP ATTRV30M are characterized by a later AO (56 years) [24].

Anticipation has to be looked in the view of other modifier factors, such as genetic modifiers. That is why, candidate genes seem to be a promising strategy to identify possible modifiers of AO.

1.7. Clinical Diagnostic

The difficulties of clinical diagnosis of FAP ATTRV30M date sixty years back, starting in the first observations of Andrade in 1939 and develops up nowadays [47].

Generally, FAP ATTRV30M can be diagnosed through a tissue biopsy indicating amyloid deposits by Congo red staining. This tissue sample may be acquired from different origins, especially from an affected organ, a nonspecific local with a 70-80% sensitivity for TTR, abdominal fat pad or salivary gland aspirate [20].

Another way to diagnose serum variant TTR protein can be detected by mass spectrometry, after immunoprecipitation with anti-TTR antibody. Approximately 90% of TTR variants exhibit the mass shift predicted by the one amino acid substitution, which allows differentiating between wild-type and TTR protein variants in serum. Nevertheless, it does not specify the local and type of amino acid substitution in *TTR* gene mutations. That is why, in this cases, DNA sequencing is usually required [24].

In some situations, where it is extremely needed a quickly response, namely in pre-natal and preimplantation diagnosis, it is necessary to apply new amplification methods with real-time detection using hybridization probes [48].

In the diagnosis of a hereditary amyloidosis, the knowledge of family history is very important, but may not be enough to establish the diagnosis, due to the existence of lack of penetrance, advanced AO and variable expression of the disease.

If a patient has a family history of TTR amyloidosis and the mutation is known, a specific DNA testing can be done to discover the concern mutation. On the other hand, if a family mutation is not known, it is necessary to carry out full *TTR* DNA sequencing [15].

1.7.1. Genetic counselling

Genetic counselling is a communication procedure that deals with the risk of occurrence of a genetic disorder in the family. In this process is important that family realize the medical facts of the disorder, understand how heredity contributes to the disorder and their putative risk in specified relatives [49].

Genetic counselling is an important strategy in FAP ATTRV30M patients with early-onset, (as the disease is autosomal dominant) [50] to prevent or delay disease progression [51]. Nevertheless, in FAP ATTRV30M with late-onset, genetic counselling is more problematic because asymptomatic carriers manifest the disease at a very advanced age and it is not possible to predict when the disease will start [50].

For families where the disease appears only in a single family member, is now available the presymptomatic testing (PST) for the *TTR* gene for relatives and the *follow up* of asymptomatic individuals carrying the *TTR* variant. Thus, this tests can be performed in referral centers and in the genetic laboratories [51].

Currently, there are two techniques that allow a FAP ATTRV30M patient have offspring with a very low risk of being carriers of the disease, which are called pre-implantation genetic diagnosis (PGD) and prenatal diagnosis (PND).

PND is a diagnostic test carried out on a developing fetus through amniocentesis, chorionic villus sampling (CVS), fetal blood sampling, collecting fetal sample in maternal circulation or ultrasound. This test is designed to detect a fetus with a chromosomal abnormality, congenital malformation or disease, and therefore allows the parents the decision to stop the pregnancy in order to prevent the birth of offspring with genetic and/or congenital abnormalities.

On the other hand, the PGD test consists in the analysis of one or two embryo cells with three days, allowing the detection of genetic diseases, and only the embryo cells without the mutation are implanted in the mother's uterus [49]. This technique has the advantage of not interrupting pregnancy.

1.8. Treatment

The treatment in the FAP ATTRV30M disease can be divided into two principal groups: the symptomatic treatment and the etiologic treatment. The first one includes the application of pacemaker to treat arrhythmia in later stages of the disease, the realization of dialysis in renal perturbations and many others such as the drug therapy; and the second one includes many strategies to stop or inhibit the amyloid deposition such as liver transplantation [30].

As TTR is mostly synthesized by the liver, the only specific treatment for FAP ATTRV30M patients is the liver transplantation, where a blockage of mutated protein synthesis occurs, leading the formation of amyloid fibrils with pathological effects to stop [20]. The first liver transplantation was realized in 1990 in Sweden [52], and since then this clinical therapeutic has been performed worldwide [53].

In spite of this treatment being the most widely accepted, there are reported cases of patients in which polyneuropathy have developed after liver transplantation, which suggests that once amyloid deposits have accumulated in organs, the process can progress to amyloid fibrils from wild-type TTR [20]. Patients with such treatment

have a prospect of longer and better quality of life, once liver transplantation eventually remove the progressive and fatal nature of the disease.

Despite the positive results achieved by the introduction of liver transplantation, some problems persist, such as limitation of availability of organs, high cost involved, the obligation to carry out immunosuppressive therapy for life and the impossibility of performing liver transplant patients but not in patients symptomatic, or patients with advanced disease [7].

There is a new therapeutic option that is used by some countries for FAP ATTRV30M treatment – Tafamidis. This drug works through stabilization of TTR tetramer and is suitable for treatment of adults with FAP ATTRV30M in the early stage [54].

Diflunisal is a commercially available nonsteroidal anti-inflammatory agent that stabilizes TTR tetramers in vitro, preventing disaggregation, monomer release, and amyloid fibril formation by misfolded TTR monomers [55].

Another type of therapy is the post-transcriptional gene silencing, which is capable to inhibit hepatic production of TTR, such as Antisense Oligonucleotides (ASOs) and Small Interfering RNA (siRNA). The first one is synthetic sequence of nucleotides that binds to the RNA which encodes the targeted protein and consequently prevents the translation. An example of this is the ISIS-TTR_{RX}, which is a second-generation antisense inhibitor of the molecular target transthyretin [56].

On the other hand, the ALN-TTR01 is a lipid nanoparticle-formulation of a small interfering RNA targeting wild-type and all mutant forms of TTR. This compound can transmit the information to the liver leading to the inhibition of TTR synthesis [56].

1.9. Genetic modifiers

The wide variability in AO of affected patients is still unexplained. Although environmental factors cannot be completely ruled out, the fact is that a large variation may still be found within a single geographical area and within the same family. The intrafamilial variability raised the hypothesis of a closely linked genetic modifier, since the mutation for itself could not explain the variability observed. Therefore, the existence of genetic modifiers within or closely linked to the *TTR locus* can bring new insights into AO variability.

Several studies have been performed regarding genetic and environmental factors in order to demystify possible modifier genes. There have already been discovered some genetic modulators that may be involved in the variability of the AO in

FAP ATTRV30M such as *amyloid P component, serum* (*APCS*), complement *C1QA* and *C1QC*, *plasma retinol-binding protein 4 (RBP4*) [57] and *apolipoprotein E* (ApoE) [58], but they only explain a small part of AO variability in FAP ATTRV30M.

In addition, some studies involving Portuguese families with FAP ATTRV30M demonstrated that exist considerable differences between males and females affected with the disease, namely that women were found to have later onset when compared to men. Accordingly, sex steroid hormones such as *androgen receptor* (AR) and *hydroxysteroid (17-β) dehydrogenase 1* (*HSD17B1*) were studied, since both affect *TTR* expression levels [57].

Our group, recently conducted some studies with *APCS, RBP4* and *AR* genes, and concluded that they are all involved in the AO variation by using a family-centered approach. Regarding sex hormone genes, namely *AR* gene, our group found for the first time significant results showing that, in the male group, the presence of a rare allele in any of the variants found is responsible for decreasing the AO. On the other hand, in the female group, it was concluded that the heterozygous genotype for almost all variants is associated with an earlier AO, with the exception of heterozygous for one variant that is associated with late AO. And interestingly, while the SNP in the female group causes an increase of AO, in the male group the opposite happens, which confirms the existence of extreme differences between genders [57].

In order to complement our previous studies and confirm that the anticipation is mainly transmitted by the mother (70%) than the father (30%) [44], our group performed studies to quantify the mitochondrial DNA (mtDNA) copy number in patients with FAP ATTRV30M. So they concluded that the highest median of mtDNA copy number was found in early-onset patients [unpublished results].

In addition, our group go forward to explore candidate genes, so Santos, D. et al. analyzed another genes but only six of them were statistically significant (*NGAL*, *BGN* (only in the female group), *MEK1*, *MEK2*, *HSP27* and *YWHAZ*) and concluded that they were significantly associated with early- and/or late-onset [unpublished data].

Besides the studies already mentioned, our group conducted a study in which the main objective was to identify genetic modifiers within the *TTR* locus, or closed linked to it, that might modulate AO, using haplotype study. Thus, a haplotype linked to V30M mutation was found and another one that confers some degree of risk to “very early” onset patients, which may partially explain the early-onset of progeny, when transmitted by the non-carrier parent (*trans*-effect) [unpublished results].

1.9.1. RAC1

The Rho family of GTPases are encoded by 20 genes which have been sub-classified into various families based on sequence homology: Rac (Rac1-3, RhoG), Rho (RhoA-C), Cdc42 (Cdc42, TC10, TCL, Chp/Wrch2, Wrch1), Rnd (Rnd1-2, Rnd3/RhoE), RhoD (RhoD, Rif/ RhoF), RhoBTB (RhoBTB1-2) and RhoH [59].

Rac plays an important role in communicating cell surface receptors to biochemical pathways regulating gene expression. This family comprises 3 isoforms: Rac1, Rac2 and Rac3, being Rac1 the most studied isoform [60].

Ras-related C3 botulinum toxin substrate 1 (*RAC1*) is located on short arm of chromosome 7 (7p22) and comprises 7 coding exons. It is ubiquitously expressed and plays a key role at the level of actin cytoskeletal organization and consequently in migration of endothelial cells, tubulogenesis, adhesion and endothelial barrier function [60]. Rac1 acts a molecular switch that associates with inactive guanosine diphosphate (GDP) state and active guanosine triphosphate (GTP) state. The regulation of GTP-GDP activation is due to the actions of nucleotide exchange factors (GEFs), GTPase activating proteins (GAPs) and Rho guanine nucleotide dissociation inhibitors (RhoGDIs) [61].

The Rho GTPases plays an important role in transcriptional activation, membrane trafficking and microtubule dynamics. Therefore, the members of this family are involved in many cellular functions as well as cell transformation and invasion, and more recently it was found that it is involved also in neuronal development [62]. It has been implied that Rac1 and Cdc42 act simultaneously to regulate some functions resulting in neurite outgrowth and growth cone development [63].

The principal mechanism of Rac function in actin cytoskeletal organization is illustrated in **Figure 7.** Rac can activate the p21-activated kinase (PAK) family of serine/threonine kinases, which exist in a latent state in the cytoplasm and play an important role in regulating actin cytoskeleton. P35 in combination with Cdk5 causes hyperphosphorylation of PAK1 and this effect results in its down-regulation. Therefore, in order to regulate the PAK activity, Rac can activate PAK and the p35/Cdk5 complex, at the same time. Once active, PAK activate LIM domain-containing kinases through phosphorylation, which in turn can inactivates cofilin, resulting in actin polymerization. On the other hand, PAK can phosphorylate stathmin/Op18 at Ser16, which impedes Op18 from binding to microtubules, and subsequently leads to stabilization of microtubules. The WAVE family members when interacting with Arp2/3 complex promote the last one to nucleate the formation of actin filaments. In the inactive state, the WAVE complex is composed by Nap125, PIR121, Abi2 and the heat-shock protein

HSPC300. When this complex is added to Rac, occurs its dissociation which promotes the release of WAVE and HSPC300 with consequent activation of the Arp2/3 complex [62].

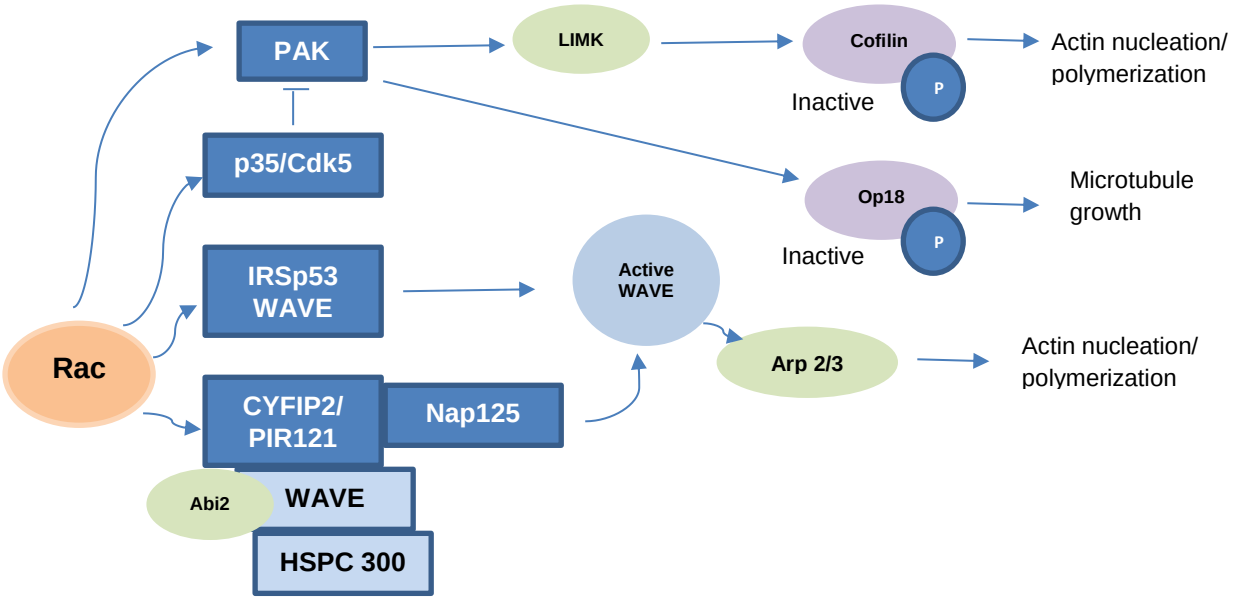

Figure 7: Implication of Rac in actin and microtubules dynamics.

Taking into account the mechanism of Rac mentioned above, this one have been identified as potential regulator of dendrite and spine development [64], namely it induces membrane ruffling and lamellipodium formation [65].

In the FAP ATTRV30M disease, there are no published studies associated with cytoskeleton damage. Nevertheless, some studies carried out by Silva et al. (unpublished data) shows that dorsal root ganglia neurons (DRG) (place near where occurs *TTR* accumulation) when treated with TTR oligomers shows cytoskeleton modifications. This effect is shown in **Figure 8** [66].

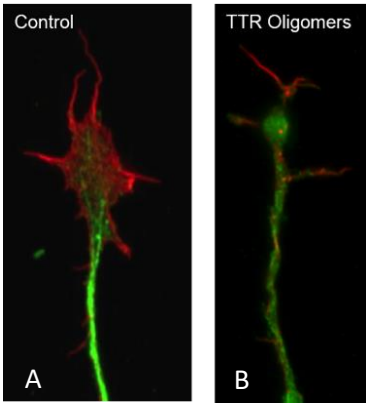

Figure 8: Growth cones of the DRG neurons in culture. (A) In a control condition, a DRG neuron has the normal actin pattern (red) composed by filopodia and lamellipodia and the microtubules (green) are well organized in bundles. (B) DRG neurons treated with *TTR* oligomers present a totally disorganized structure of the growth cone with loss of actin structure and the one that is present are organized in spikes of filopodia without lamellipodia existing in the structure. Furthermore, the disruption of the microtubules structure occurs with consequently formation of a bulb in the tip (Silva, M., unpublished results).

1.9.2. Complement *C1QA* and *C1QC*

C1 complement encompasses three subcomponents called C1q, C1r and C1s which are associated in a calcium-dependent macromolecular complex. This complex is activated through conformational changes in the collagen region of the C1q as a consequence of its binding to altered acceptor molecules [67].

The C1q human molecule with 460kDa is comprised by 18 polypeptide chains of approximately 220 residues each, called A (6A), B (6B) and C (6C) [67] and the genes for its chains are located on the short arm of chromosome 1 (1p36.3–p34.1) [68]. Each chain aforementioned has a N-terminal region flanked by collagen-like region (CLR) and a C-terminal domain (gC1q) [67]. *C1QA* and *C1QC* genes are composed by three exons, but only two are coding exons.

This complement is mostly synthesized by myeloid cells namely macrophages, monocytes and dendritic cells (DCs) and it is implicated in various cell functions such as migration of cells, adhesion, survival and differentiation [67].

It is believed that C1q could play a dual role in central nervous system (CNS), once in some cases C1q seems to be a protector factor by clearance of apoptotic cells and blebs and modulating inflammation; on the other hand, in the presence of other complement components of the classical pathway, C1q seems to develop neurotoxic inflammation [69].

Accordingly, some brain cells such as astrocytes and microglia express complement inhibitors and its function seem to be a protector factor against complement-mediated damage. In contrast, neurons are extremely susceptible to complement attack, which occurs by complement C1q binding to the membrane of neurons and it promotes the activation of classical pathway. However, unlike other brain cells, the neurons have a poor capacity to regulate the complement activation and they are susceptible to develop tissue damage. As a matter of fact, in neurodegenerative diseases, some studies demonstrated that intracellular levels of complement are increased, reinforcing what was said above [70].

In a previous study, Dardiotis et al. carried out a work with Cypriot FAP ATTRV30M carriers, and demonstrated that a polymorphism in the *C1QA* gene (genotype AA of rs172378) was associated with an earlier onset while in the *C1QC* gene, the genotype AA of rs9434 was associated with a later AO. Therefore, *C1Q* may be a modifier gene in FAP ATTRV30M and may explain the genetic anticipation in some families [68].

In order to further explore AO variation, we decided to study these three candidate genes, which seem to be also involved in FAP ATTRV30M pathways.

2. AIMS

In this dissertation we intended to determine if a group of candidate genes may act as genetic modifiers of AO in individuals with FAP-TTRV30M, since it has been observed differences in age-at-onset (AO) between different families and within the same family, in the Portuguese population

Therefore, the principal aims of this dissertation are:

- Investigate candidate genes (*RAC1*, *C1QA* and *C1QC*) as a potential genetic modifier of AO in the FAP ATTRV30M;
- *In silico* analysis of the variations found in the genes under study and its implication in gene expression and variability of AO in the FAP ATTRV30M.

3. METHODS

3.1. Sample Study

Unidade Corino de Andrade (UCA-CHP, Porto) has the largest FAP ATTRV30M registry worldwide, that was collected over 75 years and clinically well characterized. The present study is composed by 267 affected individuals with FAP ATTRV30M from 117 families currently under follow-up at UCA-CHP, which came from different geographical areas of country. The AO of each individual were established by the same group of neurologists, after the manifestation of the first symptoms.

The DNA samples of the patients were collected and stored at Centro de Genética Preditiva e Preventiva (CGPP, Porto, Portugal) biobank, authorized by CNPD (National Commission for Data Protection). It is important to note that all samples were obtained with informed consent of the individuals and all legal and ethical procedures were taken.

3.2. Genotyping of *RAC1*, *C1QA* and *C1QC*

3.2.1. DNA extraction

Genomic DNA was collected from peripheral blood leucocytes, using the QIAamp® DNA Blood Mini Kit (QIAGEN™) or from saliva using ORAGENE DNA Self Collection Kits (DNA Genotek, Ottowa, Canada) according to manufacturer's instructions.

3.2.2. Determination of DNA concentration using Nanodrop

In order to evaluate concentration and purity of genomic DNA, we extract 2 µL of each sample and analyze using the NanoDrop Spectrophotometer ND-1000 (Thermo Scientific). Purity is a parameter given by the A260/A280 ratio and concentration is expressed in µg/µL.

3.2.3. Primer design and amplification of DNA using conventional PCR

Primers were designed using Primer3Plus software and afterwards the existence of dimer and hairpin formation, as well as polymorphisms was excluded using the AutoDimer and SNPcheck softwares, respectively.

We consider all *RAC1* amplicons with a flanking region of 60 bp and *C1QA* and *C1QC* amplicons with a flanking region of 50 bp. Additionally, we consider more 100 bp for the 3'UTR. The primers' sequences that were used are presented in **Table 13** in Appendix**.**

A polymerase chain reaction (PCR) was performed to amplify genomic DNA of the patients for the three genes under study. Therefore we used a PCR Master Mix composed by H_2O, HotStarTaq DNA polymerase® (Qiagen™), and primers (forward and reverse) for each exon of every gene. The protocol that was used is presented in **Table 15** in Appendix. However, this protocol was not used for Exon 1 of *RAC1*, since this sequence is very rich in guanine (G) and cytosine (C). So it was necessary to replace the H_2O by DMSO, and the protocol is shown in **Table 14** in Appendix.

DMSO is the most commonly used enhancing agent and this facilitate strand separation[71]. This is particularly useful in sequences with high GC content because the increased hydrogen bond strength increases the difficulty of denaturing and promote the capacity of the intermolecular secondary structures to form more readily, which can compete with primer annealing [72].

Additionally, the conditions used in protocols in this step are presented in **Table 16** and **Table 17** (for *RAC1* gene), **Table 18** (for *C1QA* gene) and **Table 19** (for *C1QC* gene) in Appendix.

Posteriorly, we use the Qiaxcel® BioCalculator equipment in order to verify amplification of PCR products and to exclude possible contaminations.

3.2.4. Purification

Before proceeding to the next step is necessary to perform the purification of PCR products in order to remove unincorporated primers and dNTPs. For this purpose, we used ExoProStar™, which is able to remove these contaminants and the protocol as well as the conditions are presented in **Table 20** and **Table 21** in Appendix.

3.2.5. Sequencing

After the purification step, we proceed with the sequencing stage. The reaction was performed using Big Dye Terminator Cycle Sequencing v1.1, and the protocol as well as conditions underlying this reaction are presented in **Table 22** and **Table 23**, respectively, in Appendix. After, sequencing capillary electrophoresis was performed in ABI Prism 313xl Sequencer analyzer (Applied Biosystems ™), and further analyses of sequences were carried out using Seqscape v2.6 software (Applied Biosystems ™).

In order to confirm if the variants found are already described, we used the Alamut Mutation Interpretation Software (Interactive Biosoftware, Rouen France).
In cases of polymorphisms with uncertain genotypes or that are undescribed, these were confirmed and evaluated again by sequencing.

3.3. Statistical analysis

Since we included in the analysis several members of the same family, each patient was "nested" in his/her family. For this purpose analyses were conducted taking into account the non-independency of AO, using generalized estimating equations (GEEs).

In this model, we associate the different variants with AO (which is the dependent variable), using the most common genotype as the reference and adjusting for gender.

The unstandardized coefficient (B) corresponds to the mean AO variation observed in the individuals carrying a specific genotype when compared with the reference category. To correct for multiple testing, we applied a Bonferroni correction (α was set at 0.02 in the GEE analysis). All these statistical analyses were performed using IBM SPSS Statistics software (v.23).

3.4. *In silico* analyses

We used bioinformatics tools to perform in silico analyses such as Polyphen-2 to predict the impact of sequence variants on gene function and SNP Function Prediction (FuncPred) to predict putative SNP functional effects. In order to deepen the results obtained in FuncPred, we used other bioinformatics tools according to what we wanted to explore in more detail. Thus, to predict how likely a splicing change would occur as a result of the presence of the SNP, we used two prediction programs: ESE (Exonic Splicing Enhancer) Finder v3.0 and Human Splicing Finder (HSF) v3.0.

On the other hand to predict whether an allele change of a 3'UTR SNPs lead to a disruption or creation of novel miRNA binding sites, we used mirDIP dataset which includes twelve different datasets [73].

In addition, to predict miRNA targets in the 3' UTR and the most conserved miRNA in the WT sequence we used miRWalk.As the mirDIP software, this software integrates other eleven softwares, DIANA-microT, miRanda, mirbridge, miRDB, miRMap, miRNAMap, Pictar2, PITA, RNA 22, RNAhybrid and Targetscan.

To further explore which non-coding variants create new Transcription Factor Binding Sites that can regulate gene expression, we used rSNP-MAPPER software, taking into account TRANSFAC, MAPPER and JASPAR models.

In order to search possible gene-gene interactions, we used Multifactor Dimensionality Reduction (MDR) software (v2.0), which is a nonparametric and genetic model approach that allows to compare early and late onset patients and to identify which combinations of SNPs were involved in the disease vulnerability. Posteriorly, the significant results were corrected for multiple testing, based on 1000-fold permutation test, using MDR Permutation Testing Module (v1.0).

4. RESULTS

4.1. Descriptive analysis

Our analyses included 267 individuals with FAP ATTRV30M from 117 families, and the mean AO of our sample is 39.66 years. In the following table we can find the descriptive analysis of AO distribution, between males and females.

Table 1: AO distribution between males and females, in our sample.

		AO (years)					
Gender	**N**	**Range**	**Minimum**	**Maximum**	**Mean**	**Standard Deviation**	**Variance**
Males	134	51	21	72	38.11	13.81	190.67
Females	133	50	21	71	41.21	12.26	150.32
Total	267	51	21	72	39.66	13.13	173.34

According to the results presented in **Table 1** the mean AO in the male group was 38.11 years, whereas in the female group the mean AO was 41.21 years. So, as described in the literature, the mean AO in males is lower than in females [44].

It is important to note that it was necessary to exclude some DNA samples due to problems in the integrity and concentration of DNA.

4.2. Genotypic analysis

4.2.1. Variants presented in *RAC1* gene

Through genotyping of coding and flanking regions of the *RAC1* gene, it was possible to find 16 variants: 14 in the intronic region and 2 in exonic region, which were already described (**Table 2)**. From these, we found 8 variants significantly associated with AO (**Table 3)**.

Table 2: Variants found in *RAC1* gene

dbSNP ID	Location	Region	Effect on protein (Polyphen-2)	Frequency of common allele (HapMap: EUR-CEU)	Frequency of mutant allele (HapMap: EUR-CEU)
rs34500835	c.36-52G>A	Intron 1	N.A.	G (1.000)	A (0.000)
rs369301942	c.108-47A>G	Intron 2	N.A.	A (1.000)	G (0.000)
rs702483	c.107+27C>T	Intron 2	N.A.	C (0.540)	T (0.460)
rs3729789	c.107+39C>T	Intron 2	N.A.	C (0.995)	T (0.005)
rs3729790	c.107+40G>A	Intron 2	N.A.	G (0.803)	A (0.197)
rs836478	c.225+18C>T	Intron 3	N.A.	C (0.540)	T (0.460)
rs762492724	c.283-40G>C	Intron 4	N.A.	G (1.000)	C (0.000)
rs200639704	c.282+14T>G	Intron 4	N.A.	T (1.000)	G (0.000)
rs118039194	c.282+39C>T	Intron 4	N.A.	C (0.995)	T (0.005)
rs749782387	c.345+24A>G	Intron 5	N.A.	A (1.000)	G (0.000)
rs144607000	c.345+60C>G	Intron 5	N.A.	C (0.995)	G (0.005)
rs200222012	c.506-35G>A	Intron 6	N.A.	G (1.000)	A (0.000)
rs3729796	c.505+12T>C	Intron 6	N.A.	T (0.995)	C (0.005)
rs2303364	c.505+30T>C	Intron 6	N.A.	T (0.955)	C (0.045)
rs61753123	c.552C>G, p.Leu184Leu	Exon 7	Synonymous (Benign)	C (0.995)	G (0.005)
rs144425992	c.603G>A, p.Val201Val	Exon 7	Synonymous (Benign)	G (1.000)	A (0.000)

N.A.: Non Applicable

CEU - Caucasian population

Using Polyphen-2, we found that exonic variants had a benign impact in *RAC1* function, since we are dealing with synonymous substitutions.

Table 3: Statistically significant variants in *RAC1* gene

	Variant	B	Genotype	95% Wald Confidence Interval	P-value
	Intercept	36.64		[32.79; 40,50]	<0.001
1.	rs3729789	-9.82	CC (Reference) CT	[-15.46; -4.19]	0.001
2.	rs762492724	-10.45	GG (Reference) GC	[-14.21; -6.69]	<0.001
3.	rs200639704	-7.05	TT (Reference) TG	[-10.26; -3.83]	<0.001
4.	rs118039194	-4.40	CC (Reference) CT	[-8.58; -0.23]	0.039
5.	rs749782387	5.36	AA (Reference) AG	[1.50; 9.21]	0.006
6.	rs144607000	28.55	CC (Reference) CG	[24.79; 32.32]	<0.001
7.	rs200222012	-7.45	GG (Reference) GA	[-11.21;-3.69]	<0.001
8.	rs144425992	-14.05	GG (Reference) GA	[-17.26; -10.83]	<0.001

From the eight statistically significant polymorphisms (sequencing results can be found in Annex 2 of the Appendix), six of them (rs3729789, rs200639704, rs118039194, rs762492724, rs200222012 and rs144425992) were associated with an earlier onset, leading to a decrease in the mean AO between 4 and 14 years. The other two variants (rs749782387 and rs144607000) were associated with a later onset corresponding to an increase of 5 years and 29 years in mean AO, respectively.

After Bonferroni correction, it was found that rs118039194 was no longer statistically significant.

4.2.2. Variants found in *C1QA* gene

After genotyping of coding and flanking regions of the *C1QA* gene, we found six variants: one in intronic region and five in exonic region, which were already described (**Table 4**); two of those variants were statistically significant (**Table 5)**.

Table 4: Variants found in *C1QA* gene.

dbSNP ID	Location	Region	Effect on protein (Polyphen-2)	Frequency of common allele (HapMap: EUR-CEU)	Frequency of mutant allele (HapMap: EUR-CEU)
rs201693493	c.125 G>A p.Arg42Gln	Exon 2	Missense (Benign)	G (1.000)	A (0.000)
rs369926227	c.163+35C>A	Intron 2	N.A.	C (1.000)	A (0.000)
rs172378	c.276 A>G p.Gly92Gly	Exon 3	Synonymous (Benign)	A (0.601)	G (0.399)
rs180679721	c.295 A>C p.Ile99Leu	Exon 3	Missense (Benign)	A (1.000)	C (0.000)
rs149050968	c.525 C>T p.Ile175Ile	Exon 3	Synonymous (Benign)	C (1.000)	T (0.000)
rs765715625	c.609G>C p.Gly203Gly	Exon 3	Synonymous (Benign)	G (1.000)	C (0.000)

N.A. Non Applicable

Polyphen-2 software shows that all variants found had a benign function in *C1QA* gene, wherein two of them are missense substitutions and the other three are synonymous substitutions.

Table 5: Statistically significant variants in *C1QA* gene.

	Variant	B	Genotype	95% Wald Confidence Interval	P-value
	Intercept	39.11		[35.69; 42.53]	<0.001
1.	rs201693493	- 15.85	GG (Reference) GA	- [13.02; 18.67]	- <0.001
2.	rs149050968	- 10.13	CC (Reference) CT	- [6.97; 13.28]	- <0.001

The two statistically significant polymorphisms (rs201693493 and rs149050968) were associated with a late-onset corresponding to an increase of 16 years and 10 years in mean AO, respectively.

4.2.3. Variants identified in *C1QC* gene

When we analyzed *C1QC* gene, it was possible to find 7 variants, which were already described (**Table 6**). The exonic variant is a synonymous substitution, so it has a benign effect on protein. The 4 statistically significant polymorphisms are presented in **Table 7**.

Table 6: Variants found in *C1QC* gene

dbSNP ID	Location	Region	Effect on protein (Polyphen-2)	Frequency of common allele (HapMap: EUR-CEU)	Frequency of mutant allele (HapMap: EUR-CEU)
rs2935537	c.-3G<A	5'UTR	N.A.	G (1.000)	A (0.000)
rs74909167	c.-13-72G>A	Intron 1	N.A.	G (0.995)	A (0.005)
rs15940	c.126C>T p.Pro126Pro	Exon 2	Synonymous (Benign)	C (0.732)	T (0.268)
rs36049190	c.182-21C>G	Intron 2	N.A.	C (0.914)	G (0.086)
rs200952686	c.*5G>A	3'UTR	N.A.	G (1.000)	A (0.000)
rs201241346	c.*14G>A	3'UTR	N.A.	G (1.000)	A (0.000)
rs9434	c.*21C>A	3' UTR	N.A.	C (0.601)	A (0.399)

Table 7: Statistically significant variants in *C1QC* gene.

	Variant	B	Genotype	95% Wald Confidence Interval	P-value
	Intercept	38.60		[34.69; 42.50]	<0.001
1.	rs2935537	-6.85	GG (Reference) GA	[-11.33; -2.36]	0.003
2.	rs15940	-5.25	CC (Reference) CT/TT	[-10.19; -0.31]	0.037
3.	rs200952686	31.91	GG (Reference) GA	[28.74; 35.07]	<0.001
4.	rs201241346	-10.86	GG (Reference) GA	[-13.95; -7.77]	<0.001

Three of the statistically significant polymorphisms (rs2935537, rs15940 and rs201241346) were associated with an early-onset corresponding to a decrease

between 5 and 11 years in the mean AO, whereas the rs200952686 was associated with a late-onset with an increase of ≈32 years in mean AO. However, after the Bonferroni correction, rs15940 was no longer statistically significant.

4.3. *In silico* analyses

To unravel possible regulatory effects of the variants found, we used SNP Functional Prediction (FuncPred) software, as a first approach. This software allows us to explore possible changes at the level of transcription factors binding sites (TFBSs), miRNAs, and splicing sites.

Our analyses included the variants found but also polymorphisms in linkage disequilibrium (LD) with the ones we identified. The study of LDsnps in the human genome allows the identification and characterization of additional genetic variants in order to cover the entire genetic information of the genes studied [74].

4.3.1. TFBSs

The variants found to change TFBS are presented in **Table 8.**

Table 8: Variants predicted to alter TFBS.

Genes	SNPs	Model	TF	Score
RAC1	rs3729790	M00453	IRF-7	4.5
	rs702483	M00011	Evi-1	4.0
C1QC	rs2935537	M00531	NERF1a	3.8
	rs294180 (LDsnp rs9434)	M00732	CAT8	3.9

The rSNP-MAPPER software highlighted that rs3729790 may alter the binding of IRF-7 and rs702483 affects the consensus motifs of Evi-1, both in *RAC1* gene.

In the *C1QC* gene, we found that rs2935537 may change the binding of NERF1a and rs294180 (LDsnp rs9434) leads to alteration in CAT8.

4.3.2. miRNA target predictions

In **Table 9**, we present the miRNA target predictions for WT and minor allele of rs9374 (LDsnp rs3729790) and rs6951997 (LDsnp rs3729796) in the *RAC1* gene.

Table 9: miRNA target predictions associated to SNPs using miRDIP software.

Gene	SNPs	LDsnps	WT allele	Minor allele
RAC1	rs9374	rs3729790	hsa-miR-4427 hsa-miR-3653-3p hsa-miR-5187-3p hsa-miR-5585-3p hsa-miR-4680-3p	hsa-miR-4539
	rs6951997	rs3729796	hsa-miR-4733-3p hsa-miR-6776-3p hsa-miR-1226-3p	hsa-miR-4773

We found in *RAC1* gene two SNPs in LD with rs3729790 and rs3729796 that create a new possibly target site for hsa-miR-4539 and hsa-miR-4773, respectively.

In *C1QC* gene, FuncPred software predicted changes at the level of miRNA. However, when we used other miRNA prediction software, particularly mirDIP software, we did not find any significant results.

4.3.3. Splicing activity

Using FuncPred software was possible to find that rs172378 of *C1QA* gene causes alterations in Exonic Splicing Enhancer (ESE) or Exonic Splicing Silencer (ESS) motifs. To further explore these results, we used Human Splicing Finder (HSF) software to detect specific alterations in splice sites (**Table 10**) and ESEfinder software to detect modifications on ESE motifs (**Table 11**).

Table 10: Splice sites predictions by HSF software

Gene	SNPs	Splice Sites	WT CV	Mutant CV	Variation	Effect
C1QA	rs172378	acceptor site	75.63	46.68	-38.28	**Site Broken**
		branch point	73.89	44.26	------	**Site Broken**
	rs201693493	acceptor site	54.12	83.06	53.47	**New site**
		acceptor site	76.51	76.32	-0.25	------

CV – Consensus Value

We found two SNPs in *C1QA* gene with possibly effect in the splicing mechanism. rs172378 reduce the CV of an acceptor site (CV= 46.68) and a branch point comparing to WT (CV= 44.26). For rs201693493 we found a mild reduction in an acceptor site but more interestingly this SNP creates a significant increase in another acceptor site (CV>80) comparing to WT.

Table 11: ESE alterations predictions by ESEfinder software.

Gene	SNPs	SR Protein	Motif Threshold	WT Score	Mutant Score
C1QA	rs172378	SC35 site	2.38	3.78	5.54
		SF2/ASF site	1.96	2.29	3.22
		SF2/ASF (IgM-BRCA1) site	1.87	1.93	2.17
	rs149050968	SRp40 site	2.67	2.70	3.06

SR proteins – Ser/Arg-rich proteins

Despite both SNPS, rs172378 and rs149050968 in *C1QA*, increasing the motif score compared to WT they don't affect the SR protein sites.

For *C1QC* gene, when we analyzed variants in FuncPred software, it was possible to find splicing predictions. Nevertheless, when we analyzed the variants in prediction software splicing factors, we did not find significant results.

4.4. Summary table of result

Table 12: Summary of in silico regulatory analyses caused by variants found.

Genes	SNPs	LDsnps	TFBSs	miRNA target site		Splicing activity	
				WT	Mutant	ESE	HSF
RAC1	rs3729790	----	IRF-7	----	----	----	----
	rs9374	rs3729790	----	hsa-miR-4427 hsa-miR-3653-3p hsa-miR-5187-3p hsa-miR-5585-3p hsa-miR-4680-3p	hsa-miR-4539	----	----
	rs702483	----	Evi-1	----	----	----	-----
	rs6951997	rs3729796	----	hsa-miR-4733-3p hsa-miR-6776-3p hsa-miR-1226-3p	hsa-miR-4773	----	-----
C1QA	rs172378	----	----	----	----	SC35 SF2/ASF SF2/ASF (IgM-BRCA1)	Site Broken
	rs201693493	----	----	----	----	----	New site
	rs149050968	----	----	----	----	SRp40	----
C1QC	rs2935537	----	NERF1a	----	----	----	----
	rs294180	rs9434	CAT8	----	----	----	----

4.5. Interactions between genes

We assessed a putative *C1QA-C1QC* gene interaction using MDR and it was possible to find two significant interactions, in the group of patients with late-onset when comparing with early-onset patients. The first interaction found was between rs172378 of the *C1QA* gene and rs15940 of the *C1QC* gene, with testing balance accuracy (TBA) of 0.54 and a cross-validation consistency (CVC) of 8/10. After we performed a permutation test, we obtained a significant TBA of 0.67 and a CVC of 10/10 (p= 0.002). The second interaction found was between rs172378 of the *C1QA* gene and rs15940 plus rs74909167 of the *C1QC* gene with a TBA of 0.61 and a CVC of 10/10. After permutation test, the values mentioned above were the same and were statistically significant (p=0.04).

When we assessed the *RAC1-C1QA-C1QC* gene interactions, we did not found any significant results.

5. DISCUSSION

Age-at-onset (AO) variability is a hallmark of FAP ATTRV30M. Unlike other neurodegenerative diseases, characterized by dynamic expansions, as Huntington's disease (HD), the point mutation found in *TTR* does not determine AO variation. To explore other possible genetic factors that modulate AO in FAP ATTRV30M, we performed a family-centered approach and focused on candidate genes that can affect functional pathways in the course of disease.

5.1. Candidate genes associated with AO variability

5.1.1. *RAC1* gene

We found five significant variants in this gene associated with an early-onset. Therefore we hypothesize that these variants could be a risk factor for AO in FAP ATTRV30M patients. On the other hand, two other significant variants were associated with a late AO, which suggests that these variants may have a protective effect.

However, it is important to note that in all the variants detected, with the exception of rs3729789, the heterozygous genotype only appears once in our sample, which is in accordance with the observed genotypic frequencies of the European population in the HapMap and 1000 genomes databases, and therefore we think that this result is worth mentioning. In addition, as most variants found are rare we analyzed the heterozygous and the rare homozygous together in order to achieve a higher statistical power.

Rac has been implicated in axon growth and guidance [75], whereby has a protective function against apoptosis-mediated neurodegeneration [61]. Taking this into account, a study carried out by Silva et al using *D. melanogaster* expressing TTRV30M demonstrated that the presence of *Rac1* plays a protective factor against rough eye and severe phenotype. On the other hand, flies expressing *Rac1* dominant-negative (DN) lead to the appearance of rough eye phenotype and the expression of constitutively active form of Rac1 is fatal [66]. This finding emphasizes the importance of *Rac1* regulation during development and may explain the absence of homozygous in our sample.

In addition, dysregulation of Rho GTPases pathways (in which *RAC1* is involved) appears to underlie the etiology of several neurodegenerative diseases such as Alzheimer's disease (AD), Amyotrophic Lateral Sclerosis (ALS), Huntington's disease (HD) and Parkinson's disease (PD). Therefore, using these Rho GTPases and

more specifically *Rac1* as targets, it can be a useful therapeutic strategy for the treatment of these disorders [76].

Our results are in accordance with these findings, since rs749782387 and rs144607000 may modulate the protective role of *RAC1*, leading to a delay in disease onset. On the contrary, the other variants found seem to interfere with the active state of *Rac1*, potentially promoting GDP-binding, leading to its inactive state, and consequently to an early disease onset.

Interestingly, rs144425992 is a coding variant leading to a decrease in AO, in approximately 14 years. In fact it is interesting to note that the carrier of this variant has a very early-onset (23 years). Therefore, this variant can be a genetic risk factor for AO in FAP ATTRV30M patients.

To further explore the role of *RAC1* gene in FAP ATTRV30M, we performed *in silico* analysis.

We found that rs3729790 may lead to an alteration of the binding of the IRF-7 factor and can also interfere with the putative binding to target site of the hsa-miR-4539, when the minor allele of rs9374 (LDsnp rs3729790) is present.

Some studies demonstrated that IRF7 may interact and be acetylated by histone acetyltransferases (HATs) p300 and CBP, leading to the increase of transcriptional levels [77]. In this study, we found that rs3729790 probably leads to an early-onset, despite there is no statistical evidence to prove this fact (data not shown).

Taking into account this, we can assume that this SNP may lead to an alteration of the binding of the IRF-7 factor decreasing the transcription of *RAC1* gene. This fact is reinforced by enabling the binding of a new miRNA, hsa-miRNA-4539 which does not bind in the WT variant. These two mechanisms together possibly leads to a downregulation of *RAC1* expression contributing for an early AO.

We also found other five miRNAs with the putative binding of target sites for the WT allele of rs9374 (LDsnp rs3729790). However, the minor allele of this variant does not interfere with these miRNAs, leading us to conclude that these five miRNAs may not be related with AO variability of FAP ATTRV30M.

We also found that rs702483 may change the consensus motifs of the Evi-1 factor. The Evi-1 may activate transcription or more generally may act as a transcriptional repressor [78] dependent on DNA binding through either the first or the second zinc-finger domain [79]. It has been thought that many transcriptional repressors interact with histone deacetylases (HDACs) and contribute to tight nucleosomal packing of DNA [78]. Since this transcription factor may contribute to the decrease of transcriptional levels it can lead to a downregulation of *RAC1* expression. However, we found that this variant is associated with an increase in the mean AO in

human FAP ATTRV30M patients (data not shown) which can be due to an alteration in Evi-1 binding leading Evi-1 activate transcription which promote *RAC1* expression and consequently acting as protective modifier.

Additionally, we also found that rs6951997 (LDsnp rs3729796) is related with miRNA binding alterations as hsa-miR-4773 leads to an alteration of target binding sites. Since rs3729796 probably tend to increase mean AO, we can hypothesize that this miRNA leads to an upregulation of *RAC1* expression, and therefore plays a protective role for FAP ATTRV30M carriers.

5.1.2. *C1QA* and *C1QC* genes

In a previous study carried out in Cypriot FAP ATTRV30M patients, Dardiotis et al., 2009 found that the genotype AA of rs172378 in the *C1QA* gene was associated with an early AO. Also, in the same study, it was described that rs9434 in the *C1QC* gene modulates AO, leading to an increase in AO [68]. Therefore, these results show that these genes may be considered genetic modifiers in FAP ATTRV30M Cypriot population. Importantly, this study did not take into account family structure [80]. We found in our study two statistically significant variants in *C1QA* gene associated with late AO and we can hypothesize that these variants could be a protective factor for FAP ATTRV30M patients.

Regarding the variants found in *C1QC* gene, three are statistically significant associated with an early AO and one variant associated with a late AO.

Similarly to what was found for *RAC1* gene, for *C1QA* and *C1QC* genes, in some variants the heterozygous genotype only appears once in our sample, which is in accordance with the genotypic frequencies described in HapMap and 1000 genomes databases. However, as we found significant results for these genotypes, we considered that these results should be reported in order to assess their implications in AO variation and to unravel more data on SNPs frequency in FAP ATTRV30M Portuguese population.

In our sample we also studied the variants found in the Cypriot population, but the results obtained were not statistically significant, leading us to conclude that these polymorphisms are not related with AO variability in our population. This evidence suggests that different approaches can lead to different results. Furthermore, it is important the different genetic background between populations that can result in different disease mechanisms. Additionally, this conclusion had already been raised in a previous study of our group, when they analyzed the role of *APCS* gene in AO variability, and concluded that our results were different from those carried out by

Dardiotis et al., 2009. Furthermore, their samples' size was quite small when compared to ours, showing that increasing sample size will increase statistical power.

C1Q complement genes have demonstrated to play an important role in other neurodegenerative diseases as Alzheimer's disease (AD) and Parkinson's disease (PD) [81]. The complement system can have a double role, beneficial or damaging, depending on the associated pathophysiological mechanisms as in some cases it can lead to tissue damage [70]. On contrary, in AD it was demonstrated under some circumstances, that C1Q plays a protective role and downregulates the inflammation process, in the early stages of neuronal injury as well as amyloid induced neurotoxicity [82].

In our sample occurs a mechanism similar to that already described for Alzheimer's patients as variants found in *C1QA* gene seem to be involved in an increase in mean AO as well as the rs200952686 in the *C1QC* gene.

In contrast, the remaining variants that were found in *C1QC* gene may play a risk factor for FAP ATTRV30M patients, and possibly leads to tissue damage.

Regarding *in silico* analyses for *C1QA* gene, we found that rs172378 and rs20163493 presents significant splicing alterations.

In the region of rs172378, a weaker splice site existent in WT variant can possibly result in an inactive conformation when the mutant variant is introduced. In the analysis of rs201693493 we found a different effect, with strong active splice site in the mutant sequence, leading to the creation of a new splice site. In most cases, the variants which implies unnatural splice sites or intron retention leads to the premature termination codons (PTCs) into the mRNA, resulting in degradation by nonsense-mediated decay (NMD) and loss of function of the minor allele [83]. Thus, this may explain the association of this variant with a late-onset of the disease.

Taking into account the scores obtained in ESEfinder and the results presented in **Table 11**, we concluded that there is no disruption or ESE creation due to the variants found.

For *C1QC* gene, we found that rs2935537 may alter the binding of NERF-1a factor. NERF (new-ETS-related factor) belong to ETS-domain family which plays an important role in the transcriptional control and may act as either transcriptional activators or repressors. So, this transcription factor seems to be involved in the regulation of cellular proliferation, differentiation both in embryonic and adult development, and in tissue specificity [84].

Some studies demonstrated that NERF-1a and NERF-1b lack significant transactivation capabilities, so they act as negative regulators. In addition, one study

revealed that NERF1 act as a topoisomerase II, and both repressed basal transcription [85].

Since NERF1a contribute to the decrease of transcriptional levels, promoting a downregulation of *C1QC* gene expression, we can postulate that rs2935537 plays a risk factor for the FAP ATTRV30M patients. In this case, it seems that complement levels confer a protective effect, which is lost in the presence of the polymorphism. Thus, this variant leads to a decrease in complement levels, which according to some authors may be associated with an increase of the autoimmune response. This argument corroborates the decrease in the mean AO caused by this SNP.

We also found that rs294180 (LDsnp rs9434) may result in the alteration of the CAT8 factor binding.

Despite rs9434 is associated with a late AO in our sample (data not shown), these results were not statistically significant. However, the delayed onset of symptoms in these patients can partly be explained by the role of CAT8 factor in derepression of gluconeogenesis enzymes [86], leading to better glucose delivery in the brain [87], and it may play a protective role in FAP ATTRV30M carriers.

5.2. The importance of in silico analyses: functional implications

Deciphering the molecular basis of genetic diseases is supported by the availability of high throughput strategies for the identification of pathways associated with disease mechanism [88]. To explore functional implications of SNPs found, we performed *in silico* analyses including alterations in Transcription Factors Binding (TFBSs), miRNA target sites and splicing activity. *In silico* softwares have in the past few years emerged as an important tool to analyze the functional impact of the variants [89].

Some approaches revealed that functional TFBSs can be considered as the on/off switches for gene transcription, so the alterations in these factors can lead to an alteration in gene expression, leading to a human disease [90].

Also, mutations in exonic splicing enhancers (ESE) and exonic splicing silencers (ESS) may deregulate splicing machinery [83]. Importantly, approximately 15% of mutations that cause genetic disease affect pre-mRNA splicing. In addition, synonymous coding SNPs can disrupt (or eventually create) ESEs and ESSs; create new splice sites or reinforce cryptic ones; alter pre-mRNA secondary structures; and, possibly cause changes in RNA Polymerase II [91].

Additionally, miRNAs are being intensively studied in another neurodegenerative diseases [92], [93], [94], [95] have been the focus of many studies since it has been thought that SNPs in the 3' UTR may disrupt the miRNA-mRNA interaction or create a new target site for a miRNA, leading to the alteration of the expression of target gene **(Figure 9)** [96]. In FAP ATTRV30M, the presence of miRNA target sites in 3'UTR of *TTR* gene had already been studied [97], [98], but the results are still not evident.

All these studies reinforce the importance of the in *silico* analyses carried out in this study that was validated using different softwares, since the usage of these programs should be done with a critical view, because they can present different results from each other according to the algorithms and the thresholds used.

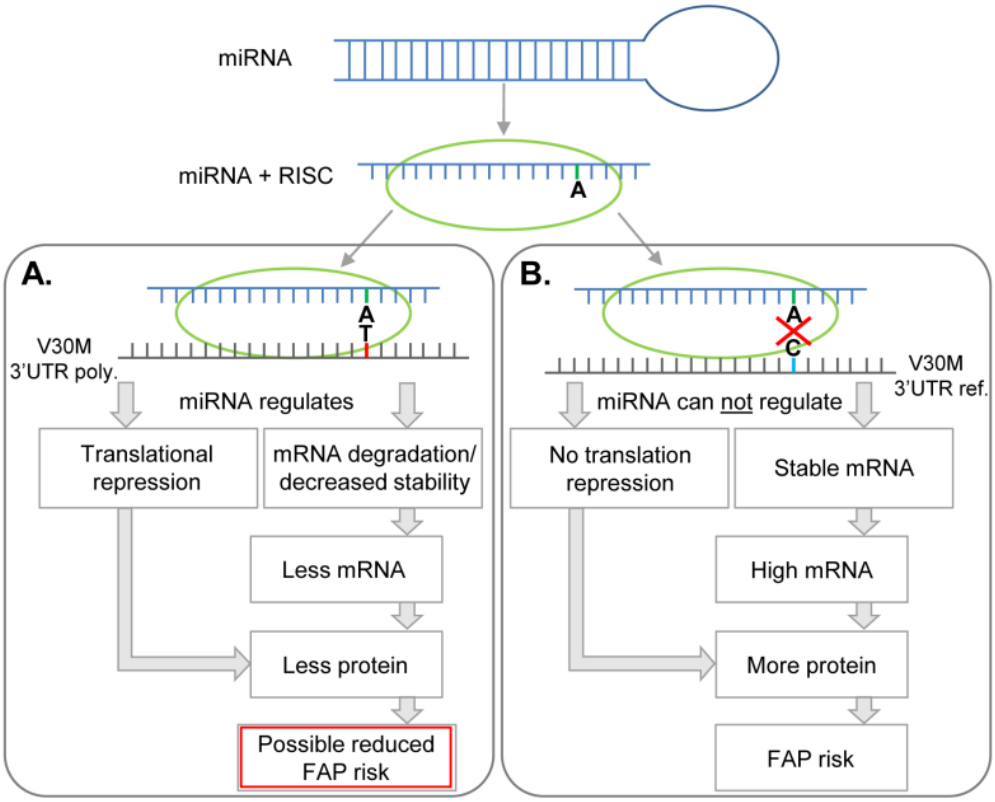

Figure 9: Regulation of gene expression mediated by miRNA: A) V30M allele with a polymorphism in 3' UTR; B) V30M allele with WT 3'UTR.

5.3. Interaction between genes

Furthermore, we also performed a MDR analysis for detection of gene-gene interactions, which is a powerful statistical tool of multilocus data reduction to improve the detection of genotypic combinations, in order to predict disease risk [99].

MDR studies are very important since we can detect high-order interaction even in the absence of statistically significant results and in small sample sizes. Additionally, combination of cross-validation and permutation testing diminishes the presence of false positive results. So, this study gives us confidence in our results and allows us to find highly significant interaction between SNPs which may act together in the modulation of AO variability FAP ATTRV30M patients [100].

We conclude that there is a strong interaction between *C1QA* and *C1QC* genes. This fact can be confirmed by the C1Q molecule structure and the three polypeptide chains, A, B and C, are the product of three distinct genes organized to form 6 triple helical strands (**Figure 10**). Chains A and B form a dimer through formation of disulfide bridges between the half-cysteine residues located at A-4 and B-4, whereas the C chain forms a disulfide-linked with the C chain of the next strand through cysteine residues found at position C-4. Posteriorly, the two dimer strands form a doublet, and three doublets are linked by non-covalent forces [101]. Therefore, each A-B dimer associates non-covalently with a C-chain to form a heterotrimeric collagen-like triple helix [102]. These data shows the molecular interaction between *C1QA* and *C1QC* reinforcing our statistical findings and showing that these two genes may act together as genetic modifiers.

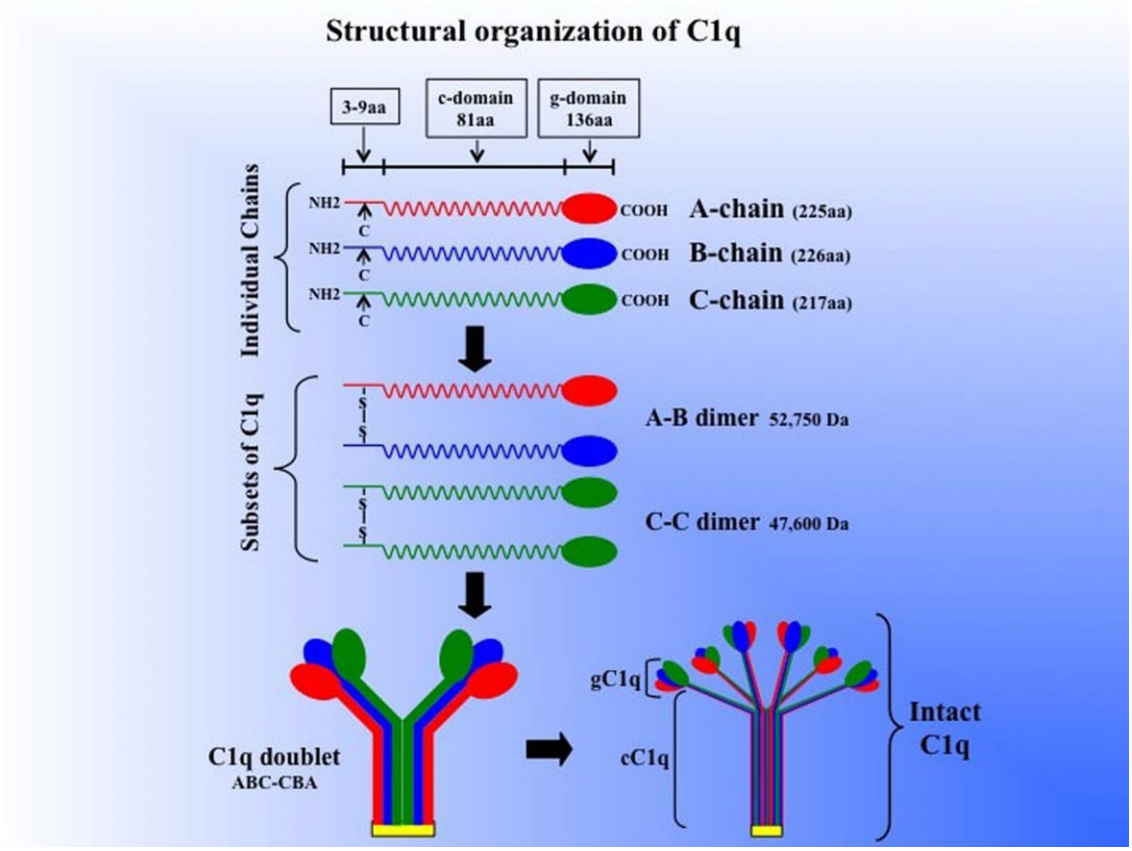

Figure 10: Structure and assembly of C1q.

In addition, we performed a MDR analysis for detection of *C1QA-C1QC-RAC1* interactions, and we conclude that *RAC1* gene do not interact neither with *C1QA* gene nor with *C1QC* gene.

We have to pay attention for the possibility of other genes and other SNPs within these genes being associated with AO variability in FAP ATTRV30M. But importantly, our study revealed several strengths: a large sample size for a rare disease and a family centered approach which prevents population stratification, as opposed to case-control study [103]; the use of GEE analysis, since corrects for familial correlations of AO and it allows a greater power to detect a statistically significant effect [104]; the inclusion of multiple testing corrections to avoid errors type I [105]; and finally, the clinical observation of all patients by the same team of neurologists, minimizing differences in the establishment of AO.

6. CONCLUSIONS

This study aimed to identify possible genetic modifiers associated to FAP ATTRV30M pathways, in order to assess their involvement in AO variability. After genotyping a large number of individuals, we found a significant association of some variants in candidate genes with AO variability.

Concerning *RAC1* gene:

- We found five significant SNPs associated with an earlier onset and two SNPs associated with a later onset in FAP ATTRV30M.
- rs3729790 may alter the binding to the IRF-7 factor, increasing the transcriptional levels, and it may act as a risk factor in FAP ATTRV30M carriers. On the other hand, rs702483 may change the consensus motif of the Evi-1 factor, which may cause a downregulation of *RAC1* expression, so it may act as a protective factor in FAP ATTRV30M disease.
- miRNAs also seem interfere with the putative binding of target sites: the hsa-miR-4539 related with rs9374 (LDsnp rs3729790) and the hsa-miR-4773 related with rs6951997 (LDsnp rs3729796).
- Although there is no statistical evidence regarding association with AO to support the results obtained by *in silico* analysis, our results open new perspectives that deserve to be further explored since new disease targets can be unraveled.

Regarding complement C1Q genes:

- For *C1QA* gene we found two SNPs associated with a late-onset, and for *C1QC* gene three SNP associated with an early-onset, and one SNP was associated with a late-onset.
- For *C1QA* gene we found that rs172378 was implicated in the inactivation of an acceptor site and branch point site. On the contrary, the analysis of rs201693493 showed the creation of a new splice site, with probability to alter splicing activity in this gene.
- For *C1QC* we can conclude that two variants *(rs29*35537 and rs294180 (LDsnp 9434)) caused alterations in the NERF1a and CAT8 TFBs, respectively. The first one contribute to the decrease of transcriptional levels and it plays a possible risk factor for the FAP ATTRV30M patients, while the second was involved in expression of a large number of genes required for metabolism of nonfermentable carbon sources.

With this study, we confirmed the role of *C1Q* complement genes in FAP ATTRV30M Portuguese patients and we revealed, for the first time, the role of *RAC1* gene as a genetic modifier of AO variability. Importantly, we found very interesting and unreported results in the *in silico* analysis since alterations in these mechanisms can lead to dysregulation of gene expression, which can have an impact in AO and phenotypic variability.

We concluded that there is a strong interaction between *C1QA* and *C1QC* genes showing that these two genes may act together as genetic modifiers.

In conclusion, our results are important to understand the differences in AO between members of the same family unraveling disease mechanisms which may have implications in genetic counselling and in the discovery of new biomarkers.

7. FUTURE PERSPECTIVES

As futures perspectives we intend to:

- Increase our sample to draw conclusions about the impact of rare polymorphisms in AO variation in FAP ATTRV30M.
- Study the genetic interaction of *RAC1* gene with *TTR* to further explore the results found in *Drosophila* models.
- Confirm *in vivo* studies if the presence of the new TFBs (IRF-1 and Evi-1) found in the *RAC1* gene have a significant impact on disease progression.
- Study the possible presence of rs9374 (LDsnp rs3729790) and rs6951997 (LDsnp rs3729796) in a larger number of FAP ATTRV30M carriers and study the association of the putative target site for hsa-mir-4539 and hsa-mir-4773 (respectively) with the phenotype;
- Apply the information about *C1QA* mRNA splicing alterations in therapeutics, like antisense oligonucleotides;
- Perform a SNP profile for *RAC1, C1QA* and *C1QC* genes in livers of patients already transplanted to understand the differences in disease progression found in these patients;

8. BIBLIOGRAPHY

1. Fandrich, M., *On the structural definition of amyloid fibrils and other polypeptide aggregates.* Cell Mol Life Sci, 2007. **64**(16): p. 2066-78.
2. Plante-Bordeneuve, V. and G. Said, *Familial amyloid polyneuropathy.* Lancet Neurol, 2011. **10**(12): p. 1086-97.
3. Lobato, L. and A. Rocha, *Transthyretin amyloidosis and the kidney.* Clin J Am Soc Nephrol, 2012. **7**(8): p. 1337-46.
4. Hund, E., et al., *Transthyretin-associated neuropathic amyloidosis. Pathogenesis and treatment.* Neurology, 2001. **56**(4): p. 431-5.
5. Sipe, J.D., et al., *Nomenclature 2014: Amyloid fibril proteins and clinical classification of the amyloidosis.* Amyloid, 2014. **21**(4): p. 221-4.
6. Westermark, P., et al., *A primer of amyloid nomenclature.* Amyloid, 2007. **14**(3): p. 179-83.
7. Luís, M.d.L.S., *Polineuropatia Amiloidótica Familiar do tipo Português: do artigo original ao futuro*, in *Sinapse: Publicação da Sociedade Portuguesa de Neurologia*, J. Barros, Editor. 2006: Porto.
8. Andrade, C., *A peculiar form of peripheral neuropathy; familiar atypical generalized amyloidosis with special involvement of the peripheral nerves.* Brain, 1952. **75**(3): p. 408-27.
9. Andrade, C., et al., *The genetic aspect of the familial amyloidotic polyneuropathy. Portuguese type of paramyloidosis.* Humangenetik, 1969. **7**(2): p. 163-75.
10. Becker, P.E., et al., *[Paramyloidosis of Peripheral Nerves in Portugal].* Z Mensch Vererb Konstitutionsl, 1964. **37**: p. 329-64.
11. Costa, P.P., A.S. Figueira, and F.R. Bravo, *Amyloid fibril protein related to prealbumin in familial amyloidotic polyneuropathy.* Proc Natl Acad Sci U S A, 1978. **75**(9): p. 4499-503.
12. Saraiva, M.J., et al., *Amyloid fibril protein in familial amyloidotic polyneuropathy, Portuguese type. Definition of molecular abnormality in transthyretin (prealbumin).* J Clin Invest, 1984. **74**(1): p. 104-19.
13. Tawara, S., et al., *Identification of amyloid prealbumin variant in familial amyloidotic polyneuropathy (Japanese type).* Biochem Biophys Res Commun, 1983. **116**(3): p. 880-8.
14. Hellman U, L.H., Westermark P, Stafberg C, Nahi H, Tachlinski S, Guggi M, Flogegard F, Hamid M, Escher S, Suhr O, *A genealogical and clinical study of the phenotypical variation within the swedish transthyretin His88Arg (p.His108Arg) amyloidosis family.* European Journal of Medical Genetics, 2015. **58**(4): p. 211-215.

15. Benson, M.D. and J.C. Kincaid, *The molecular biology and clinical features of amyloid neuropathy.* Muscle Nerve, 2007. **36**(4): p. 411-23.
16. Misumi Y, U.M., Obayashi K, Jono H, Su Y, Yamashita T, Ohshima T, Ando Y, Uchino M, *Relationship between amyloid deposition and intracellular structural changes in familial amyloidotic.* Human Pathology, 2012. **43**(1): p. 96-104.
17. Blake, C.C., et al., *Strjcture of human plasma prealbumin at 2-5 A resolution. A preliminary report on the polypeptide chain conformation, quaternary structure and thyroxine binding.* J Mol Biol, 1974. **88**(1): p. 1-12.
18. Hou, X., M.I. Aguilar, and D.H. Small, *Transthyretin and familial amyloidotic polyneuropathy. Recent progress in understanding the molecular mechanism of neurodegeneration.* FEBS J, 2007. **274**(7): p. 1637-50.
19. Saraiva, M.J., et al., *Transthyretin deposition in familial amyloidotic polyneuropathy.* Curr Med Chem, 2012. **19**(15): p. 2304-11.
20. Barreiros, A.P., P.R. Galle, and G. Otto, *Familial amyloid polyneuropathy.* Dig Dis, 2013. **31**(1): p. 170-4.
21. Estefania Azevedo, P.F.S., Fernando Palhano, Carolina A. Braga and Debora Foguel, *Transthyretin-Related Amyloidoses: A Structural and Thermodynamic Approach*, in *Amyloidosis*, D. Feng, Editor. 2013, In Tech: Rio de Janeiro, Brazil.
22. Swiecicki, P.L., et al., *Hereditary ATTR amyloidosis: a single-institution experience with 266 patients.* Amyloid, 2015. **22**(2): p. 123-31.
23. Hirai, T., et al., *Transthyretin-related familial amyloid polyneuropathy: evaluation of CSF enhancement on serial T1-weighted and fluid-attenuated inversion recovery images following intravenous contrast administration.* AJNR Am J Neuroradiol, 2005. **26**(8): p. 2043-8.
24. Ando, Y., et al., *Guideline of transthyretin-related hereditary amyloidosis for clinicians.* Orphanet J Rare Dis, 2013. **8**: p. 31.
25. Rapezzi, C., et al., *Disease profile and differential diagnosis of hereditary transthyretin-related amyloidosis with exclusively cardiac phenotype: an Italian perspective.* Eur Heart J, 2013. **34**(7): p. 520-8.
26. Almeida, M.R., et al., *Comparative studies of two transthyretin variants with protective effects on familial amyloidotic polyneuropathy: TTR R104H and TTR T119M.* Biochem Biophys Res Commun, 2000. **270**(3): p. 1024-8.
27. Gustavsson, A., et al., *Amyloid fibril composition and transthyretin gene structure in senile systemic amyloidosis.* Lab Invest, 1995. **73**(5): p. 703-8.

28. Conceição, I., *Clínica e História Natural da Polineuropatia Amiloidótica Familiar*, in *Sinapse: Publicação da Sociedade Portuguesa de Neurologia*, J. Barros, Editor. 2006: Porto.
29. Amaral, J.M.B.d., *Insights in Familial Amyloidotic Polyneuropathy Portuguese Type (Val30Met) Salivary Glands*. 2011, Instituto de Ciências Biomédicas Abel Salazar: Porto.
30. Ando, Y. and O.B. Suhr, *Autonomic dysfunction in familial amyloidotic polyneuropathy (FAP).* Amyloid, 1998. **5**(4): p. 288-300.
31. Lobato, L., *Portuguese-type amyloidosis (transthyretin amyloidosis, ATTR V30M).* J Nephrol, 2003. **16**(3): p. 438-42.
32. Ando, Y., M. Nakamura, and S. Araki, *Transthyretin-related familial amyloidotic polyneuropathy.* Arch Neurol, 2005. **62**(7): p. 1057-62.
33. Julião, O.F., *Aspectos clínicos da neuropatia amilóide. Considerações a propósito de 14 casos.* Rev Paul Med, 1963. **63**: p. 349-352.
34. Nakazato, M., et al., *Structurally abnormal transthyretin causing familial amyloidotic polyneuropathy in Sweden.* Clin Chim Acta, 1987. **167**(3): p. 341-2.
35. Araki, S., et al., *Polyneuritic amyloidosis in a Japanese family.* Arch Neurol, 1968. **18**(6): p. 593-602.
36. Grateau, G., et al., *Late-onset familial amyloid polyneuropathy with the TTR Met 30 mutation in France.* Clin Genet, 1993. **43**(3): p. 143-5.
37. Reines, J.B., et al., *Epidemiology of transthyretin-associated familial amyloid polyneuropathy in the Majorcan area: Son Llatzer Hospital descriptive study.* Orphanet J Rare Dis, 2014. **9**: p. 29.
38. Benson, M.D. and F.E. Dwulet, *Identification of carriers of a variant plasma prealbumin (transthyretin) associated with familial amyloidotic polyneuropathy type I.* J Clin Invest, 1985. **75**(1): p. 71-5.
39. Liu, J.Y., et al., *Clinical and histopathological features of familial amyloidotic polyneuropathy with transthyretin Val30Ala in a Chinese family.* J Neurol Sci, 2011. **304**(1-2): p. 83-6.
40. Holt, I.J., et al., *Molecular genetics of amyloid neuropathy in Europe.* Lancet, 1989. **1**(8637): p. 524-6.
41. Bonaiti, B., et al., *Parent-of-origin effect in transthyretin related amyloid polyneuropathy.* Amyloid, 2009. **16**(3): p. 149-50.
42. Conceicao, I., *Clinical features of TTR-FAP in Portugal.* Amyloid, 2012. **19 Suppl 1**: p. 71-2.

43. Sousa, A., et al., *Genetic epidemiology of familial amyloidotic polyneuropathy (FAP)-type I in Povoa do Varzim and Vila do Conde (north of Portugal).* Am J Med Genet, 1995. **60**(6): p. 512-21.
44. Lemos, C., et al., *Overcoming artefact: anticipation in 284 Portuguese kindreds with familial amyloid polyneuropathy (FAP) ATTRV30M.* J Neurol Neurosurg Psychiatry, 2014. **85**(3): p. 326-30.
45. Sousa, A., *Epidemiologia Genética da Polineuropatia Amiloidótica Familiar*, in *Sinapse: Publicação da Sociedade Portuguesa de Neurologia*, J. Barros, Editor. 2006: Porto.
46. Bonaiti, B., et al., *TTR familial amyloid polyneuropathy: does a mitochondrial polymorphism entirely explain the parent-of-origin difference in penetrance?* Eur J Hum Genet, 2010. **18**(8): p. 948-52.
47. Coelho, T., *Diagnóstico Clínico da Polineuropatia Amiloidótica Familiar*, in *Sinapse: Publicação da Sociedade Portuguesa de Neurologia*, J. Barros, Editor. 2006: Porto.
48. Costa, P.P.e., *Diagnóstico Molecular da Polineuropatia Amiloidótica Familiar*, in *Sinapse: Publicação da Sociedade Portuguesa de Neurologia*, J. Barros, Editor. 2006: Porto.
49. Sirpa Soini, D.I., Violeta Anastasiadou, Ségolène Aymé, Suzanne Braga, Martina Cornel, Domenico A. Coviello, Gerry Evers-Kiebooms, Joep Geraedts, et al, *The interface between assisted reproductive technologies and genetics: technical, social, ethical and legal issues.* European Journal of Human Genetics, 2006. **14**: p. 588-645.
50. Adams, D., *Hereditary and acquired amyloid neuropathies.* J Neurol, 2001. **248**(8): p. 647-57.
51. Obici, L., et al., *Recommendations for presymptomatic genetic testing and management of individuals at risk for hereditary transthyretin amyloidosis.* Curr Opin Neurol, 2016. **29 Suppl 1**: p. S27-35.
52. Holmgren, G., et al., *Biochemical effect of liver transplantation in two Swedish patients with familial amyloidotic polyneuropathy (FAP-met30).* Clin Genet, 1991. **40**(3): p. 242-6.
53. Okamoto, S., et al., *Liver transplantation for familial amyloidotic polyneuropathy: impact on Swedish patients' survival.* Liver Transpl, 2009. **15**(10): p. 1229-35.
54. Said, G., S. Grippon, and P. Kirkpatrick, *Tafamidis.* Nat Rev Drug Discov, 2012. **11**(3): p. 185-6.
55. Berk, J.L., et al., *Repurposing diflunisal for familial amyloid polyneuropathy: a randomized clinical trial.* JAMA, 2013. **310**(24): p. 2658-67.

56. Adams, D., et al., *FAP neuropathy and emerging treatments.* Curr Neurol Neurosci Rep, 2014. **14**(3): p. 435.
57. Santos, D., et al., *Variants in RBP4 and AR genes modulate age at onset in familial amyloid polyneuropathy (FAP ATTRV30M).* Eur J Hum Genet, 2015.
58. Soares, M.L., et al., *Susceptibility and modifier genes in Portuguese transthyretin V30M amyloid polyneuropathy: complexity in a single-gene disease.* Hum Mol Genet, 2005. **14**(4): p. 543-53.
59. DeGeer, J. and N. Lamarche-Vane, *Rho GTPases in neurodegeneration diseases.* Exp Cell Res, 2013. **319**(15): p. 2384-94.
60. Tan, W., et al., *An essential role for Rac1 in endothelial cell function and vascular development.* FASEB J, 2008. **22**(6): p. 1829-38.
61. Marei, H. and A. Malliri, *Rac1 in human diseases: The therapeutic potential of targeting Rac1 signaling regulatory mechanisms.* Small GTPases, 2016: p. 1-25.
62. Govek, E.E., S.E. Newey, and L. Van Aelst, *The role of the Rho GTPases in neuronal development.* Genes Dev, 2005. **19**(1): p. 1-49.
63. Hall, A., *Rho GTPases and the actin cytoskeleton.* Science, 1998. **279**(5350): p. 509-14.
64. Nakayama, A.Y., M.B. Harms, and L. Luo, *Small GTPases Rac and Rho in the maintenance of dendritic spines and branches in hippocampal pyramidal neurons.* J Neurosci, 2000. **20**(14): p. 5329-38.
65. Kozma, R., et al., *Rho family GTPases and neuronal growth cone remodelling: relationship between increased complexity induced by Cdc42Hs, Rac1, and acetylcholine and collapse induced by RhoA and lysophosphatidic acid.* Mol Cell Biol, 1997. **17**(3): p. 1201-11.
66. Silva, M.O.d., *Dissecting the role of cytoskeleton remodelling in a Drosophila model of TTR-induced neurodegeneration*, in *Departamento de Química e Bioquímica*. 2015, Faculdade de Ciências da Universidade do Porto: Porto.
67. Nayak, A., et al., *The non-classical functions of the classical complement pathway recognition subcomponent C1q.* Immunol Lett, 2010. **131**(2): p. 139-50.
68. Dardiotis, E., et al., *Complement C1Q polymorphisms modulate onset in familial amyloidotic polyneuropathy TTR Val30Met.* J Neurol Sci, 2009. **284**(1-2): p. 158-62.
69. Lambris, J.D., ed. *Current topics in complement*. 1 ed. Advances in Experimental Medicine and Biology 2006, Springer: USA. 153-166.

70. van Beek, J., K. Elward, and P. Gasque, *Activation of complement in the central nervous system: roles in neurodegeneration and neuroprotection.* Ann N Y Acad Sci, 2003. **992**: p. 56-71.
71. Susan Frackman, G.K., Dan Simpson and Doug Storts, *Betaine and DMSO: Enhancing Agents for PCR.* Promega, 1998. **65**: p. 27.
72. Amelia Hardjasa, M.L., Kevin Ma and Hang Yu, *Investigating the Effects of DMSO on PCR Fidelity Using a Restriction Digest-Based Method.* Journal of Experimental Microbiology and Immunology, 2010. **14**: p. 161-164.
73. Pio, G., et al., *Integrating microRNA target predictions for the discovery of gene regulatory networks: a semi-supervised ensemble learning approach.* BMC Bioinformatics, 2014. **15 Suppl 1**: p. S4.
74. Kim, K.J., et al., *SNP identification, linkage disequilibrium, and haplotype analysis for a 200-kb genomic region in a Korean population.* Genomics, 2006. **88**(5): p. 535-40.
75. Ng, J., et al., *Rac GTPases control axon growth, guidance and branching.* Nature, 2002. **416**(6879): p. 442-7.
76. Stankiewicz, T.R. and D.A. Linseman, *Rho family GTPases: key players in neuronal development, neuronal survival, and neurodegeneration.* Front Cell Neurosci, 2014. **8**: p. 314.
77. Mancino, A. and G. Natoli, *Specificity and Function of IRF Family Transcription Factors: Insights from Genomics.* J Interferon Cytokine Res, 2016. **36**(7): p. 462-9.
78. Vinatzer, U., et al., *The leukaemia-associated transcription factors EVI-1 and MDS1/EVI1 repress transcription and interact with histone deacetylase.* Br J Haematol, 2001. **114**(3): p. 566-73.
79. Hirai, H., et al., *Oncogenic mechanisms of Evi-1 protein.* Cancer Chemother Pharmacol, 2001. **48 Suppl 1**: p. S35-40.
80. Depboylu, C., et al., *Possible involvement of complement factor C1q in the clearance of extracellular neuromelanin from the substantia nigra in Parkinson disease.* J Neuropathol Exp Neurol, 2011. **70**(2): p. 125-32.
81. Carbutt, S., et al., *Variation in complement protein C1q is not a major contributor to cognitive impairment in Parkinson's disease.* Neurosci Lett, 2015. **594**: p. 66-9.
82. Fonseca, A.J.T.a.M.I., *The Double-Edged Flowe: Roles of complement protein C1q in neurodegenerative diseases*, in *Current Topics in Complement*, J.D. Lambris, Editor. 2006, Springer: California.

83. Faustino, N.A. and T.A. Cooper, *Pre-mRNA splicing and human disease.* Genes Dev, 2003. **17**(4): p. 419-37.
84. Sharrocks, A.D., et al., *The ETS-domain transcription factor family.* Int J Biochem Cell Biol, 1997. **29**(12): p. 1371-87.
85. Brou, C., et al., *Sequence-specific transactivators counteract topoisomerase II-mediated inhibition of in vitro transcription by RNA polymerases I and II.* Nucleic Acids Res, 1993. **21**(17): p. 4011-8.
86. Biddick, R.K., G.L. Law, and E.T. Young, *Adr1 and Cat8 mediate coactivator recruitment and chromatin remodeling at glucose-regulated genes.* PLoS One, 2008. **3**(1): p. e1436.
87. J. H. Exton, L.E.M., L. S. Jefferson, Ellen H. A. Wong, Naomi Friedmann, T. B. Miller and C. R. Park, *The Hormonal Control of Hepatic Gluconeogenesis*, in *Recent progress in hormone research* E.B. Astwood, Editor. 1970, Academic Press: New York.
88. Murray, D., et al., *In silico gene expression analysis--an overview.* Mol Cancer, 2007. **6**: p. 50.
89. Pillay, K.S., *Gene mutations: Understanding the significance using in silico analysis.* J Pract Cardiovasc Sci, 2015. **1**(1): p. 58-59.
90. Mathelier, A., W. Shi, and W.W. Wasserman, *Identification of altered cis-regulatory elements in human disease.* Trends Genet, 2015. **31**(2): p. 67-76.
91. Caceres, J.F. and A.R. Kornblihtt, *Alternative splicing: multiple control mechanisms and involvement in human disease.* Trends Genet, 2002. **18**(4): p. 186-93.
92. Sethupathy, P. and F.S. Collins, *MicroRNA target site polymorphisms and human disease.* Trends Genet, 2008. **24**(10): p. 489-97.
93. Wang, G., et al., *Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein.* Am J Hum Genet, 2008. **82**(2): p. 283-9.
94. Laguette, M.J., et al., *Sequence variants within the 3'-UTR of the COL5A1 gene alters mRNA stability: implications for musculoskeletal soft tissue injuries.* Matrix Biol, 2011. **30**(5-6): p. 338-45.
95. Amin, A.S., et al., *Variants in the 3' untranslated region of the KCNQ1-encoded Kv7.1 potassium channel modify disease severity in patients with type 1 long QT syndrome in an allele-specific manner.* Eur Heart J, 2012. **33**(6): p. 714-23.
96. Olsson, M., *Familial Amyloidosis with Polyneuropathy: Studies of genetic factors modifying the phenotype of the disease*, in *Department of Public Health and Clinical Medicine* 2010: Sweden.

97. Olsson, M., et al., *A possible role for miRNA silencing in disease phenotype variation in Swedish transthyretin V30M carriers.* BMC Med Genet, 2010. **11**: p. 130.
98. Norgren, N., et al., *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**(11): p. e49981.
99. Motsinger, A.A. and M.D. Ritchie, *Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene-gene interactions in human genetics and pharmacogenomics studies.* Hum Genomics, 2006. **2**(5): p. 318-28.
100. Lemos, C., et al., *BDNF and CGRP interaction: implications in migraine susceptibility.* Cephalalgia, 2010. **30**(11): p. 1375-82.
101. Ghebrehiwet, B., et al., *The C1q family of proteins: insights into the emerging non-traditional functions.* Front Immunol, 2012. **3**.
102. Reid, K.B., *Chemistry and molecular genetics of C1q.* Behring Inst Mitt, 1989(84): p. 8-19.
103. Bolton, E.C., et al., *Cell- and gene-specific regulation of primary target genes by the androgen receptor.* Genes Dev, 2007. **21**(16): p. 2005-17.
104. Macintyre, G., et al., *is-rSNP: a novel technique for in silico regulatory SNP detection.* Bioinformatics, 2010. **26**(18): p. i524-30.
105. Hahn, L.W., M.D. Ritchie, and J.H. Moore, *Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions.* Bioinformatics, 2003. **19**(3): p. 376-82.

9. APPENDIX

Annex 1

Table 13: Primers sequence (Forward and Reverse) of each exon of each gene.

Genes	Exons	Primers	Sequences
RAC1	**Exon 1**	Primer Forward	TGCAGTTTTCCTCAGCTTTG
	Exon 1	Primer Reverse	AGGGGATCGAGAGGTTCATG
	Exon 2	Primer Forward	TGCTAAGTATGTGATGTATATGCC
	Exon 2	Primer Reverse	GTTCATATGAGGATTAAAGATGGC
	Exon 3	Primer Forward	TTCTTGGCACACCTTCTCTAGG
	Exon 3	Primer Reverse	AATGATGGCAATATAAGCAAGTCC
	Exon 4	Primer Forward	GTGTTTCAGTGGCATTTGTCATC
	Exon 4	Primer Reverse	GCAAAACCCCAACCTTGAGTG
	Exon 5	Primer Forward	TCTGACCACACCACTGGTAGAC
	Exon 5	Primer Reverse	GGCTACAGTGAGCTCTGATGTC
	Exon 6	Primer Forward	GAGGTTAGGTGTCTGGCATGAG
	Exon 6	Primer Reverse	GACCCCACTCTCTTCTGATCACAC
	Exon 7	Primer Forward	GATAAAGACACGATCGAGAAACTG
	Exon 7	Primer Reverse	AATTACTGATTGGTTCAAAAAATGG
C1QA	**Exon 2**	Primer Forward	TATCATTGTGTGCATGGGACTC
	Exon 2	Primer Reverse	ACAAACTGGAGAGGATTCATGG
	Exon 3.1	Primer Forward	TTCATTGCCCTTTATCCCATAG
	Exon 3.1	Primer Reverse	CCTGGAAGGTGAAGTAGTAGTAGCC
	Exon 3.2	Primer Forward	ACACGGTCATCACCAACCAG
	Exon 3.2	Primer Reverse	TCAGCAGACACAGACACAGATG
C1QC	**Exon 2**	Primer Forward	GAAGAAAAGGCCCCACCATC
	Exon 2	Primer Reverse	CCAGACACACCGTGAGAGTCC
	Exon 3.1	Primer Forward	CCTAGGGAGATAGCCCATCTATGAG
	Exon 3.1	Primer Reverse	ATGCCCACCATGTCGTAGTAGTC
	Exon 3.2	Primer Forward	CCAAAACCAATCAGGTCAACTC
	Exon 3.2	Primer Reverse	AGGAAGGAGAGAACCCATGG

Table 14: PCR Protocol for *RAC1* (Exon 1).

Reagents	Volume per sample
DMSO	0.60 µL
HotStar Master Mix (Qiagen)	3.15 µL
Primer Fw 10µM	0.75 µL
Primer Rw 10µM	0.75 µL
DNA (10 ng)	1 µL
Total	6.25 µL

Table 15: PCR Protocol for *RAC1* (Exon 2-7), *C1QA* and *C1QC.*

Reagents	Volume per sample
H_2O	1.12 µL
HotStar Master Mix (Qiagen)	3.13 µL
Primer Fw 10µM	0.5 µL
Primer Rw 10µM	0.5 µL
DNA (10 ng)	1 µL
Total	6.25 µL

Table 16: Conditions of PCR protocol for *RAC1* gene (Exon 1).

Number of Cycles	Temperature	Time	Phases
1	95°C	15 min	Initial Denaturation
	95°C	45 sec	Denaturation
35	59ºC	1 min 30 sec	Annealing
	72°C	1 min	Extension
1	72°C	10 min	Final Extension
-	15°C	∞	-

Table 17: Conditions of PCR protocol for *RAC1* (Exon 2-7).

Number of Cycles	Temperature	Time	Phases
1	95°C	15 min	Initial Denaturation
	95°C	45 sec	Denaturation
30	58/61°C *	1 min	Annealing
	72°C	1 min	Extension
1	72°C	10 min	Final Extension
-	15°C	∞	

*We use 58°C for exon 2 of *RAC1* and the remaining exons were amplified at 61°C.

Table 18: Conditions of PCR protocol for *C1QA* gene.

Number of Cycles	Temperature	Time	Phases
1	95°C	15 min	Initial Denaturation
	95°C	45 sec	Denaturation
30	61°C	1 min	Annealing
	72°C	1 min	Extension
1	72°C	10 min	Final Extension
-	15°C	∞	

Table 19: Conditions of PCR protocol for *C1QC* gene.

Number of Cycles	Temperature	Time	Phases
1	95°C	15 min	Initial Denaturation
	95°C	45 sec	Denaturation
30	61/63°C *	1 min	Annealing
	72°C	1 min	Extension
1	72°C	10 min	Final Extension
-	15°C	∞	

* We use 61°C for exon 2 and exon 3.1 and for exon 3.2 we use 63ºC.

Table 20: Protocol for PCR products purification (*RAC1, C1QA* and *C1QC).*

Reagents	Volume per sample
PCR product	2 µL
Exostar	0.5 µL
Total	2.5 µL

Table 21: Conditions for PCR products purification (*RAC1, C1QA* and *C1QC*).

Temperature	Time	Phases
37°C	15 min	Enzymatic Digestion
85°C	15 min	Enzymatic Inactivation
15°C	∞	

Table 22: Sequencing protocol for three genes (*RAC1, C1QA* and *C1QC).*

Reagents	Volume per sample
Purified PCR product	2.5 µL
H_2O	5.0 µL
Primer (Fw/Rv)	0.5 µL
BigDye ® Terminator v1.1 – 1:1	2.0 µL
Total	10 µL

Table 23: Conditions for sequencing protocol for three genes (*RAC1, C1QA* and *C1QC*).

Number of Cycles	Temperature	Time	Phases
1	95°C	5 min	Initial Denaturation
	96°C	10 sec	Denaturation
35	50°C	5 sec	Annealing
	60°C	4 min	Extension
-	15°C	∞	

Annex 2

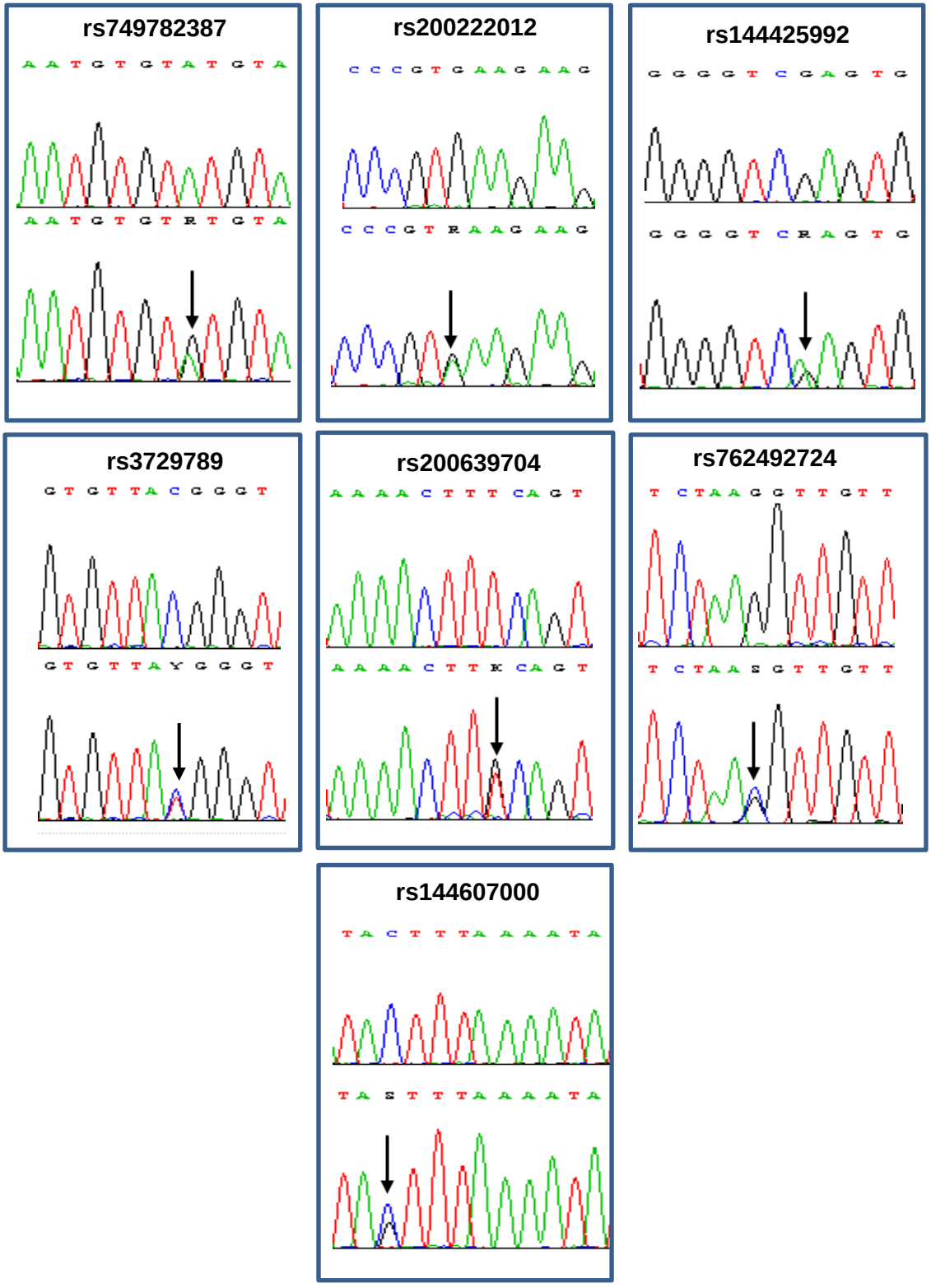

Figure 11: Statistically significant variants for *RAC1* gene.

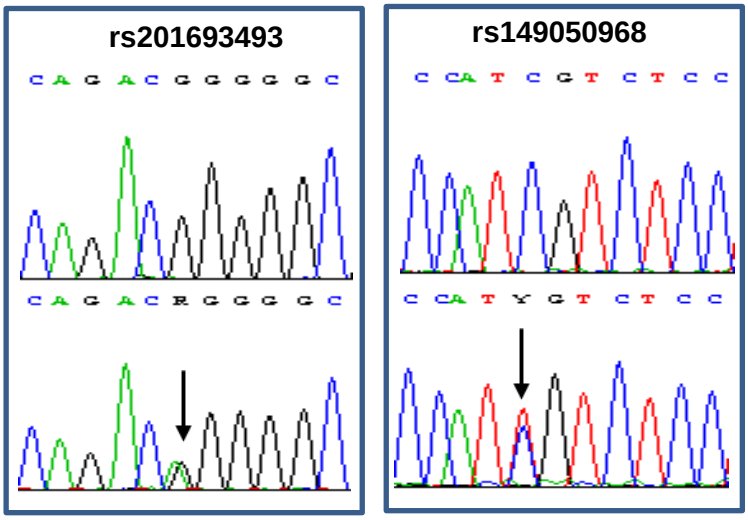

Figure 12: Variants statistically significant for *C1QA* gene.

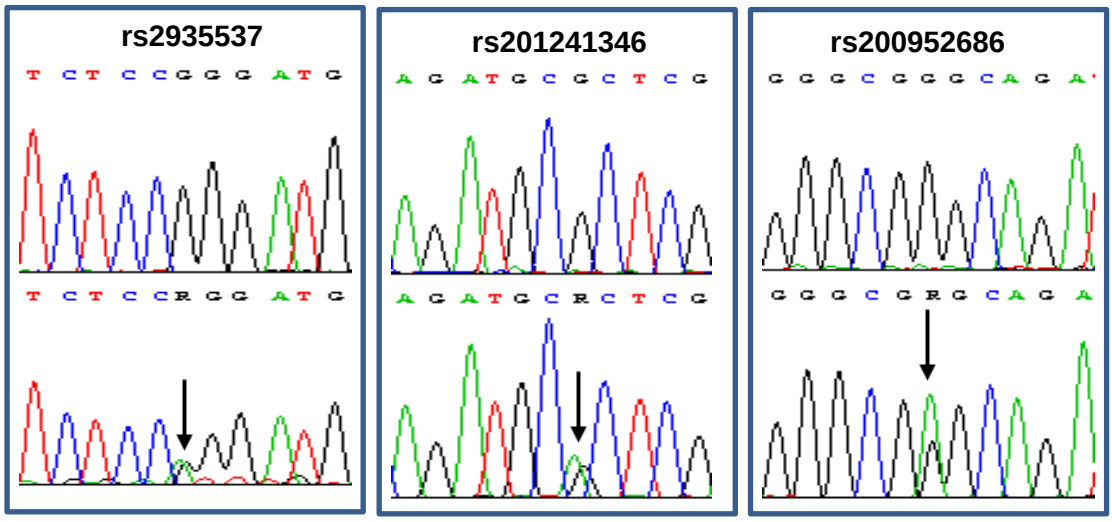

Figure 13: Variants statistically significant for *C1QC* gene.