

Aus dem  
Department für Diagnostische Labormedizin  
der Universität Tübingen  
Institut für Medizinische Genetik und angewandte Genomik

**Die Analyse von Einzelnukleotid-Polymorphismen in  
APOE und CAST als modifizierende Faktoren der  
Spinocerebellären Ataxie Typ 3**

Inaugural-Dissertation  
zur Erlangung des Doktorgrades  
der Medizin

der Medizinischen Fakultät  
der Eberhard Karls Universität  
zu Tübingen

vorgelegt von  
Burger, Rahel Maria

2023

Dekan: Professor Dr. B. Pichler

1. Berichterstatter: Professor Dr. O. Rieß

2. Berichterstatterin: Privatdozentin Dr. K. Brockmann

Tag der Disputation: 08.11.2023

## *Inhaltsverzeichnis*

<b>Abbildungsverzeichnis</b>	<b>4</b>
<b>Tabellenverzeichnis</b>	<b>5</b>
<b>Abkürzungsverzeichnis</b>	<b>7</b>
<b>1 Einleitung</b>	<b>9</b>
1.1 Die Spinozerebelläre Ataxie Typ 3: Eine seltene Erkrankung	9
1.2 Der genetische Hintergrund der SCA3	9
1.3 Ein klinischer Überblick	11
1.4 Die Pathologie der SCA3	13
1.5 Therapeutische Ansätze für die SCA3	14
1.6 Kandidatengene mit Einfluss auf das Erkrankungsalter der SCA3	15
1.7 Einzelnukleotidpolymorphismen	16
1.8 APOE : Apolipoprotein E	17
1.8.1 rs429358 und rs7412 in APOE	18
1.9 CAST: Calpastatin	20
1.9.1 rs1559085 und rs27852 in CAST	20
1.10 Fragestellung der Arbeit: Ein Ausblick	22
<b>2 Material</b>	<b>23</b>
2.1 Patienten - Kohorte und DNA – Proben	23
2.1.1 EUROSCA (Proben aus Mittel- & Ost-Europa)	23
2.1.2 EuSAge (Proben aus Israel und Peru)	24
2.2 Primer	25
2.3 Enzyme	25
2.4 Chemikalien und Reagenzien	26
2.5 Lösungen und Puffer	26
2.6 Fertige Reagenzienkits	26

<b>2.7</b>	<b>Geräte</b>	<b>27</b>
<b>2.8</b>	<b>Verbrauchsmaterialien</b>	<b>27</b>
<b>3</b>	<b>Methoden</b>	<b>28</b>
<b>3.1</b>	<b>Amplifizierung der Gen-Abschnitte</b>	<b>28</b>
3.1.1	Polymerase-Kettenreaktion (PCR)	28
3.1.2	Allelspezifische PCR	29
3.1.3	Amplifizierung von rs429358 und r27412 in APOE	30
3.1.4	Amplifizierung von rs1559085 und rs27852 in CAST	31
<b>3.2</b>	<b>Schmelzkuvenanalyse</b>	<b>32</b>
3.2.1	Analyse im Light Cycler 480 II	33
<b>3.3</b>	<b>Agarose - Gelelektrophorese</b>	<b>34</b>
3.3.1	Analytische Agarose – Gelelektrophorese	35
<b>3.4</b>	<b>Aufreinigung von PCR – Produkten</b>	<b>35</b>
<b>3.5</b>	<b>Konzentrationsbestimmung von DNA in Post-PCR-Proben</b>	<b>37</b>
3.5.1	Konzentrationsbestimmung mit dem Photometer	37
<b>3.6</b>	<b>Sanger – Sequenzierung</b>	<b>39</b>
3.6.1	Sequenzierungs – PCR	40
3.6.2	Ethanol – Fällung	42
<b>3.7</b>	<b>Fragmentlängenanalyse mittels Kapillarelektrophorese</b>	<b>44</b>
3.7.1	PCR von CAG – Repeats im ATXN3	45
3.7.2	Verdünnung der PCR – Produkte	45
3.7.3	Versetzung der Proben mit Puffer und Größenmarker	46
3.7.4	Fragmentanalyse im Beckmann Coulter CEQ 8000	46
<b>3.8</b>	<b>Statistische Auswertung</b>	<b>47</b>
3.8.1	Hardy – Weinberg – Verteilung	47
3.8.2	Lineare Regressionsanalyse	47
3.8.3	Familienfaktor	48
3.8.4	T-Test	49
3.8.5	Varianzanalyse (ANOVA)	49
3.8.6	Graphen	49
<b>4</b>	<b>Ergebnisse</b>	<b>50</b>

<b>4.1</b>	<b>Analyse der Kohorte</b>	<b>51</b>
<b>4.2</b>	<b>Fragmentlängen von ATXN3 in EUSage - Patientenproben</b>	<b>51</b>
<b>4.3</b>	<b>Einzelnukleotid-Polymorphismen in APOE und ihr Einfluss auf das Erkrankungsalter der SCA3</b>	<b>53</b>
4.3.1	Die Einzelnukleotid – Polymorphismen	53
4.3.2	Primerauswahl	54
4.3.3	Etablierung der Methodik	55
4.3.4	Verteilung der ermittelten Genotypen in rs429358 und rs7412	59
4.3.5	Verteilung nach dem Hardy-Weinberg-Gleichgewicht	60
4.3.6	Signifikante Abweichungen der Verteilung von Genotypen in SCA3 Proben	61
4.3.7	Regressionsanalyse der Polymorphismen und ihr Einfluss auf das Erkrankungsalter	62
4.3.8	Die Verteilung der Haplotypen von APOE in SCA3 - Patienten	64
4.3.9	Der Einfluss von APOE Haplotypen auf das Erkrankungsalter	65
4.3.10	Der Einfluss von APOE $\epsilon$ 2 auf das Erkrankungsalter	67
4.3.11	Prädiktive Faktoren für das Erkrankungsalter	68
<b>4.4</b>	<b>Einzelnukleotid-Polymorphismen in CAST und ihr Einfluss auf das Erkrankungsalter der SCA3</b>	<b>69</b>
4.4.1	Primerauswahl	70
4.4.2	Etablierung der Methodik	70
4.4.3	Die Verteilung der Genotypen von rs1559085 in CAST	74
4.4.4	Chi – Quadrat – Test: Keine signifikanten Abweichungen der Verteilung von Genotypen in SCA3 Proben	75
4.4.5	Der Einfluss von rs1559085 in CAST auf das Erkrankungsalter	76
<b>4.5</b>	<b>Zusammenfassung der Ergebnisse</b>	<b>77</b>
<b>5</b>	<b>Diskussion</b>	<b>79</b>
<b>5.1</b>	<b>CAG-Wiederholungen als Einflussfaktor auf das Erkrankungsalter</b>	<b>80</b>
<b>5.2</b>	<b>Einzelnukleotid-Polymorphismen in APOE als Einflussfaktoren auf das Erkrankungsalter an SCA3</b>	<b>81</b>
5.2.1	Abweichung vom Hardy-Weinberg-Gesetz	83
5.2.2	Die Verteilung der APOE Isoformen bleibt unverändert	85
5.2.3	Ein nachgewiesener Einfluss von APOE Haplotypen auf das Erkrankungsalter	85
5.2.4	Mögliche Gründe für abweichende Ergebnisse	86

<b>5.3</b>	<b>Diskussion des High- Resolution- Melting als Methode zur Analyse von Einzelnukleotid-Polymorphismen</b>	<b>88</b>
<b>5.4</b>	<b>Einzelnukleotid-Polymorphismen in CAST und ihr Einfluss auf die Erkrankung der SCA3</b>	<b>90</b>
5.4.1	Die Etablierung von HRM für die Genotypisierung von Einzelnukleotid-Polymorphismen in CAST	91
5.4.2	Vereinbarung von Genotypen in CAST mit dem Hardy-Weinberg-Gleichgewicht	92
5.4.3	Kein signifikanter Einfluss von CAST rs1559085 auf das Erkrankungsalter der SCA3	93
5.4.4	Schwächen der Analyse von CAST rs1559085 im Zusammenhang mit der Kohorte an SCA3-Patienten	94
<b>6</b>	<b>Schlussfolgerung und Ausblick</b>	<b>94</b>
<b>7</b>	<b>Zusammenfassung</b>	<b>98</b>
<b>8</b>	<b>Erklärung zum Eigenanteil</b>	<b>111</b>
<b>9</b>	<b>Veröffentlichungen</b>	<b>112</b>
<b>10</b>	<b>Danksagung</b>	<b>112</b>

## **Abbildungsverzeichnis**

<i>Abbildung 1: Ein umgekehrt proportionaler Zusammenhang zwischen der Anzahl an CAG – Wiederholungen und dem Erkrankungsalter der SCA3.</i>	16
<i>Abbildung 2: Allelspezifische PCR mit Bindung unterschiedlicher Primer je nach Vorliegen eines Einzelnukleotid-Polymorphismus.</i>	29
<i>Abbildung 3: Schmelzkurvenanalyse im Light-Cycler.</i>	34
<i>Abbildung 4: Aufreinigung von PCR – Produkten mit dem QIAquick PCR Purification Kit.</i>	36
<i>Abbildung 5: Bestimmung der DNA – Konzentration mithilfe einer Gel- Elektrophorese.</i>	39
<i>Abbildung 6: Beispiel eines Elektropherogramms</i>	44
<i>Abbildung 9: Gel – Elektrophorese nach MgCl<sub>2</sub> Titration.</i>	57
<i>Abbildung 10: Die verschiedenen Schmelzkurven der Genotypen von rs429358 in APOE.</i>	58
<i>Abbildung 11: Schmelzkurvenverläufe nach der Reaktionsoptimierung.</i>	59
<i>Abbildung 12: Beispiel zur Berechnung der Allelfrequenz</i>	60
<i>Abbildung 13: Regressionsanalyse der Einzelnukleotid-Polymorphismen und ihr Einfluss auf das Erkrankungsalter.</i>	63
<i>Abbildung 14: Regressionsanalyse und Darstellung des Effekteinflusses der APOE Haplotypen auf das Erkrankungsalter.</i>	65
<i>Abbildung 15: Verteilung des Erkrankungsalters bei ε2 Allel-Trägern und Nicht-Trägern.</i>	67
<i>Abbildung 16: Gel – Elektrophorese mit 1,5 % Agarose-Gel nach einer PCR mit verschiedenen Primerkonstellationen.</i>	71
<i>Abbildung 17: High-Resolution-Melting von rs27852.</i>	73
<i>Abbildung 18: High-Resolution-Melting von CAST.</i>	74
<i>Abbildung 19: Regressionsanalyse des Einflusses von rs1559085 auf das Erkrankungsalter.</i>	76

## **Tabellenverzeichnis**

<b>Tabelle 1: Die fünf bekannten Phänotypen der SCA3 weisen unterschiedliche klinische Symptomatik auf.....</b>	<b>12</b>
<b>Tabelle 2: Die Eigenschaften der verschiedenen APOE Allele und ihre bekannten klinischen Assoziationen.....</b>	<b>19</b>
<b>Tabelle 3: Auflistung der Sequenzen aller verwendeten Primer.....</b>	<b>25</b>
<b>Tabelle 4: Auflistung der verwendeten Enzyme und ihrer Hersteller.....</b>	<b>25</b>
<b>Tabelle 5: Auflistung verwendeter Chemikalien und Reagenzien und ihre Hersteller.....</b>	<b>26</b>
<b>Tabelle 6: Auflistung der verwendeten Lösungen und Puffer und ihre Zusammensetzungen .....</b>	<b>26</b>
<b>Tabelle 7: Auflistung der verwendeten Reagenziensets und ihre Hersteller.....</b>	<b>26</b>
<b>Tabelle 8: Liste der verwendeten Geräte mit der entsprechenden Bezeichnung und ihrer Hersteller....</b>	<b>27</b>
<b>Tabelle 9: Auflistung der verwendeten Verbrauchsmaterialien und die entsprechenden Hersteller .....</b>	<b>27</b>
<b>Tabelle 10: Auflistung der verwendeten Primer für die Amplifikation von APOE .....</b>	<b>30</b>
<b>Tabelle 11: PCR – Programm für eine allelspezifische PCR.....</b>	<b>31</b>
<b>Tabelle 12: Eingesetzte Primer für die Amplifizierung des CAST-Gens .....</b>	<b>31</b>
<b>Tabelle 13: PCR – Mix für die Amplifikation von APOE.....</b>	<b>40</b>
<b>Tabelle 14: PCR – Programm für die Amplifizierung von APOE.....</b>	<b>40</b>
<b>Tabelle 15: PCR-Mix für die Amplifikation von CAST.....</b>	<b>41</b>
<b>Tabelle 16: PCR – Programm für die Amplifikation von CAST.....</b>	<b>41</b>
<b>Tabelle 17 a: Zusammensetzung eines PCR – Mix für eine Sequenzierungs – PCR.....</b>	<b>41</b>
<b>Tabelle 17 b: Liste der verwendeten Primer für die Sequenzierungs-PCR.....</b>	<b>40</b>
<b>Tabelle 18: PCR – Programm im Thermocycler nach Angaben der Protokollbeilage .....</b>	<b>42</b>
<b>Tabelle 19: PCR – Mix für die CAG-Repeat Analyse im ATXN3 .....</b>	<b>45</b>
<b>Tabelle 20: PCR – Programm im Thermocycler für die CAG – Repeat Analyse im ATXN3.....</b>	<b>45</b>
<b>Tabelle 21: Verdünnungsreihe der PCR – Produkte in eine 1:200 Verdünnung .....</b>	<b>46</b>
<b>Tabelle 22: Ansatz zur CAG – Repeat – Analyse im ATXN3.....</b>	<b>46</b>
<b>Tabelle 23: Einzelnukleotid – Polymorphismen in APOE.....</b>	<b>53</b>
<b>Tabelle 24: Liste der Primer für APOE die auf ihre Effektivität und Funktionalität getestet wurden.....</b>	<b>54</b>
<b>Tabelle 25: Verteilung der Genotypen von rs429358 und rs7412 in APOE nach der Typisierung .....</b>	<b>59</b>
<b>Tabelle 26: erwartete Verteilung der Genotypen von rs429358 und rs7412 in APOE .....</b>	<b>60</b>
<b>Tabelle 27: Vergleich der typisierten Genotypen in Proben von SCA3 Betroffenen und einer Verteilung nach dem Hardy- Weinberg – Gesetz.....</b>	<b>61</b>
<b>Tabelle 28: Verminderung des Erkrankungsalters der SCA3 in Jahren durch verschiedene Genotypen..</b>	<b>64</b>
<b>Tabelle 29: Liste der möglichen Haplotypen.....</b>	<b>64</b>
<b>Tabelle 30: T-Test der Abweichung der Mittelwerte des Erkrankungsalters in Abhängigkeit der verschiedenen Haplotypen.....</b>	<b>66</b>

<i>Tabelle 31: Statistische Analyse des Einflusses der APOE Haplotypen auf das Erkrankungsalter .....</i>	<i>66</i>
<i>Tabelle 32: Einzelnukleotid-Polymorphismen in APOE als prädiktive Faktoren für das Erkrankungsalter der SCA3.....</i>	<i>69</i>
<i>Tabelle 34: Liste der Primer für CAST.....</i>	<i>70</i>
<i>Tabelle 35: Verteilung der ermittelten Genotypen von rs1559085 in CAST.....</i>	<i>74</i>
<i>Tabelle 36: Erwartete Verteilung der Genotypen.....</i>	<i>75</i>
<i>Tabelle 37: Vergleich der Verteilung der Genotypen .....</i>	<i>75</i>
<i>Tabelle 38: Zusammenfassung der untersuchten Einzelnukleotid-Polymorphismen .....</i>	<i>78</i>

## Abkürzungsverzeichnis

SCA3	Spinocerebelläre Ataxie Typ 3
MJD	Machado Joseph Disease
ADCA	autosomal-dominante cerebelläre Ataxie
CAG	Cytosin-Adenosin-Guanin
TGC	Thymin-Guanin-Cytosin
CGC	Cytosin-Guanin-Cytosin
<i>ATXN3</i>	Ataxin-3 Gen
<i>ATXN2</i>	Ataxin-2 Gen
<i>APOE</i>	Apolipoprotein E Gen
<i>CAST</i>	Calpastatin Gen
PolyQ	Polyglutamin
SBMA	Spinobulbäre Muskelatrophie
DRPLA	Dentatorubrale-Pallidolysiale Atrophie
SARA	Scale for the assessment and rating of ataxia
MRT	Magnet-Resonanz-Tomographie
SNP	Single-Nucleotide-Polymorphismus
LDL	Low-Density-Lipoprotein
CEPH	<i>Centre d'Étude du Polymorphisme Humain</i>
EUROSCA	European integrated Project on spinocerebellar ataxias
EuSage	Europäisch-südamerikanische Zusammenarbeit zur Identifizierung altersabhängiger genetischer modifizierender Faktoren der SCA3
DNA	Desoxyribonucleic Acid
PCR	Polymerase-Chain-Reaction

AS-PCR	Allel-specific Polymerase-Chain-Reaction
ASO	Antisense – Oligonukleotide
MAF	Minor allel frequency
NIH	National Institutes of Health
ANOVA	Analysis of Variance
NLM	National Library of Medicine
HRM	High-Resolution-Melting
NEB	New England Biolabs
JMP	Jump
GWAS	Genome wide Association Study

# 1 Einleitung

## 1.1 Die Spinozerebelläre Ataxie Typ 3: Eine seltene Erkrankung

Die Spinozerebelläre Ataxie Typ 3 (SCA3), auch Machado-Joseph-Krankheit (MJD) genannt, gehört zu der heterogenen Gruppe der autosomal-dominant vererbten Ataxien (ADCA) (Sullivan et al, 2019). Mit einer Prävalenz von weniger als 3 Betroffenen pro 100.000 gehört sie *per definitionem* zu den Seltenen Erkrankungen (Nguengang et al, 2019). Bei Vorliegen einer starken regionalen Varianz gilt die SCA3, mit einem Anteil von 15-45% an allen hereditären Ataxien weltweit, dennoch als die häufigste autosomal-dominant vererbte Ataxie (Costa et al, 2012). Im Verlauf der letzten Jahre hat die Auseinandersetzung mit Seltenen Erkrankungen und ihre diagnostische Einordnung einen wachsenden Stellenwert erlangt. Durch eine Vielzahl an zur Verfügung stehender Methoden wurde die Diagnostik dieser Erkrankungen erheblich erleichtert. Insbesondere Erkrankungen mit Multiorganbeteiligung werden jedoch häufig immer noch nicht früh genug erkannt. Betroffene haben daher in der Regel einen langen Weg zur richtigen Diagnose und zahlreiche Krankenhausaufenthalte vor sich (Ferreira, 2019). Diese Arbeit soll ihren Teil dazu beitragen, das Verständnis, um die Pathophysiologie der SCA3 zu erweitern und modifizierende genetische Einflussfaktoren der Erkrankung zu identifizieren. Mit den Erkenntnissen dieser Arbeit soll ein Beitrag zur besseren medizinischen Versorgung von Betroffenen der SCA3 geleistet werden.

## 1.2 Der genetische Hintergrund der SCA3

Die SCA3 wird durch eine Verlängerung einer Cytosin-Adenin-Guanin (CAG) Trinukleotid-Sequenz verursacht. Bereits in den frühen 90er Jahren konnte die krankheitsauslösende Mutation im *MJD 1 Gen* (heute *Ataxin-3-Gen (ATXN3)*) auf dem Chromosom 14.q32.1 durch Kawaguchi et al. (1994) identifiziert werden. Sie führt zu einer pathologischen Expansion der CAG-Wiederholungen im Exon 10 von *ATXN3*. CAG-Triplets codieren auf Proteinebene für eine Glutaminsequenz, eine Verlängerung an CAG-Triplett-Wiederholungen resultiert demnach in einer verlängerten Polyglutaminsequenz (PolyQ) im kodierten Ataxin-3 Protein.

Neben der SCA3 machen weitere spinozerebelläre Ataxien (SCA1, SCA2, SCA6, SCA7, und SCA17) und die Spinobulbäre Muskelatrophie (SBMA), die Dentatorubrale-Pallidolysiale Atrophie (DRPLA) und Chorea Huntington die Gruppe der, aktuell bekannten, sogenannten PolyQ-Erkrankungen aus (Klockgether et al., 2019). Bis auf die x-chromosomal vererbte SBMA, sind alle diese Erkrankung autosomal-dominant vererbbar.

Unter Zusammenschau mehrerer Studien von 1994 – 2017 gilt eine Anzahl von >45 CAG-Wiederholungen als pathogenetische Ursache der Erkrankung (zusammengefasst in Costa et al, 2012). Betroffene, die eine Anzahl von 45 – 55 CAG-Wiederholungen aufweisen, zeigen häufig eine inkomplette Penetranz der klinischen Symptomatik, während eine Anzahl von >55 CAG-Wiederholungen mit einer vollständigen Ausprägung der Symptomatik in Zusammenhang gebracht wird (McLoughlin et al., 2020). Die Anzahl der CAG-Wiederholung korreliert dabei direkt mit der Schwere der Erkrankung.

Bei der klinischen Beobachtung mehrerer Generationen von erkrankten Familienmitgliedern kann bei der SCA3 das Phänomen der Antizipation beobachtet werden, d.h. Betroffene der jüngeren Generation weisen ein früheres Erkrankungsalter auf und zeigen eine stärker ausgeprägte klinische Symptomatik als ihre erkrankten Eltern. Das Phänomen der Antizipation ist ein typisches Merkmal von Trinukleotid-Repeat-Erkrankungen und wird vermutlich durch die Instabilität der Tripletverlängerung und Interaktion beider Allele des *ATXN3* Gens verursacht (Igarashi et al, 1996). Es ließ sich nachweisen, dass eine Instabilität der Trinukleotid-Wiederholungs-Sequenz vor allem bei paternalen Transmission vorliegt (Løkkegaard, 1998).

Trotzdem die SCA3 eine autosomal-dominante Erkrankung ist, gibt es Familien in denen homozygote Individuen mit zwei mutierten Allelen vorkommen. In diesem Fall ist ein „Gendosis-Effekt“ zu beobachten. Individuen mit zwei mutierten Allelen erkranken wesentlich früher und schwerer als Betroffene mit heterozygotem Genstatus (Lerer et al, 1996). *De novo* Mutationen sind bei der SCA3 nicht bekannt.

### **1.3 Ein klinischer Überblick**

Erstmals beschrieben wurde die SCA3 als unbekannter Symptomkomplex unterschiedlicher Bewegungsauffälligkeiten in Familien mit Abstammung von den Azoren (Nakano et al, 1972). Betroffene entwickeln typischerweise zwischen dem 20. und 50. Lebensjahr eine progrediente Gangataxie (Paulson, 1998 überarbeitet 2015). Das durchschnittliche Erkrankungsalter liegt bei 36 Jahren (Kieling et al, 2007) mit extremen Ausreißern zwischen 5 und 70 Jahren. Neben der charakteristischen Gangataxie äußert sich die SCA 3 in einer vielschichtigen klinischen Ausprägung. Trotz zahlreicher klinischer Studien und Tiermodelle konnte bisher kein eindeutiger Symptomkomplex beschrieben werden. Eine externe Ophthalmoplegie, sowie eine unspezifische Hirnstammsymptomatik wurden allerdings überdurchschnittlich häufig beobachtet (Rana et al, 2016).

Ein besserer Überblick über die Vielfalt an klinischer Symptomatik kann durch die Differenzierung von fünf klinischen Subtypen der Erkrankung geschaffen werden (Tabelle 1). Diese werden je nach Erkrankungsalter und klinischer Manifestation unterteilt (Riess et al., 2008). Die größte Subgruppe (Typ 2) erkrankt zwischen dem 20. und 50. Lebensjahr und weist die charakteristischen Symptom-Trias einer fortschreitenden zerebellären Ataxie, einer externen Ophthalmoplegie sowie Pyramidenbahnzeichen in Form eines pathologischen Reflexstatus auf. Die Gruppe, der an Subtyp 1 Erkrankten, entwickelt Symptome bereits sehr früh, im Alter von < 20 Jahren und zeichnet sich durch eine rasche Progredienz aus. Zudem weisen Betroffene in diesem Fall häufig einen hohen Grad an Spastik, Rigor und Dystonie auf. Der dritte Subtyp erkrankt im durchschnittlichen Alter von 40-75 und leidet neben einer Gangataxie vorrangig an einer peripheren Symptomatik wie Muskelatrophie und Polyneuropathie. Der seltenste Subtyp (Typ 4) fällt durch eine charakteristische „Parkinson – ähnliche“ Symptomatik auf (Tuite et al.,1995). Die Definition einer fünften Subgruppe wurde im Nachhinein hinzugefügt, da es in seltenen Fällen Betroffene gibt, die eine komplette spastische Paraparese aufweisen, ohne weitere Symptomatik (Wang et al., 2009). Gruppenübergreifend lassen sich zudem klinische Symptome wie Gewichtsverlust, das Auftreten eines Restless-Legs-Syndroms (Paulson, 2007)

und in Einzelfällen kognitive Einschränkungen und Verhaltensveränderungen beobachten (Kawai et al., 2004).

**Tabelle 1: Die fünf bekannten Phänotypen der SCA3 weisen unterschiedliche klinische Symptomatik auf. Tabellarischer Überblick über die phänotypischen Ausprägungen der fünf bekannten Subtypen der SCA3 (nach Coutinho und Andrade 1978, Wang et al. 2009)**

Bezeichnung	Erkrankungsalter	Symptome
Typ I	< 20 Jahre	<ul style="list-style-type: none"> <li>▪ rasche Progredienz</li> <li>▪ pyramidale Symptome (Spastik)</li> <li>▪ extrapyramidale Symptome (Dystonie)</li> </ul>
Typ II	20 – 50 Jahre	<ul style="list-style-type: none"> <li>▪ häufigster Subtyp</li> <li>▪ zerebelläre Ataxie</li> <li>▪ Ophthalmoplegie</li> <li>▪ pathologischer Reflexstatus</li> </ul>
Typ III	40 – 75 Jahre	<ul style="list-style-type: none"> <li>▪ periphere Symptomatik (Polyneuropathie und Muskelatrophie)</li> </ul>
Typ IV	altersunabhängig	<ul style="list-style-type: none"> <li>▪ seltenster Subtyp</li> <li>▪ parkinsonoide Symptomatik</li> </ul>
Typ V	altersunabhängig	<ul style="list-style-type: none"> <li>▪ Einzelfälle</li> <li>▪ komplette spastische Paraparese</li> </ul>

Für die klinische Beurteilung der Erkrankung entwickelten Schmitz-Hübsch *et al.* (2007) ein Bewertungssystem für die charakteristischen Aspekte einer ataktischen Erkrankung. Dieses System wird international eingesetzt und bewertet die Betroffenen in acht Kategorien nach Punkten. Im sogenannten SARA (Scale for the assessment and rating of ataxia)-Score werden die Kategorien Stand (0-6 Punkte), Gang (0-8 Punkte), Sitz (0-4 Punkte), Sprachmotorik (0-6 Punkte), Dysmetrie (0-4 Punkte), Tremor (0-4 Punkte) und Koordination (0-4 Punkte), sowie Diadochokinese (0-4 Punkte), semiquantitativ erfasst. Insgesamt können 40 Punkte erreicht werden. Bei Überprüfung der Koordination und Motorik wird darauf geachtet beide Körperhälften getrennt voneinander zu untersuchen. Anhand der Punkteanzahl kann anschließend die klinische Ausprägung der Erkrankung objektiviert werden. Je höher die erreichte Punktzahl, desto schwerer ist der Betroffene erkrankt.

Insbesondere in der präklinischen Phase der Erkrankung weist der SARA Score allerdings eine geringe Sensitivität auf. Zudem ist die Vergabe des Scores stark vom Untersucher abhängig. Da man davon ausgeht, dass neuroprotektive und krankheitslimitierende therapeutische Maßnahmen insbesondere in der präklinischen Phase der Erkrankung greifen (Rubinsztein und Orr, 2016) wird viel daran gesetzt objektive Biomarker zur Einschätzung der Erkrankung zu finden.

#### **1.4 Die Pathologie der SCA3**

Durch die Verlängerung der CAG – Wiederholungen kommt es bei der SCA3 zu expandierten Polyglutaminsequenzen im Ataxin-3-Protein. Diese lagern sich vor allem in den Zellkernen von Neuronen ab. Es wird davon ausgegangen, dass dadurch neurotoxische Prozesse entstehen, die in der Folge zu einem Verlust der Neurone und einer Hirnatrophie führen (Costa und Paulson, 2012). Die genauen toxischen Mechanismen des mutierten Ataxin-3-Proteins sind bis heute nicht vollständig verstanden (Ramani et al, 2017).

Mittels Magnet-Resonanz-Tomographie (MRT) lassen sich Korrelate zwischen spezifischen anatomischen Veränderungen der Gehirnstruktur und der Erkrankung der SCA3 gut darstellen. Als häufigste Korrelation fällt eine signifikante Erweiterung des vierten Ventrikels auf (Murata et al.,1998). Im Vergleich von Betroffenen und Kontrollprobanden kann zudem eine ausgeprägte Hirnatrophie umschriebener Bereiche beobachtet werden (Yoshizawa et al.,2003). Dabei sind vor allen Dingen der Pons, der Vermis, die zerebellären Hemisphären und Basalganglien betroffen (Etchebehere et al., 2001). Eine Volumenverminderung des Putamen und Nucleus caudatus wird ebenfalls häufig beobachtet (Klockgether et al.,1998). Die Progredienz der Atrophie des Zerebellums und des Hirnstamms wird von Eichler et al. (2011) mit der Länge der CAG-Wiederholungs-Sequenz und dem Alter des Patienten in einen direkten Zusammenhang gebracht. Auch makroskopisch findet man in Gehirnen betroffener Patienten eine verringerte Pigmentierung der Substantia nigra, sowie eine sichtbare Atrophie von Zerebellum, Pons, Medulla oblongata und verschiedenen Hirnnervenkernen (Rub et al., 2008). Trotzdem ist bis heute kein eindeutiges anatomisch pathologisches Korrelat zur SCA3 bekannt. Mehrere Studien konnten zeigen, dass die Erkrankung einen über das Zerebellum hinausreichenden Einfluss auf das komplette zentrale Nervensystem hat und auch die zerebralen Hemisphären, der Thalamus und die einzelnen Lobi eine Atrophie aufweisen können (De Oliveira et al., 2010, D`Abreu et al., 2011).

Zudem wurden funktionelle Auffälligkeiten im Zerebellum von SCA3 – Erkrankten mittels MR - Spektroskopie überprüft. Dabei konnten Veränderungen des Stoffwechsels, insbesondere des Glukose-Stoffwechsels und eine

Beeinträchtigung der axonalen Funktionalität sichtbar gemacht werden (D'Abreu et al., 2009). Ein Verlust dopaminerger Zellen im Striatum symptomatischer SCA3 Patienten konnte ebenfalls nachgewiesen und insbesondere mit dem „Parkinson-ähnlichen“ Erkrankungstyp in Verbindung gebracht werden (Wüllner et al., 2005).

### **1.5 Therapeutische Ansätze für die SCA3**

Für die zahlreichen Symptome einer SCA3 gibt es mehrere Möglichkeiten zur ausschließlich symptomatischen Therapie. Betroffene mit parkinsonoiden Symptomen oder einem Restless-Legs-Syndrom profitieren beispielsweise von dopaminergen Medikamenten (Buhmann et al., 2003).

Bis zum aktuellen Zeitpunkt sind keine kurativen Therapien zur Behandlung der SCA3 bekannt. In der Vergangenheit gab es zahlreiche Studien mit dem Ziel eine Therapie für diese obligat progressive und innerhalb von 10-20 Jahren zum Tod führende Erkrankung zu finden. Dabei wurden sowohl pharmakologische als auch genetische Ansätze erprobt.

Auf Grundlage eines von Schmitt et al. (2007) gefundenen Zusammenhangs zwischen dem Ausbleiben phänotypischer Merkmale einer SCA3 und dem Fehlen des *ATXN3* Genes scheint die Suppression von *ATXN3* eine vielversprechende therapeutische Möglichkeit zu sein. Es wurden im Rahmen von Tierversuchen bereits mehrere Mechanismen zur Ausschaltung des Genes erprobt. Antisense – Oligonukleotide (ASO) scheinen dabei besonders geeignet, um das *ATXN3* Gen bei betroffenen Mäusen auszuschalten. Durch die Behandlung mit ASOs konnte eine Akkumulation von Polyglutamin-Aggregaten in Neuronen vermindert werden. Mäuse, die über einen längeren Zeitraum behandelt wurden, zeigten weniger motorische Symptome (McLoughlin et al., 2018).

Zwei Jahre später (2009) konnten Boy et al. mittels eines Mausmodells mit humanen Ataxin-3 mit expandierter Polyglutaminsequenz (77 CAG – Wiederholungen) ebenfalls einen Effekt durch das Ausschalten von *ATXN3* nachweisen. Dabei wurde in Mäusen mit progredient neurodegenerativem Phänotyp die Expression des humanen Ataxin-3 ausgeschaltet. Trotz zuvor

eindeutigen neurologischen Symptomen (cerebelläre Symptomatik, Hyperaktivität, Gewichtsverlust) konnten die erkrankten Mäuse fünf Monate später nicht mehr von gesunden Kontrollmäusen unterschieden werden.

Diese Ergebnisse zeigen, dass die Reduktion der Expression von pathologischem Ataxin-3 einen vielversprechenden Ansatz zu Behandlung der SCA3 darstellt. Die Verbesserung der Symptomatik scheint dabei unmittelbar vom Zeitpunkt der Behandlung abzuhängen. So zeigten die Mäuse im Tiermodell von Boy et al. (2009) nur bei therapeutischen Maßnahmen vor Eintreten irreversibler neurologischer Schäden und einer Behandlung über einen längeren Zeitraum von mindestens fünf Monaten eine deutliche Verbesserung der Symptome.

### **1.6 Kandidatengene mit Einfluss auf das Erkrankungsalter der SCA3**

Für alle PolyQ-Erkrankungen kann ein negativer Zusammenhang zwischen der Anzahl an CAG-Wiederholungen und dem Erkrankungsalter nachgewiesen werden (Abbildung 1). Der Korrelationskoeffizient liegt dabei in der Regel bei knapp 0,5 ( $r^2=0,48$ ) (Maciel et al., 1995). Für etwa 50% der Betroffenen kann der Zeitpunkt der Erkrankung demnach nicht allein durch die Anzahl der CAG-Wiederholungen zuverlässig eingeschätzt werden. Durch diese große Varianz ist davon auszugehen, dass es weitere genetische Faktoren gibt, die einen Einfluss auf den Zeitpunkt und die Schwere der Erkrankung haben (van de Warrenburg et al., 2005). Mithilfe der Identifikation und Charakterisierung weiterer genetischer Einflussfaktoren der Erkrankung könnte man den Verlauf der Erkrankung besser verstehen und die Diagnostik präsymptomatischer Betroffener erheblich verbessern.

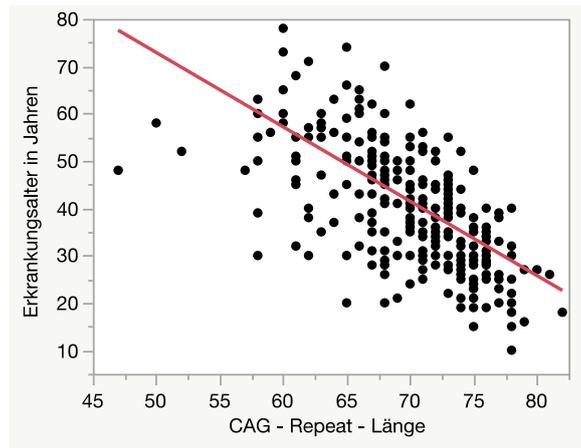


Abbildung 1: Ein umgekehrt proportionaler Zusammenhang zwischen der Anzahl an CAG – Wiederholungen und dem Erkrankungsalter der SCA3. Die Korrelation des Erkrankungsalters mit der Anzahl an CAG-Wiederholungen ergibt einen negativen Zusammenhang zwischen den beiden Variablen. Je größer die Anzahl an CAG-Wiederholungen, desto geringer ist das Erkrankungsalter. Mit einem Korrelationskoeffizienten von 0,48 bleibt die Varianz des Erkrankungsalters für etwa 50% der Betroffenen unerklärt.

Innerhalb der letzten Jahre wurden bereits mehrere Gene als mögliche Einflussfaktoren auf das Erkrankungsalter der SCA3 analysiert. Dabei erhielten Einzelnukleotid-Polymorphismen einen großen Stellenwert auf der Suche nach genetischen Modifikatoren der Erkrankung.

### 1.7 Einzelnukleotidpolymorphismen

Als Einzelnukleotidpolymorphismen oder SNP (single nucleotide polymorphism) werden Varianten eines einzelnen Basenpaares in einem definierten Genabschnitt bezeichnet. Sie bilden die kleinste interindividuelle vererbare genetische Variante im menschlichen Genom und werden häufig zur Erklärung interindividueller Unterschiede herangezogen. Abgegrenzt zu einer Punktmutation wird ein Einzelnukleotid – Polymorphismus durch seine deutlich größere Verbreitung in der Allgemeinbevölkerung mit einer Häufigkeit von über einem Prozent (Brookes, 1999).

Die meisten Einzelnukleotid-Polymorphismen sind funktionslos („Silent Mutation“). Sie befinden sich in kodierenden Bereichen (Exons) und nicht kodierenden Bereichen (Introns) eines Genes, teilweise auch intergenetisch. In Einzelfällen können sie dabei zu einer Missense- oder Nonsense-Mutation führen. Diese können zu einem funktionellen Austausch von Aminosäuren führen oder einen Einfluss auf das Transkriptionsverhalten der regulatorischen

Genabschnitte haben. (Arneemann, 2019 a). Ein Individuum wird als heterozygot bezeichnet, wenn an einem Genlocus zwei verschiedene Allele vorliegen. Homozygote Individuen weisen an einem Genlocus zwei identische Allele auf. Nach aktuellem Forschungsstand sind mehr als 3 Millionen Einzelnukleotid - Polymorphismen im menschlichen Genom bereits identifiziert worden (International HapMap Consortium, 2007). Durch intensive Forschungen wird für die nächsten Jahre eine deutlich steigende Tendenz der identifizierten Einzelnukleotid-Polymorphismen angenommen. Diese Tendenz spiegelt sich in der Vielzahl an Einzelnukleotid-Polymorphismen wider, die bereits im Zusammenhang mit der SCA3 untersucht wurden und mit Krankheitsmerkmalen in Verbindung gebracht werden konnten.

Als Kandidatengene kamen bisher das *ATXN3* selbst und *ATXN2* in Frage. In einigen Fällen konnte zwar ein Einfluss auf das Erkrankungsalter nachgewiesen werden, der Effekt belief sich dabei allerdings meist auf eine minimale Verbesserung der Erklärung des Erkrankungsalters von einem Prozent (zusammengefasst in Mattos et al., 2018). Dies lässt vermuten, dass der Einflussfaktor mit dem größten Effekt neben der CAG-Repeat-Verlängerung bisher noch unbekannt ist.

Als weitere mögliche Kandidatengene werden in dieser Arbeit *APOE* und *CAST* genauer analysiert. Dabei werden Einzelnukleotid-Polymorphismen der Gene in SCA3 Patienten typisiert und anschließend mit dem Erkrankungsalter korreliert.

### **1.8 APOE : Apolipoprotein E**

Das *Apolipoprotein E (APOE) Gen* wurde schon für einige neurodegenerative Erkrankungen als modifizierender Faktor analysiert (Peng et al., 2014). Es ist der wichtigste Ligand für den LDL-Rezeptor und somit maßgeblich am Transport, Speicher und Metabolismus von Lipiden im Blutplasma und Liquor beteiligt (Mahley, 1988). Hauptsächlich wird *APOE* in der Leber synthetisiert, es befindet sich aber auch im peripheren und zentralen Nervensystem, sowie dem Gehirn.

Es gibt drei klinisch relevante Isoformen von *APOE* bedingt durch drei verschiedene Allele:  $\epsilon_2$ ,  $\epsilon_3$  und  $\epsilon_4$ . Das  $\epsilon_1$ -Allel gilt als „missing Allel“ und kommt in der Regel nicht vor. In der Normalbevölkerung ist das  $\epsilon_3$  mit über 78% am

weitesten verbreitet und das  $\epsilon 2$  mit einer Allelfrequenz von 7% das seltenste Allel (vgl. Tabelle 2). Die drei Isoformen unterscheiden sich dabei jeweils lediglich durch eine Aminosäure aufgrund von Einzelnukleotid-Polymorphismen in Codon 112 oder 158.

In mehreren Studien wurde das  $\epsilon 4$  Allel bereits als wichtigster Risikofaktor für eine frühe Alzheimer Erkrankung charakterisiert (Liu et al., 2013, Strittmatter et al., 1993). Seither gilt das *APOE* als wichtiger Gegenstand medizinischer Forschungen, insbesondere im Bereich der neurologischen Erkrankungen. Das *APOE*  $\epsilon 2$  wurde bisher immer mit einem erhöhten Arterioskleroserisiko in Verbindung gebracht. Im Jahr 2011 konnten Bettencourt et al. allerdings einen vermindernden Effekt des *APOE*  $\epsilon 2$  Allels auf das Erkrankungsalter der SCA3 nachweisen. Sie konnten in einer Fallgruppe von 200 SCA3-Probanden bei Vorliegen eines  $\epsilon 2$  Allels eine Verringerung des Erkrankungsalters um 5 Jahre feststellen. 2014 überprüften Zhou et al. diese Entdeckung in einer größeren Kohorte mit 403 Probanden. Dabei konnte der angenommene Effekt von *APOE*  $\epsilon 2$  nicht bestätigt werden.

### **1.8.1 rs429358 und rs7412 in *APOE***

Die Einzelnukleotid - Polymorphismen rs429358 und rs7412 befinden sich in Codon 112 und Codon 158 des Exon 4 von *APOE*. Durch die verschiedenen Allelvarianten (homozygot/heterozygot) kommt es zu einem Aminosäureaustausch (Missense-Variante) auf Proteinebene, der einen fraglichen Einfluss auf das Erkrankungsalter der SCA3 haben könnte.

Der Einzelnukleotid-Polymorphismus rs429358 äußert sich in einer Variation in Codon 112. Dadurch kann hier das Basentriplett Thymin-Guanin-Cytosin (TGC) durch einen Basenaustausch zu Cytosin-Guanin-Cytosin (CGC) werden. Im Rahmen der Translation führt der Austausch einer einzelnen Base an dieser Stelle zur Transkription eines Arginins anstelle von Cystein. Für den Einzelnukleotid-Polymorphismus rs7412 gilt das gleiche Prinzip in Codon 158. Hier liegt normalerweise das Basentriplett Cytosin-Guanin-Cytosin (CGC) vor, welches durch einen einfachen Basenaustausch mit Thymin zur Translation eines Cystein anstelle von Arginin führt.

Die Varianten mit dem höchsten Aufkommen in der Allgemeinbevölkerung sind rs429358 T/T homozygot und rs7412 C/C homozygot. Diese Konstellation entspricht der Isoform  $\epsilon 3$  und resultiert auf Proteinebene in einem Cystein in Codon 112 (rs429358) und einem Arginin in Codon 158 (rs7412). Durch die oben beschriebenen Variationen der Basentriplets entstehen die Isoformen  $\epsilon 2$  und  $\epsilon 4$  welche entweder zwei Cystein oder zwei Arginin auf Proteinebene aufweisen (Abbildung 2).

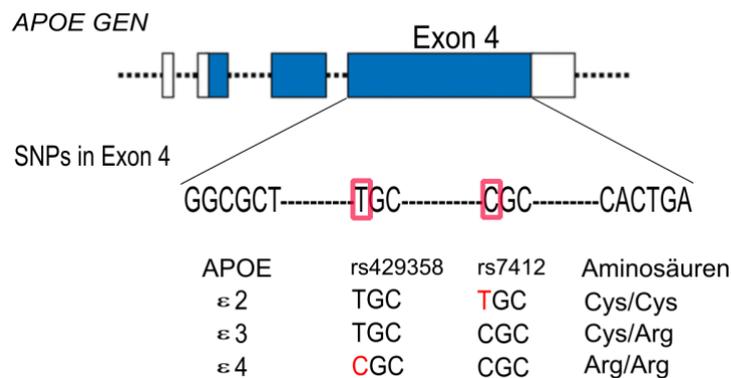


Abbildung 2: **Zwei Polymorphismen in APOE resultieren in drei klinisch relevanten Allelen des Proteins Apolipoprotein E.** Schematische Darstellung des APOE Gens und seinen drei klinisch relevanten Allelen. Das APOE Gen beinhaltet vier Exons, die auf dem langen Arm des Chromosom 19 liegen. In Codon 112 und Codon 158 in Exon 4 befinden sich zwei Polymorphismen, die für vier verschiedene Allele des APOE verantwortlich sind. Das Allel  $\epsilon 1$  gilt als „missing Allel“ und ist der Übersicht halber in dieser Abbildung nicht aufgeführt.

Im menschlichen Genom können diese Allele sowohl homozygot ( $\epsilon x/\epsilon x$ ), als auch heterozygot ( $\epsilon x/\epsilon y$ ) vorliegen. Die resultierenden Missense-Varianten führen zu einer veränderten Affinität des APOE zum LDL-Rezeptor und werden vereinzelt mit klinischen Krankheitsbildern in Verbindung gebracht (Tabelle 2)

Tabelle 2: **Die Eigenschaften der verschiedenen APOE Allele und ihre bekannten klinischen Assoziationen.** Auflistung der APOE Allele im menschlichen Genom resultierend aus zwei Einzelnukleotid-Polymorphismen in Exon 4. Neben einer Missense – Variation der Aminosäuren, weisen die verschiedenen Allele auch eine unterschiedlich stark ausgeprägte Affinität zum LDL – Rezeptor auf. Für die Allele  $\epsilon 2$  und  $\epsilon 4$  sind zudem Zusammenhänge zwischen bestimmten Erkrankungen und dem Vorliegen des jeweiligen Allels bekannt. (Tabelle nach Philipps,2014)

Allel	Affinität zum LDL-Rezeptor (in %)	Vorkommen in der Normalbevölkerung (in %)	Klinische Assoziation
$\epsilon 2$	1	7	Hyperlipoproteinämie Typ III
$\epsilon 3$	100	78	keine bekannt
$\epsilon 4$	100	14	Alzheimer, kardiovaskuläre Erkrankungen

## **1.9 CAST: Calpastatin**

Das *CAST* Gen codiert für das Protein Calpastatin. Calpastatin ist das einzige bekannte endogene Protein, das Kalzium abhängige Cystein Proteasen (Calpain) spezifisch hemmen kann (Hübener et al., 2013). In mehreren Studien konnte gezeigt werden, dass die Expression der Huntingtin oder Ataxin-3-Proteine in voller Länge, unabhängig von der Größe der Polyglutaminsequenz, eine deutlich geringere Toxizität aufweisen als PolyQ – Fragmente (z.B. Martindale et al, 1998).

Aufgrund dieser Beobachtung wurde die Hypothese aufgestellt, dass erst die Fragmentierung der Polyglutamin enthaltenden Proteine zur Manifestation einer PolyQ – Erkrankung führt („Hypothese der toxischen Fragmente“).

Für die SCA3 ist der krankheitsauslösende Effekt durch die Fragmentierung von Ataxin-3 sehr kontrovers diskutiert. Haacke et al. konnten 2007 nachweisen, dass die Fragmentierung durch zwei Calpain Proteasen erfolgt. 2012 konnten Simoes et al. im Mausmodell zeigen, dass die Inhibierung von Calpain Proteasen eine signifikante Verbesserung der neurodegenerativen Symptome der SCA3 bewirkt. Dabei korrelierte der Spiegel an Calpain negativ mit der Konzentration an Calpastatin. Nach diesen Erkenntnissen stellt Calpastatin als einziger bekannter endogener Inhibitor von Calpain Proteasen (Takano et al., 2005) einen vielversprechenden Ansatzpunkt für zukünftige Therapiemöglichkeiten einer SCA3 dar.

### **1.9.1 rs1559085 und rs27852 in CAST**

Die Einzelnukleotid-Polymorphismen rs1559085 und rs27852 befinden sich jeweils in nicht codierenden Bereichen (chr5:96742998/chr5:96710335) des *CAST* Gens auf Chromosom 5. Veränderungen einer Base führen daher an dieser Stelle nicht unmittelbar zu einer Veränderung auf Proteinebene. Allerdings wurden bereits zahlreiche Einzelnukleotid-Polymorphismen in intrinsischen Regionen identifiziert, welche einen Einfluss auf die Morphologie und Funktion des translatierten Genes aufwiesen.

Der als rs1559085 bezeichnete Polymorphismus betrifft das Nukleotid an Stelle 96742998 auf dem Chromosom 5. In 87% der Fälle befindet sich an dieser Position in *CAST* eine Adenin-Base (A), mit einer Häufigkeit von 12% kann es allerdings zu einem Austausch von Adenin zu Guanin (G) kommen (dbSNP, NIH). Allen et al. konnten 2010 für den Polymorphismus rs1559085 in *CAST* einen signifikanten Zusammenhang ( $p=0,1$ ) zwischen dem Vorliegen der Variante und einer Parkinson-Erkrankung nachweisen. Diese Ergebnisse lassen einen möglichen Effekt des Polymorphismus in neurodegenerativen Triplet-Erkrankungen vermuten.

Der rs27852 befindet sich ebenfalls auf Chromosom 5, 32.7kB entfernt von rs1559085. An dieser Position können vier Basen vorkommen. In der Normalbevölkerung kommt Adenin (A) mit einer Häufigkeit von 31% vor und Cytosin (C) mit einer Häufigkeit von knapp 68% (dbSNP,NIH). In seltenen Fällen können anstelle von A oder C auch Guanin (G) oder Thymin (T) vorliegen. Mit dem Polymorphismus rs27852 konnten bisher keine Erkrankungen assoziiert werden.

### **1.10 Fragestellung der Arbeit: Ein Ausblick**

Trotzdem die krankheitsverursachende Mutation der SCA3 bereits seit mehr als 25 Jahren bekannt ist, sind viele Aspekte der Erkrankung noch nicht vollständig verstanden. In spezialisierten Forschungsgruppen mit internationalen Kooperationen werden immer mehr Erkenntnisse hinzugewonnen mit dem übergeordneten Ziel den Pathomechanismus der Erkrankung besser zu verstehen und Therapiemöglichkeiten für Betroffene zu entwickeln.

Ein Aspekt der SCA3 der bis heute nicht verstanden ist, ist die große Spanne des Erkrankungsalters zwischen 5-70 Jahren. Mit einem durchschnittlichen Erkrankungsalter von 20-50 Jahren erkranken die meisten Betroffenen in der ersten Lebenshälfte und erfahren dadurch eine erhebliche Verkürzung ihrer Lebenszeit. Seit nunmehr 25 Jahren ist eine negative Korrelation zwischen der individuellen Anzahl an CAG-Wiederholungen und dem Erkrankungsalter bekannt. Diese Korrelation kann, je nach Kohorte, für 45-70 % der Betroffenen einen möglichen Vorrausagewert haben. Für eine große Anzahl an Trägern der Mutation ist eine alleinige Bestimmung der CAG-Wiederholungen demnach kein aussagekräftiges Kriterium, um den Zeitpunkt und Verlauf der Erkrankung vorauszusagen.

Im Rahmen dieser Arbeit werden Einzelnukleotid-Polymorphismen in *APOE* und *CAST* analysiert, um zu überprüfen, ob diese einen Einfluss auf das Erkrankungsalter haben. Diese beiden Gene wurden ausgewählt, da sie *in vivo* vielversprechende Eigenschaften mit potenziellem Einfluss auf die Erkrankung aufweisen. Während *APOE* bereits intensiv im Rahmen neurodegenerativer Erkrankungen erforscht wurde, ist der Einfluss von *CAST* auf das Erkrankungsalter der SCA3 ein ganz neuer Ansatzpunkt.

## **2 Material**

### **2.1 Patienten - Kohorte und DNA – Proben**

Im Rahmen dieser Arbeit wurden insgesamt 346 DNA-Proben von Betroffenen der SCA 3 analysiert. Mit dem gemeinsamen Ziel der Erforschung von seltenen Erkrankungen, haben sich weltweit Forschungsgruppen zusammengetan. Daher existieren große Datenbanken mit DNA-Proben und klinischen Informationen zu den Betroffenen. Die Proben dieser Arbeit wurden aus zwei verschiedenen Datenbanken rekrutiert. Diese wurden aus dem Patientenblut isoliert und liegen zur Analyse in 3:200 Verdünnung (3 µl DNA/300 µl H<sub>2</sub>O) bei - 20°C vor. Für die Etablierung der Typisierungsmethoden wurden Kontroll-Proben gesunder Probanden aus dem *Centre d'Étude du Polymorphisme Humain* (CEPH) verwendet. Ein Ethikvotum liegt unter der Projektnummer 046/2013B02 vor.

#### **2.1.1 EUROSCA (Proben aus Mittel- & Ost-Europa)**

Im Jahr 2008 entstand ein europäisches Projekt zur Erforschung genetischer Hintergründe der autosomal-dominant vererbten Ataxien (SCA) mit dem Ziel einer Therapieentwicklung. Mit dieser Absicht schlossen sich 22 Forschungsgruppen aus 9 verschiedenen europäischen Ländern zusammen und generierten Patientenproben für die weltweit größte Sammlung von DNA Proben SCA Betroffener (EUROSCA). Dank dieser Proben-Sammlung ist es möglich, auf eine große Anzahl an SCA3 Patienten-Proben zuzugreifen und diese für eine genetische Untersuchung zu rekrutieren. In einer separaten Datenbank wurden zusätzlich klinische Daten, das Erkrankungsalter und familiäre Zusammenhänge der Proben gesammelt. Diese Situation erleichtert die Analyse genetischer Einflüsse auf die Erkrankung in einem statistisch relevanten Umfang. In dieser Arbeit wurden insgesamt 295 Proben der EUROSCA Kohorte auf zwei Einzelnukleotid-Polymorphismen (rs429358 und rs7412) im *APOE* und zwei (rs1559085 und rs27852) im *CAST* Gen analysiert. Zusätzlich standen uns freundlicherweise aus vorherigen Forschungsarbeiten, Informationen über die individuelle Anzahl der CAG-Wiederholungen in *ATXN3*, zur Verfügung und konnten in die statistische Auswertung mit einbezogen werden.

### **2.1.2 EuSAge (Proben aus Israel und Peru)**

Im Rahmen dieser Arbeit wurden zusätzliche Patientenproben aus einer zweiten Datenbank (EuSAge) analysiert. Die EuSAge Kohorte umfasst Proben aus Israel und Peru. Dabei wurden unter Zusammenarbeit mit Kooperationspartnern aus den jeweiligen Ländern DNA-Proben von betroffenen Familien gesammelt und uns für die Analyse von Einzelnukleotid-Polymorphismen in *APOE* (rs429358, rs7412) und *CAST* (rs1559085, rs27852) zur Verfügung gestellt. Aus Israel wurden 35 Proben analysiert, darunter befanden sich drei Betroffene im präsymptomatischen Stadium (SARA 0) und eine Kontrollperson mit zwei betroffenen Schwestern. Aus Peru wurden 16 Proben aus fünf Familien analysiert. Es standen klinische Informationen (SARA Score) und die Information über das individuelle Erkrankungsalter der Betroffenen zur Verfügung. Die Anzahl der CAG-Wiederholungen in *ATXN3* wurde ebenfalls zur Verfügung gestellt. In einer stichprobenhaften Qualitätsanalyse konnte jedoch eine erhebliche Diskrepanz zwischen den erhobenen Werten der Kooperationspartner und der Kontrollanalyse in der Diagnostik in Tübingen festgestellt werden. Um eine statistische Auswertung nicht zu verfälschen, wurde für alle Proben aus der EUSAge Kohorte eine erneute Analyse der CAG-Wiederholungen durchgeführt.

## 2.2 Primer

*Tabelle 3: Auflistung der Sequenzen aller verwendeten Primer. F = Vorwärts-, R = Rückwärtsprimer. Alle Sequenzen sind in 5' – 3' Richtung aufgeführt.*

Primer	Primer – Sequenz (5'–3')
APOE FO	ACTGACCCCGGTGGCGGAGGA
APOE RO	CAGGCGTATCTGCTGGGCCTGCTC
APOE 112 F	GGCGCGGACATGGAGGACgGC
APOE 112 R	GCGGTAAGCCTCTCACGTAATTCT
APOE 158 F	CGATGCCGATGACCTGCAGAcGC
APOE 158 R	CCCGGCCTGGTACTACTGCCAGtCA
APOE 112 F-C	gcgggcagggcggcGCGCGGACATGGAGGACGTGC
APOE 112 F-T	gcgggcGCGCGGACATGGAGGACGTGT
APOE 112 R	CGCCGCGGTAAGCCTCTCACGTAATTCT
APOE 158 F-C	gcgggcagggcggcGCCGATGACCTGCAGAAGC
APOE 158 F-T	gcgggcGCCGATGACCTGCAGAAGT
APOE 158 R	CTCGCGGGCCCCGGCCTGGTA
APOE Seq - F	GACCATGAAGGAGTTGAAGGCCTAC
CAST rs27852-F1	TGTCACCAGGAATAGAGCACT
CAST rs27852-R1	CTGTAAGCCTCTCACGTAATTCT
CAST rs27852-F2	GACTAATGTACAGCAATAACGG
CAST rs27852-R2	AAGAGGGAAGGAGGTCTGTG
CAST rs1559085-F1	GCATTGCCTCCTTGTACCTT
CAST rs1559085-F2	CAGAAGCACCCAGAACCCTT
CAST rs1559085-R1	GGGGATAGTTACATGACTGACTT

Alle aufgelisteten Primer wurden von Metabion International AG, Martinsried bezogen. Die Konzentration der gelieferten Primer betrug 100 pmol/µl. Aus diesen wurden Lösungen mit einer Konzentration von 10 pmol/µl hergestellt. Die Verdünnung erfolgte mit sterilem H<sub>2</sub>O (Ampuwa) im Verhältnis 1:10 (90 µl Ampuwa + 10 µl Primer). Die Stocklösungen wurden nach einem laborinternem Kennzeichnungssystem mit einer Buchstaben-Nummern-Kombination versehen und wie die Arbeitslösungen bei -20 °C gelagert.

## 2.3 Enzyme

*Tabelle 4: Auflistung der verwendeten Enzyme und ihrer Hersteller*

Enzyme	Hersteller
One Taq DNA Polymerase	New England BioLabs GmbH, Frankfurt am Main, D
FastStart Polymerase	New England BioLabs GmbH, Frankfurt am Main, D
HotStart Polymerase	New England BioLabs GmbH, Frankfurt am Main, D

Alle Enzyme wurden nach Herstellerangaben bei -20°C gelagert und zusammen mit dem entsprechenden Puffern des Herstellers verwendet.

## 2.4 Chemikalien und Reagenzien

Tabelle 5: Auflistung verwendeter Chemikalien und Reagenzien und ihre Hersteller

Bezeichnung	Hersteller
Ampuwa (steriles Wasser für Injektionszwecke)	Fresenius Kabi AG, Bad Homburg, D
Desoxynukleotide (dNTPs) 10mM	Thermo Fisher Scientific, Karlsruhe, D
GenomeLab DNA Size Standard Kit – 600	Beckman Coulter, Beverley, MA, USA
Glycogen	Beckman Coulter, Beverley, MA, USA
Lambda/HindIII DNA Größenmarker	Thermo Fisher Scientific, Karlsruhe, D
pUC19/Msp3I DNA Größenmarker	Thermo Fisher Scientific, Karlsruhe, D

## 2.5 Lösungen und Puffer

Tabelle 6: Auflistung der verwendeten Lösungen und Puffer und ihre Zusammensetzungen

Puffer	Zusammensetzung
DNA Ladepuffer 6x	10 mM Tris-HCl (pH 7.6), 0,03 % Bromphenolblau, 0,03 % Xylencyanol FF, 60 % Glycerol und 60 mM EDTA
TBE Puffer 10x	890 mM Tris 890 mM Borsäure 20mM EDTA (pH 8,0)
3M Natriumacetat (pH 5,2)	3M Natriumacetat mit Essigsäure 100% auf pH 5,2 titriert
100mM Na <sub>2</sub> EDTA (pH 8,0)	100mM Na <sub>2</sub> EDTA mit 5M NaOH auf pH 8,0 titriert

Der TBE Puffer wurde im Verhältnis 1:10 verdünnt und als TBE 1x Puffer verwendet. Die Lagerung erfolgte mit Ausnahme des DNA Ladepuffers 6x (4°C) bei Raumtemperatur.

## 2.6 Fertige Reagenzienkits

Tabelle 7: Auflistung der verwendeten Reagenziensets und ihre Hersteller

Kit	Hersteller
GenomeLab DTCS Quick Start Kit	Beckman Coulter, Beverley, MA, USA
LightCycler 480 High resolution Melting Master Kit	Roche Diagnostics, Mannheim, D
QIAquick PCR Purification Kit	Qiagen GmbH, Hilden, D

## 2.7 Geräte

*Tabelle 8: Liste der verwendeten Geräte mit der entsprechenden Bezeichnung und ihrer Hersteller*

Gerät	Bezeichnung und Hersteller
Gelelektrophoresekammer	Bio-rad Laboratories GmbH, München, D
LightCycler	LightCycler 480 II, Roche Diagnostics GmbH, Mannheim, D
Photometer	Ultraspec 3100 pro Amersham Biotech Bioscience Europe GmbH, Freiburg, D
Sequenzierer	CEQ 2000 Genetic Analysis System, Beckman Coulter, Beverly, MA, USA
Thermo Cycler	G-storm, AlphaMetrix Biotech GmbH, Rödermark, D
Vakuumzentrifuge	Vakuumzentrifuge Concentrator 5301, Eppendorf AG, Hamburg, D

## 2.8 Verbrauchsmaterialien

*Tabelle 9: Auflistung der verwendeten Verbrauchsmaterialien und die entsprechenden Hersteller*

Material	Hersteller
Easy Peel Thermo Scientific	Thermo Fisher Scientific, Karlsruhe, D
LightCycler 480 Multiwell Plate 384	Roche Diagnostics, Mannheim, D
Reaktionsgefäß, 1,5 ml	Eppendorf, Hamburg, D
Reaktionsgefäß, 0,5 ml	Eppendorf, Hamburg, D
Sequenzierplatte 96-well	Greiner bio-one, Frickenhausen, DE

## 3 Methoden

### 3.1 Amplifizierung der Gen-Abschnitte

#### 3.1.1 Polymerase-Kettenreaktion (PCR)

Die Polymerase – Kettenreaktion (PCR) ist eine enzymatische Reaktion, die zur *in vitro* Vervielfältigung von DNA verwendet wird und erstmal 1985 von Saiki *et al.* beschrieben wurde. Die Spezifität der PCR wird durch chemisch synthetisierte Oligonukleotide (Primer) gewährleistet. Diese binden komplementär an den Randbereich der zu amplifizierenden Sequenz und begrenzen die Reaktion auf die erwünschte Zielregion (Ramesh *et al.*, 1992). Jedwede PCR- Reaktion benötigt eine zu vervielfältigende DNA als Vorlage (*Template*), mindestens zwei Primer, Nucleotide und eine DNA – Polymerase (Garibyan *et al.* 2013). Die Amplifikation läuft an beiden DNA – Einzelsträngen in 5' zu 3' Richtung ab und durchläuft grundsätzlich drei verschiedene Temperatur – Phasen.

Die erste Phase der Denaturierung findet bei 95°C statt und dient der Auftrennung der Wasserstoffbrückenbindungen zwischen den Basen der zu amplifizierenden DNA. Nach diesem Schritt liegen zwei Einzelstränge einer normalerweise doppelsträngigen DNA vor (Petruska *et al.*, 1988). In der darauffolgenden Phase wird das Reaktionsgemisch auf eine spezifische Annealing – Temperatur herabgekühlt, um die Anlagerung der Primer an die DNA-Einzelstränge zu ermöglichen. Diese Temperatur liegt im Normalfall zwischen 55°C und 72°C und ist abhängig von der Basensequenz und Länge der synthetisierten Primer. Die optimale Annealing – Temperatur eines Primers liegt ca. 10°C unter dessen spezifischer Schmelztemperatur (Ramesh *et al.*, 1992). Im letzten Schritt wird das Reaktionsgemisch in der Regel auf 72°C erhitzt um die Phase der Extension zur ermöglichen. Dabei synthetisiert die Polymerase, ausgehend vom 3' Ende der Primer, eine komplementäre Kopie des DNA-Einzelstranges (Saiki *et al.*, 1989). Jede Kopie der Zielsequenz dient in weiteren Durchläufen der beschriebenen Reaktionsschritte als DNA-Template, somit kommt es zu einer exponentiellen Vervielfältigung der Zielsequenz (Garibyan *et al.*, 2013).

### 3.1.2 Allelspezifische PCR

Die allelspezifische PCR (AS-PCR) stellt eine effektive und kostengünstige Methode zur Detektion von bekannten Einzelnukleotid-Polymorphismen dar. Erstmals beschrieben von Newton *et al.* (1989) wurde die Methode seither stetig weiterentwickelt.

Eine allelspezifische PCR verlangt die Synthese von spezifischen Primern, die eine Amplifikation durch die Polymerase nur erlauben, wenn die Base am 3' Ende exakt komplementär an entsprechender Stelle der Template-DNA binden kann (Darawi *et al.*, 2013). Nach der Methode von Chen (2016) wurden zur Identifizierung von einem Einzelnukleotid-Polymorphismus zwei Vorwärts-Primer verwendet, die am 3' Ende jeweils spezifisch an eine Variante des Polymorphismus binden und ein Rückwärts-Primer der beide Varianten amplifiziert. Um anschließend eine Differenzierung der Polymorphismen durchführen zu können, müssen PCR-Produkte mit unterschiedlichen Eigenschaften entstehen. Durch einen GC-reichen Anhang am 5' Ende der allelspezifischen Primer entstehen PCR - Produkte unterschiedlicher Größe mit unterschiedlichen Schmelzeigenschaften (Abbildung 2).

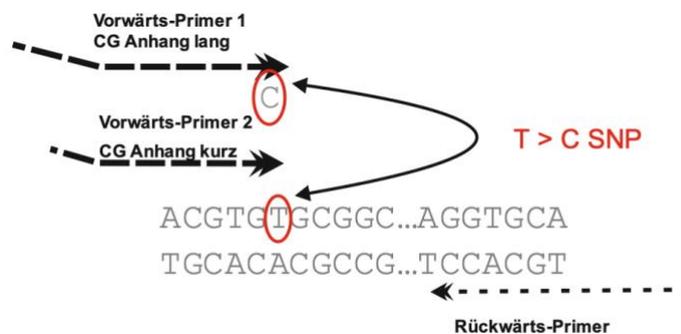


Abbildung 2: **Allelspezifische PCR mit Bindung unterschiedlicher Primer je nach Vorliegen eines Einzelnukleotid-Polymorphismus.** Darstellung einer allelspezifischen PCR mit designten Primern. Während der PCR entstehen Produkte unterschiedlicher Länge. Dies ist bedingt durch den individuellen 5' GC-Anhang der Primer. In Abhängigkeit des vorliegenden Polymorphismus bindet nur ein Vorwärts-Primer komplementär. Verhält sich der Polymorphismus heterozygot, binden beide Primer und es entstehen zwei Produkte.

### 3.1.3 Amplifizierung von rs429358 und r27412 in APOE

Der Reaktionsansatz für die Analyse von Einzelnukleotid-Polymorphismen im APOE mittels einer allelspezifischen PCR setzte sich pro 1,0 µl DNA aus 5,0 µl High-Resolution-Master-Mix, jeweils 0,2 µl beider Vorwärts-Primer, 0,4 µl des Rückwärts-Primers und 1,2 µl MgCl<sub>2</sub> zusammen. Mit 2,0 µl H<sub>2</sub>O für PCR-Reaktionen (aus Light Cycler 480 High Resolution Melting Master Kit) wurde das Reaktionsvolumen pro DNA-Probe auf 10µl gebracht. Pro Polymorphismus und DNA-Probe musste ein PCR-Reaktionsmix angesetzt werden.

Tabelle 10: Auflistung der verwendeten Primer für die Amplifikation von APOE zur anschließenden Analyse von den Einzelnukleotiden rs429358 und rs7412. Alle Primer sind in 5' – 3' Richtung aufgelistet. F= Vorwärts- und R= Rückwärtsprimer.

Primer	Sequenz
APOE 112 F-C (10pmol/µl)	gcgggcagggcggcGCGCGGACATGGAGGACGTGC
APOE 112 F-T (10pmol/µl)	gcgggcGCGCGGACATGGAGGACGTGT
APOE 112 R (10pmol/µl)	CGCCGCGGTACTGCACCAGG
APOE 158 F-C(10pmol/µl)	gcgggcagggcggcGCCGATGACCTGCAGAAGC
APOE 158 F-T(10pmol/µl)	gcgggcGCCGATGACCTGCAGAAGT
APOE R(10pmol/µl)	CTCGCGGGCCCCGGCCTGGTA

Die Reaktion wurde im Light Cycler 480 in einer 384-er PCR-Platte durchgeführt. Das Programm „Gene Scanning II“ wurde für eine allelspezifische PCR mit anschließender Schmelzkurvenanalyse modifiziert (siehe Tabelle 12 a und b).

Zunächst findet über 10 Minuten eine Präinkubationsphase bei 95°C statt. Diese dient der Aktivierung der Polymerase. Anschließend wird bei der Analyse von APOE 45 Mal, bei der von CAST 50 Mal, ein Amplifikationszyklus durchlaufen. Dieser beinhaltet die drei Phasen einer PCR-Reaktion. Die Denaturierung findet für 10 Sekunden bei 95°C statt, die Annealing-Phase für 15 Sekunden bei 65-53°C und eine Extension für 15 (APOE) oder 30 (CAST) Sekunden bei 72°C. Während der Annealing-Phase wird das Prinzip einer Touchdown-PCR verwendet. Diese bietet sich an, wenn man die genaue Annealing-Temperatur eines spezifischen Primers nicht kennt. Dabei wird eine Spanne von (hier) 12°C im Bereich der zu erwartenden Schmelztemperatur der Primer verwendet. Pro Zyklus wird die Annealing-Temperatur um 0.5°C gesenkt.

Die Dauer der Annealing-Phase wird nach der Größe der zu erwartenden PCR-Produkte abgeschätzt (Tabelle 11).

Tabelle 11: **PCR – Programm für eine allelspezifische PCR.** Verwendet wurde der Light Cycler 480 Instrument II, ein Hochdurchsatz – System zur Detektion und graphischen Darstellung von Echtzeit – PCR – Produkten. \* = ungefähre Dauer der Annealing – Phase =  $\frac{\text{Größe der PCR-Produkte in bp}}{25} + 10$

Reaktionsschritt	Anzahl d. Zyklen	Dauer	Temperatur	ΔTemperatur (°C)	
<b>Präinkubation</b>	1x	10 min	95°C	4.8°/s	
<b>Amplifikation</b>	45x	10 sek	95°C	4.8°/s	
		Touchdown (0.5°C/Zyklus)	15 sek*	65°C-53°C	2.5°/s
		15 sek	72°C	4.8°/s	

Im Anschluss an die Amplifikation der Zielsequenz findet eine Abschmelzung der PCR-Produkte statt. Hier ist ein Unterschied im Schmelzverhalten der verschiedenen PCR-Produkte von Bedeutung.

### 3.1.4 Amplifizierung von rs1559085 und rs27852 in CAST

Für die Genotypisierung von rs1559085 in CAST erfolgte zunächst eine PCR mit anschließender Schmelzkurvenanalyse im Light-Cycler. Für den PCR-Mix wurde das Light Cycler 480 High Resolution Melting Master Kit verwendet. Dabei wurden pro 1,0 µl DNA-Probe, 5,0 µl High-Resolution-Master-Mix, jeweils 0,2 µl Primer und 1,2 µl 25 mM MgCl<sub>2</sub> angesetzt. Mit 2,0 µl H<sub>2</sub>O für PCR-Reaktionen (aus Light Cycler 480 High Resolution Melting Master Kit) wurde das Reaktionsvolumen pro DNA-Probe auf 10µl gebracht. Für die Analyse multipler DNA-Proben wurde ein Mastermix entsprechend der oben genannten Mengenangaben erstellt.

Tabelle 12: **Eingesetzte Primer für die Amplifizierung des CAST-Gens im Bereich von rs1559085 und rs27852.** Aufführung der Primersequenz in 5' – 3' Richtung.

Primer	Sequenz
CAST rs1559085-F1 (10pmol/µl)	GCATTGCCTCCTTGTTACCTT
CAST rs1559085-R1 (10pmol/µl)	GGGGATAGTTACATGACTGACTT
CAST rs27852-F2 (10pmol/µl)	GCACTAATGTACAGCAATAACGG
CAST rs27852-R2 (10pmol/µl)	AAGAGGGAAGGAGGTCTGTG

Die Proben wurden auf eine 384-er Platte (Thermo Fisher Scientific, DE, Krefeld) aufgetragen und das PCR-Programm wurde im Light-Cycler 480 durchlaufen. Eine Analyse der Daten erfolgte anschließend nach der Methodik der Schmelzkurvenanalyse.

### **3.2 Schmelzkurvenanalyse**

Erstmals durchgeführt in den späten 90er Jahren (Lay und Wittwer, 1997) stellt die Methodik der Schmelzkurvenanalyse heute eine beliebte und effektive Methode zur Detektion genetischer Varianten dar (Simko, 2016). Dabei handelt es sich um eine Post-PCR-Methode, die demnach erst nach der Durchführung einer PCR zum Einsatz kommt. Insbesondere nach der Einführung von fluoreszierenden Farbstoffen und Geräten die eine hochauflösende Echtzeit-Analyse von Schmelzkurven erlauben, gehört die Schmelzkurvenanalyse zu einer Standardmethode bei der Genotypisierung von DNA-Proben (Wittwer, 2003). Eine hochauflösende Schmelzkurvenanalyse ermöglicht die Detektion von genetischen Varianten im Bereich der amplifizierten Zielsequenz, ohne eine zeit- und kostenaufwendige Sequenzierung (Liew *et al.*, 2004).

Um einen Unterschied im Schmelzverhalten der PCR – Produkte sichtbar machen zu können, wird während der Amplifikation der Zielsequenz ein fluoreszierender Farbstoff eingesetzt. Dieser Farbstoff bindet spezifisch an doppelsträngige DNA und weist im ungebundenen Zustand nur eine sehr geringe Fluoreszenz auf (Wittwer *et al.*, 2003). Im Anschluss an die Amplifikation werden die entstandenen PCR – Produkte unter einer langsamen Erhitzung auf 95°C zunehmend denaturiert. Während dieser Phase der Abschmelzung von PCR-Produkten kommt es durch die Auftrennung von DNA-Doppelsträngen zur Abnahme der Fluoreszenz des verwendeten Farbstoffes. Unter Auftragen der Fluoreszenz der PCR - Produkte gegen die Temperatur entsteht eine charakteristische Schmelzkurve. Der individuelle Verlauf der Schmelzkurve ist dabei abhängig von der Länge der PCR – Produkte, ihrem Gehalt an Guanin (G) und Cytosin (C) und von der Komplementarität der DNA – Doppelstränge (Reed *et al.*, 2007).

### 3.2.1 Analyse im Light Cycler 480 II

Im Anschluss an die allelspezifische Amplifikation der Zielsequenz wurde eine Schmelzkurvenanalyse im Light Cycler 480 Instrument II durchgeführt. Hierfür wurden die entstanden PCR-Produkte ohne Unterbrechung des PCR – Programms (s. Tabelle 12b) nochmals für 1 Minute auf 95°C erhitzt und im direkten Anschluss für 1 Minute auf 40°C abgekühlt. Dieser Reaktionsschritt gewährleistet eine Formierung der entstanden komplementären Kopien der Zielsequenz zu Doppelsträngen.

Im weiteren Verlauf wurde der Reaktionsansatz über einen Zeitraum von 1°C pro Sekunde von 60°C auf 95°C erhitzt, um eine Denaturierung der Doppelstränge herbeizuführen und ein spezifisches Schmelzverhalten detektieren zu können.

*Tabelle 12: fortlaufendes Programm im Light Cycler 480 Instrument II zur Detektion genetischer Varianten in der Zielsequenz. Eine initiale Erhitzung und anschließende Abkühlung der PCR – Produkte, unmittelbar nach der Amplifikation gewährleistet eine sichere Ausbildung von DNA – Doppelsträngen. Eine anschließende schrittweise Erhitzung der Doppelstränge auf 95°C führt zum Aufbrechen von Doppelsträngen und eine Veränderung des Fluoreszenz-Verhalten des verwendeten Farbstoffes.*

Reaktionsschritt	Temperatur	Dauer	$\Delta$ Temperatur (°C)
	95°C	1 min	4.4°/s
	40°C	1 min	2.2°/s
<b>Schmelzkurve</b>	60°C-95°C	----	1°/s
<b>Kühlung</b>	40°C	30s	2.2°/s

Für die Analyse des charakteristischen Schmelzverhalten ist die Fluoreszenz des verwendeten Farbstoffes von Bedeutung. Der Light Cycler 480 Instrument II trägt das Fluoreszenz - Verhalten der PCR – Produkte gegen die Temperatur auf und ermöglicht eine graphische Visualisierung des individuellen Schmelzverhaltens der Produkte (Abbildung 3 A) bei einer Temperaturveränderung von 65°C bis 95°C in Intervallen von 1°C pro Sekunde. Durch die I. mathematische Ableitung der Funktion kann die Veränderung der Fluoreszenz mit Maximum und Minimum dargestellt werden. Diese graphische Darstellung erlaubt eine eindeutige Differenzierung der entstanden PCR – Produkte (vgl. Abbildung 3 C). Eine weitere Möglichkeit der Analyse besteht in der Darstellung der Unterschiede in der Fluoreszenz im Schmelzverlauf. Diese wird von der Light Cycler Software „Gene Scanning“ mit Hilfe eines Bland-Altman-Plots anschaulich dargestellt. Dabei werden alle Proben bezüglich ihrer Zu- und Abnahme der Fluoreszenz im

Verlauf der Temperaturerhöhung analysiert und aufgezeichnet (Abbildung 3 B). Anhand eines Bland-Altman-Plots ist es möglich insbesondere heterozygote Proben von homozygoten zu unterscheiden.

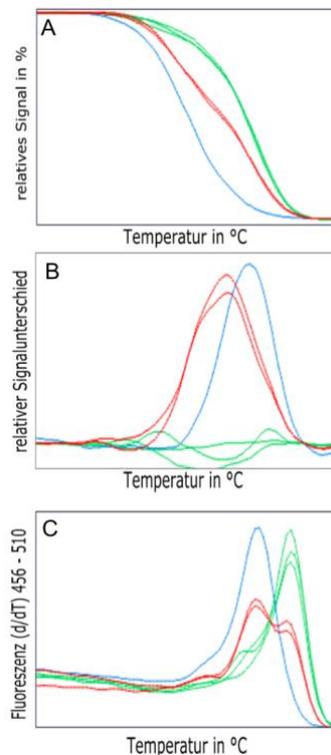


Abbildung 3: **Schmelzkurvenanalyse im Light-Cycler.** Beim High – Resolution – Melting wird im Light Cycler die Änderung des Fluoreszenzverhaltens eines Farbstoffes detektiert. Dabei kann während des Schmelzvorgangs ein Unterschied in der Fluoreszenz der Proben festgestellt werden. A) zeigt die unterschiedlichen Schmelzkurvenverläufe von homozygoten und heterozygoten Proben nach Amplifikation mittels allelspezifischer Primer. In C) ist die Fluoreszenz gegen die Schmelztemperatur aufgetragen. Die Schmelzkurven sind dabei in ihrer 1. mathematischen Ableitung dargestellt. Es sind deutlich drei verschiedene Kurvenverläufe zu erkennen. Diese können als drei mögliche Genotypen eines Polymorphismus interpretiert werden. Hier entspricht die grüne Kurve dem homozygoten Genotyp C/C, die blaue Kurve dem homozygoten Genotyp T/T und blau dem heterozygoten T/C. B) zeigt die Unterschiede des Fluoreszenzverhaltens in Form eines Bland-Altman-Plots. Dabei sind heterozygote Proben am deutlichsten anhand ihrer starken Fluoreszenzzunahme im Schmelzverlauf zu erkennen (rot). Homozygote Proben sind hingegen nicht eindeutig zu differenzieren (blau).

### 3.3 Agarose - Gelelektrophorese

Die Elektrophorese ist eine gängige Methode zur Auftrennung von geladenen Substanzen wie Proteinen und Nukleinsäuren (DNA). Dabei wirkt Agarose als stabilisierendes Medium mit einem erheblichen Einfluss auf die Wanderungsgeschwindigkeit der Moleküle. Agarose ist ein ungeladenes, lineares Molekül und zeigt wenig Interaktion mit anderen Molekülen. Durch ein Schmelzen von Agarose bilden ihre linearen Moleküle Helices, die wiederum zu Suprafasern polymerisieren. Es entsteht ein Agarose – Netzwerk mit Poren von

100 - 300 nm Durchmesser. Die Größe der Poren hängt von der Konzentration der Agarose im Gel ab. Eine Auftrennung der geladenen Substanzen wird durch den Aufbau eines elektrischen Feldes gewährleistet. Nach Beladen der Gelkammern, läuft die Elektrophorese in einer Elektrophorese-Kammer ab. Die angelegte Spannung und Dauer der Elektrophorese hängen direkt von der Konzentration der Agarose und der Größe der PCR-Fragmente ab. Bei einer Gelelektrophorese wird stets ein Größenstandard als Marker verwendet. (Adkins et al., 1996)

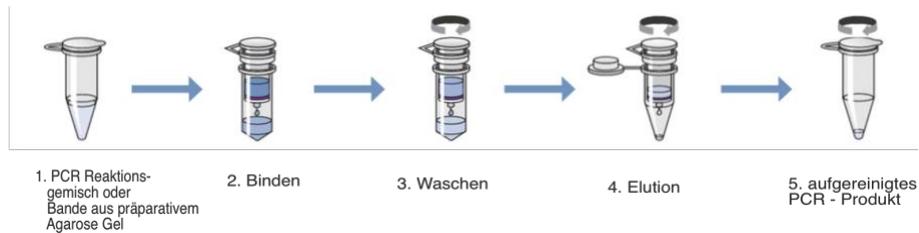
### **3.3.1 Analytische Agarose – Gelelektrophorese**

Für die Elektrophorese der linearen DNA – Fragmente im Größenbereich von 55 – 65 Basenpaaren (bp) wurde ein Agarose – Konzentration von 2,5 – 3 % (w/v) in 1 x TBE Elektrophoresepuffer verwendet. Für eine Visualisierung unter UV – Licht wurden 2 – 5 µl Ethidiumbromid hinzugefügt. Die DNA wurde mit 0,2 Volumen 6x Ladepuffer gemischt und aufgetragen. Als Größenstandard diente pUC 19 DNA die durch MspI vollständig verdaut und in gelöst wurde. Die Gelelektrophorese wurde je nach Größe des Gels bei einer Spannung von 90 - 120 V über 30 Minuten durchgeführt.

### **3.4 Aufreinigung von PCR – Produkten**

Nach Ablauf einer PCR Reaktion liegen neben den erwünschten PCR Produkten mit der Zielsequenz auch alle weiteren Bestandteile einer PCR- Reaktion weiterhin im Gemisch vor. Um Primer, dNTPs und Salzurückstände zu entfernen führt man eine Aufreinigung der PCR – Produkte durch. Insbesondere vor Durchführung einer Sanger-Sequenzierung ist dieser Schritt notwendig.

Die Aufreinigung wurde mit dem QIAquick PCR Purification Kit von Qiagen nach Protokoll des Herstellers durchgeführt. Die einzelnen Reaktionsschritte werden im Folgenden aufgeführt (Abbildung 4).



**Abbildung 4: Aufreinigung von PCR – Produkten mit dem QIAquick PCR Purification Kit.** Darstellung eines vereinfachten Prinzips der Aufreinigung von PCR – Produkten bei Verwendung eines QIAquick PCR Purification Kits (QIAGEN, Düsseldorf, DE). Die einzelnen Schritte der Aufreinigung sind schematisch dargestellt. Zunächst wird dem Post – PCR – Reaktionsgemisch ein Bindungs – Puffer aus dem Kit hinzugegeben. Anschließend wird das Gemisch abzentrifugiert. Dabei bleiben die Nukleinsäuren in der Membran der verwendeten Säule hängen. Nach Verwerfen des Überstandes wird das Produkt auf der Membran mit der entsprechenden Lösung aus dem Kit (Wasch-Puffer) von Rückständen und Nebenprodukten gereinigt. Dafür wird das Gemisch erneut zentrifugiert. Zuletzt wird das gewonnene Produkt aus der Membran der Säule gelöst. Hierfür wird der Lösungs-Puffer aus dem Kit verwendet. Unter Zentrifugation können die Nukleinsäuren aus der Membran in die Pufferlösung übergehen.

Das QIAquick PCR Purification Kit enthält alle benötigten Chemikalien und Materialien. Lediglich Eppendorfgefäße zur Aufbewahrung der gereinigten Produkte müssen aus dem Laborbestand verwendet werden.

Zunächst wurde pro PCR-Gemisch eine Zentrifugationssäule mit Sammelgefäß verwendet. Auf eine Zentrifugationssäule wurde das gesamte Volumen des PCR-Gemisches gegeben und das 5 x Volumen an isopropanolhaltigem Bindungspuffer (PB) hinzugegeben. Anschließend wurde die Zentrifugationssäule für 1 Minute bei 13.000 rpm zentrifugiert. Der Überstand wurde verworfen und die Säule erneut im gleichen Sammelgefäß platziert. Im nächsten Schritt folgte eine Waschung der, auf der Membran der Zentrifugationssäule, gebundenen PCR-Produkte. Dafür wurde jeweils 750 µl Ethanol-Puffer (PE) auf eine Säule gegeben und die Gefäße im Anschluss erneut für 1 Minute zentrifugiert. Nach Verwerfen des Überstandes und erneuter Zentrifugation, um alle Restbestandteile an Ethanol zu entfernen, wird die Säule in einem frischen Eppendorfgefäß platziert. Im letzten Schritt, der Lösung, wurden durch Zugabe eines Lösungs-Puffers (EB) und anschließender Zentrifugation, die PCR-Produkte aus der Membran der Zentrifugationssäule gelöst. Die Konzentration der aufgereinigten PCR-Produkte variiert je nach verwendeter Menge an Elutions-Puffer. Um die Konzentration zu erhöhen wurde ein Volumen von 30 µl Elutions-Puffer verwendet und die Säule nach auftragen des Puffers für 1 Minute stehen gelassen.

Um die Qualität der Aufreinigung beurteilen zu können, wurden die aufgereinigten Produkte im Anschluss anhand einer 1 % Agarose Gel-Elektrophorese auf Nebenprodukte und Primerdimere analysiert und die Konzentration der enthaltenen Nukleinsäuren ermittelt.

### **3.5 Konzentrationsbestimmung von DNA in Post-PCR-Proben**

Die Konzentrationsbestimmung von DNA in einem Post-PCR-Gemisch dient der Qualitätsanalyse der durchgeführten PCR und gibt Anhalt darüber wieviel Nanogramm (ng) DNA pro  $\mu\text{l}$  des Gemisches vorliegen. Für die Bestimmung oder Abschätzungen können verschiedene Methoden verwendet werden. Diese unterscheiden sich in ihrer Spezifität und werden situationsabhängig unterschiedlich verwendet.

#### **3.5.1 Konzentrationsbestimmung mit dem Photometer**

Zur Bestimmung der Konzentration von DNA in Post-PCR-Proben mittels Photometer werden die Proben und eine entsprechende Referenz-Substanz benötigt. Bei der Referenz-Substanz handelt es sich in der Regel um den Puffer in dem das auf gereinigte PCR-Produkt vorliegt. Das Photometer wurde zunächst kalibriert und auf die Messung von doppelsträngiger DNA (dsDNA) eingestellt. Auf einer entsprechenden Cuvette (Eppendorf  $\mu\text{Cuvette}$ , 1 mm Schichtdicke) wurde 1  $\mu\text{l}$  Elutionspuffer aufgetragen und als Referenz auf den Wert Null gesetzt. Im Folgenden konnte jede Probe einzeln mit je 1  $\mu\text{l}$  auf ihren Gehalt an Nukleinsäuren analysiert werden. Die Konzentration der Probe wird in einem Photometer anhand der Absorption von Licht berechnet. Dabei wird die Absorption durch den negativen Logarithmus des Verhältnisses, von dem in die Probe eintretenden zum austretenden Licht berechnet. Die ermittelte Absorption hängt dabei von verschiedenen Faktoren der Probe ab. Diese werden im Lambert-Beer'schen -Gesetz zusammengefasst (Schichtdicke der Cuvette, Probenkonzentration und einem für die Proben spezifischen Absorptionskoeffizienten). Dabei ist der Absorptionskoeffizient nicht nur spezifisch für eine Probe, sondern ebenso für eine bestimmte Wellenlänge. Der Absorptionskoeffizient für Nukleinsäuren gilt aus diesem Grund nur für eine Wellenlänge von 260 Nanometer (nm). Um Kontaminationen der Proben

auszuschließen kann zusätzlich die Absorption bei 280 nm (Wellenlänge für Proteine und Phenole) und 230 nm (Wellenlänge für organische Verbindungen) berechnet werden und entsprechend über das Verhältnis der Absorption eine Aussage über die Reinheit der Probe getroffen werden.  $A_{260} / A_{280}$  sollte bei reinen Proben zwischen 1,8 und 2,0 liegen,  $A_{260} / A_{230}$  entsprechend bei  $> 2,0$ .

### **2.6.2 Konzentrationsbestimmung mittels Gel-Elektrophorese**

Um die DNA – Konzentration einer Probe abzuschätzen kann auch eine herkömmliche Gelelektrophorese durchgeführt werden. 0,1 Volumen der Post-PCR-Proben werden hierfür auf ein 1 % Agarose-Gel aufgetragen und bei 90 Volt für 30 Minuten aufgetrennt. Bei dieser Methode ist der gewählte Marker als Größenstandard bedeutend. Nach der Auftrennung im Gel kann die Konzentration der PCR – Produkte mit der Konzentration des Markers gleicher Größe verglichen werden. Für diese Methode müssen die Eigenschaften und das Verhalten des gewählten Markers bekannt sein (pUC19 DNA/Mspl (Hball) Marker 23). Die Eigenschaften können dem „User Guide“ des Markers entnommen werden. Abbildung 5 zeigt ein Beispiel für die Konzentrationsbestimmung mithilfe einer Gel-Elektrophorese.

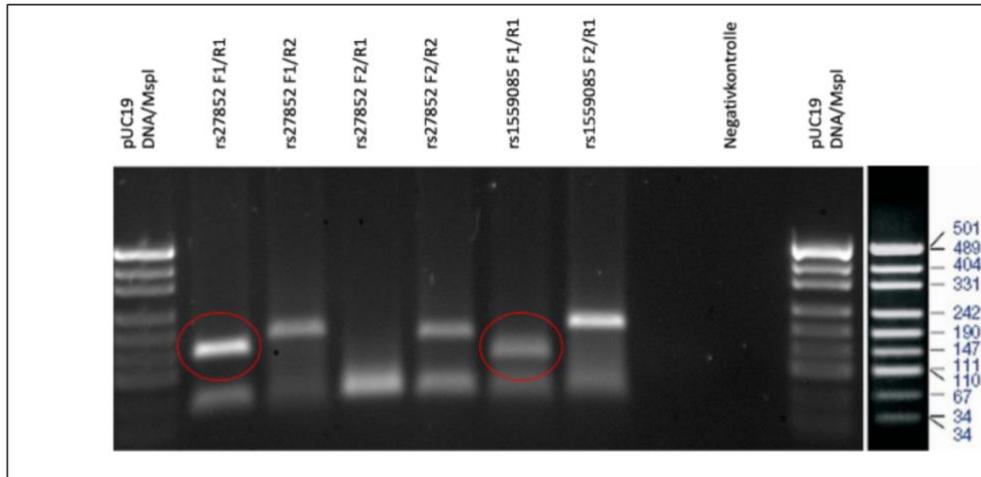


Abbildung 5: **Bestimmung der DNA – Konzentration mithilfe einer Gel- Elektrophorese.** Exemplarische Darstellung zur Verdeutlichung einer Konzentrationsabschätzung mithilfe einer Gel – Elektrophorese. Hier wird die Stärke der Bande von Probe 4 (markiert) von 325 bp mit der Signalstärke der 331 bp Markerbande (markiert) verglichen. Laut „User Guide“ entspricht die 331 bp Bande bei voller Signalstärke 61,6 ng/5  $\mu$ l. Bei ungefähr gleicher Signalstärke entspricht die Konzentration der Probe 4 daher 61,1 ng/aufgetragenes Volumen. Für die Konzentrationsangabe der Probe in ng/ $\mu$ l wird in diesem Fall 61,1 ng durch das Volumen der aufgetragenen Probe dividiert.

### 3.6 Sanger – Sequenzierung

Die Methode der Sanger – Sequenzierung, auch Didesoxymethode nach Sanger, ist eine bewährte Methode zur Analyse der Basenabfolge in einem DNA – Strang und wurde in den 1970er Jahren entwickelt. In den darauffolgenden Jahren wurde diese Methode weiterentwickelt und angepasst, sodass sie noch heute eine wichtige Möglichkeit zur Analyse von Basenabfolgen darstellt. Grundlage dieser Methode ist die hemmende Wirkung von Didesoxy-Triphosphaten (ddNTP) auf die DNA - Polymerase I (Sanger *et al*,1977). Der Einbau eines ddNTP anstelle eines einfachen Desoxyribonukleosidtriphosphats (dNTP) führt aufgrund der fehlenden Hydroxy-Gruppe am 3' Ende zum Kettenabbruch. Zur Durchführung dieser Methode muss ein Sequenzbereich des zu analysierenden DNA-Stranges bekannt sein. Dieser dient als Gegenstrang für einen Oligonukleotid-Primer. Es folgt eine PCR mit dNTPs und unterschiedlich fluoreszenzmarkierten ddNTPs. Durch die Konkurrenz von dNTPs und ddNTPs um die komplementäre Bindungsstelle erfolgt ein Kettenabbruch nach dem Zufallsprinzip. Statistisch gesehen sollte so nach dem Durchlaufen mehrerer PCR-Zyklen an jeder Stelle des DNA-Stranges zu einem beliebigen Zeitpunkt der Einbau eines fluoreszierenden ddNTPs erfolgen. In einer anschließenden

Elektrophorese werden die einzelnen Synthese-Produkte der Größe nach aufgetrennt und beim Passieren eines Lasers in ihrer Fluoreszenz angeregt. Die Identifikation der Basenabfolge ist daraufhin durch die Aufzeichnung eines Elektropherogramms möglich.

### 3.6.1 Sequenzierungs – PCR

Der zu sequenzierende DNA-Bereich wurde in einer vorhergegangenen PCR amplifiziert.

#### a) PCR zur Amplifikation von APOE für eine anschließende Sequenzierung

*Tabelle 13: PCR – Mix für die Amplifikation von APOE. Der amplifizierte Bereich des APOE Gens sollte beide Einzelnukleotid-Polymorphismen umfassen, um diese in einer anschließenden Sequenzierung detektieren zu können. Zudem sollte eine entsprechende PCR möglichst viel Amplifikat und wenig Nebenbanden liefern.*

PCR Mix	Menge
DNA	1,0 µl
Qiagen Puffer (10x)	2,0 µl
dNTP (2mM)	0,4 µl
RO (10 pmol/µl)	1,0 µl
FO (10pmol/µl)	1,0 µl
8% DMSO	1,6 µl
Q – Solution (5x)	4,0 µl
HotStart Taq Polymerase	0,125 µl
MgCl <sub>2</sub> (25mM)	1,2 µl
H <sub>2</sub> O (Ampuwa)	7,6 µl
Gesamtvolumen	20 µl

*Tabelle 14: PCR – Programm für die Amplifizierung von APOE. Für eine möglichst effektive Amplifikation von APOE ohne Nebenprodukte wählten wir eine sehr spezifische HotStart Taq Polymerase aus. Entsprechend musste das PCR – Programm im Thermocycler angepasst werden.*

Reaktionsschritt	Anzahl der Zyklen	Dauer des Zyklus (min)	Temperatur (°C)
<b>Präinkubation</b>	1x	15 min	95°C
<b>Amplifikation</b>	35x	1 min	94°C
		1 min	61°C
		1 min	72°C
<b>Elongation</b>	1x	10 min	72°C

## b) PCR zur Amplifikation von CAST für eine anschließende Sequenzierung

*Tabelle 15: PCR-Mix für die Amplifikation von CAST. Die Amplifikation der entsprechenden Genabschnitte von CAST zur anschließenden Sequenzierung und Typisierung der beiden Einzelnukleotid-Polymorphismen erfolgte in zwei getrennten PCR – Ansätzen, da die beiden Polymorphismen zu weit voneinander entfernt liegen. Die Zusammensetzung des PCR – Mix variiert nicht. Es wurden lediglich zwei verschiedene Primerpaare verwendet. \* = verschiedene Primerpaare zur Amplifikation von den Genabschnitten, in denen die entsprechenden Polymorphismen vorliegen.*

PCR Mix	Menge
DNA	1,0 µl
One Taq Puffer (5x)	4,0 µl
dNTP (2mM)	0,4 µl
rs27852-F1 / rs1559085-F2 (10 pmol/µl)*	1,0 µl
rs27852-R1 / rs1559085-R1 (10pmol/µl)*	1,0 µl
One Taq Polymerase	0,125 µl
H <sub>2</sub> O (Ampuwa)	11,3 µl
Gesamtvolumen	20 µl

*Tabelle 16: PCR – Programm für die Amplifikation von CAST Abschnitten im Thermocycler.*

Reaktionsschritt	Anzahl der Zyklen	Dauer des Zyklus (min)	Temperatur (°C)
<b>Präinkubation</b>	1x	1 min	94°C
<b>Amplifikation</b>	35x	1 min	94°C
		1 min	63°C
		1 min	68°C
<b>Elongation</b>	1x	6 min	68°C

Anschließend erfolgte eine Aufreinigung der PCR – Produkte und eine Konzentrationsbestimmung. Das entstandene Amplifikat diente als Vorlage für die folgende Sequenzierungs - PCR.

*Tabelle 17 a: Zusammensetzung eines PCR – Mix für eine Sequenzierungs – PCR mit Cyler Sequencing Kit (Beckmann Coulter, DE, Krefeld) nach der Protokollbeilage. Einzelne Mengen wurden leicht abgeändert um das Gesamtvolumen von 10 µl nicht zu überschreiten. Die Mengenangaben verstehen sich pro DNA – Probe und müssen bei Analyse multipler Proben entsprechend der Mengenangaben angepasst werden. \* = das Volumen der eingesetzten DNA ist abhängig von der ermittelten Konzentration und der Größe des zu analysierenden DNA – Abschnittes. Entsprechend der Empfehlungen aus der Protokollbeilage des Kits wurden DNA – Mengen zwischen 1- 4 µl eingesetzt.\*\* = das H<sub>2</sub>O - Volumen ergibt sich aus der Differenz zwischen bereits eingesetzten Volumina und dem Gesamtvolumen von 10 µl.*

PCR -Mix	Menge
DTCS Quick Start Master Mix	4 µl
SEQ-Primer (10pmol/µl)	0,5 µl
DNA (Amplifikat)	1-4 µl *
H <sub>2</sub> O (Ampuwa)	1,5 – 4,5 µl **
Gesamtvolumen	10 µl

Tabelle 17 b: **Liste der verwendeten Primer für die Sequenzierungs-PCR** von APOE und CAST. Die Primer wurden in 5'-3' Richtung aufgelistet.

Gen	Primer	Sequenz
APOE	FO (10pmol/μl)	ACTGACCCCGGTGGCGGAGGA
CAST (rs1559085)	rs1559085-R1 (10pmol/μl)	GGGGATAGTTACATGACTGACTT
CAST (rs27852)	rs27852 -R1 (10pmol/μl)	CTGTAAGCCTCTCACGTAATTCT

Im DTCS Quick Start Master Mix sind eine thermostabile Polymerase und Pyrophosphatase enthalten, sowie dNTPs und CEQ WellRED farbmarkierte ddNTPs. Für die Sequenzierungs-PCR wurde nur jeweils ein vorwärts (APOE) oder rückwärts (CAST) Primer verwendet. Aufgrund der farbmarkierten ddNTPs wurde der PCR-Mix unter möglichst geringem Lichteinfluss angesetzt. Die Reaktion wurde anschließend in einem Thermocycler durchgeführt. Dabei wurde folgender Programmablauf (nach Protokollangaben im Kit) ausgewählt.

Tabelle 18: **PCR – Programm im Thermocycler nach Angaben der Protokollbeilage im Cycle Sequencing Kit zur Amplifizierung mit farbmarkierten ddNTPs.** Die Veränderung des Gesamtvolumens muss im Gerät angegeben werden.

Reaktionsschritt	Anzahl d. Zyklen	Dauer	Temperatur
<b>Denaturierung</b>	30 x	20 sek	96°C
<b>Annealing</b>		20 sek	50°C
<b>Elongation</b>		4 min	60°C

### 3.6.2 Ethanol – Fällung

Im Anschluss an die Sequenzierungs-PCR wurden die die PCR-Produkte mittels Ethanol-Fällung von den Chemikalien des Reaktionsgemisches getrennt. Die Fällung erfolgte nach Protokollbeilage des DTCS Quick Start Kits.

Zunächst wurde eine Stopp-Lösung vorbereitet. Diese setzte sich pro Probe aus 1 μl Natriumacetat (3 M, pH 5.2) und 1 μl Dinatrium-Ethylendiamintetraessigsäure (0,1 M, pH 8.0), sowie 0,5 μl Glycogen (20mg/ml) zusammen. Jedem Post-PCR-Reaktionsgemisch wurden 2,5 μl Stopp-Lösung hinzugefügt und sofort gemischt. Die weiteren Reaktionsansätze erfolgten auf Eis und lichtgeschützt.

Im nächsten Schritt wurden zu jedem Reaktionsgemisch 60 μl kaltes 95% Ethanol direkt aus dem -20°C Gefrierschrank hinzugegeben. Anschließend erfolgte eine unmittelbare Zentrifugation (14.000 rpm) bei 4°C für 30 Minuten. Nach Ablauf der Zentrifugation wurde der flüssige Überstand abpipettiert, dabei

sollte ein sichtbares Pellet nicht berührt werden. Der nächste Reaktionsschritt wurde zweimal durchgeführt. Jedes Pellet wurde mit 200 µl kaltem 70 % Ethanol direkt aus dem Gefrierschrank versetzt und anschließend für jeweils 15 Minuten bei 4°C zentrifugiert (14.000 rpm). Der Überstand wurde verworfen und der beschriebene Reaktionsschritt wiederholt.

Im Anschluss an die dritte Zentrifugation wurde der Überstand erneut verworfen und die Gefäße mit geöffnetem Deckel in einen Vakuum-Trockner gestellt. Für mindestens 10 Minuten wurden hier unter Vakuum die letzten Flüssigkeitsspuren von den DNA-Pellets entfernt. Zuletzt sollte ein weißes Pellet in jedem Reaktionsgefäß sichtbar sein. Dieses wurde in je 40 µl SLS (Sample-Loading-Solution)-Puffer gelöst und für mindestens 60 Minuten bei 4°C inkubiert.

Anschließend wurden die Proben im Beckmann Coulter CEQ 8000 sequenziert und mit dem entsprechenden Analyse System ausgewertet. Dafür wurden die Proben auf eine Proben Platte (12 x 8 Reihen) aufgetragen und, um eine Verdampfung der Proben zu verhindern, mit einem Tropfen Mineralöl bedeckt. Leere Probengefäße einer Reihe wurden mit SLS-Puffer aufgefüllt. In einer separaten Platte wurden Probengefäße entsprechend der Belegung der Probenplatte mit einem Laufpuffer befüllt. Die Bedingungen der Kapillar – Elektrophorese wurden auf eine Kapillartemperatur von 50°C, einer Spannung von 4,5 kV und einer Analysezeit von 110 Minuten festgelegt (Programm LFR-a). Die Ergebnisse der Sequenzierung wurden anschließend mittels des GenomeLab Systems analysiert und als Elektropherogramm dargestellt (Abbildung 6).

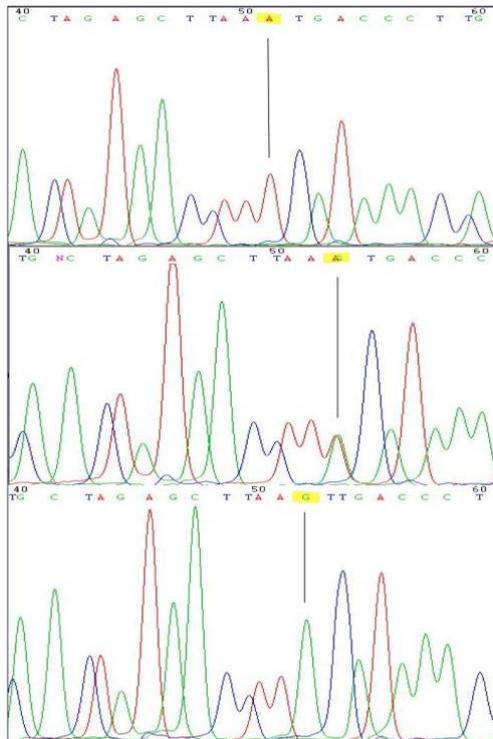


Abbildung 6: **Beispiel eines Elektropherogramms** als Darstellungsmöglichkeit einer Basenabfolge in einem bestimmten DNA – Strang. Die farbmarkierten ddNTPs ermöglichen eine Identifikation der jeweils eingebauten Base und eine Auftrennung der unterschiedlichen Stränge nach Größe gibt Aufschluss über die Abfolge der Basen im DNA – Strang.

### 3.7 Fragmentlängenanalyse mittels Kapillarelektrophorese

Die Fragmentlängenanalyse mittels Kapillarelektrophorese ist eine beliebte Methode zur Diagnostik von Triplet-Repeat-Erkrankungen (z.B. SCA3). Die Methode beruht auf dem Prinzip einer Gelelektrophorese in Form einer elektrophoretischen Analyse mittels Lasersystem. Nach der Amplifizierung definierter Gen-Abschnitte unter Verwendung eines fluoreszierenden Primers, werden die DNA-Einzelstränge nach der Denaturierung in einer Kapillarelektrophorese der Größe nach aufgetrennt und in einem lasergestützten Detektionssystem analysiert. Durch den Zusatz eines ebenfalls farbmarkierten Größenstandards können die PCR-Produkte in einem entsprechenden Software-Programm in ihrer Größe definiert werden. (Arnemann, 2019 b).

### 3.7.1 PCR von CAG – Repeats im ATXN3

Zunächst wurde unter Verwendung eines fluoreszenzmarkierten vorwärts-Primers und eines rückwärts-Primers der Bereich der CAG-Repeats im ATXN3 eingegrenzt und spezifisch amplifiziert. Für die Reaktion wurde eine Fast Start Taq Polymerase eingesetzt und der zugehörige Fast Start Puffer in zehnfacher Konzentration. Für das optimale Gelingen der PCR war das Zusetzen von Q-Solution erforderlich. Aufgrund der Verwendung eines farbmarkierten Primers, wurde der PCR-Mix unter möglichst geringem Lichteinfluss angesetzt.

*Tabelle 19: PCR – Mix für die CAG-Repeat Analyse im ATXN3. Die Mengenangaben verstehen sich pro DNA – Probe und müssen für einen Mastermix bei der Analyse multipler Proben entsprechend angepasst werden.*

PCR -Mix	Menge
DNA	3 µl
Fast Start Puffer (10x)	2,5 µl
dNTPs (2mM)	0,5 µl
fluoreszenzmarkierter vorwärts Primer (10pmol/µl)	1 µl
rückwärts Primer (10pmol/µl)	1 µl
Fast Start Taq	0,2 µl
Q - Solution	4,0 µl
H <sub>2</sub> O (Amouwa)	12,8 µl
Gesamtvolumen	25 µl

Die PCR erfolgte in einem Thermocycler. Die Reaktionsbedingungen wurden an die Verwendung der Fast Start Taq Polymerase angepasst. Um diese zu aktivieren erfolgte eine fünfminütige Präinkubation bei 94°C.

*Tabelle 20: PCR – Programm im Thermocycler für die CAG – Repeat Analyse im ATXN3.*

Reaktionsschritt	Anzahl d. Zyklen	Dauer	Temperatur
<b>Präinkubation</b>	1x	5 min	94°C
<b>Amplifikation</b>	30 x	1 min	94°C
		1 min	60°C
		2 min	72°C
<b>Elongation</b>	1x	10 min	72°C

### 3.7.2 Verdünnung der PCR – Produkte

Für die anschließende Fragmentlängenanalyse im Beckmann Coulter CEQ 8000 wurden die PCR-Produkte in einer Verdünnungsreihe 1:200 verdünnt. Zunächst wurde eine 1:20 Verdünnung angesetzt (Schritt 1, Tabelle 7), diese wurde in einem folgenden Schritt in die erforderliche 1:200 Verdünnung umgewandelt.

Dafür wurden die 10 µl für den Schritt 2 (Tabelle 7) aus der bereits entstanden 1:20 Verdünnung entnommen. Nach jedem Schritt erfolgte durch vorsichtiges auf – und abpipettieren eine gute Durchmischung der Verdünnung.

*Tabelle 21: Verdünnungsreihe der PCR – Produkte in eine 1:200 Verdünnung. Für die 1:200 Verdünnung erfolgte im 2. Schritt die Verwendung von 10 µl „PCR – Produkt“ aus der 1:20 Verdünnung. Das H<sub>2</sub>O sollte Reinstwasser (MilliQ) sein.*

Schritt	Verdünnung	PCR - Produkt	H <sub>2</sub> O
1	1:20	5 µl	95 µl
2	1:200	10 µl	90 µl

### 3.7.3 Versetzung der Proben mit Puffer und Größenmarker

Im nächsten Schritt erfolgte die Versetzung der zu analysierenden Proben mit einem Ladepuffer (SLS) und einem D1-markierten Größenstandard (600bp).

*Tabelle 22: Ansatz zur CAG – Repeat – Analyse im ATXN3. Verwendet wird ein D1 – markierter Größenstandard bis 600 bp. Der Ladepuffer (SLS) entspricht dem der Sanger – Sequenzierung von Beckmann Coulter (DE, Krefeld). Das zuvor verdünnte PCR – Produkt (1:200) wird eingesetzt. Pro zu analysierende Probe entsteht ein Gesamtvolumen von 40 µl. Zur Analyse multipler DNA – Proben empfiehlt sich das Ansetzen eines SLS – Marker – Mastermixes unter Beachtung unten angegebener Mengen.*

Analyse – Mix	Menge
SLS	34,5 µl
Größenstandard (600 bp)	0,5 µl
verdünntes PCR – Produkt (1:200)	5 µl
Gesamtvolumen	40 µl

### 3.7.4 Fragmentanalyse im Beckmann Coulter CEQ 8000

Zur Fragmentlängenanalyse wurden die Proben auf eine 12 x 8 Well - Probenplatte (Beckmann Coulter, DE, Krefeld) aufgetragen und mit einem Tropfen Mineralöl bedeckt. Auf der entsprechenden Pufferplatte wurden nach Belegungsplan der Probenplatte die Wells jeweils vollständig mit Laufpuffer befüllt. Die Beladung des Beckmann Coulter erfolgte nach Angaben des Gerätes. Ein digitaler Belegungsplan wurde erstellt, um die Zuordnung der Proben zu erleichtern. Für die Analyse wurde das Programm „Frag 3\_60 min“ ausgewählt. Dieses entspricht einer Kapillarelektrophorese bei einer Kapillartemperatur von 50°C, einer Spannung von 4,5kV und einer Analysezeit von 60 Minuten. Die Ergebnisse der Analyse wurden im GenomeLab System analysiert und graphisch als Peaks in einer Aufzeichnung des Fluoreszenzsignals gegen die Nukleotidlänge dargestellt.

### **3.8 Statistische Auswertung**

Die statistische Analyse der Ergebnisse erfolgte im Programm JMP® 15.0.0 (SAS Institute Inc.).

#### **3.8.1 Hardy – Weinberg – Verteilung**

Um die Normalverteilung der ermittelten Genotypen zu überprüfen, wurde die erhobene Verteilung, mit der einer idealen Verteilung nach dem Hardy-Weinberg-Gleichgewicht verglichen. Das Hardy-Weinberg-Prinzip, welches von dem Mathematiker G.H.Hardy und dem Mediziner Dr. med. W. Weinberg beschrieben wurde besagt, dass in einer idealen Population mit zufälliger Paarung und gleicher Verteilung der Geschlechter, die generelle Verteilung der Genotypen über Generationen hinweg stabil bleibt (Stern, 1943).

Die Berechnung erfolgte nach der Hardy-Weinberg-Gleichung  $p^2 + 2pq + q^2 = 1$  bei Betrachtung eines Genlocus mit den zwei möglichen Allelausführungen A und a.  $p^2$  entspricht dabei der relativen Häufigkeit an Proben mit den Allelen „AA“,  $q^2$  der Häufigkeit von „aa“ und  $2pq$  entspricht der relativen Häufigkeit an heterozygoten Proben mit den Allelen Aa. Die Werte für die „Minor Allele Frequency“ (MAF) der jeweiligen Einzelnukleotid-Polymorphismen, wurden der Datenbank „dbSNP Short Genetic Variations“ (NIH) entnommen. Dabei entsprach die  $MAF = q^2$

#### **3.8.2 Lineare Regressionsanalyse**

Die lineare Regressionsanalyse ist ein relevantes mathematisches Verfahren, mit dem es gelingt, einen linearen Zusammenhang zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen zu bilden. Die Merkmale können dabei quantitativ oder qualitativ sein. Die Qualität der Regression hängt jedoch wesentlich von der Qualität der einfließenden Variablen ab. Die Berechnung einer logistischen Regression erfolgt annäherungsweise über die Formel  $y = a + bx$ . Dabei entspricht b der Steigung der Geraden und a dem y-Achsenabschnitt. (Weiß, 2018)

Um den Einfluss der Einzelnukleotid-Polymorphismen auf das Erkrankungsalter statistisch zu überprüfen, wurde zunächst über eine bivariate Regressionsanalyse ein signifikanter Zusammenhang ( $p < 0,0001$ ) zwischen der

Anzahl an CAG-Wiederholungen und dem Erkrankungsalter bestätigt. Daraufhin wurde der Effekt der verschiedenen Genotypen der Einzelnukleotid-Polymorphismen in einem bivariaten Modell von Erkrankungsalter und der Anzahl an CAG-Wiederholungen einzeln überprüft und eine Vorhersagegleichung für die Ermittlung des Erkrankungsalters unter Berücksichtigung der Polymorphismen erstellt.

Über die Kategorie *Analysieren* und die Funktion *Modell anpassen* konnte eine Regression der Daten erstellt werden, bei der die verschiedenen Allele oder Genotypen als Variablen auf der *x-Achse* in einem Diagramm parallel dargestellt werden konnten. Dabei wurde das Erkrankungsalter als stetige Variable auf dem *y-Achsenabschnitt* dargestellt. Durch die Funktion *Gewichtung* konnte der zuvor berechnete *family factor* als stetige Variable in die Berechnung miteinbezogen werden. Der *family factor* entspricht dabei dem Kehrwert der Anzahl an typisierten Familienmitgliedern. Als Variable *x* wurden je nach Fragestellung einzelne Allele (z.B.  $\epsilon 2$ ) oder die verschiedenen Genotypen (z.B.  $\epsilon 2/\epsilon 3$ ) als nominale Variable eingesetzt. Um eine statistisch relevante Aussage bezüglich des Einflusses der einzelnen Genotypen treffen zu können, wurden die Daten auf der *x-Achse* jeweils mit der individuellen Anzahl an CAG-Wiederholungen gekreuzt. Als *Charakter* der Analyse wurde die Darstellung *Gewöhnliche kleinste Quadrate* gewählt mit der Betonung *Effekteinfluss*. Mittels dieser Analyse wurde jeweils eine Regressionsgerade der Variablen erstellt, sowie der Effektzusammenhang analysiert.

### **3.8.3 Familienfaktor**

Für die statistische Analyse der Daten wurde ein Familienfaktor definiert. Dieser entspricht dem Kehrwert der Anzahl an typisierten Mitgliedern einer Familie. Durch die Ermittlung eines solchen Wertes, werden alle verwandten Probanden in der späteren Auswertung als eine einzelne Person gewertet. Eine Verfälschung der Ergebnisse durch eine familiäre Häufung eines seltenen Genotyps kann durch diese Maßnahme verhindert werden.

#### **3.8.4 T-Test**

Mittels eines t-Testes zum Vergleich von zwei Populationen wurde überprüft ob bei dem Vorliegen eines bestimmten Genotypes der Einzelnukleotid-Polymorphismen ein signifikanter Unterschied im Erkrankungsalter nachweisbar ist. Dafür wurden die Probanden in die Gruppen Träger=1 und Nicht-Träger=0 eines bestimmten Allels eingeteilt.

#### **3.8.5 Varianzanalyse (ANOVA)**

Die einfaktorielle Varianzanalyse (ANOVA) bietet eine Möglichkeit der statistischen Analyse von mehr als zwei Stichproben unabhängiger Variablen. In diesem Fall stellt der T-Test keine geeignete Methodik dar. Eine ANOVA wird durchgeführt, um signifikante Unterschiede zwischen unabhängigen Stichproben zu analysieren. Dabei gelten gewisse Voraussetzungen, um die Ergebnisse einer ANOVA verwenden zu können.

- Die Stichproben sollten unabhängig voneinander erhoben sein
- Die Daten sollten normalverteilt sein
- Es besteht eine Varianzhomogenität der Stichproben

Diese Bedingungen sollten vor Durchführung einer ANOVA überprüft werden.

Es erfolgte die Varianzanalyse der verschiedenen Genotypen der jeweiligen Polymorphismen in APOE und CAST und eine Überprüfung der Normalverteilung.

#### **3.8.6 Graphen**

Eine graphische Darstellung der Daten statistischer Analysen bietet sich an, um dem Betrachter eine schnelle Erfassung der Daten zu ermöglichen. Für die Darstellung verschiedener Datensätze bieten sich unterschiedliche Diagramme an.

Um die Übereinstimmung zweier Variablen darzustellen, bietet sich ein Streudiagramm an. Diese Form der Darstellung wählten wir zur Darstellung des Zusammenhangs von der individuellen CAG – Länge und dem Erkrankungsalter, sowie zur anschließenden linearen Regression des Einflusses der Einzelnukleotid-Polymorphismen auf das Erkrankungsalter.

Zum Vergleich verschiedener Ausprägungen einer oder mehrere Variablen bietet sich die Darstellung mittels eines (gruppierten) Balkendiagramms an. Besonders zur Visualisierung der Verteilung der verschiedenen Genotypen wählten Balkendiagramme. Ebenso geeignet zur Darstellung der verschiedenen Genotypen erschienen und Kreisdiagramme. Die Häufigkeitsverteilung einer Variablen wird idealerweise mittels eines Histogramms dargestellt.

Unter der Funktion „Graphik erstellen“ konnten verschiedene graphische Darstellungsweisen ausgewählt werden.

## 4 Ergebnisse

Das individuelle Erkrankungsalter der SCA3 wird multifaktoriell beeinflusst und ist seit der Identifikation der krankheitsauslösenden Mutation in den frühen 90er Jahren immer wieder Gegenstand aktueller Forschungen. Dabei ist die Anzahl der CAG-Wiederholungen im expandierten Allel nicht nur der am besten erforschte, sondern auch der stärkste Einflussfaktor (de Mattos et al., 2018). Die bekannte inverse Korrelation zwischen der Anzahl an CAG-Wiederholungen und dem Erkrankungsalter weist allerdings nur einen Korrelationskoeffizienten von ca. 0,5 auf und lässt dabei den individuellen Erkrankungszeitpunkt für die Hälfte der Betroffenen unerklärt. Im Verlauf der letzten Jahrzehnte wurden bereits mehrere mögliche genetische Einflussfaktoren auf das Erkrankungsalter der SCA3 analysiert und einzelne als einflussreich identifiziert.

Als mögliche Modifikatoren des Erkrankungsalters wurden in dieser Promotion Einzelnukleotid-Polymorphismen in den Genen *APOE* und *CAST* analysiert. *APOE* war bereits im Zusammenhang mit anderen neurodegenerativen Erkrankungen Gegenstand mehrerer Studien. *CAST* ist im Zusammenhang mit PolyQ-Erkrankungen noch nicht als Modifikator des Erkrankungsalters untersucht worden, zeigte sich aber in einer Studie von Hübner et al. (2013) als möglicher Einflussfaktor auf die Pathogenese der SCA3 und vielversprechender Ansatzpunkt für eventuelle zukünftige Therapiemöglichkeiten.

Es wurden jeweils zwei Einzelnukleotid-Polymorphismen in einem Patientenkollektiv von 346 Betroffenen aus zwei Kohorten (EUROSCA und

EuSage) mittels High-Resolution-Melting typisiert. Im Rahmen der statistischen Analyse wurde ein Zusammenhang zwischen dem Auftreten der genetischen Variante und dem Erkrankungsalter der SCA3 überprüft.

#### **4.1 Analyse der Kohorte**

In der folgenden Auswertung wurden die Kohorten EUROSCA und EuSage gemeinsam betrachtet um eine ausreichend große Kohortenstärke zu erreichen. Wir konnten insgesamt 346 Proben von SCA3 Betroffenen typisieren. Darunter einige Proben von Verwandten. Um die Ergebnisse durch eine familiäre Häufung von seltenen Allelen nicht zu verfälschen, wurden alle analysierten Proben von Verwandten durch die Erhebung des *family factors* in der statistischen Auswertung als eine einzelne Probe betrachtet.

Das durchschnittliche Erkrankungsalter aller typisierten Betroffenen lag bei 40 Jahren. Das maximale Erkrankungsalter betraf 78 Jahre und das minimale Erkrankungsalter 10 Jahre. Diese Ergebnisse entsprechen den aktuellen Kenntnissen der Varianz des Erkrankungsalters an SCA3. Der Median der CAG-Repeat-Längen betrug 71 Wiederholungen. Diese Beobachtung ist vereinbar mit einem medianen Erkrankungsalter von 40 Jahren. In einer Regressionsanalyse des Erkrankungsalters in Jahren in Abhängigkeit von der Anzahl an CAG-Wiederholungen ergab sich eine negative Korrelation signifikanter Relevanz ( $p < 0,001^*$ ). Für die Berechnung des Erkrankungsalters der Kohorte ergab sich unter der Berücksichtigung der Variable *Anzahl an CAG-Wiederholungen* folgende Formel:

*Formel 1: Formel zur Berechnung des Erkrankungsalters der SCA3 nach einer Regressionsanalyse des Erkrankungsalters in Jahren in Abhängigkeit von der individuellen Anzahl an CAG-Wiederholungen.*

$$\text{Erkrankungsalter} = 151,43 - 1,571 \times \text{CAG} - \text{Repeat} - \text{Länge (expandiertes Allel)}$$

#### **4.2 Fragmentlängen von ATXN3 in EUSage - Patientenproben**

Im Rahmen dieser Arbeit wurden Patientenproben aus zwei verschiedenen Kohorten analysiert. Dabei wurde in die statistische Analyse auch immer die individuelle Anzahl an CAG-Wiederholungen mit einbezogen. Für die Proben der EUROSCA-Kohorte wurden uns freundlicherweise die Ergebnisse aus vorangegangenen Arbeiten zur Verfügung gestellt. Die Proben der EuSage-Kohorte

wurden in Peru und Israel bereits bezüglich ihrer Anzahl an CAG-Wiederholungen analysiert. Wir führten eine stichprobenhafte Qualitätsprüfung durch. Diese erfolgte in der medizinischen Diagnostik in Tübingen. Dabei fiel eine deutliche Abweichung der Ergebnisse von >3 CAG-Triplets auf. Unter den vier Proben aus Israel wichen zwei mit > 4 Wiederholungen von der Analyse in Israel ab. Alle vier Proben aus Peru wurden in Tübingen mit + 3 Wiederholungen analysiert.

Aufgrund der deutlichen Abweichungen in der Qualitätskontrolle entschieden wir uns dafür eine erneute Analyse der CAG-Wiederholungen für alle Proben aus Israel und Peru mittels einer Fragmentlängenanalyse von *ATXN3* durchzuführen. Die Analyse der Proben erfolgte in zwei Durchgängen. Proben mit einem eindeutigen Signal im ersten Durchgang wurden nicht wiederholt. Abweichende Ergebnisse im ersten und zweiten Durchgang wurden gemittelt und mit dem Ergebnis des Kooperationspartners verglichen. Nach Durchführung der unter 3.7.4 beschriebenen Methodik zeigten sich zwei Peaks im Analysefenster. Unter Verwendung von der Diagnostik Tübingen definierter Analyseparameter wurde die Anzahl der CAG-Wiederholungen im gesunden Allel automatisch ermittelt. Die Anzahl der CAG-Wiederholungen im betroffenen Allel wurde anhand der Anzahl von Nukleotiden in eine Anzahl an Wiederholungen umgerechnet. Für die Umrechnung wurde uns die entsprechende Umrechnungstabelle der Diagnostik zur Verfügung gestellt.

Während der Fragmentanalyse ergab sich für einige Proben kein eindeutiges Muster. Dabei wurden Signale detektiert, die nicht eindeutig einem Allel zuzuordnen waren und ein untypisches Aussehen aufwiesen. Proben mit diesem Ergebnis wurden in einem zweiten Durchlauf erneut analysiert. Nach der Analyse aller CAG-Wiederholungen der Proben aus Israel und Peru wurden die Ergebnisse, mit denen der Kooperationspartner verglichen. Unter den 35 Proben aus Israel ergab sich für 15 Proben ein abweichendes Ergebnis von > 3 Wiederholungen. Dies entspricht einer Bestätigung von 57% der in Israel ermittelten Ergebnisse. Für die 16 Proben aus Peru zeigte sich eine deutlich geringe Übereinstimmung. Für 11 der 16 Proben ergab sich ein Ergebnis mit

einer Abweichung von > 3 Wiederholungen. Nur 32% der Ergebnisse aus Peru konnten mittels einer Fragmentlängenanalyse bestätigt werden.

### 4.3 Einzelnukleotid-Polymorphismen in *APOE* und ihr Einfluss auf das Erkrankungsalter der SCA3

#### 4.3.1 Die Einzelnukleotid – Polymorphismen

Im Rahmen der vorliegenden Arbeit wurden zwei Einzelnukleotid-Polymorphismen im *APOE* Gen analysiert. *APOE* kodiert für das Apolipoprotein E, ein multifunktionales Protein mit einer zentralen Rolle für den Metabolismus und Transport von Lipiden im Plasma und Liquor. Apolipoproteine welche ein Lipid gebunden haben können an verschiedene Rezeptoren der Zelloberfläche binden, dort tragen sie zur Stabilisierung der Zellmembran bei und sind an Reparaturmechanismen bei Zellschädigung beteiligt. Das Protein kommt in drei verschiedenen Isoformen vor, welche durch die zwei Einzelnukleotid-Polymorphismen in *APOE* und die drei resultierenden Allele  $\epsilon_2, \epsilon_3, \epsilon_4$  bedingt werden. Das  $\epsilon_4$  Allel ist dabei das am besten untersuchte der drei Varianten. Es konnte in der Vergangenheit mit der Alzheimer-Erkrankung assoziiert werden. Dabei korrelierten das Auftreten und die Anzahl an  $\epsilon_4$  Allelen mit dem Zeitpunkt der Erkrankung. Das  $\epsilon_2$  Allel gilt im Zusammenhang mit Alzheimer als protektiver Faktor (Zhao et al., 2018).

Tabelle 23: *Einzelnukleotid – Polymorphismen in APOE mit potentiellm Einfluss auf das Erkrankungsalter an SCA3*

Polymorphismus	Position	Variante	Aminosäure
rs 429358	chr19:44908684	T>C	Cys>Arg
rs 4712	chr19:44908822	C>T	Arg>Cys

Die Polymorphismen rs429358 auf Codon 112 und rs4712 auf Codon 158 liegen mit einem Abstand von nur 138 Basenpaaren relativ eng zusammen und ergeben zusammen einen Haplotypen. Beide Polymorphismen kommen als homozygote Varianten mit T/T oder C/C vor, sowie in der heterozygoten Variante T/C. Bei dem rs429358 ist das Vorkommen von T/T deutlich häufiger, die MAF für C liegt bei 15%. Beim rs7412 überwiegt das Vorkommen von C/C, während die MAF von T bei 8% liegt. In Folge der Einzelnukleotid-Polymorphismen kommt es zu

einer Missense Variante auf Aminosäureebene. Liegt im rs429358 ein C anstelle eines T vor codiert das Triplet für Arginin anstelle von Cystein. Bei Vorliegen eines T anstelle eines C im rs7412 codiert das Triplet wiederum für Cystein anstelle von Arginin.

Aufgrund ihrer eng beieinander liegenden Positionen konnten die beiden Polymorphismen im Rahmen einer Sanger-Sequenzierung zusammen analysiert werden. Zusätzlich erfolgte, bedingt durch das gekoppelte Vererbungsmuster, eine Analyse des Haplotypes in *APOE*. Der Haplotyp ergibt sich durch Allel<sup>rs429358</sup>/Allel<sup>rs7412</sup>.

#### 4.3.2 Primerauswahl

Die Auswahl der Primer zur Analyse der Einzelnukleotid-Polymorphismen in *APOE* erfolgte nach einer ausgiebigen Literaturrecherche in der „National Library of Medicine“ (NLM). Die Methode einer allelspezifischen PCR mit anschließendem High-Resolution-Melting nach Chen et al. (2016) erwies sich laut Recherche als besonders vielversprechend für die Typisierung großer Kohorten. Bei der Auswahl der Primer beriefen wir uns auf die Angaben im Paper. Zunächst überprüften wir die beschriebene Methodik auf ihre Reproduzierbarkeit und Validität. Im Verlauf passten wir die Versuchsbedingungen an unsere Laborbedingungen an.

Tabelle 24: Liste der Primer für *APOE* die auf ihre Effektivität und Funktionalität getestet wurden

Methoden	Primer	Primer – Sequenz
Allelspezifische – PCR	112 F-C	5'-gcgggcagggcggcGCGCGGACATGGAGGACGTGC-3'
	112 F-T	5'-gcgggcGCGCGGACATGGAGGACGTGT-3'
	112 R	5'-CGCCGCGGTACTGCACCAGG-3'
	158 F-C	5'-gcgggcagggcggcGCCGATGACCTGCAGAAGC-3'
	158 F-T	5'-gcgggcGCCGATGACCTGCAGAAGT-3'
	158 R	5'-CTCGCGGGCCCCGGCCTGGTA -3'
	Seq - F	5'- GACCATGAAGGAGTTGAAGGCCTA C-3'

### **4.3.3 Etablierung der Methodik**

Die Methodik der allelspezifischen-PCR beruht auf der Methode der Schmelzpunktanalyse nach Wang et al. (2005). Dabei werden für jeden SNP zwei vorwärts-Primer designt, die am 3'Ende komplementär an jeweils einen der SNP-Nukleotide binden. Zusätzlich wird ein gemeinsamer rückwärts-Primer für beide SNP entworfen. Der rückwärts-Primer wird dabei nicht mehr als 20 bp vom SNP entfernt platziert, um ein möglichst kleines PCR-Produkt zu erhalten. Um eine eindeutige Differenzierung der Produkte in der Schmelzkurvenanalyse zu ermöglichen, werden die vorwärts-Primer jeweils mit Anhängen von G und C unterschiedlicher Länge versehen. Um die Erfolgsrate und Genauigkeit der Ergebnisse zu verbessern, wird ein langer GC-Anhang von 14 bp am 5' Ende des Primers angehängt, der am 3' Ende die Base mit der höheren Schmelztemperatur aufweist. G oder C haben eine höhere Schmelztemperatur als A oder T. Entsprechend wird der kurze GC-Anhang an den Primer mit T am 3' Ende gehängt.

Die Primer für den rs429358 werden hier als 112-F-T/112-F-C und 112-R bezeichnet. Die Bezeichnung 112 entspricht dabei dem Codon, in welchem der SNP liegt, F steht für forward/vorwärts und das Ende T oder C entspricht der komplementären Base zum SNP. Die Systematik der Bezeichnung lässt sich auf rs7412 in Codon 158 übertragen.

#### Etablierung im Thermocycler

Die Etablierung der Methodik erfolgte schrittweise. Zunächst überprüften wir die Funktionalität der Primer im Thermocycler. Als PCR-Vorlage nutzen wir CEPH-DNA mit bekanntem Genotyp. Der PCR-Ansatz enthielt jeweils 3 Primer (2 vorwärts, 1 rückwärts), dNTPS (2mM), One-Taq-Puffer (5x) und H<sub>2</sub>O. Die Amplifikation erfolgte mittels One-Taq-Polymerase.

Um die Spezifität der PCR zu erhöhen, erfolgte die Amplifikation mittels Touchdown-PCR. Dabei wird die Annealing-Temperatur für die ersten Zyklen der PCR ein paar Grad höher eingestellt als die Schmelztemperatur der verwendeten Primer. Die höheren Temperaturen zu Beginn der PCR destabilisieren Primerdimere und erhöhen somit die Spezifität der PCR-Produkte. Um zu verhindern, dass die höhere Annealing-Temperatur durch eine schnellere

Ablösung des Primers von der DNA-Vorlage eine adäquate Amplifikation stört, wird pro Zyklus die Annealing-Temperatur um 1°C erniedrigt. Sobald die Annealing-Temperatur erreicht ist, wird diese für die restlichen Zyklen aufrechterhalten. Die PCR erfolgte für die ersten zehn Zyklen mit schrittweise abnehmender Annealingtemperatur von 67°C startend. Die letzten 25 Zyklen liefen mit einer Annealing-Temperatur von 57°C. Nach der Durchführung zeigten sich keine spezifischen Produkte für beide SNPs.

Im nächsten Schritt erfolgte die Etablierung von Zusätzen. Dabei wurde der oben beschriebene PCR-Ansatz verwendet und mit unterschiedlichen Zusätzen versehen. Es wurden die Zusätze 2% Formamid, 5% DMSO und 16% Betain ausprobiert. Dabei zeigten sich mit dem Zusatz von 16% Betain PCR-Produkte mit dem geringsten Hintergrundsignal. Vereinzelt entstanden Produkte von 190bp die am ehesten einer Kombination eines vorwärts- und rückwärts-Primers entsprachen. Um diesen Effekt zu unterbinden, führten wir im weiteren Etablierungsverlauf eine PCR mit Temperaturgradienten durch, um die optimale Annealing-Temperatur der Primer zu eruieren. Für den Temperaturgradienten wählten wir den Temperaturbereich von 57,9°C bis 67,2°C. Bei dem Prinzip des Temperaturgradienten werden einzelne Spalten des Thermocyclers nur auf eine bestimmte Temperatur erhitzt. Die Temperatur nimmt von links nach rechts in unterschiedlich großen Temperaturschritten zu. Dabei liegen im Bereich der geringsten und der höchsten Annealing-Temperatur die kleinsten Temperaturunterschiede. Im Bereich von 60-64°C beträgt der Temperaturunterschied pro Spalte ca. 1°C. Der Temperaturgradient ergab eine Temperatur von >64°C als positiv modifizierenden Faktor für die Spezifität der Reaktion. Eine Amplifikation fand in allen Fällen statt, bei zunehmender Annealing-Temperatur nahm allerdings das Hintergrundsignal ab.

Unter den etablierten Bedingungen nach den oben beschriebenen Versuchsabläufen, zeigte sich die Funktionalität der Primer. Da die Methodik für eine Typisierung von SNPs mittels HRM entworfen wurde, erfolgte im nächsten Schritt eine Etablierung im Light Cycler.

### Etablierung im Light – Cycler

Die Etablierung der Reaktion erfolgte im Light – Cycler Roche 480. Dabei wurden die Chemikalien aus dem High Resolution Melting Master Kit verwendet und die Konzentrationen der eingesetzten Primer nach den Herstellerangaben auf 0,2  $\mu\text{M}$  reduziert. Als PCR-Vorlage verwendeten wir ebenfalls CEPH-DNA mit bekanntem Genotyp.

Um die Spezifität der Reaktion zu erhöhen, erfolgte eine Titration von  $\text{MgCl}_2$  zu den Reaktionsansätzen. Dabei wurde die Konzentration von  $\text{MgCl}_2$  zwischen 2 mM und 3mM auf ihren Einfluss überprüft. Die Titration erfolgte in Schritten von 0,5 mM. Eine Konzentration von < 2,5 mM erwies sich dabei als unzureichend, ab 2,5 mM  $\text{MgCl}_2$  konnte einer Steigerung der Produktrate festgestellt werden (Abbildung 9).

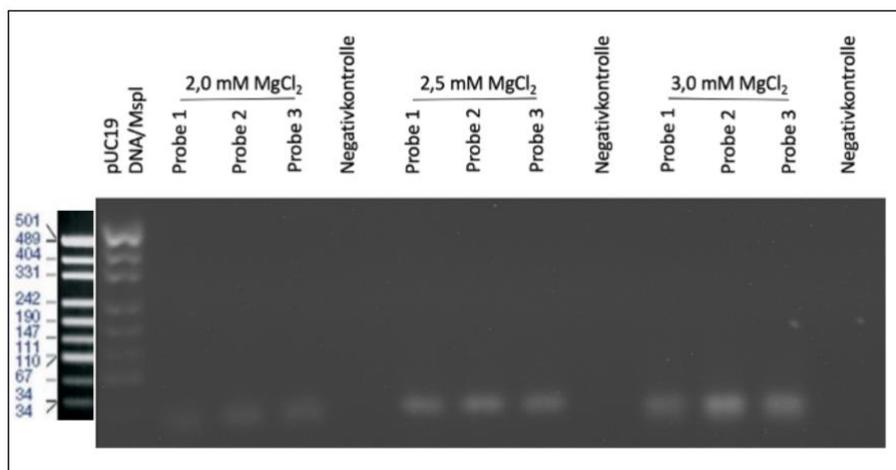


Abbildung 7: **Gel – Elektrophorese nach  $\text{MgCl}_2$  Titration.** Es wurden jeweils drei CEPH DNA-Proben mit bekanntem Genotyp unter mit dem Zusatz verschiedenen  $\text{MgCl}_2$  -Konzentration getestet. Hier wurden die Primer 112-F-C, 112-F-T und 112-R verwendet. Von links nach rechts sind die PCR Produkte unter steigender  $\text{MgCl}_2$  Konzentration sichtbar. Mit zunehmender Konzentration an  $\text{MgCl}_2$  steigt die Signalstärke der Produkte. Die Unterschiede in der Länge an bp durch die Amplifikation durch die Bindung der unterschiedlich langen Primer fällt so gering aus, dass er in der Gel-Elektrophorese nicht eindeutig erkennbar ist.

Mittels High-Resolution-Melting im Light-Cycler 480 konnten anschließend drei verschiedene Kurvenverläufe detektiert werden. Bei zuvor bekanntem Genotyp der verwendeten CEPH-DNA konnten die spezifischen Kurvenverläufe jeweils einem Polymorphismus zugeordnet werden. Nach der beschriebenen Etablierung im Light-Cycler erfolgte die Typisierung von rs429358 unter den beschriebenen Bedingungen (Abbildung 10). Für eine eindeutige Zuordnung der

spezifischen Schmelzkurven zu den verschiedenen Genotypen erfolgte eine Sanger-Sequenzierung exemplarischer Proben.

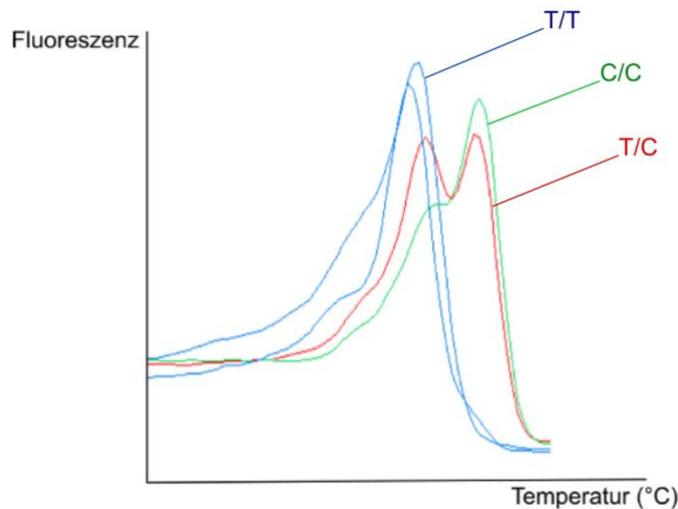


Abbildung 8: **Die verschiedenen Schmelzkurven der Genotypen von rs429358 in APOE.** Die blaue Kurve zeigt den Schmelzkurvenverlauf einer Probe mit dem homozygoten Genotyp T/T, die grüne Kurve den Schmelzkurvenverlauf einer homozygoten Probe C/C und die rote Kurve mit zwei Schmelzmaxima den Kurvenverlauf einer heterozygoten Probe T/C.

#### Optimierung der Methodik zur Analyse von rs7412 in APOE

Für den Polymorphismus rs7412 stellte sich die Differenzierung zwischen homozygoten Proben mit dem Genotyp T/T und heterozygoten Proben mit dem Genotyp T/C zunächst als schwierig dar. Der Schmelzkurvenverlauf heterozygoter Proben war lediglich durch eine leichte „Schulter“ von dem Verlauf homozygoter Proben zu unterscheiden und erschwerte somit eine zuverlässige Typisierung anhand des Schmelzverhaltens (Abbildung 11A). Durch eine Anpassung der Primerkonzentrationen von 1:2 der Primer für 158-F-T:158-F-C zeigten heterozygote Proben im Schmelzkurvenverlauf ein deutliches zweites Schmelzmaximum (Abbildung 11B).

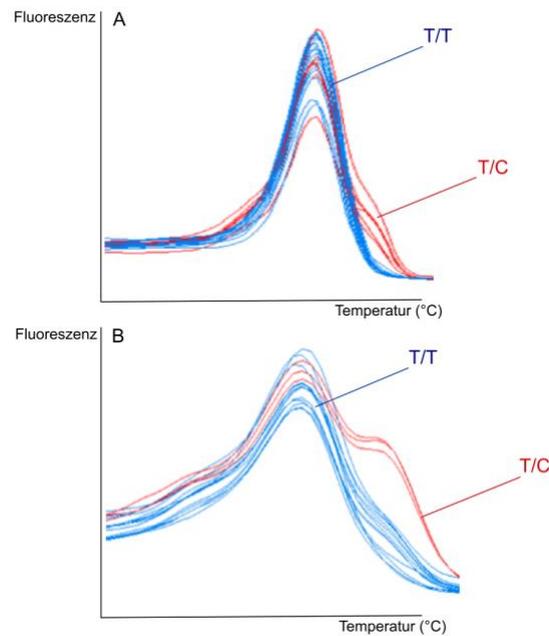


Abbildung 9: **Schmelzkurvenverläufe nach der Reaktionsoptimierung.** A) Schmelzkurvenverlauf für rs7412 in APOE von homozygoten Proben T/T (blau) und heterozygoten Proben (rot) vor der Anpassung der Primerkonzentrationen. B) Schmelzkurvenverlauf für rs7412 in APOE nach der Anpassung der Primerkonzentrationen in einem Verhältnis von 158-F-T:158-F-C = 1:2.

#### 4.3.4 Verteilung der ermittelten Genotypen in rs429358 und rs7412

346 Patienten konnten erfolgreich für rs429358 und rs7412 genotypisiert werden. Dabei wurde zwischen Proben der EUROSCA Kohorte und der EuSAge Kohorte (Israel und Peru) unterschieden. Aus der EUROSCA Kohorte wurden 295 Proben typisiert. Aus der EUSAGE Kohorte wurden insgesamt 51 Proben typisiert, 35 aus Israel und 16 aus Peru.

Tabelle 25: **Verteilung der Genotypen von rs429358 und rs7412 in APOE nach der Typisierung**

Polymorphismus	rs 429358		rs 7412	
	EUROSCA	EuSAge	EUROSCA	EuSAge
T/T	235	44	7	0
C/T	56	6	42	0
C/C	4	1	246	51
Gesamt	295	51	295	51

### 4.3.5 Verteilung nach dem Hardy-Weinberg-Gleichgewicht

Mittels der Hardy-Weinberg-Gleichung konnte ermittelt werden, welche Verteilung der Genotypen unter idealen Bedingungen zu erwarten gewesen wären. Dafür wurde in der Hardy-Weinberg-Gleichung  $p^2 + 2pq + q^2 = 1$  als Wert für die Variable  $q$  die Allelfrequenz des Genotyps eingesetzt, welcher jeweils seltener vorkommt. Der Wert dieser sogenannten Minor-Allelfrequenz (MAF) wurde der Datenbank „ALFA allele frequency“ der National Library of Medicine (NLM) entnommen. Da die Allelfrequenzen in unterschiedlichen Populationen nicht unbedingt gleich verteilt sind, wird die, der zu analysierenden Population zugehörige, Allelfrequenz eingesetzt. In diesem Fall wurden die Frequenzen für eine europäische (European Sub für EUROSCA) und eine südamerikanische Subgruppe (Latin American 1 Sub für EuSAge) verwendet.

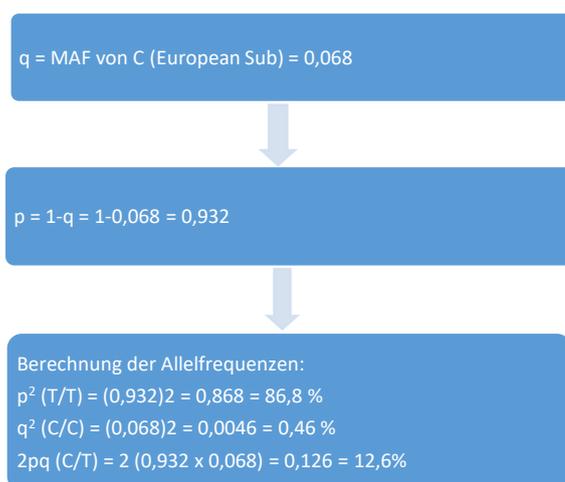


Abbildung 10: **Beispiel zur Berechnung der Allelfrequenz** von rs429358 in APOE in einer idealen europäischen Kohorte nach dem Hardy-Weinberg-Gleichgewicht  $p^2 + 2pq + q^2 = 1$ .

Tabelle 26: **erwartete Verteilung der Genotypen von rs429358 und rs7412 in APOE** in einer idealen Kohorte laut Hardy-Weinberg

Polymorphismus	rs 429358		rs 7412	
	EUROSCA	EuSAge	EUROSCA	EuSAge
T/T	255 (86,6%)	41 (79%)	2 (0,6%)	0 (0,5%)
C/T	38 (12,6%)	10 (19,5%)	44 (15%)	7 (13,2%)
C/C	2 (0,46%)	0 (1,2%)	249 (84,4%)	44 (86,3%)
Gesamt	295	51	295	51

#### 4.3.6 **Signifikante Abweichungen der Verteilung von Genotypen in SCA3 Proben**

Im Vergleich der Verteilung der Genotypen in den von uns typisierten Patientenproben und der erwarteten Verteilung nach Hardy-Weinberg wurden beide Kohorten zusammen betrachtet. Dabei blieben die relativen Häufigkeiten der einzelnen Genotypen in ihrem Verhältnis zueinander unverändert. Für rs429358 wurde, entsprechend der in der Normalbevölkerung beobachteten Verteilung, T/T (n= 279) > C/T (n= 62) > C/C (n=5) typisiert. Es zeigte sich jedoch um den Faktor 1,2 häufiger das Vorkommen eines heterozygoten Genotyps T/C in SCA3 Patienten Proben. Ebenso konnte widererwartend fünfmal der Genotyp C/C typisiert werden, dies entspricht einem um den Faktor 2,5 erhöhten Vorkommen in der Patientenkohorte. Für den Polymorphismus rs7412 wurde C/C (n=297) > C/T (n=42) > T/T (n=7) typisiert. Hier zeigte sich im Vergleich mit der erwarteten Verteilung nach Hardy-Weinberg, um den Faktor 3,5 häufiger der Genotyp T/T in Patienten-Proben. Mithilfe eines Chi-Quadrat-Tests wurde überprüft, ob die Verteilung der Genotypen in der Gruppe des SCA3 Patienten signifikant von der erwarteten Verteilung nach Hardy-Weinberg abweicht. Dabei zeigte sich für beide Polymorphismen ein p-Wert <0,05. Daher konnten wir einen signifikanten Unterschied im Vorkommen der Genotypen innerhalb der von uns typisierten Kohorte an SCA3-Patienten im Vergleich mit der generellen Verteilung in der Normalbevölkerung annehmen (vgl. Tabelle 27).

*Tabelle 27: Vergleich der typisierten Genotypen in Proben von SCA3 Betroffenen und einer Verteilung nach dem Hardy- Weinberg – Gesetz*

Polymorphismus	rs 429358		rs 7412	
	SCA3- Proben	Hardy-Weinberg	SCA3- Proben	Hardy-Weinberg
T/T	279	296	7	2
C/T	62	48	42	51
C/C	5	2	297	293

#### **4.3.7 Regressionsanalyse der Polymorphismen und ihr Einfluss auf das Erkrankungsalter**

Nach der Analyse von rs429358 und rs7412 in den Patientenproben und der Feststellung des vorliegenden Genotyps wurde eine Regressionsanalyse zur Überprüfung des Einflusses des jeweiligen Einzelnukleotid-Polymorphismus auf das Erkrankungsalter durchgeführt. Das Erkrankungsalter der Patienten wurde dabei als stetige Y-Variable und die individuelle Anzahl der CAG-Wiederholungen im expandierten Allel als stetige X-Variable eingesetzt. Abhängig vom ermittelten Genotyp für rs429358 und rs7412 als nominale Variable, wurden unter Berücksichtigung eines „Familienfaktors“ drei Regressionsgeraden dargestellt.

Der „Familienfaktor“ wurde anhand der Anzahl typisierter Mitglieder einer Familie berechnet. Drei typisierter Patientenproben einer Familie erhielten dabei jeweils den „Familienfaktor“  $\frac{1}{3}$ , vier typisierte Proben den Faktor  $\frac{1}{4}$  usw. In der anschließenden statistischen Analyse wurden dadurch alle typisierten Mitglieder einer Familie als eine Patientenprobe in die Analyse einbezogen. Dieses Vorgehen berücksichtigt die Wahrscheinlichkeit der Vererbung eines Genotyps innerhalb einer Familie und verhindert somit die Verfälschung einer statistischen Analyse durch die Typisierung von verwandten Patienten

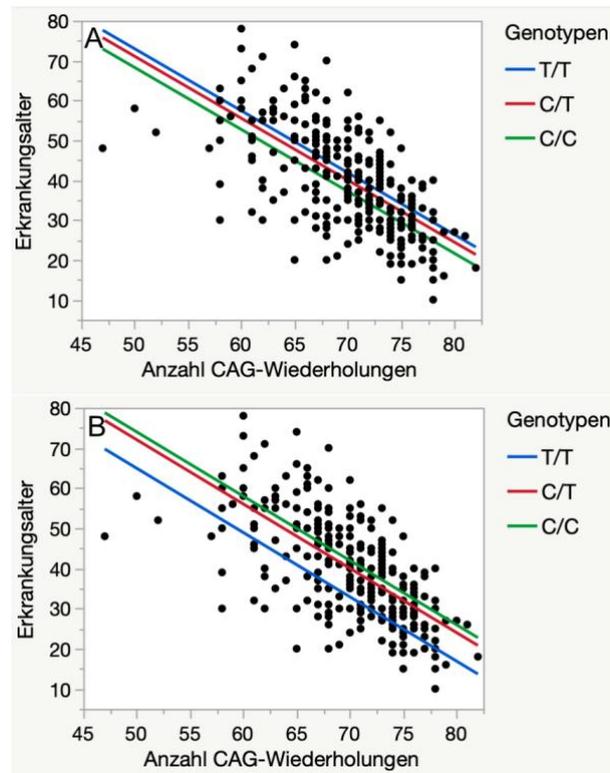


Abbildung 11: **Regressionsanalyse der Einzelnukleotid-Polymorphismen und ihr Einfluss auf das Erkrankungsalter.** Es wurde jeweils eine Regressionsanalyse getrennt für beide untersuchten Polymorphismen durchgeführt. Für rs429358 zeigt sich in der Regressionsanalyse in Abbildung A nach Anpassung der Geraden für den Genotyp C/C eine Verminderung der Erkrankungsalters. In der Regressionsanalyse des Polymorphismus rs7412 zeigt sich in Abbildung B eine deutliche Verminderung des Erkrankungsalters bei Vorliegen eines T/T Genotyps.

In den Abbildungen ist ein Einfluss der verschiedenen Genotypen von rs429358 und rs7412 auf das Erkrankungsalter zu erkennen. Für rs429358 stellt sich der Genotyp C/C als größter Einfluss auf das Erkrankungsalter dar und für rs7412 der Genotyp T/T. Beide Genotypen sind für den einzelnen Polymorphismus der jeweils seltenste Genotyp. Der heterozygote Genotyp T/C zeigt in beiden Polymorphismen ebenfalls einen leicht verringerten Effekt (-1,4 / -1,9 Jahre) auf das Erkrankungsalter. In absoluten Zahlen in Jahren verringert sich das Erkrankungsalter bei Vorliegen des Genotyp rs7412 T/T am stärksten um 6,2 Jahre (vgl. Tabelle 28). Bei einer Anzahl von 7 Patientenproben mit Genotyp rs7412 T/T und einem Standardfehler von 4,4 ist dieses Ergebnis allerdings kritisch zu bewerten.

*Tabelle 28: Verminderung des Erkrankungsalters der SCA3 in Jahren durch verschiedene Genotypen in rs429358 und rs7412 bezogen auf das Erkrankungsalter bei Vorliegen des jeweils am häufigsten vorkommenden Genotyps (T/T für rs429358 und C/C für rs7412).*

Polymorphismus	Genotyp	Erkrankungsalter (in Jahren)
rs429358	T/T	40,9
	T/C	- 1,4
	C/C	- 3,2
rs7412	C/C	41,01
	T/C	- 1,9
	T/T	- 6,2

#### **4.3.8 Die Verteilung der Haplotypen von APOE in SCA3 - Patienten**

Nach der Feststellung eines Effektes der einzelnen Polymorphismen stellte sich eine Überprüfung des Haplotypes als Einflussfaktor auf das Erkrankungsalter an SCA3 als sinnvoll dar. Der Haplotyp von rs429358 und rs7412 ergibt vier mögliche Allele  $\epsilon_1$ – $\epsilon_4$ , die entweder in homozygoter oder heterozygoter Form vorliegen können. Dabei wird das  $\epsilon_1$  Allele als „missing Allel“ bezeichnet und kommt in der Regel nicht vor.

In einer gesunden Normalbevölkerung gibt es eine bestimmte Verteilung an Haplotypen, dadurch gibt es Allele, die deutlich häufiger vorkommen als andere. Weltweit geht man von einer Verteilung von ungefähr 60 - 90% homozygoten und heterozygoten  $\epsilon_3$ -Allel-Trägern, zwischen 10 und 20%  $\epsilon_4$ -Allel-Trägern und 0-20%  $\epsilon_2$  Allel-Trägern aus (Singh et al, 2006). Das Vorkommen der Allele variiert stark und erlaubt daher nur eine Aussage über eine grobe Häufigkeitsverteilung in einer europäischen oder südamerikanischen Vergleichskohorte.

*Tabelle 29: Liste der möglichen Haplotypen von rs429358 und rs7412 und die resultierenden Allele von APOE und ihre relativen Häufigkeiten in einer normalverteilten Population nach Singh et al., 2007.*

rs429358	rs7412	Allel	Häufigkeit
C	T	$\epsilon_1$	0%
T	T	$\epsilon_2$	0-20%
T	C	$\epsilon_3$	60-90%
C	C	$\epsilon_4$	10-20%

In der Gruppe der analysierten SCA3 Patienten sind 96%  $\epsilon_3$ -Träger (n=332), 19,4 %  $\epsilon_4$ -Träger (n=67) und 14,6 %  $\epsilon_2$ -Träger (n=51). Die Bestimmung der Allele erfolgte nach dem Prinzip Nukleotid<sup>rs429358</sup>/Nukleotid<sup>rs7412</sup>. Unter Berücksichtigung der ermittelten Genotypen wurde der jeweilige Haplotyp der Patienten-Probe in

der Ergebnistabelle aufgeführt. Die Analyse der Verteilung erfolgte mittels statistischer Auszählung bezogen auf die Anzahl aller typisierter Proben. Die Ergebnisse entsprechen einer Verteilung vergleichbar mit den weltweiten Häufigkeiten der jeweiligen Allele (vgl. Tabelle 29).

#### 4.3.9 Der Einfluss von APOE Haplotypen auf das Erkrankungsalter

Die Analyse des Einflusses der Haplotypen fand mittels einer Regressionsanalyse vom Erkrankungsalter der Patienten und der individuellen Anzahl an CAG-Wiederholungen unter Einbezug der verschiedenen Allele statt (Abbildung 13). Die Haplotypen wurden dabei weiter differenziert in homo- und heterozygote Allel-Varianten. Das APOE Gen befindet sich im menschlichen Genom jeweils auf zwei Chromosomen, einem maternalen und einem paternalen. Dabei kann sich der Haplotyp auf dem paternalen Chromosom, von dem auf dem maternalen Chromosom unterscheiden. Durch die Ermittlung der Genotypen bei der Typisierung der Patientenproben, ließ sich im Verlauf bestimmen, ob das jeweilige Allel homo- oder heterozygot vorliegt. Durch den individuellen „Familienfaktor“ wurden alle typisierten Proben einer Familie als eine Patienten-Probe betrachtet.

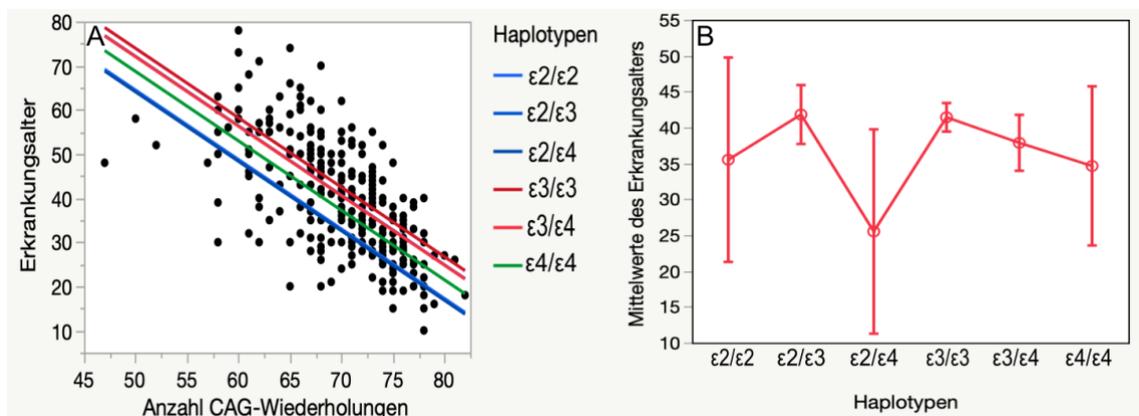


Abbildung 12: **Regressionsanalyse und Darstellung des Effekteinflusses der APOE Haplotypen auf das Erkrankungsalter.** A: Bei Betrachtung der Regressionsanalyse wird ein vermindernder Effekt von  $\epsilon 2$  auf das Erkrankungsalter deutlich. Dabei scheint es keinen Unterschied zu machen, ob das  $\epsilon 2$  Allel heterozygot oder homozygot vorliegt. B: Bei der Darstellung der Mittelwerte wird eine deutliche Abweichung bei dem Vorliegen eines  $\epsilon 2/\epsilon 4$  Haplotyps sichtbar. In einem durchgeführten t-Test erweist sich die Abweichung der Mittelwerte zwischen den Haplotypen  $\epsilon 2/\epsilon 3, \epsilon 2/\epsilon 4$  und  $\epsilon 3/\epsilon 3$  als signifikant relevant.

Durch die Regressionsanalyse wurde ein vermindernder Einfluss durch das Vorliegen von mindestens einem  $\epsilon 2$ -Allel auf das Erkrankungsalter deutlich (Abbildung 14). Bei einer durchschnittlichen Anzahl von 71 CAG-Wiederholungen in SCA3-Patienten lag das durch die Regressionsgerade ermittelte Erkrankungsalter für  $\epsilon 2$ -Träger mit ungefähr 30 Jahren ca. 10 Jahre unter dem von homozygoten  $\epsilon 3$ -Trägern. Nach der Darstellung des Effekteinflusses der verschiedenen Haplotypen auf das Erkrankungsalter führten wir einen t-Test der Abweichungen der Mittelwerte des Erkrankungsalters bezogen auf die jeweiligen Haplotypen. Dabei zeigte sich ein signifikanter Unterschied zwischen  $\epsilon 2/\epsilon 3$  und  $\epsilon 2/\epsilon 4$ , sowie zwischen  $\epsilon 3/\epsilon 3$  und  $\epsilon 2/\epsilon 4$  (Tabelle 30).

*Tabelle 30: T-Test der Abweichung der Mittelwerte des Erkrankungsalters in Abhängigkeit der verschiedenen Haplotypen. Haplotypen die nicht über den gleichen Buchstaben verbunden sind, unterscheiden sich signifikant. Dies trifft auf die Haplotypen  $\epsilon 2/\epsilon 4$  und  $\epsilon 2/\epsilon 3$ , sowie  $\epsilon 2/\epsilon 4$  und  $\epsilon 3/\epsilon 3$  zu.*

Haplotyp		Mittelwert des Erkrankungsalter (in Jahren)
e2/e3	<b>A</b>	41,8
e3/e3	<b>A</b>	41,4
e3/e4	A B	37,8
e2/e2	A B	35,5
e4/e4	A B	34,6
e2/e4	<b>B</b>	25,5

Aufgrund dieser Beobachtung überprüften wir die Signifikanz des modifizierenden Einflusses der APOE Haplotypen auf das Erkrankungsalter. Dabei zeigte sich, dass die individuellen Haplotypen in kombinierter Betrachtung mit der individuellen Anzahl an CAG – Repeats mit einem p – Wert von  $< 0,05$  einen signifikanten Einfluss auf das Erkrankungsalter der SCA3 haben (Tabelle 31).

*Tabelle 31: Statistische Analyse des Einflusses der APOE Haplotypen auf das Erkrankungsalter. Die Tabelle fasst die p-Werte der einzelnen Parameter zusammen und zeigt, dass der APOE Haplotyp nur in Verbindung mit der individuellen CAG-Repeat-Länge einen signifikanten Einfluss auf das Erkrankungsalter hat.*

Parameter	p-Wert
Anzahl CAG-Wiederholungen	0,001*
Anzahl CAG-Wiederholungen + APOE Haplotyp	0,03*

#### 4.3.10 Der Einfluss von APOE $\epsilon 2$ auf das Erkrankungsalter

Da im Laufe der bisherigen Auswertung ein möglicher modifizierender Einfluss des  $\epsilon 2$  Allels auf das Erkrankungsalter beobachtet wurde, führten wir in einem weiteren Schritt eine genauere statistische Analyse der  $\epsilon 2$  Träger durch. Dafür wurden alle typisierten Patienten in Nicht – Träger (0) und Träger (1) unterteilt. Dabei wurde nicht zwischen homozygoten oder heterozygoten Merkmalsträgern unterschieden. Durch die getrennte Betrachtung von Merkmalsträgern und nicht-Merkmalsträgern konnte ein verringernder Einfluss auf das Erkrankungsalter durch das Vorliegen eines APOE  $\epsilon 2$  Allels von 2 Jahren nachgewiesen werden (Abbildung 15). Bei einem p-Wert von 0,09 kann dieser Einfluss zwar nicht als statistisch signifikant gewertet werden, allerdings gibt der Wert eine Tendenz des Einflusses von  $\epsilon 2$  an. Es wurden 44 Patienten-Proben als  $\epsilon 2$ -Träger identifiziert, dadurch fällt die Stichprobe relativ gering aus. Zudem zeigt die Gruppe der  $\epsilon 2$ -Träger eine große Varianz des Erkrankungsalters. Dabei liegt das früheste Erkrankungsalter der  $\epsilon 2$ -Träger mit 35,5 Jahren deutlich unter dem frühesten Erkrankungsalter anderer Allel-Träger.

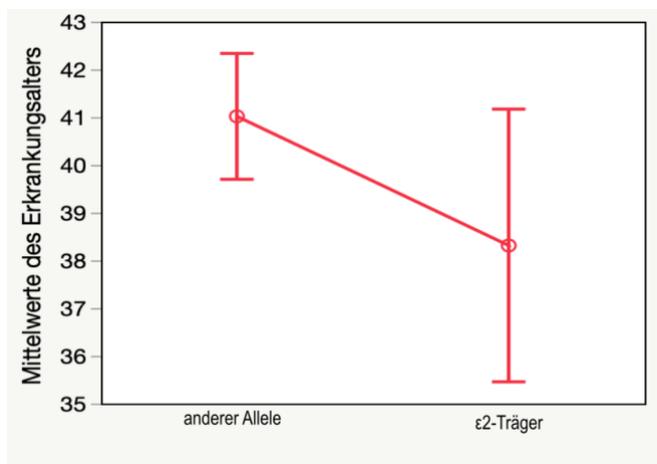


Abbildung 15: Verteilung des Erkrankungsalters bei  $\epsilon 2$  Allel-Trägern und Nicht-Trägern. 44 Proben wurden als Träger von  $\epsilon 2$  typisiert. Das mittlere Erkrankungsalter bei  $\epsilon 2$  Trägern liegt mit 38,3 Jahren 2 Jahre unter dem von Trägern anderer Allele. In einem anschließendem t-Test zur Überprüfung der Signifikanz der Abweichungen, konnte kein signifikanter Unterschied nachgewiesen werden. Somit musste die Alternativhypothese „APOE  $\epsilon 2$  Allel Träger erkranken signifikant früher an SCA3“ verworfen werden.

Für die Berechnung des individuellen Erkrankungsalters einer betroffenen Person unter der Berücksichtigung des Vorliegens eines APOE ε2 Allels ergibt sich folgende Voraussagefunktion:

*Formel 2: Berechnung des Erkrankungsalters der SCA3 unter Berücksichtigung von APOE ε2. Unter Einbezug des Merkmals „APOE ε2 Träger“ in die Berechnung des individuellen Erkrankungsalters ergibt sich folgende annäherungsweise Prognoseformel:*

$$151,7985 + e2 - \text{Träger/anderes Allel} + (-1,5886 \times \text{Anzahl CAG - Wiederholungen})$$

$$\left( \begin{array}{l} e2: -1,3525 \\ \text{Anderes Allel: } 1,352 \end{array} \right)$$

#### 4.3.11 Prädiktive Faktoren für das Erkrankungsalter

Um den einzelnen prädiktiven Wert der Einzelnukleotidpolymorphismen für das Erkrankungsalter zu überprüfen, wurde mittels statistischer Analyse der Vorhersagewert (r) der Anzahl an CAG-Wiederholungen ermittelt. Bei den von uns typisierten 346 Proben lag der Voraussagewert für das Erkrankungsalter durch die individuelle Anzahl an CAG-Wiederholungen bei 40%. Durch die Hinzunahme des Einzelnukleotid-Polymorphismuses rs429358 verbesserte sich der Voraussagewert mit einer signifikanten Relevanz (0,024\*) um 2%. Der Polymorphismus rs7412 ergab in Kombination mit der individuellen Anzahl an CAG-Wiederholungen keine signifikante Verbesserung des prädiktiven Wertes. Bei der Ermittlung des r-Wertes unter Betrachtung der individuellen Anzahl an CAG-Wiederholungen + des APOE Haplotypes (rs429358 und rs7412) ergab sich eine Verbesserung um 5%.

*Tabelle 32: Einzelnukleotid-Polymorphismen in APOE als prädiktive Faktoren für das Erkrankungsalter der SCA3. Zunächst wurde der prädiktive Wert durch die alleinige Betrachtung der individuellen CAG-Repeat-Länge bestimmt. Dieser lag in der untersuchten Kohorte bei 40%. Der relative niedrige prädiktive Wert spricht für eine sehr heterogene Kohorte mit großen Varianzen und starken Ausreißern. Durch die Hinzunahme der einzelnen Polymorphismen konnte der prädiktive Wert durch den rs429358 um 2% signifikant verbessert werden. Die Kombination von rs429358 und rs7412 ergab sogar eine Verbesserung des Wertes um 5%.*

Parameter	r <sup>2</sup>	Δr <sup>2</sup>	p – Wert
CAG – Repeat - Länge	0,40		< 0,0001*
+ rs429358	0,42	0,02	0,024*
+ rs7412	0,42	0,02	0,14
+ rs429358 and rs7412	0,45	0,05	0,03*

#### 4.4 Einzelnukleotid-Polymorphismen in CAST und ihr Einfluss auf das Erkrankungsalter der SCA3

Im Rahmen einer Literaturrecherche in der *National Library of Medicine* (NLM) ergaben sich zwei weitere Polymorphismen im Calpastatin (CAST) Gen als potenzielle Modifikatoren des Erkrankungsalters der SCA3. Calpastatin ist der natürliche Gegenspieler von Calpain, einer calcium-abhängigen Protease, die bereits als wichtiger Einflussfaktor mit vielen neurodegenerativen Erkrankungen wie Parkinson, Alzheimers oder Multiple Sklerose in Zusammenhang gebracht wurde (Allen et al., 2010). Shukla *et al.* (2006) konnten zeigen, dass die Überexpression von Calpain einen Einfluss auf neuronale Abbauprozesse haben könnte. Die Hemmung von Calpain in Parkinson-Tiermodellen führte wiederum nachweislich zu einer Prävention von neurologischen Defiziten (Crocker et al., 2003). Diese Erkenntnisse qualifizieren genetische Varianten von CAST als mögliche Einflussfaktoren auf neurodegenerative Erkrankungen im Allgemeinen, und im Rahmen dieser Arbeit, SCA3 im Speziellen.

Allen et al. (2010) konnten zeigen, der Polymorphismus rs1559085 signifikant häufig im Zusammenhang mit einer Parkinson-Erkrankung vorliegt. 32 663 Basenpaare entfernt liegt rs27852. Dieser Polymorphismus wurde in einem weiteren Datensatz an Parkinson-Proben ebenfalls mit der Erkrankung in Verbindung gebracht. Beide Polymorphismen rs27852 und rs1559085 sind Intron Varianten auf Chromosom 5 und haben somit keinen direkten Effekt auf Aminosäureebene. Allerdings ist ein funktioneller Einfluss der Varianten nicht auszuschließen (vgl. Tabelle 33).

*Tabelle 33: Einzelnukleotid – Polymorphismen in CAST mit potenziellem Einfluss auf das Erkrankungsalter an SCA3. Die beiden Polymorphismen liegen über 32 Tsd. Basenpaare voneinander entfernt auf Chromosom 5. Als Intron-Varianten haben sie keinen direkten Einfluss auf Aminosäuren-Ebene, ein funktioneller Einfluss ist allerdings nicht auszuschließen.*

Polymorphismus	Position	Variante	Aminosäure
rs 27852	chr5:96710335	A>C/G/T	Intron-Variante
rs 1559085	chr5:96742998	A>G	Intron-Variante

#### 4.4.1 Primerauswahl

Für die Typisierung der Einzelnukleotid-Polymorphismen in *CAST* wurden sieben verschiedene Primer mithilfe eines Primer-Sign-Tools ([eurofinngenomics.eu](http://eurofinngenomics.eu)) entworfen. Bei dem Designen der Primer wurde darauf geachtet, dass die Tendenz zur Dimerbildung möglichst gering ausfällt. Die Primer flankieren jeweils den Bereich des Genes, in dem der entsprechende Polymorphismus vorliegt. Für den Polymorphismus rs27852 ergaben sich zwei mögliche Primerpaare, während sich für rs1559085 nur ein sinnvoller rückwärts-Primer designen ließ. In Tabelle 34 sind die entsprechenden Primer aufgelistet. Rückwärts-Primer (Rx) sind in 5'-3'-Richtung aufgelistet, binden während der PCR aber in 3'-5'-Richtung am komplementären Strang.

Tabelle 34: Liste der Primer für *CAST* die auf ihre Effektivität und Funktionalität überprüft wurde. Primer mit der Bezeichnung „F“ sind vorwärts-Primer, die Bezeichnung „R“ markiert rückwärts-Primer.

Primer	Sequenz
rs27852-F1	5' TGCACCAGGAATAGAGCACT 3'
rs27852-R1	5' CTGTAAGCCTCTCACGTAATTCT 3'
rs27852-F2	5' GCACTAATGTACAGCAATAACGG 3'
rs27852-R2	5' AAGAGGGAAGGAGGTCTGTG 3'
rs1559085-F1	5' GCATTGCCTCCTTGTTACCTT 3'
rs1559085-F2	5' CAGAAGCACCCAGAACCCTT 3'
rs1559085-R1	5' GGGATAGTTACATGACTGACTT 3'

#### 4.4.2 Etablierung der Methodik

##### Primeretablierung im Thermocycler

Die Etablierung und Überprüfung der Primerpaare auf ihre Funktionalität und Effektivität erfolgte schrittweise. Zunächst erfolgte eine Touchdown-PCR im Thermocycler. Dabei wurden die Primer in 10pmol/μl Konzentration in den Kombinationen F1/R1, F1/R2, F2/R1 und F2/R2 für rs27852, sowie F1/R1 und F2/R1 für rs1559085 eingesetzt. Als Polymerase wurde One-Taq (NEB) verwendet. Die PCR durchlief 35 Zyklen. Davon 10 Zyklen mit abnehmender Annealing-Temperatur, bei 65°C beginnend. Pro Zyklus nahm die Annealing-Temperatur um 1°C ab. Anschließend erfolgte die PCR für 25 Zyklen bei einer Annealing-Temperatur von 55°C.

Die Spezifität und Effektivität der entworfenen Primer konnten unter den beschriebenen Bedingungen erfüllt werden. Einzig für das Primerpaar rs27852 F2/R1 ergab sich kein spezifisches Amplikon in ausreichender Menge. Alle Primerkombinationen zeigten eine Tendenz zur Primerbildung. (vgl. Abbildung 17).

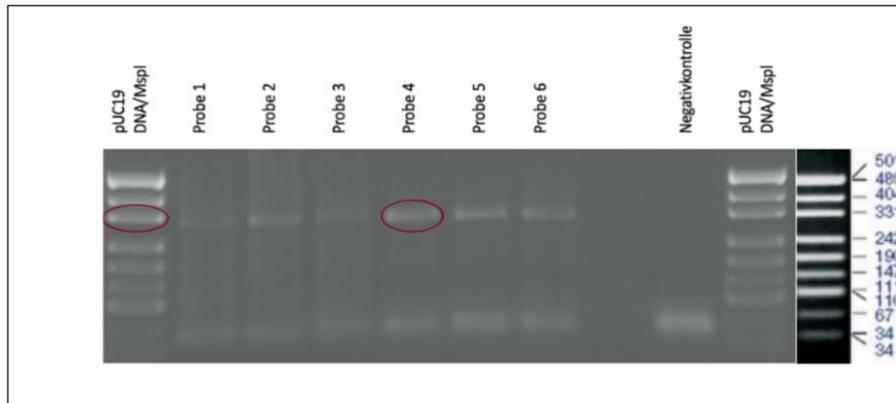


Abbildung 16: **Gel – Elektrophorese mit 1,5 % Agarose-Gel nach einer PCR mit verschiedenen Primerkonstellationen** zur Amplifikation von CAST-Genabschnitten. Zum Vergleich wurde rechts im Bild eine pUC19 DNA/MspI Vergleichsbande mit Größenangabe (in bp) eingefügt. Das Primerpaar rs27852 F2/R1 ergab kein spezifisches Produkt. Alle Primerkombinationen zeigten eine Tendenz zur Primerbildung. Die Kombinationen rs27852-F1/R1 und rs1559085-F1/R1 ergeben das jeweils kleinste spezifische Produkt (vgl. rote Markierung) und eignen sich am ehesten zur anschließenden Analyse im Light – Cycler.

### Etablierungsversuch im Light – Cycler

Im nächsten Schritt erfolgte die Etablierung im Light-Cycler für eine Analyse mittels High-Resolution Melting. Hierfür wurden die Primerpaare ausgewählt, die nach der PCR im Thermocycler das kleinste spezifische Produkt ergaben. Dies entspricht den Kombinationen F1/R1 für rs27852 und F1/R1 für rs1559085 (vgl. Abbildung 17). Die Etablierung erfolgte für beide Polymorphismen getrennt voneinander. Als DNA-Template wurde CEPH-DNA eingesetzt. Für den PCR-Mix wurde der Melting-Master-Mix (Roche) und 25mM MgCl<sub>2</sub> verwendet. Die Amplifikation erfolgte nach dem Programm „Gene Scanning“ (Light Cycler 480) mit leichten Modifikationen. Die Amplifikation erfolgte bei 72°C für 16 Sekunden. Die Berechnung und Anpassung der Zeit erfolgte nach der Schätzungsformel:  $\frac{150 \text{ pb}}{25} + 10 \text{ Sekunden}$ . Die Schmelzkurve wurde mit einer Steigerungsrate von 1°C/Sekunde beginnend bei 65°C aufgetragen. Nach Durchlauf der Amplifikation

zeigte sich kein Signal in der Schmelzkurvenanalyse. In einer Kontrolle mittels Gel-Elektrophorese konnten ebenfalls keine Produkte nachgewiesen werden. Im Rahmen umfassender Etablierungsversuche erfolgte zunächst eine  $MgCl_2$ -Titration. Hierbei zeigte sich nur für rs1559085 bei einer höheren Konzentration von 3-3,5 mM  $MgCl_2$  ein leichtes Produktsignal in der anschließenden Gelelektrophorese. Die Umstellung der Reaktionsbedingungen auf eine Amplifikationszeit von 20 Sekunden bei 72°C blieb ebenfalls erfolglos.

#### Etablierung von Reaktionsbedingungen für rs27852

Eine weitere Anpassung des PCR-Protokolls mit einer Touchdown-Annealing-Phase von 65-53°C und 15 Sekunden pro Zyklus, sowie einer Elongation bei 72°C für 30 Sekunden bei einer insgesamt erhöhten Anzahl von 50 Zyklen, ergab weiterhin kein Produkt. Um die Möglichkeit auszuschließen, dass die ausbleibende Amplifikation mit den verwendeten Chemikalien zusammenhängt, wählten wir einen Versuchsansatz mit den Chemikalien aus dem Thermocycler und dem Zusatz Resolight. Das PCR-Protokoll wurde entsprechend angepasst. Es erfolgte eine Präinkubation für 1 Minute bei 94°C, die Elongation fand bei 68°C für jeweils 15 Sekunden statt und die Schmelzkurve wurde von 60-95°C aufgetragen. Unter diesen Bedingungen zeigten sich spezifische Produkte für rs27852, allerdings auch sehr prominente Primerdimere. Mittels „Multiple Primer Analyzer“ von Thermo Scientific führten wir eine Analyse möglicher Dimerformationen der Primer durch. Dabei zeigte sich für F1 eine deutliche Tendenz einer Dimerbildung des Primers mit sich selbst, sowie multiple Cross-Dimer der verschiedenen Primer untereinander. Um diese zu minimieren, erfolgte die Ermittlung einer optimalen Annealing-Temperatur mittels Temperaturgradienten-PCR. Dabei erwies sich die PCR unter einer Temperatur von 62°C am effektivsten. In einem weiteren Schritt konnten wir unter der Verwendung verschiedener Zusätze mittels GC-Enhancer das Auftreten von Primerdimeren fast gänzlich unterbinden. In einem Ansatz mit One-Taq-Polymerase (NEB), 20% GC-Enhancer, Resolight und einer festen Annealing-Temperatur von 60°C konnten schließlich einzelne spezifische Kurvenverläufe im High-Resolution-Melting detektiert werden (Abbildung 18), in einer anschließenden Sanger-Sequenzierung konnten diese allerdings keinen

spezifischen Genotypen zugeordnet werden. Im zeitlichen Rahmen dieser Arbeit erwies sich die Etablierung optimaler Reaktionsbedingungen für rs27852 daher als erfolglos.

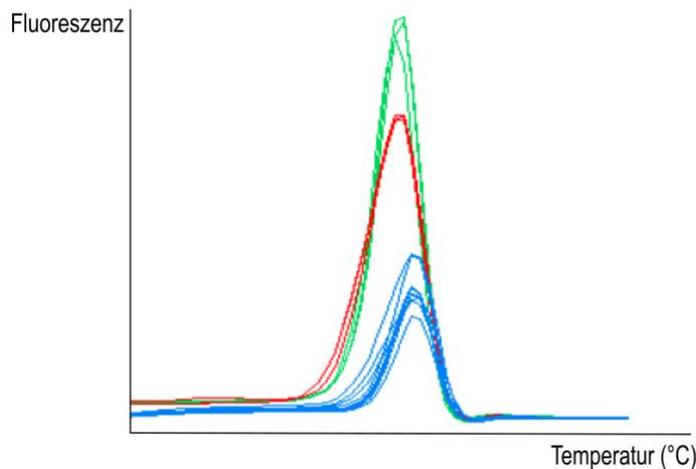


Abbildung 17: **High-Resolution-Melting von rs27852**. Die Etablierung von Reaktionsbedingungen zur erfolgreichen Typisierung von rs27852 erwies sich als kompliziert. Neben einzelnen unspezifischen Produkten entstanden in erster Linie Primerdimere. Auch nach verschiedenen Versuchen die Reaktionsbedingungen zu optimieren, konnte keine eindeutige Typisierung des Polymorphismus durchgeführt werden.

#### Etablierung von Reaktionsbedingungen für rs1559085

Die Änderung des PCR-Protokolls in eine Touchdown-PCR mit Annealingtemperaturen von 65-53°C und 15 Sekunden pro Zyklus, sowie einer Elongation bei 72°C für 30 Sekunden bei einer insgesamt erhöhten Anzahl von 50 Zyklen, ergab für rs1559085 spezifische Produkte in der Gelelektrophorese. Im Light-Cycler konnte ein vermehrtes Fluoreszenzsignal mit einem Maximum zwischen 75°C-80°C nachgewiesen werden. In der Darstellung der Schmelzkurvenverläufe zwischen 76°C-80°C zeigten sich drei verschiedene Kurvenverläufe. Dabei waren zwei Kurvenverläufe eindeutig voneinander zu unterscheiden, während die Darstellung des Unterschieds der anderen zwei Kurvenverläufe nur im Bland-Altman-Plot gelang. Um zu bestätigen, dass die Kurvenverläufe unterschiedlichen Genotypen von rs1559085 entsprechen, führten wir eine Sanger-Sequenzierung exemplarischer Proben durch. Dabei konnten die Kurven als entsprechende Kurvenverläufe für die drei möglichen Genotypen validiert werden (Abbildung 18).

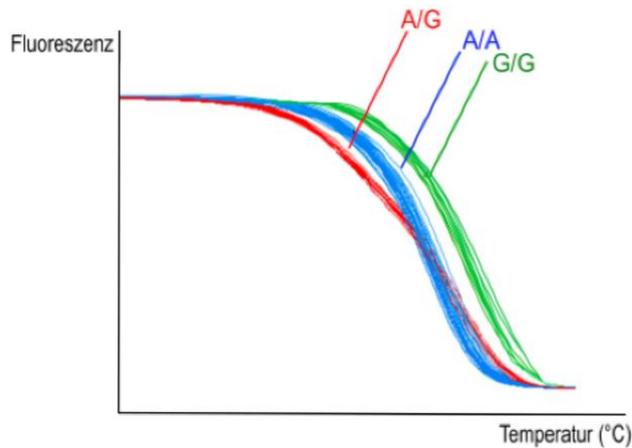


Abbildung 18: **High-Resolution-Melting von CAST.** Es ergeben sich nach dem High – Resolution – Melting drei unterschiedliche Schmelzkurvenverläufe für den Polymorphismus rs1559085 in CAST. Mittels Sanger – Sequenzierung konnten die verschiedenen Kurvenverläufe den drei Genotypen von CAST zugeordnet werden. Dabei entspricht in die blaue Kurve dem Genotyp A/A, die grüne Kurve dem Genotyp G/G und die rote Kurve dem heterozygoten Genotyp A/G. Anhand des Schmelzkurvenverlaufes ist eine eindeutige Differenzierung der Genotypen möglich.

#### 4.4.3 Die Verteilung der Genotypen von rs1559085 in CAST

Die Typisierung der Patientenproben ergab für insgesamt 316 Patienten einen eindeutigen Genotyp für die Einzelnukleotid-Polymorphismen rs1559085 in CAST. Dabei wurde zwischen Proben der EUROSCA Kohorte und der EUSAGE Kohorte unterschieden. Aus der EUROSCA Kohorte wurden 271 Proben eindeutig typisiert. Aus der EUSAGE Kohorte wurden insgesamt 45 Proben typisiert, 29 aus Israel und 16 aus Peru. Tabelle 35 zeigt die Verteilung der ermittelten Genotypen von rs1559085.

Tabelle 35: **Verteilung der ermittelten Genotypen von rs1559085 in CAST**

Polymorphismus	rs 1559085	
Kohorte	EUROSCA	EUSAGE
A/A	204	41
A/G	59	4
G/G	8	0
Gesamt	271	45

Die ermittelte Verteilung nach unserer Typisierung wurde mit der erwarteten Verteilung nach dem Hardy-Weinberg-Gesetz verglichen. Die Berechnung erfolgte nach der bekannten Formel. Die MAF der für rs1559085 entnommen wir der NLM. In Tabelle 36 wird die erwartete Verteilung der Genotypen dargestellt.

Tabelle 36: **Erwartete Verteilung der Genotypen** nach der Berechnung durch das Hardy – Weinberg – Gesetz in einer idealen Population.

Polymorphismus	rs 1559085	
Kohorte	EUROSCA	EUSAGE
A/A	202	35
A/G	64	10
G/G	5	0
Gesamt	271	45

Dabei zeigt sich lediglich für die EUSAGE Kohorte eine Abweichung zwischen der Anzahl der typisierten A/G Genotypen und der erwarteten Anzahl nach Hardy-Weinberg um einen Faktor  $> 2$ . Um zu überprüfen ob diese Abweichung signifikant ist, führten wir einen Chi-Quadrat-Test durch.

#### 4.4.4 **Chi – Quadrat – Test: Keine signifikanten Abweichungen der Verteilung von Genotypen in SCA3 Proben**

Für die Durchführung des Chi-Quadrat-Tests wurden die Kohorten im Folgenden als eine Gesamtkohorte zusammengefasst. Dabei zeigt sich im direkten Vergleich der Verteilung der Genotypen in der von uns typisierten Kohorte und einer idealen Verteilung nach Hardy-Weinberg keine Abweichung  $> 10$ . Die stärkste Abweichung zeigt sich für den Genotyp A/G (vgl. Tabelle 37).

Tabelle 37: **Vergleich der Verteilung der Genotypen** in typisierten SCA3 Patientenproben und der idealen Verteilung nach dem Hardy – Weinberg – Gesetz.

Polymorphismus	rs 1559085	
Kohorte	SCA3- Proben	Hardy-Weinberg
A/A	245	237
A/G	63	74
G/G	8	5

Ein p-Wert von 0,16 im Chi-Quadrat-Test bestätigte statistisch, dass in der Kohorte von SCA3-Patienten keine signifikante Abweichung der Verteilung an Genotypen gegenüber einer idealen Population nach Hardy-Weinberg besteht. Demnach kommt der Genotyp A/G in der von uns typisierten Kohorte von 316 Patienten zwar vergleichsweise häufiger auf, durch die geringe Anzahl an Proben ist dieser Effekt jedoch am ehesten als zufällige Schwankung einzuordnen.

#### 4.4.5 Der Einfluss von rs1559085 in CAST auf das Erkrankungsalter

Nach der Typisierung von rs1559085 in CAST wurde mittels des Statistikprogramms Jump (JMP) eine Regressionsanalyse zur Überprüfung des Einflusses auf das Erkrankungsalter durchgeführt. Dabei wurde das Erkrankungsalter als Y-Variable eingesetzt und die individuelle Anzahl an CAG-Wiederholungen als X-Variable. Zwischen dem Erkrankungsalter und der individuellen Anzahl an CAG-Wiederholungen besteht eine negative Korrelation von signifikanter Relevanz ( $<0,0001^*$ ). Durch die Hinzunahme des CAST Genotyp als weitere X-Variable konnte ein Einfluss des rs1559085 auf das Erkrankungsalter festgestellt werden (Abbildung 20).

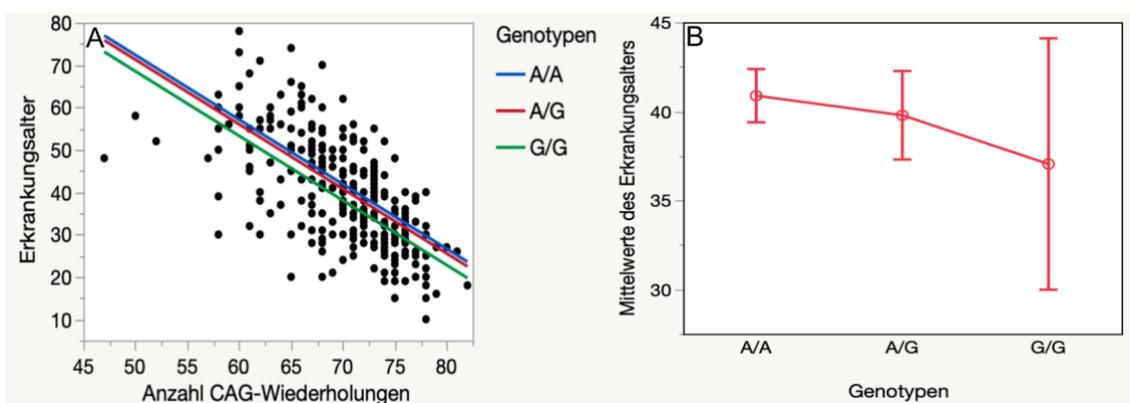


Abbildung 19: **Regressionsanalyse des Einflusses von rs1559085 auf das Erkrankungsalter.** A: Dargestellt ist eine Regressionsanalyse des Erkrankungsalters bezogen auf die individuelle Anzahl an CAG-Wiederholungen mit einer Geradenanpassung für die jeweiligen Genotypen von rs1559085. Es zeigt sich, dass es bei Vorliegen eines homozygoten G/G Genotyps zu einer Verringerung des Erkrankungsalters im Vergleich zu den Genotypen A/A und A/G kommt. B: Darstellung der Mittelwerte des Erkrankungsalters bezogen auf die unterschiedlichen Genotypen. Dabei zeigt sich eine Verringerung des mittleren Erkrankungsalters für G/G bei deutlich größerer Varianz des Erkrankungszeitpunktes. Dabei liegt das Erkrankungsalter weiterhin zwischen 30 und 45 Jahren und somit deutlich unter dem statistisch durchschnittlichen Erkrankungsalter von 70 Jahren.

Die Regressionsanalyse des Erkrankungsalters erfolgte bezogen auf die individuelle Anzahl an CAG-Wiederholungen mit einer Geradenanpassung für die jeweiligen Genotypen von rs1559085. Dabei zeigte sich ein durchschnittlich verringertes Erkrankungsalter für den Genotyp G/G. Die Ermittlung der Mittelwerte des Erkrankungsalters der verschiedenen Genotypen ergab für G/G ein mittleres Erkrankungsalter von 37 Jahren und somit eine Abweichung von 3,8 Jahren vom Genotyp A/A (mittleres Erkrankungsalter 40,8 Jahre). Die Überprüfung der statistischen Signifikanz dieser Beobachtung und des alleinigen Einflusses von rs1559085 ergab einen p-Wert von 0,47.

Erfolgte die Korrelationsanalyse des Erkrankungsalters bezogen auf die Anzahl an CAG-Wiederholungen in Kombination mit dem Genotyp von rs1559085 verbesserte den p-Wert auf 0,25. Im Rahmen einer Varianzanalyse (ANOVA) konnte der Einfluss von G/G auf das Erkrankungsalter mit einer Verringerung des Erkrankungsalters bestätigt werden. Die Abweichungen zeigten mit einem p-Wert von 0,1 keinen Anhalt für die Annahme einer statistischen Relevanz.

#### **4.5 Zusammenfassung der Ergebnisse**

Einen Teil der Proben für diese Arbeit bekamen wir aus Israel und Peru zugeschickt. Um eine valide statistische Analyse mit den Daten zu den Proben durchführen zu können, wurden diese 51 Proben nicht nur für die Einzelnukleotid-Polymorphismen typisiert, sondern auch ihre individuelle Anzahl an CAG-Wiederholungen analysiert. Eine Fragmentlängenanalyse für ATXN3 wurde für diese Proben in den Laboren der Kooperationspartner zwar durchgeführt, bei einer stichprobenhaften Überprüfung durch die Diagnostik im Uniklinikum Tübingen ergaben sich jedoch deutlich Abweichungen von mehr als 3 Wiederholungen. Dies veranlasste uns dazu alle Proben erneut zu analysieren.

In den Kohorten konnten drei Polymorphismen in zwei Genen mittels High-Resolution-Melting im Light-Cycler typisiert werden. Für die beiden Polymorphismen rs429358 und rs7412 in APOE wurde die Methodik einer allelspezifischen PCR mit anschließendem High-Resolution-Melting gewählt, der rs1559085 in CAST wurde mittels Schmelzkurvenanalyse typisiert. Für den Polymorphismus rs27852 in CAST konnte keine zuverlässige Methodik zur Typisierung etabliert werden (zusammengefasst in Tabelle 38).

Das Erkrankungsalter der typisierten Kohorten lag im Durchschnitt bei 40,5 Jahren, die Anzahl der CAG-Wiederholungen betrug durchschnittlich 70,5. Nach der statistischen Analyse ergab sich in Kombination mit der individuellen Anzahl an CAG-Wiederholungen ein signifikanter Einfluss des APOE Haplotyps auf das Erkrankungsalter (p-Wert:0,037\*). Zudem konnte nachgewiesen werden, dass bei  $\epsilon 2$  Allel Trägern das durchschnittliche Erkrankungsalter im Vergleich zu homozygoten  $\epsilon 3$  Allel Trägern im Schnitt um ca. 10 Jahre verringert ist. Homozygote  $\epsilon 2$  Träger bildeten dabei das Extrem mit dem geringsten

Erkrankungsalter und homozygote  $\epsilon 3$  Träger das Extrem mit dem höchsten Erkrankungsalter. Heterozygote Träger von  $\epsilon 2$  erkrankten im Vergleich mit allen anderen Genotypen im Schnitt 2,4 Jahre früher.

Bei der individuellen Betrachtung der Einzelnukleotid-Polymorphismen in APOE konnte zudem gezeigt werden, dass der Einbezug der Genotypen von rs429358 in die Analyse, den prädiktiven Wert für das Erkrankungsalter signifikant um zusätzliche 2% verbessert. Der rs7412 schien einzeln betrachtet keinen prädiktiven Mehrwert zu haben. Die Kombination beider Polymorphismen als Haplotyp verbessert den prädiktiven Wert für das Erkrankungsalter um 5%.

Der Einfluss von rs1559085 in CAST auf das Erkrankungsalter konnte mit einem p-Wert von 0,47 nicht als signifikant festgestellt werden. Anhand einer Regressionsanalyse des Erkrankungsalters und der Anzahl an CAG-Wiederholungen mit Anpassung der Geraden für die einzelnen Genotypen von rs1559085 konnte allerdings eine deutliche Verminderung des Erkrankungsalters bei Vorliegen eines homozygoten G/G Genotyp um durchschnittlich – 3,8 Jahre im Vergleich mit dem häufigeren Genotyp A/A festgestellt werden. Ebenso zeigte eine Analyse der durchschnittlichen Anzahl an CAG-Wiederholungen in Abhängigkeit der Genotypen von rs1559085, dass bei Vorliegen eines G/G Genotyps die Anzahl der Wiederholungen im Schnitt um + 3 erhöht ist. Dieser Zusammenhang ist mit einem p-Wert von 0,25 nicht signifikant, zeigt jedoch eine mögliche Tendenz eines Zusammenhangs zwischen dem Genotyp von rs1559085 in CAST und der Anzahl an CAG – Wiederholungen.

*Tabelle 38: Zusammenfassung der untersuchten Einzelnukleotid-Polymorphismen, und ihr Einfluss auf das Erkrankungsalter*

Einzelnukeotidpolymorphismus	Gen	Methodik	Einfluss auf das Erkrankungsalter
rs429358	APOE	Typisierung mit HRM und Sequenzierung	<u>signifikant*</u>
rs7412		Typisierung mit HRM und Sequenzierung	nicht signifikant
rs1559085	CAST	Typisierung mit HRM und Sequenzierung	nicht signifikant
rs27852		Schwache Amplifikation im HRM, keine Typisierung möglich	nicht nachgewiesen
Anzahl der CAG – Wiederholungen	ATXN3	Fragmentlängen-Analyse im Beckman Coulter	<u>signifikant*</u>

## 5 Diskussion

Die SCA3 ist eine autosomal-dominant vererbte Erkrankung, die durch eine Expansion von CAG-Wiederholungen im *ATXN3* Gen hervorgerufen wird. Dabei kommt es zu einer Verlängerung der kodierten Polyglutaminsequenz im Ataxin-3 Protein (Takiyama et al. 1993; Kawaguchi et al. 1994). Die Tendenz der Akkumulation von Polyglutamin führt zu multiplen, bisher nicht vollständig verstandenen, neurotoxischen Prozessen und einer progredienten neurodegenerativen Symptomatik betroffener Patienten (Matos, de Almeida und Nobrega 2019). Betroffene Patienten weisen klinisch sehr heterogene Phänotypen auf und erkranken im Durchschnitt innerhalb ihrer 4. Lebensdekade mit einer Varianz von 5 – 70 Jahren. Es ist bekannt, dass ein direkter Zusammenhang zwischen der Länge der CAG-Wiederholungen und der Schwere und dem Zeitpunkt der Erkrankung besteht (van de Warrenburg et al., 2005). Dabei bleibt jedoch für etwa die Hälfte der Betroffenen unklar in welchem Alter und wie schwer sie wahrscheinlich erkranken werden. Um den Erkrankungsprozess dieser Patienten besser zu verstehen, wurden in der Vergangenheit bereits einige Gene bezüglich ihres Einflusses auf das Erkrankungsalter der SCA3 überprüft. Vor diesem Hintergrund beschäftigt sich diese Promotion vor allem mit der Identifizierung möglicher Einflussfaktoren auf das individuelle Erkrankungsalter der SCA3 mit dem übergeordneten Ziel eines besseren Krankheitsverständnis und der Charakterisierung neuer Prognosemarker. Die Analyse möglicher Einflussfaktoren auf das Erkrankungsalter erfolgte mittels High-Resolution-Melting, Fragmentlängenanalyse und Sanger-Sequenzierung und bezog sich auf jeweils zwei Einzelnukleotid-Polymorphismen im *APOE* und *CAST* Gen, sowie der Anzahl von CAG-Wiederholungen im *ATXN3* in einer europäischen und südamerikanischen Kohorte.

## 5.1 CAG-Wiederholungen als Einflussfaktor auf das Erkrankungsalter

Im Rahmen einer großen Metaanalyse genetischer Einflussfaktoren der SCA3 von Mattos et al.(2018) wurden die in diesem Zusammenhang untersuchten Gene zusammengetragen. Die Analysen bezogen sich auf den Zeitraum von 1994 bis 2017 und geben einen guten Überblick über die bis dato 2018 gewonnenen Erkenntnisse. Als wichtigster Einflussfaktor auf das Erkrankungsalter gilt nach zahlreichen Studien weiterhin die Länge der CAG-Expansion im mutierten *ATXN3*. Der signifikante Einfluss der CAG-Länge wurde bereits mehrfach bewiesen und gilt mit einem Voraussagewert bis zu 60% als wichtigster Prognosefaktor. Neben dem mutierten *ATXN3* wurde in einzelnen Studien auch der Einfluss von dem nicht-expandierten *ATXN3* überprüft, dabei konnte keine signifikante Verbesserung des Voraussagewertes nachgewiesen werden. Als weiterer wichtiger Einflussfaktor gilt vielmehr die Ethnie. Durch Einbezug der Herkunft der Betroffenen in die statistische Analyse konnte der Voraussagewert im Durchschnitt um 9% verbessert werden.

Auch im Rahmen dieser Arbeit ergab die Korrelation der individuellen CAG-Repeat-Länge mit dem Erkrankungsalter eine eindeutige negative Korrelation mit einem Korrelationsfaktor von 0,45. Dabei wurden zur Gewährleistung einer ausreichenden Kohortenstärke die EUROSCA und EuSAGE Kohorte zusammengefasst und unabhängig ihres ethnischen Hintergrundes als homogene Gruppe analysiert. Den größten Anteil der typisierten Proben bezogen wir aus der EUROSCA Kohorte. 51 der typisierten Proben kamen aus Peru und Israel und gehören zur EuSAGE Kohorte. Die Proben dieser Kohorte wurden in den Ländern der Kooperationspartner aufgereinigt und in Puffer gelagert. Der Versand erfolgte ungekühlt per Expressversand. Außerdem wurde in den jeweiligen Kooperationslaboren bereits eine Analyse der Anzahl an CAG-Wiederholungen in *ATXN3* durchgeführt. Um eine Verfälschung der Ergebnisse durch abweichende Ergebnisse bei der Analyse der CAG-Wiederholungen vorzubeugen, führten wir eine Qualitätskontrolle der zugesandten Proben durch. Dabei mussten wir feststellen, dass eine deutliche Abweichung von > 3 CAG-Wiederholungen in einem Großteil der Stichproben detektiert wurden. Dies veranlasste uns dazu alle 51 zugesandten Proben erneut mittels

Fragmentlängenanalyse im *ATXN3* zu analysieren. Leider konnten wir dabei nur für 57% der Proben aus Israel die bisherigen Ergebnisse der Analyse bestätigen. Bei den Proben aus Peru gelang uns die Bestätigung der Ergebnisse sogar nur für 32%. Im Rahmen der Analyse von CAG-Wiederholungen von *ATXN3* wurden in Absprache mit der Diagnostik der Humangenetik und angewandten Genomik in Tübingen, Abweichungen von weniger als 3 Wiederholungen als bestätigt angesehen.

Aufgrund der deutlichen Abweichungen setzten wir uns mit den jeweiligen gewählten Methoden zur Analyse der CAG-Expansionen in *ATXN3* auseinander. In Peru und Israel wurden die Expansionen mittels Gel-Elektrophorese detektiert, während wir eine Kapillarelektrophorese durchführten. Beide Methoden sind gut etabliert und werden standardisiert angewandt. Allerdings scheint die Gel-Elektrophorese etwas ungenauer in der Bestimmung der Anzahl an CAG-Triplets zu sein und erweist sich naturgemäß als fehleranfälliger durch menschliches Versagen. Ein weiterer Grund für die starke Abweichung der Analyseergebnisse könnte eine mögliche Beschädigung der Proben während des ungekühlten Versands aus Brasilien darstellen. Die statistische Auswertung der Daten erfolgte mit den Ergebnissen der von uns durchgeführten Kapillarelektrophorese.

## **5.2 Einzelnukleotid-Polymorphismen in APOE als Einflussfaktoren auf das Erkrankungsalter an SCA3**

Einzelnukleotid-Polymorphismen in APOE wurden bereits im Zusammenhang mit mehreren neurodegenerativen Erkrankungen als Risiko-, Prognose- und Einflussfaktor überprüft (zusammengefasst in Verghese et al.,2013). Das APOE  $\epsilon 4$  Allel gilt dabei als signifikanter Risikofaktor für ein frühes Erkrankungsalter an Alzheimer (Evidenzklasse A). Über 60% aller an Alzheimer erkrankten Patienten sind Träger von mindestens einem APOE  $\epsilon 4$  Allel (Riedel et al.,2016). Homozygote  $\epsilon 4$  Träger weisen ein besonders hohes Erkrankungsrisiko auf und zeigen eine deutliche Tendenz zur Akkumulation von  $\beta$ Amyloid im Gehirn. Der Nachweis von entsprechenden Ablagerungen im Gehirn gelingt bei  $\epsilon 4$  heterozygoten Trägern im Durchschnitt 10-15 Jahre später als bei  $\epsilon 4/\epsilon 4$

Genotypen (Jack et al.,2015). Zudem wurde nachgewiesen, dass Träger von APOE  $\epsilon$ 4 nach Schädel-Hirn-Traumata ein deutlich schlechteres Outcome aufweisen als Träger der anderen APOE Allele (Crawford et al., 2002). Im Gegensatz dazu scheint das Vorliegen eines APOE  $\epsilon$ 2 Allels einen präventiven Einfluss im Zusammenhang eines späteren Erkrankungsalters an Alzheimer zu haben. Allerdings erkrankten Träger eines APOE  $\epsilon$ 2 Allels statistisch häufiger und früher an Parkinson (Zarepari et al., 2002). Außerdem konnte im Zusammenhang mit APOE  $\epsilon$ 2 eine höhere Prävalenz an frontotemporaler Demenz (Verpillat et al.,2002) und einem früheren Erkrankungsalter an Huntington (Kehoe et al.,1999) nachgewiesen werden. Im Rahmen einigen Studien wurde zudem bereits der Einfluss von APOE Allelen auf das Erkrankungsalter der progredient neurodegenerativen Erkrankung der SCA3 diskutiert. Dabei wurde vor allem der Einfluss des  $\epsilon$ 2 Allels auf das Erkrankungsalter der SCA3 genauer untersucht. Bettencourt et al. kamen dabei 2011 zu dem Schluss, dass das Vorliegen eines APOE  $\epsilon$ 2 Allels den Phänotyp der SCA3 beeinflusst und das Risiko einer früheren Erkrankung erhöht. In der von ihnen untersuchten Kohorte erkrankten Träger des APOE  $\epsilon$ 2 Allels im Durchschnitt 5 Jahre früher als die Kontrollprobanden. Allerdings konnte durch das Einbeziehen des Vorliegens eines APOE  $\epsilon$ 2 Allels der prädiktive Wert für das Erkrankungsalter der SCA3 nur um einen Prozentpunkt verbessert werden. Vermutlich aufgrund einer geringen Kohorte an APOE  $\epsilon$ 2 Trägern (n=20) konnte in diesem Fall eine statistische Relevanz nachgewiesen werden. Eine weitere Arbeit von Peng et al. (2014) ergab ähnliche Ergebnisse. Die Arbeitsgruppe um Zhou et al. setzte sich 2014 ebenfalls mit dem Einfluss von APOE  $\epsilon$ 2 auf das Erkrankungsalter der SCA3 auseinander und kam zu dem Schluss, dass die verschiedenen APOE Allele keinen signifikanten Einfluss auf das Erkrankungsalter der SCA3 haben. Zudem äußerten sie kritisch, dass der positiv nachgewiesene Effekteinfluss in vorherigen Studien durch eine zu geringe Kohortenstärke zu Stande kam. Eine weitere Erklärung für die unterschiedlichen Ergebnisse könnte die Herkunft der verschiedenen Kohorten sein. Während Bettencourt et al. eine europäische und südamerikanische Kohorte mit 192 Betroffenen und Peng et al. eine Kohorte mit 155 Betroffenen vom chinesischen

Festland untersuchten, bezogen sich die Erkenntnisse von Zhou et al. auf eine Kohorte von 403 chinesischen Betroffenen.

In dieser Arbeit wurden insgesamt 346 Betroffene SCA3 Patienten einer hauptsächlich europäischen Kohorte auf zwei Einzelnukleotid-Polymorphismen in *APOE* typisiert. Die Polymorphismen rs429358 und rs7412 liegen in Exon 4 von *APOE* 138 Basenpaare voneinander entfernt. In ungefähr 70% der Fälle befindet sich an Position von rs429358 in Codon 112 ein Thymin (T) und an Position von rs7412 in Codon 158 ein Cytosin (C). Diese Kombination entspricht dem *APOE* Allel  $\epsilon 3$ . Bei dem  $\epsilon 2$  Allel befindet sich anstelle von C in Codon 158 ein T (rs7412). Auf Proteinebene entspricht dies einer Missensevariante bei der Cystein anstelle von Argenin translatiert wird. Das Allel *APOE*  $\epsilon 4$  hingegen entsteht bei Vorliegen eines Einzelnukleotid-Polymorphismus in Codon 112 (rs429358) bei dem ein C anstelle eines T auf Proteinebene zur Missensvariante eines Argenins anstelle von Cystein führt.

### **5.2.1 Abweichung vom Hardy-Weinberg-Gesetz**

Nach der Typisierung der Genotypen von rs429358 und rs7412 stellte sich im Chi-Quadrat-Test eine signifikante Abweichung der beobachteten Genotypen von der erwarteten Verteilung nach dem Hardy-Weinberg-Gesetz heraus. In beiden untersuchten Stichproben-Gruppen fanden sich in der Regel mehr homozygote Proben für das seltenere Allel.

Die Abweichungen der typisierten Ergebnisse von den erwarteten wirft zunächst einige Fragen auf. Um zu beurteilen, welchen Wert dieses Ergebnis in der folgenden Auswertung bekommt, erfolgte die kritische Auseinandersetzung mit dem Chi-Quadrat-Test und seiner Aussagekraft. Zunächst sollte für einen zuverlässigen Chi-Quadrat-Test die erwartete Häufigkeit jeder Zelle über 5 sein (Hopper, 1971). Diese Gegebenheit ist in dem Fall des homozygoten Genotyps für das seltenere Allel nach dem Hardy-Weinberg-Gesetz nicht gegeben (rs429358 (C/C)=0 und rs7412 (T/T)=2). Zudem ist bei allen Signifikanzen die Stichprobengröße von Bedeutung. Bei größeren Stichproben erlangen kleinere Abweichungen eine höhere Signifikanz. Im Allgemeinen sagt die Signifikanz

einer Abweichung im Chi-Quadrat-Test nichts über die Ausprägung und Stärke des Effekts aus.

Unter der Annahme, dass *APOE* Genotypen eventuell einen Einfluss auf das Erkrankungsalter haben könnte, verfolgen wir im weitesten Sinne auch die Fragestellung, ob die verschiedenen Genotypen auch einen funktionellen Einfluss auf die Pathophysiologie der Erkrankung nehmen können. So könnte eine signifikante Abweichung der typisierten Proben auch einen Hinweis auf einen deutlich größeren Effekt von *APOE* auf die Erkrankung der SCA3 geben. Die Erfüllung des Hardy-Weinberg-Gleichgewichts beruht auf der Annahme, dass keine Unterschiede zwischen den Genotypen in Hinblick auf Fruchtbarkeit oder Überlebensfähigkeit vorliegen, eine zufällige Paarung erfolgte (Panmixie), eine unendliche große Population zugrunde liegt und keine Mutationen oder Selektionen aufgetreten sind (Hartl und Clark, 1997). Sollte ein bestimmtes *APOE* Allel allerdings einen maßgeblichen Einfluss auf die Pathophysiologie der SCA3 nehmen, hätte dies in Folge einer Selektion Einfluss auf das gehäufte Auftreten einzelner Genotypen in der Kohorte von SCA3 Patienten. Zudem ist die untersuchte Kohorte begrenzt und bezieht sich auf teilweise genetisch verwandte Proben. Eine weitere Erklärung für Abweichungen vom Hardy-Weinberg-Gleichgewicht ist der sogenannte Wahlund-Effekt. Dieser besagt, dass räumlich und zeitlich isolierte Teilpopulationen verhindern, dass eine vollständige genetische Durchmischung der Gesamtpopulation stattfindet (Wahlund 1928). Setzt man nun unterschiedliche genetische Kompositionen der Teilpopulationen voraus, führt dies zu einem Mangel an heterozygoten Individuen im Vergleich zu der erwarteten Anzahl in der Gesamtbevölkerung durch genetische Effekte wie Inzucht und genetischer Drift. Ein ausgeprägtes Heterozygotiedefizit wurde in der von uns untersuchten Kohorte nicht festgestellt. Zudem scheint der Wahlund-Effekt wenig wahrscheinlich, da die untersuchte Kohorte Proben aus Mitteleuropa und Südamerika umfasst.

Die Möglichkeit, dass die bevorzugte Amplifikation eines Allels einen Einfluss auf die Typisierungsergebnisse gehabt haben kann, besteht. Bei der Etablierung der Methodik zeigte sich für den rs7412 zunächst eine stärkere Amplifikation durch den allelspezifischen-Primer für T/T. Durch Anpassungen der

Primerkonzentrationen konnte dieser Nachteil allerdings erfolgreich ausgeglichen werden. Allerdings könnte eine verstärkte Amplifikation durch den allelspezifischen Primer für T das gehäufte Auftreten von T/T homozygoten Proben für rs7412 erklären. Eine weitere Möglichkeit wäre der Einfluss durch Nullallele. Bei Nullallelen handelt es sich um Allele, die durch fehlerhafte PCR-Prozesse ihre genetischen Eigenschaften verlieren oder gar nicht amplifiziert werden. Dies führt in der Folge zu Fehlinterpretationen und einer erhöhten Anzahl an falsch-homozygoten Proben. In der von uns untersuchten Stichprobe müsste für rs429358 T dabei das Nullallel darstellen. Dies scheint bei einem Anteil von 80% T/T typisierten Proben sehr unwahrscheinlich.

Abschließend konnte nicht vollständig erklärt werden, warum die typisierten Ergebnisse von den erwarteten nach dem Hardy-Weinberg-Gesetz abweichen.

### **5.2.2 Die Verteilung der APOE Isoformen bleibt unverändert**

Bei der Betrachtung der APOE Isoformen durch die Kombination der Genotypen beider Polymorphismen fällt wiederum eine Verteilung auf, die den ungefähren Anteilen in der Allgemeinbevölkerung entspricht.

In der Allgemeinbevölkerung sind bis zu 60-90 % Träger von APOE  $\epsilon_3$ , zwischen 10 und 20%  $\epsilon_4$  – Allel – Träger und 0 – 20 %  $\epsilon_2$  Allel – Träger (Singh et al, 2007). In der von uns untersuchten Kohorte wurde diese Verteilung widergespiegelt. Auch die von Zhou et al., Bettencourt et al. und Peng et al. untersuchten Kohorten zeigten keinen Unterschied in der relativen Häufigkeit der verschiedenen APOE Allele bei SCA3 Betroffenen. Die mögliche Annahme, dass das Vorliegen eines bestimmten APOE Allels in direktem Kausalzusammenhang mit der Erkrankung der SCA3 stehen könnte und in einer Gruppe von Erkrankten häufiger vorliegt wurde durch diese Erkenntnisse mehrfach widerlegt.

### **5.2.3 Ein nachgewiesener Einfluss von APOE Haplotypen auf das Erkrankungsalter**

In der Gruppe der typisierten Patienten wurde ein statistischer Zusammenhang zwischen den verschiedenen APOE Genotypen und dem individuellen Erkrankungsalter überprüft. Vorherige Arbeiten zu dieser Thematik kamen zu

unterschiedlichen Ergebnissen und gaben einen Anlass diesen Zusammenhang in einer möglichst großen und heterogenen Kohorte zu überprüfen. Die von uns untersuchte Kohorte setzte sich aus 295 europäischen DNA-Proben (EUROSCA) und 51 Proben einer gemischten Kohorte aus Israel und Peru (EuSAge) zusammen. Dabei konnten wir einen Einfluss des *APOE* Genotyps auf das Erkrankungsalter der SCA3 nachweisen. Wir kamen zu dem Ergebnis, dass die Erkrankung bei *APOE*  $\epsilon$ 2 Trägern im Vergleich mit allen anderen Genotypen im Schnitt 2 Jahre früher auftritt. Bei der Überprüfung des Einflusses der einzelnen Einzelnukleotid-Polymorphismen konnten wir nachweisen, dass der Genotyp des rs429358 eine signifikante Verbesserung des Voraussagewertes des Erkrankungsalters um 2 Prozentpunkte bewirkt. In Kombination mit dem Polymorphismus rs7412 konnte der prädiktive Wert sogar um 5 Prozentpunkte gesteigert werden. Bei der alleinigen Analyse des prädiktiven Wertes der individuellen Länge an CAG-Wiederholungen ergab sich ein Voraussagewert von 40% ( $r^2 = 0,40$ ), unter der Zusammenschau der Länge der CAG-Wiederholungen und dem *APOE* Haplotyp konnte die Varianz des Erkrankungsalters somit für 45% der Kohorte erklärt werden. Im Vergleich mit anderen Studien ist ein  $r^2$  Wert von 0,45 relativ gering und lässt die Varianz des Erkrankungsalters für 55% der überprüften Kohorte weiterhin unerklärt. Der von Bettencourt et al. 2011 und Peng et al. 2014 nachgewiesene Einfluss von *APOE*  $\epsilon$ 2 auf das Erkrankungsalter wurde als deutlich ausgeprägter beschrieben. Sie konnten nachweisen, dass ein Zusammenhang zwischen dem Vorliegen eines *APOE*  $\epsilon$ 2 und der Verringerung des Erkrankungsalters von 4-5 Jahren besteht. Zudem konnten Bettencourt et al. unter Hinzunahme des *APOE*  $\epsilon$ 2 Status die Varianz des Erkrankungsalters für 60,9% ihrer Kohorte erklären. Wir konnten die Beobachtung, dass *APOE*  $\epsilon$ 2 einen negativen Einfluss auf das Erkrankungsalter der SCA3 hat zwar bestätigen, allerdings zeichnete sich der tatsächliche Effekt in unserer Kohorte deutlich geringer ab. Dies könnte verschiedene Gründe haben.

#### **5.2.4 Mögliche Gründe für abweichende Ergebnisse**

Der nachgewiesene signifikante Effekt des *APOE*  $\epsilon$ 2 Status auf das Erkrankungsalter der SCA3 wurde zuvor von Bettencourt et al. (2011) und Peng

et al. (2014) beschrieben und von Zhou et al. (2014) kritisch überprüft. Dabei untersuchten Peng et al. und Zhou et al. (2014) chinesische Kohorten und Bettencourt et al. (2011) eine gemischte europäische und südamerikanische Kohorte. Während Zhou et al. in einer größeren Kohorte von 403 Patienten den nachgewiesenen Effekt von *APOE*  $\epsilon 2$  auf das Erkrankungsalter nicht bestätigen konnte, konnten wir in unserer Arbeit einen signifikanten Einfluss des *APOE* Status auf das Erkrankungsalter feststellen. Damit konnten wir in einer deutlich größeren Kohorte von 346 Patienten, den zuvor beschriebenen Effekt des *APOE*  $\epsilon 2$  in einer europäischen und südamerikanischen Patientengruppe bestätigen. Dennoch zeigte sich der Einfluss in der von uns untersuchten Kohorte deutlich geringer als in der kleineren Vergleichskohorte von Bettencourt et al. (199 Patienten). Ein ähnlicher Effekt wurde zuvor bereits bei Studien beobachtet, die den Einfluss von *APOE* auf das Erkrankungsalter an Huntington untersuchten. Auch dieser Effekt wurde sehr kritisch überprüft und diskutiert, dabei wurde ein Einfluss des *APOE* Status stets in den kleineren Kohorten beobachtet und in größer angelegten Studien widerlegt (Keheo et al., 1999 und Saft et al., 2004).

In unserer Arbeit könnte der schwächer nachgewiesene Effekt von *APOE*  $\epsilon 2$  durch die Verteilung der anderen Genotypen zustande gekommen sein. Während bei Bettencourt et al. nur um die 20 Patienten den Haplotyp  $\epsilon 2/\epsilon 3$  aufwies und keiner den Haplotyp  $\epsilon 2/\epsilon 2$ , waren in unsere Studie mit ca. 40 Patienten doppelt so viele Träger von  $\epsilon 2/\epsilon 3$  und 7 Patienten Träger des Haplotyps  $\epsilon 2/\epsilon 2$ . Diese Verteilung könnte in den Studien mit kleineren Kohorten zu einer verzerrten Einschätzung des mittleren Erkrankungsalters geführt haben. Zudem wies unsere Kohorte 6 Ausreißer mit einem Erkrankungsalter unter 20 Jahren auf. Dadurch wies die von uns untersuchte Kohorte bereits ein relativ geringes mittleres Erkrankungsalter von 40,5 Jahren auf.

Ein weiterer Effekt, der die unterschiedlichen Ergebnisse der verschiedenen Studien erklären könnte, ist der Einfluss der Herkunft und Epigenetik auf den *APOE* Status. Im Jahr 2020 wurde erstmals eine Studie über monozygotische Zwillinge mit SCA3 veröffentlicht. Diese wiesen die gleiche Länge an CAG-Wiederholungen im ATXN3 auf und exprimierten die gleichen bekannten

genetischen Einflussfaktoren (rs709930/rs910369 in ATXN3, APOE  $\epsilon$ 2 Allele und rs7969300 in ATXN2). Im Rahmen der Studie wurde der Einfluss epigenetischer Faktoren auf das Erkrankungsalter überprüft. Dafür wurde die DNA Methylierung der betroffenen Zwillinge mit der eines gesunden Geschwisterkindes verglichen. Trotz übereinstimmender genetischer Einflussfaktoren erkrankten die Zwillinge unterschiedlich schwer und zu verschiedenen Zeitpunkten. Das Ausmaß der DNA Methylierung unterschied sich bei den Zwillingen deutlich. Da sie ähnlichen Umwelteinflüssen ausgesetzt waren, scheint diese Erkenntnis ein Hinweis auf einen möglichen Einfluss von epigenetischen Faktoren auf den Verlauf der Erkrankung zu sein (Zhao et al., 2020).

Zudem wurden in allen bisher beschriebenen Studien unterschiedliche Kohorten untersucht. Zhou et al. (2014) stellten die Theorie auf, dass eine genetische Heterogenität der Kohorten eine Erklärung für die verschiedenen Erkenntnisse sein könnte. Je nach Herkunft der Patienten könnten andere funktionale Polymorphismen in engen Zusammenhang mit dem APOE Genotyp  $\epsilon$ 2/ $\epsilon$ 3 stehen, die eine Varianz im Erkrankungsalter erklären könnten.

### **5.3 Diskussion des High- Resolution- Melting als Methode zur Analyse von Einzelnukleotid-Polymorphismen**

Die Technik des High-Resolution Melting stellt eine geeignete Methode zur Analyse von Einzelnukleotid-Polymorphismen dar. Im Vergleich mit anderen Methoden zur Genotypisierung von Patientenproben, bietet sie einige Vorteile. Da im Light-Cycler sowohl die Amplifikation des entsprechenden Genabschnittes als auch das anschließende High-Resolution-Melting stattfinden kann, ohne dass man das Reaktionsgefäß nochmal öffnen müsste, ist die Gefahr einer Kontamination gering (Reed et al., 2007). Das HRM ist zudem wesentlich schneller durchführbar als vergleichbare Methoden zur Analyse von Polymorphismen, die das Auftragen auf ein Gel erfordern. Da keine anschließenden Trennverfahren nötig sind, ergibt sich zudem eine signifikante Kostenersparnis.

Durch das Hinzufügen fluoreszierender Marker beim Ansetzen des PCR-Mixes können im HRM zudem heterozygote Proben mit einer Sensitivität von bis zu

100% detektiert und von homozygoten Proben unterschieden werden. Allerdings kann es vorkommen, dass homozygote Proben anhand ihres Schmelzkurvenverhaltens nicht eindeutig voneinander abzugrenzen sind. Dies würde in einem nächsten Schritt das Mixen der entsprechenden Proben mit einer Probe bekannten Genotyps erfordern. Bei der Detektion von Einzelnukleotid-Polymorphismen wird die Sensitivität der Methodik insbesondere durch relativ kleine PCR-Fragmente gesteigert. Die Produkte zum anschließenden HRM sollten dabei nicht größer als 80-100 bp sein. Zudem scheint die Variante eines A zu G Basentausches den größten sichtbaren Effekt im HRM hervorzurufen (Reed und Wittwer,2004).

Neben den vielen Vorteilen existieren auch Faktoren, die eine Genotypisierung mittels HRM erschweren können. Ein Nachteil bei der Typisierung von Einzelnukleotid-Polymorphismen ist zum Beispiel, dass das HRM keine Multiplex-Methodik erlaubt. Das heißt, dass pro Durchgang meist nur ein Polymorphismus detektiert werden kann. Zudem zeigten Studien, dass ein 10% Risiko besteht, Proben anhand des Schmelzverhaltens im HRM falsch zu typisieren (Tindall et al,2009). Neben der Gefahr, dass unterschiedliche homozygote Proben sich in ihrem Schmelzkurvenverlauf nicht unterscheiden, besteht auch das Risiko unterschiedliche heterozygote Proben unter Umständen nicht eindeutig differenzieren zu können. Unterschiede im Schmelzverhalten von Proben mit dem gleichen Genotyp können zum Beispiel durch unterschiedliche chemische Zusammensetzungen der DNA-Proben zu Stande kommen. Typisierte Proben sollten daher in einer standardisierten Konzentration und in einem homogenen Puffer vorliegen. Ein weiterer wichtiger Faktor für eine effektive Analyse ist zudem die Auflösung des Instruments (Herrmann et al.,2007). Herrmann et al. untersuchten 2007 die Fähigkeit von verschiedenen zum HRM geeigneten Instrumenten auf ihre Sensitivität bezogen auf die Detektion von heterozygoten Proben anhand des Schmelzkurvenverlaufes. Während homozygote Proben in der Regel durch unterschiedliche Schmelztemperaturen ( $T_m$ ) unterschieden werden, lassen sich heterozygote Proben durch einen andersartigen Kurvenverlauf differenzieren (Zhou et al., 2005).

Die von uns analysierten Amplifikationsprodukte wiesen eine Größe von 54-120 bp auf. Durch den Einsatz kürzerer Fragmente wird die Auflösung des HRM verbessert (Slomka et al.,2017). Ein Einsatz von Fragmenten >200 bp birgt die Gefahr einer fehlerhaften Typisierung. Je größer das Amplifikationsprodukt ist, desto geringer ist der Einfluss eines einzelnen Basenaustausches auf die Schmelztemperatur des gesamten Genabschnittes (Liew et al,2007).

Die Typisierung von APOE erfolgte nach der Durchführung einer allelspezifischen PCR. Durch die Amplifizierung der entsprechenden Genabschnitte mit unterschiedlichen Primern, wurde die Schmelztemperatur der verschiedenen Genotypen zusätzlich verschoben. Dies erlaubte eine eindeutige optische Trennung der einzelnen Kurvenverläufe. Bei einer Größe der Fragmente von 54 – 72 bp konnte eine zuverlässige Typisierung der Genotypen erfolgen. Die Kurvenverläufe wurden anhand einer Sanger-Sequenzierung exemplarischer Proben validiert.

#### **5.4 Einzelnukleotid-Polymorphismen in CAST und ihr Einfluss auf die Erkrankung der SCA3**

In der aktuellen Literatur wird der Einfluss von Calpastatin (CAST) auf neurodegenerative Erkrankungen vermehrt diskutiert. Pankratz et al. fanden 2009 im Rahmen einer GWA-Studie von Patienten mit familiärem Parkinson einen Zusammenhang zwischen Einzelnukleotid-Polymorphismen in *CAST* und der neurodegenerativen Erkrankung Parkinson. Dieser Effekt wurde 2010 von Allen et al., für den Polymorphismus rs1559085 sogar als signifikant nachgewiesen. In einer unabhängigen Studie von Zhang et al. (2013) konnte diese Beobachtung in einer chinesischen Kohorte von 615 Probanden nicht bestätigt werden. Möglicherweise bedingt durch eine sehr geringe Häufigkeit des Polymorphismus rs1559085 konnte ein signifikanter Einfluss auf die Erkrankung nicht nachgewiesen werden, allerdings konnten auch Zhang et al. (2013) einen beeinflussenden Trend des Polymorphismus beobachten.

*CAST* kodiert für Calpastatin, ein Protein das  $Ca^{2+}$  abhängig die Calpain Protease hemmt. Der Wirkmechanismus von Calpastatin ist komplex und weiterhin nicht vollständig verstanden. Es konnte allerdings gezeigt werden, dass eine

pharmakologische Unterdrückung von Calpastatin eine Überaktivierung von Calpain nach sich zieht und dadurch den Zelltod von dopaminergen Neuronen induziert (Crocker et al.,2003). Es konnte zudem nachgewiesen werden, dass auch Ataxin-3 von Calpain proteolytische gespalten wird (Hübener et al.,2013). Die Spaltung von Ataxin-3 ist Voraussetzung für den nukleozytoplasmatischen Transport und die Formation von Aggregaten. Theoretisch verringert Calpastatin durch die Hemmung von Calpain die Ablagerung von Ataxin-Aggregaten in Neuronen (Haacke et al.,2007) und somit eine stärkere und schnellere Progression des neurologischen Phänotyps der SCA3 (Hübener et al.,2013). Auf Grundlage dieser Erkenntnisse deutet vieles darauf hin, dass Calpastatin einen vielversprechenden Faktor bei der Forschung nach therapeutischen Ansätzen für die SCA3 darstellt.

Bisher konnten keine Polymorphismen in *CAST* mit der SCA3 in Verbindung gebracht werden. Da bereits ein Zusammenhang zwischen dem Polymorphismus rs1559085 und einer neurodegenerativen Erkrankung beobachtet werden konnte, wollten wir einen Einfluss auf das Erkrankungsalter der SCA3 überprüfen. Als Intron Variante hat der Polymorphismus rs1559085 keinen direkten Effekt auf der Aminosäureebene des Proteins, ein funktioneller Einfluss des Basenaustausches ist jedoch nicht auszuschließen. Der Polymorphismus rs27852 stellt ebenfalls eine Intron Variante im Bereich des *CAST* Gens dar. Im Rahmen unserer Überlegungen, stellte der rs27852 ebenfalls einen vielversprechenden Kandidaten für die Überprüfung modifizierender Faktoren auf die SCA3 dar. Im Zusammenhang mit Parkinson wurde dieser Polymorphismus bereits untersucht, einen signifikanten modifizierenden Effekt konnte man dabei nicht nachweisen (Zhang et al.,2013).

#### **5.4.1 Die Etablierung von HRM für die Genotypisierung von Einzelnukleotid-Polymorphismen in *CAST***

In dieser Arbeit wurden die Einzelnukleotid-Polymorphismen in *CAST* mittels HRM anhand unterschiedlicher Schmelztemperaturen typisiert. Für den Polymorphismus rs1559085 gelang eine Etablierung der Methodik und eine Validierung der Ergebnisse mittels anschließender Sanger-Sequenzierung. Für

rs27852 konnten keine eindeutig unterschiedlichen Kurvenverläufe oder Schmelztemperaturen detektiert werden. Es besteht die Möglichkeit, dass weitere, bisher unbekannte Einzelnukleotid-Polymorphismen die Genotypisierung mittels HRM erschwert haben. Slomka et al, diskutierten 2017 verschiedene Problematiken, die eine Typisierung von DNA mittels HRM erschweren könnten. Darunter auch die Situation einer genetischen Variation im Bereich eines Primers oder weiteren Einzelnukleotid-Polymorphismen im Amplikon. Zudem betrug die Länge des DNA-Fragments zur Analyse des rs27852 150 bp. Eine Reduktion der Fragmentgröße hätte die Analyse ggf. erleichtern können. Im Rahmen der Etablierungsversuche erfolgte die genaue Analyse des Amplikons, dabei ergaben sich keine Hinweise auf bekannte Polymorphismen. Zudem erfolgte eine Analyse des Schmelzverhaltens mithilfe einer Umelt Animation. Mögliche weitere Störfaktoren der PCR stellt die chemikalische Zusammensetzung der Proben dar. Um eine Störung der Reaktion durch die Light-Cycler Chemikalien auszuschließen, versuchten wir die Reaktion mittels Taq-Polymerase und Resolight zu etablieren. Auch diese Etablierungsversuche scheiterten. Die Methode des HRM konnte für die Analyse von rs27852 nicht etabliert werden.

#### **5.4.2 Vereinbarung von Genotypen in CAST mit dem Hardy-Weinberg-Gleichgewicht**

Für den Polymorphismus rs1559085 konnte die Typisierung mittels HRM erfolgreich etabliert werden. Die Verteilung der beobachteten Genotypen wich dabei nur vereinzelt von der errechneten Verteilung in der Normalbevölkerung ab. Der Chi-Quadrat-Test mit einem p-Wert von 0,16 zeigte, dass keine signifikante Abweichung der Verteilung von Genotypen in einer Kohorte von SCA3-Betroffenen im Vergleich mit einer idealen Population nach Hardy-Weinberg vorliegt. Diese Beobachtung unterstützt die Annahme, dass die Typisierung mittels HRM erfolgreich und spezifisch durchgeführt werden konnte.

#### **5.4.3 Kein signifikanter Einfluss von CAST rs1559085 auf das Erkrankungsalter der SCA3**

Im Rahmen der statistischen Analyse konnte kein signifikanter Einfluss des Polymorphismus rs1559085 auf das Erkrankungsalter der SCA3 nachgewiesen werden. Im Rahmen einer Regressionsanalyse mit Geradenangleichung für die verschiedenen Genotypen deutete sich eine Verminderung des Erkrankungsalters bei Vorliegen eines G/G Genotyps zwar an, eine Überprüfung der statistischen Relevanz ergab allerdings einen p-Wert von 0,47. Unter Hinzunahme der individuellen Anzahl an CAG-Wiederholungen als weiterer Faktor der Regressionsanalyse verbesserte sich der p-Wert auf 0,25. Die tendenzielle Verbesserung des p-Wertes unterstreicht den starken Einfluss der Anzahl an CAG-Wiederholungen auf das Erkrankungsalter der SCA3. Der Aminosäureaustausch von Adenin zu Guanin scheint jedoch keine signifikante Beeinträchtigung der Funktionsweise von Calpastatin zu bewirken. Aufgrund der aktuellen Studienlage zur Funktionsweise von Calpastatin ist diese Beobachtung überraschend. In der aktuellen Literatur wird Calpastatin eng mit der Pathogenese der SCA3 in Verbindung gebracht (Haacke et al., 2007). Der von uns untersuchte Polymorphismus rs1559085 wurde in der Vergangenheit bereits bezüglich seines Einflusses auf die Pathogenese von Parkinson analysiert. Dabei wurde in einer Studie von Allen et al., 2010 ein signifikanter Zusammenhang zwischen dem Auftreten von Parkinson und dem Polymorphismus rs1559085 in CAST mittels Mantel-Haenszel-Test detektiert. Eine vorangegangene Studie von Pankratz et al., 2009 konnte diesen Effekt nicht darstellen. Beide Studien erfolgten nach dem Design einer „Genomewide Association Study (GWAS)“. Ein möglicher signifikanter Einfluss von rs1559085 auf die Pathogenese von Parkinson würde den Polymorphismus auch als einen vielversprechenden Modifikator der SCA3 charakterisieren. Der Zusammenhang zwischen dem Polymorphismus und SCA3 wurde bisher nicht überprüft. Um zu verstehen, warum rs1559085 wiedererwartend keinen relevanten Einfluss auf die SCA3 nimmt, scheint die funktionelle Analyse des Polymorphismus ein vielversprechender Gegenstand zukünftiger Forschung zu sein.

#### **5.4.4 Schwächen der Analyse von CAST rs1559085 im Zusammenhang mit der Kohorte an SCA3-Patienten**

In der untersuchten Kohorte konnte keine statistische Signifikanz des Einflusses von rs1559085 auf das Erkrankungsalter nachgewiesen werden. Dennoch zeigte sich in der Gruppe der typisierten Proben mit Genotyp G/G ein im Durchschnitt um 9 Jahre vermindertes Erkrankungsalter. Da dieser starke absolute Effekt nicht als signifikant eingestuft wurde, liegt am ehesten an der geringen Anzahl an Probanden mit Genotyp G/G. Von allen analysierten Proben wurden sechs als homozygot G/G typisiert. Dies entspricht der Verteilung nach dem Hardy-Weinberg-Gleichgewicht und der relativen Häufigkeit des Polymorphismus in der Normalbevölkerung. Um einen relevanten Einfluss der genetischen Variante auf die Erkrankung der SCA3 untersuchen zu können, sollte die Kohorte wesentlich größer gewählt werden und mindestens N=20 Probanden mit dem Genotyp G/G enthalten. Da Allen et al. (2010) einen signifikanten Zusammenhang von rs1559085 und der neurodegenerativen Erkrankung Parkinson in einer deutlich größeren Kohorte im Rahmen einer GWA – Studie nachweisen konnten, liegt ein möglicher modifizierender Einfluss von rs1559085 auf SCA3 nahe und sollte trotz der im Rahmen dieser Arbeit nicht nachweisbaren Effekte weiterhin Gegenstand zukünftiger Forschung bleiben.

## **6 Schlussfolgerung und Ausblick**

Im Rahmen dieser Arbeit konnte für drei Polymorphismen in zwei Genen die Methodik des High-Resolution-Melting zur Genotypisierung etabliert werden. Die untersuchten Einzelnukleotid-Polymorphismen konnten in der Normalbevölkerung (CEPH-DNA) und in 346 Patientenproben nachgewiesen werden. Im Allgemeinen wichen die ermittelten Allelfrequenzen dabei nicht signifikant von den bisher ermittelten in der aktuellen Literatur ab.

Da der Umfang einer Patientenkohorte in der Regel einen limitierenden Faktor in Bevölkerungsstudien darstellt, wurden im Rahmen dieser Arbeit die Ergebnisse von zwei unterschiedlichen Kohorten gemeinsam betrachtet. Die statistische Analyse erfolgte dabei unabhängig von der ethnischen Zugehörigkeit der Patienten. Der Großteil der Patientenproben stammte ohnehin aus einer Kohorte

aus Mitteleuropa, vereinzelt wurden aber auch Proben aus Peru in die statistische Analyse mit einbezogen. Mit 346 Patientenproben bildet unsere Kohorte ein deutlich größeres Kollektiv als die meisten vergleichbaren Studiendesigns. Durch die Analyse fast ausschließlich europäischer Proben kann ein populationsspezifischer Effekt jedoch nicht gänzlich ausgeschlossen werden. Die Ergebnisse dieser Arbeit müssten in einer weiteren, unabhängigen Kohorte bestätigt werden. Zudem erfolgte die Überprüfung des Einflusses der Polymorphismen ausschließlich an Betroffenen. Um eine relevante Modifikation durch den jeweiligen überprüften Polymorphismus nachweisen zu können, sollte ein Vergleich mit gesunden Kontrollprobanden erfolgen.

Die Anzahl der typisierten Proben unterscheidet sich für die Gene *APOE* und *CAST*. Für einige Proben konnte trotz wiederholter Analysen und Positivkontrollen kein eindeutiger Genotyp festgelegt werden. Da es sich in diesen Fällen meist um die gleichen Patientenproben handelte, ist dieser Zustand am ehesten auf eine mangelnde DNA-Qualität zurückzuführen. Die Patientenproben wurden teilweise bereits seit einigen Jahren bei  $-20^{\circ}\text{C}$  gelagert und wiederholt aufgetaut.

Im *APOE* Gen konnten zwei Einzelnukleotid-Polymorphismen typisiert werden. Beide konnten im Vergleich mit einer idealen Population nach dem Hardy-Weinberg-Gesetz signifikant häufiger bei SCA3 Betroffenen nachgewiesen werden. In der Allgemeinbevölkerung sind drei gängige *APOE* Isoformen bekannt. Diese werden durch die kombinierte Betrachtung der beiden Polymorphismen rs429358/rs7412 charakterisiert. Es gibt drei klinisch relevante Isoformen von *APOE* bedingt durch drei verschiedene Allele:  $\epsilon 2$ ,  $\epsilon 3$  und  $\epsilon 4$ . Das  $\epsilon 1$ -Allel gilt dabei als „missing Allel“ und kommt in der Regel nicht vor. In der Normalbevölkerung ist das  $\epsilon 3$  mit über 78% am weitesten verbreitet und das  $\epsilon 2$  mit einer Allelfrequenz von 7% das seltenste Allel. Die drei Isoformen unterscheiden sich dabei jeweils lediglich durch eine Aminosäure aufgrund von Einzelnukleotid-Polymorphismen in Codon 112 oder 158. In mehreren Studien wurde das  $\epsilon 4$  Allel bereits als wichtigster Risikofaktor für eine frühe Alzheimer Erkrankung charakterisiert (Liu et al., 2013, Strittmatter et al, 1993). Es existieren

mehrere mögliche Theorien zur funktionellen Beeinflussung neurodegenerativer Erkrankungen durch *APOE* Allele. Dabei konnte bisher keine eindeutig belegt werden. Nachgewiesen ist, dass *APOE* im menschlichen Gehirn proteolytisch fragmentiert wird (Muñoz et. al, 2019), daher kommt dem Verständnis der Fragmentierung von *APOE* eine besondere Rolle bei der Frage des möglichen funktionellen Einflusses der verschiedenen Genotypen auf die Erkrankung der SCA3 zu. Spezifische *APOE*  $\epsilon$ 4 Fragmente konnten beispielsweise in einen direkten Zusammenhang mit der Bildung von  $\beta$ Amyloid-Plaques und der Hyperphosphorylierung von Tau gebracht werde (Brecht et. al, 2004). Im Jahr 2011 konnten Bettencourt et al. einen vermindernden Effekt des *APOE*  $\epsilon$ 2 Allels auf das Erkrankungsalter der SCA3 nachweisen. Im Rahmen einer Studie von Raposo et. al (2017) konnte zudem eine Verminderung des Erkrankungsalters von 10 Jahren im Zusammenhang mit *APOE*  $\epsilon$ 2 bei gleichzeitigem Vorliegen einer IL6 Variante nachgewiesen werden. Diese Ergebnisse geben Anhalt, dass ein möglicher funktioneller Einfluss von *APOE*  $\epsilon$ 2 auf inflammatorische Zytokine bestehen könnte. Zhang et. al (2011) untersuchten diesen Zusammenhang bereits für einzelne Zytokine und konnten nachweisen, dass *APOE* die Sekretion von TNF und IL1B unterdrückt. *APOE*  $\epsilon$ 2 wurde dabei mit dem geringsten Sekretionslevel assoziiert. Eine mögliche funktionelle Beeinflussung des Erkrankungsalters durch die verschiedenen *APOE* Allele könnte über den Einfluss auf inflammatorische Zytokine zwar erklärt werden, grundsätzlich bleibt diese Frage jedoch bisher weitestgehend ungeklärt.

Die Verteilung der *APOE* Isoformen weicht in der Gruppe der von uns typisierten SCA3 Probanden grundsätzlich nicht signifikant von den bisher ermittelten Häufigkeiten in bekannten Studien ab (vgl. Zhou et al., Peng et al., 2014). Im Rahmen dieser Arbeit konnte ein statistisch signifikanter Einfluss der *APOE* Haplotypen auf das Erkrankungsalter nachgewiesen werden. Dabei erlangt das *APOE*  $\epsilon$ 2 Allel eine besondere Bedeutung. Wir konnten nachweisen, dass Betroffene mit einem *APOE*  $\epsilon$ 2 Allel in der analysierten Kohorte im Schnitt 2 Jahre früher erkrankten als Betroffene mit anderen Allelen. Zudem konnte mithilfe des individuellen *APOE* Allelstatus 5% mehr der Varianz des individuellen

Erkrankungsalters erklärt werden. Ein genauer Zusammenhang von *APOE* und der *SCA3* bleibt allerdings weiterhin unbekannt und sollte aufgrund dieser Beobachtungen weiterhin wichtiger Bestandteil zukünftiger Forschungen bleiben. Insbesondere eine funktionelle Analyse des Einflusses von unterschiedlichen Einzelnukleotid-Polymorphismen wie eine Bestätigung in einer Zellkultur, stellen einen vielversprechenden Gegenstand in der Erforschung des Pathomechanismus der *SCA3* dar.

Von den beiden anfänglich identifizierten Polymorphismen im *CAST* Gen, ließ sich nur für rs1559085 erfolgreich eine Methodik zur Genotypisierung etablieren. Im Rahmen dieser Arbeit konnte kein signifikanter Einfluss des Polymorphismus auf das Erkrankungsalter nachgewiesen werden. Es konnte lediglich ein Trend der Verminderung des Erkrankungsalters durch den Genotyp C/C beobachtet werden. Dieser betrug im Durchschnitt sogar 9 Jahre. Durch eine sehr kleine Stichprobe von 6 Probanden mit nachgewiesenem homozygotem C/C Genotyp blieb die Signifikanz der Beobachtung aus. Aufgrund der bisherigen Erkenntnisse über die Funktionsweise von Calpastatin und die theoretische Beteiligung an der Pathogenese der *SCA3* stellt der rs1559085 weiterhin einen vielversprechenden Ansatzpunkt für zukünftige Forschungen dar. In gezielten Studien zur Überprüfung des vermuteten Einflusses, sollte die Stichprobe an Probanden mit dem seltenen Genotyp C/C deutlich erhöht werden.

Die Faktoren, die in dieser Arbeit untersucht wurden, stellen potenzielle Einflussfaktoren der *SCA3* dar. Vermutlich existieren viele weitere genetische, epigenetische und Umweltfaktoren, die eine Erklärung der Varianz des Erkrankungsalters der *SCA3* ermöglichen. Einflussfaktoren auf das Erkrankungsalter zu identifizieren und charakterisieren ist auch in Zukunft entscheidend, um Erkenntnisse über den Pathomechanismus der Erkrankung zu gewinnen und mögliche Ansätze für zukünftige kurative Therapiemöglichkeiten zu erkennen.

## 7 Zusammenfassung

Die Spinocerebelläre Ataxie Typ 3 ist eine autosomal-dominant vererbare neurodegenerative Erkrankung und gehört zu der heterogenen Gruppe der autosomal-vererbaren Ataxien. Unter ihnen ist sie mit einem weltweit variablen Anteil von 15-45% die häufigste Form der Ataxie. Mit einer Prävalenz von weniger als 3 Erkrankten pro 100.000 Einwohnern gehört sie dennoch zu der Gruppe der Seltenen Erkrankungen. Auslöser der Erkrankung ist eine Verlängerung der CAG-Wiederholungs-Sequenz im *ATXN3*-Gen. Auf Proteinebene codiert diese Sequenz für Glutamin, dementsprechend entsteht bei einer Mutation eine verlängerte Polyglutaminsequenz im Ataxin-3. Bei einer Anzahl von >45 CAG-Wiederholungen kommt es laut aktueller Literatur zum klinischen Vollbild der Erkrankung. Die Erkrankung präsentiert sich in ihrer klinischen Symptomatik sehr unterschiedlich, zu den typischen Symptomtrias gehören dabei eine progrediente Gangataxie, eine externe Ophthalmoplegie sowie ein pathologischer Reflexstatus. In der klinischen Ausprägung wird die Erkrankung, abhängig vom dominierenden Symptom und dem Erkrankungsalter in fünf Subtypen unterteilt.

Der Großteil der Betroffenen erkrankt im mittleren Alter zwischen 20-50 Jahren. Dabei zeigen sich auch extreme Ausreißer des Erkrankungsalters von 5-70 Jahren. Eine Erklärung für diese große Varianz bietet die negative Korrelation des Erkrankungsalters mit der individuellen Anzahl an CAG-Wiederholungen im mutierten *ATXN3*-Gen. Je nach Größe der Kohorte ist diese Korrelation für 40-75% der Betroffenen eindeutig. Für einen Großteil der Mutationsträger ist eine alleinige Bestimmung der Anzahl an CAG-Wiederholungen demnach kein aussagekräftiges Kriterium, um den Zeitpunkt und Verlauf der Erkrankung vorauszusagen.

Diese Arbeit beschäftigt sich damit, weitere genetische Einflussfaktoren auf das Erkrankungsalter der SCA3 zu identifizieren und ihren tatsächlichen Einfluss statistisch zu überprüfen. Wir wählten dafür vier Einzelnukleotidpolymorphismen in zwei Kandidatengenen aus der Literatur aus. Mittels allelspezifischer PCR und anschließender Sanger-Sequenzierung konnten wir für zwei SNPs in APOE (rs429358 und rs7412) im High-Resolution-Melting charakteristische

Kurvenverläufe für die einzelnen Genotypen identifizieren und 346 Patientenproben analysieren. Mittels Schmelzkurvenanalyse konnten wir in der gleichen Kohorte einen SNP (rs1559085) in CAST typisieren. Die Etablierung einer Typisierungsmethode für einen zweiten SNP (rs27852) in CAST blieb leider erfolglos.

Die Ergebnisse der statistischen Auswertung zeigten, dass APOE einen signifikanten Einfluss auf das Erkrankungsalter der SCA3 haben kann. Dabei erwies sich das APOE  $\epsilon$ 2 Allel als größter Einflussfaktor. In einer ANOVA Analyse von Merkmalsträgern und Nicht-Merkmalsträgern konnte eine Verringerung des Erkrankungsalters von durchschnittlich 2 Jahren bei Vorliegen eines  $\epsilon$ 2 Allels festgestellt werden. Bei der individuellen Betrachtung der SNPs in APOE konnte gezeigt werden, dass durch die Analyse von rs429358 der prädiktive Wert für das Erkrankungsalter signifikant um zusätzliche 2% verbessert werden kann. Der rs7412 schien einzeln betrachtet keinen prädiktiven Mehrwert zu haben. Die Kombination beider Polymorphismen als Haplotyp verbessert den prädiktiven Wert für das Erkrankungsalter jedoch um 5%.

Die Analyse des SNP rs1559085 in CAST ergab mit einem p-Wert  $> 0,5$  keinen signifikanten Einfluss auf das Erkrankungsalter der SCA3. Allerdings konnte eine deutliche Verminderung des Erkrankungsalters bei Vorliegen eines homozygoten G/G Genotyp um durchschnittlich  $- 3,8$  Jahre im Vergleich mit dem häufigeren Genotyp A/A festgestellt werden. Ebenso zeigte eine Analyse der durchschnittlichen Anzahl an CAG-Wiederholungen in Abhängigkeit der Genotypen von rs1559085, dass bei Vorliegen eines G/G Genotyps die Anzahl der Wiederholungen im Schnitt um  $+ 3$  erhöht ist. Dieser Zusammenhang ist mit einem p-Wert von  $0,25$  zwar nicht signifikant, zeigt jedoch eine mögliche Tendenz eines Zusammenhangs zwischen dem Genotyp von rs1559085 in CAST und der Anzahl an CAG – Wiederholungen.

Mit 346 Patientenproben bildet die in dieser Arbeit analysierte Kohorte ein deutlich größeres Kollektiv ab als die meisten vergleichbaren Studiendesigns. Durch die Verwendung fast ausschließlich europäischer Proben darf jedoch auch ein populationsspezifischer Effekt nicht gänzlich ausgeschlossen werden. Die

Ergebnisse dieser Arbeit sollten in einer weiteren, unabhängigen Kohorte bestätigt werden. Außerdem erfolgte die Überprüfung des Einflusses der Polymorphismen ausschließlich an Betroffenen. Um eine relevante Modifikation durch den jeweiligen überprüften Polymorphismus definitiv nachweisen zu können, sollte ein Vergleich mit gesunden Kontrollprobanden erfolgen.

Unter Einbezug der gewonnenen Erkenntnisse über mögliche populationsspezifische Bias und Einflüsse der DNA-Qualität auf durchgeführte Analysen, legen die Ergebnisse dieser Doktorarbeit nahe, dass mit den verschiedenen Haplotypen von APOE weitere genetische Modifikatoren auf das Erkrankungsalter der SCA3 existieren. Unter Einbezug dieser genetischen Information könnte der Vorrausagewert über das Erkrankungsalter um 5% gesteigert werden. In einzelnen Kohorten könnte somit für bis zu 75% eine zuverlässige Voraussage über das Erkrankungsalter von Mutationsträgern getätigt werden. Auch wenn dabei im Bestfall weiterhin für  $\frac{1}{4}$  der Betroffenen keine verlässliche Aussage über den Zeitpunkt ihrer Erkrankung getätigt werden kann, so zeigen die methodischen Ansätze und Ergebnisse dieser Arbeit eine Möglichkeit genetische Einfluss der SCA3 zu identifizieren. Höchstwahrscheinlich existieren viele weitere genetische, epigenetische und Umweltfaktoren, die eine Erklärung der Varianz des Erkrankungsalters der SCA3 ermöglichen würden. Diese zu identifizieren und charakterisieren ist auch in Zukunft entscheidend, um Erkenntnisse über den Pathomechanismus der Erkrankung zu gewinnen und mögliche Ansätze für zukünftige kurative Therapiemöglichkeiten zu erkennen.

## Literaturverzeichnis

**Adkins S und Burmeister M (1996):** "Visualization of DNA in agarose gels as migrating colored bands: applications for preparative gels and educational demonstrations." *Anal Biochem.*; 240(1):17-23.

**Allen AS und Satten GA (2010):** "SNPs in CAST are associated with Parkinson disease: a confirmation study." *Am J Med Genet B Neuropsychiatr Genet.*; 153B(4):973-979

**Arnemann J (2019) a:** „Lexikon der Medizinischen Laboratoriumsdiagnostik“ Springer, Berlin: 2177

**Arnemann, J (2019) b:** "Fragmentlängenanalyse." *Lexikon der Medizinischen Laboratoriumsdiagnostik.* Springer, Berlin: 902-902.

**Bettencourt C, Raposo M, Kazachkova N, Cymbron T, Santos C, Kay T, Vasconcelos J, Maciel P, Donis KC, Saraiva-Pereira ML, Jardim LB, Sequeiros J, Lima M (2011):** "The APOE  $\epsilon$ 2 allele increases the risk of earlier age at onset in Machado-Joseph disease." *Arch Neurol.*;68(12):1580-1583.

**Boy J, Schmidt T, Wolburg H, Mack A, Nuber S, Böttcher M, Schmitt I, Holzmann C, Zimmermann F, Servadio A, Riess O (2009):** "Reversibility of symptoms in a conditional mouse model of spinocerebellar ataxia type 3." *Hum Mol Genet.*;18(22):4282-4295.

**Brecht WJ, Harris FM, Chang S (2004):** "Neuron-specific apolipoprotein e4 proteolysis is associated with increased tau phosphorylation in brains of transgenic mice." *J Neurosci Off J Soc Neurosci.*;24:2527–2534.

**Brookes AJ(1999):** "The essence of SNPs." *Gene.*;234(2):177-186.

**Buhmann C, Bussopulos A, Oechsner M (2003):** "Dopaminergic response in Parkinsonian phenotype of Machado-Joseph disease." *Movement Disorders*; 18(2):0885-3185

**Chen CH (2016):** "Development of a Melting Curve-Based Allele-Specific PCR of Apolipoprotein E (APOE) Genotyping Method for Genomic DNA, Guthrie Blood Spot, and Whole Blood." *PLoS One.*;11(4):e0153593

**Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993):** "Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families." *Science.*;261(5123):921-923.

**Costa Mdo C und Paulson HL (2012):** "Toward understanding Machado-Joseph disease." *Prog Neurobiol.*;97(2):239-257

**Coutinho P ,Andrade C (1978):** "Autosomal dominant system degeneration in Portuguese families of the Azores Islands. A new genetic disorder involving cerebellar, pyramidal, extrapyramidal and spinal cord motor functions." *Neurology*; 28(7), 703–709.

**Crawford FC, Vanderploeg RD, Freeman MJ, Singh S, Waisman M, Michaels L, Abdullah L, Warden D, Lipsky R, Salazar A, Mullan MJ (2002):** "APOE genotype influences acquisition and recall following traumatic brain injury." *Neurology*.;58(7):1115-1118.

**Crocker SJ, Smith PD, Jackson-Lewis V (2003):** "Inhibition of calpains prevents neuronal and behavioral deficits in an MPTP mouse model of Parkinson's disease." *J Neurosci*.;23(10):4081-4091.

**D'Abreu A, França M Jr, Appenzeller S, Lopes-Cendes I, Cendes F (2009):** "Axonal dysfunction in the deep white matter in Machado-Joseph disease." *J Neuroimaging*.;19(1):9-12.

**D'Abreu A, França M Jr, Yasuda CL, Campos BA, Lopes-Cendes I, Cendes F (2012):** "Neocortical atrophy in Machado-Joseph disease: a longitudinal neuroimaging study." *J Neuroimaging*.;22(3):285-291.

**Darawi MN., Ai-Vyrn C., Ramasamy K., Hua PP., Pin TM., Kamaruzzaman SB., Majeed AB (2013):** "Allele-specific polymerase chain reaction for the detection of Alzheimer's disease-related single nucleotide polymorphisms." *BMC medical genetics*;14(27)

**dbSNP,NIH <https://www.ncbi.nlm.nih.gov/snp>**

**de Mattos EP, Leotti VB, Soong BW, Raposo M, Lima M, Vasconcelos J, Fussiger H, Souza GN, Kersting N, Furtado GV, Saute JAM, Camey SA, Saraiva-Pereira ML, Jardim LB (2019):** "Age at onset prediction in spinocerebellar ataxia type 3 changes according to population of origin." *Eur J Neurol*.;26(1):113-120.

**de Oliveira MS, D'Abreu A, França MC Jr, Lopes-Cendes I, Cendes F, Castellano G (2012):** "MRI-texture analysis of corpus callosum, thalamus, putamen, and caudate in Machado-Joseph disease." *J Neuroimaging*.;22(1):46-52.

**Duarte-Silva S, Maciel P (2018):** "Pharmacological Therapies for Machado-Joseph Disease." *Adv Exp Med Biol*.;1049: 369-394.

**Etchebehere EC, Cendes F, Lopes-Cendes I, Pereira JA, Lima MC, Sansana CR, Silva CA, Camargo MF, Santos AO, Ramos CD, Camargo EE (2001):** "Brain single-photon emission computed tomography and magnetic resonance imaging in Machado-Joseph disease." *Arch Neurol*.;58(8):1257-1263.

**Eichler L, Bellenberg B, Hahn HK, Köster O, Schöls L, Lukas C (2011):** "Quantitative assessment of brain stem and cerebellar atrophy in spinocerebellar ataxia types 3 and 6: impact on clinical status." *AJNR Am J Neuroradiol.*;32(5):890-897.

**Ferreira CR. (2019):** "The burden of rare diseases." *American journal of medical genetics*;179(6): 885–892.

**Garibyan L, Avashia N (2013):** "Polymerase chain reaction." *J Invest Dermatol.*;13:1-4.

**Haacke A, Hartl FU, Breuer P (2007):** "Calpain inhibition is sufficient to suppress aggregation of polyglutamine-expanded ataxin-3." *Journal of Biological Chemistry*;282(26): 18851-18856.

**Hartl DL, and Clark AG (1997):** "Principles of population genetics.";116: 288

**Herrmann MG, Durtschi JD, Wittwer CT, Voelkerding KV (2007):** "Expanded instrument comparison of amplicon DNA melting analysis for mutation scanning and genotyping." *Clin Chem.*;53(8):1544-1548.

**Hübener J, Weber JJ, Richter C, Honold L, Weiss A, Murad F, Breuer P, Wüllner U, Bellstedt P, Paquet-Durand F, Takano J, Saido TC, Riess O, Nguyen HP (2013):** "Calpain-mediated ataxin-3 cleavage in the molecular pathogenesis of spinocerebellar ataxia type 3 (SCA3)." *Hum Mol Genet.*;22(3):508-518.

**Hopper, AG (1971):** "Chiquadrat-Test." *Einführung in die Statistische Qualitätskontrolle. Lehr- und Handbücher der Ingenieurwissenschaften, Birkhäuser, Basel*;30:152-153.

**Igarashi S, Takiyama Y, Cancel G, Rogaeva EA, Sasaki H, Wakisaka A, Zhou YX, Takano H, Endo K, Sanpei K, Oyake M, Tanaka H, Stevanin G, Abbas N, Dürr A, Rogaev EI, Sherrington R, Tsuda T, Ikeda M, Cassa E, Nishizawa M, Benomar A, Julien J, Weissenbach J, Wang GX, Agid Y, St George-Hyslop PH, Brice A, Tsuji S (1996):** "Intergenerational instability of the CAG repeat of the gene for Machado-Joseph disease (MJD1) is affected by the genotype of the normal chromosome: implications for the molecular mechanisms of the instability of the CAG repeat." *Hum Mol Genet.*;5(7):923-932.

**International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wray MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD,**

Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J (2007): "A second generation human haplotype map of over 3.1 million SNPs." *Nature*.;449(7164):851-861.

**Jack CR Jr**, Wiste HJ, Weigand SD, Knopman DS, Vemuri P, Mielke MM, Lowe V, Senjem ML, Gunter JL, Machulda MM, Gregg BE, Pankratz VS, Rocca WA, Petersen RC (2015): "Age, Sex, and APOE  $\epsilon$ 4 Effects on Memory, Brain Structure, and  $\beta$ -Amyloid Across the Adult Life Span." *JAMA Neurol*.;72(5):511-519.

**Kawai Y**, Takeda A, Abe Y, Washimi Y, Tanaka F, Sobue G (2004): "Cognitive impairments in Machado-Joseph disease." *Arch Neurol*.;61(11):1757-1760.

**Kawaguchi Y**, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I (1994): "CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1." *Nat Genet*.;8(3):221-228

**Kehoe P**, Krawczak M, Harper PS, Owen MJ, Jones AL (1999): "Age of onset in Huntington disease: sex specific influence of apolipoprotein E genotype and normal CAG repeat length." *J Med Genet*.;36(2):108-111.

**Kieling C**, Prestes PR, Saraiva-Pereira ML, Jardim LB (2007): "Survival estimates for patients with Machado-Joseph disease (SCA3)." *Clin Genet*.;72(6):543-545.

**Klockgether T**, Mariotti C, Paulson HL (2019): "Spinocerebellar ataxia." *Nat Rev Dis Primers*.11;5(1):24.

**Klockgether T, Skalej M, Wedekind D, Luft AR, Welte D, Schulz JB, Abele M, Bürk K, Laccone F, Brice A, Dichgans J (1998):** "Autosomal dominant cerebellar ataxia type I. MRI-based volumetry of posterior fossa structures and basal ganglia in spinocerebellar ataxia types 1, 2 and 3." *Brain.*;121 (9):1687-1693.

**Lay MJ and Wittwer CT (1997):** "Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR." *Clinical Chemistry*;43(12): 2262-2267.

**Lerer I, Merims D, Abeliovich D, Zlotogora J, Gadoth N (1996):** "Machado-Joseph disease: correlation between the clinical features, the CAG repeat length and homozygosity for the mutation." *Eur J Hum Genet.*;4(1):3-7.

**Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E, Wittwer CT (2004):** "Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons." *Clinical chemistry*;50(7):1156-1164.

**Liew M, Seipp M, Durtschi J, Margraf RL, Dames S, Erali M, Voelkerding K, Wittwer CT (2007):** "Closed-Tube SNP Genotyping Without Labeled Probes: A Comparison Between Unlabeled Probe and Amplicon Melting." *American Journal of Clinical Pathology*;127(3):341–348,

**Liu CC, Liu CC, Kanekiyo T, Xu H, Bu G (2013):** "Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy." *Nat Rev Neurol.*;9(2):106-118

**Løkkegaard T, Nielsen JE, Hasholt L, Fenger K, Werdelin L, Tranebjaerg L, Lauritzen M, Colding-Jørgensen E, Grønbech-Jensen M, Henriksen OA, Sørensen SA (1998):** „Machado-Joseph disease in three Scandinavian families." *J Neurol Sci.*;156(2):152-157.

**Maciel P, Gaspar C, DeStefano AL (1995):** "Correlation between CAG repeat length and clinical features in Machado-Joseph disease." *Am J Hum Genet.*;57(1):54-61.

**Mahley RW (1988):** "Apolipoprotein E: cholesterol transport protein with expanding role in cell biology." *Science.*;240(4852):622-630.

**Matos CA, de Almeida LP, Nóbrega C (2019):** "Machado-Joseph disease/spinocerebellar ataxia type 3: lessons from disease pathogenesis and clues into therapy." *J Neurochem.*;148(1):8-28.

**Martindale D, Hackam A, Wiczorek A, Ellerby L, Wellington C, McCutcheon K, Singaraja R, Kazemi-Esfarjani P, Devon R, Kim SU, Bredesen DE, Tufaro F, Hayden MR (1998):** "Length of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates." *Nat Genet.*;18(2):150-154.

**McLoughlin HS, Moore LR, Chopra R, Komlo R, McKenzie M, Blumenstein KG, Zhao H, Kordasiewicz HB, Shakkottai VG, Paulson HL (2018):** "Oligonucleotide therapy mitigates disease in spinocerebellar ataxia type 3 mice." *Ann Neurol.*;84(1):64-77.

**McLoughlin HS, Moore LR, Paulson HL (2019):** "Pathogenesis of SCA3 and implications for other polyglutamine diseases." *Neurobiol Dis.*;134:104635.

**Muñoz SS, Garner B, Ooi L (2019):** "Understanding the Role of ApoE Fragments in Alzheimer's Disease." *Neurochem.*; 44: 1297–1305.

**Murata Y, Yamaguchi S, Kawakami H, Imon Y, Maruyama H, Sakai T, Kazuta T, Ohtake T, Nishimura M, Saida T, Chiba S, Oh-i T, Nakamura S (1998):** "Characteristic magnetic resonance imaging findings in Machado-Joseph disease." *Arch Neurol.*;55(1):33-37.

**Nakano KK, Dawson DM, Spence A (1972):** "Machado disease. A hereditary ataxia in Portuguese emigrants to Massachusetts." *Neurology.*;22(1):49-55.

**Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C., Kalsheker N, Smith JC, Markham AF (1989):** "Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS)." *Nucleic acids research*; 17(7):2503–2516.

**Nguengang Wakap S, Lambert DM, Olry A (2020):** "Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database." *Eur J Hum Genet.*;28(2):165-173.

**Pankratz N, Wilk JB, Latourelle JC, DeStefano AL, Halter C, Pugh EW, Doheny KF, Gusella JF, Nichols WC, Foroud T, Myers RH,PSG-PROGENI and GenePD Investigators, Coordinators and Molecular Genetic Laboratories (2009):** "Genomewide association study for susceptibility genes contributing to familial Parkinson disease." *Human genetics*;124(6): 593–605.

**Paulson H, Shakkottai V (1998, updated 2020):** "Spinocerebellar Ataxia Type 3." In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, Amemiya A, editors. *GeneReviews*® [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2021.

**Paulson HL (2007):** "Dominantly inherited ataxias: lessons learned from Machado-Joseph disease/spinocerebellar ataxia type 3." *Semin Neurol*;27(2):133-142.

**Peng H, Wang C, Chen Z, Sun Z, Jiao B, Li K, Huang F, Hou X, Wang J, Shen L, Xia K, Tang B, Jiang H (2014):** "APOE  $\epsilon$ 2 allele may decrease the age at onset in patients with spinocerebellar ataxia type 3 or Machado-Joseph disease from the Chinese Han population." *Neurobiol Aging.*;35(9):e15-8.

**Petruska J, Goodman MF, Boosalis MS, Sowers LC, Cheong C, Tinoco I Jr (1988):** "Comparison between DNA melting thermodynamics and DNA polymerase fidelity." *Proc Natl Acad Sci U S A.*;85(17):6252-6256.

**Phillips, Michael C (2014):** "Apolipoprotein E isoforms and lipoprotein metabolism." *IUBMB Life*;66(9):616-623

**Ramani B, Panwar B, Moore LR, Wang B, Huang R, Guan Y, Paulson HL (2017):** "Comparison of spinocerebellar ataxia type 3 mouse models identifies early gain-of-function, cell-autonomous transcriptional changes in oligodendrocytes." *Hum Mol Genet.*;26(17):3362-3374.

**Ramesh R, Munshi A, Panda SK (1992):** "Polymerase chain reaction." *Natl Med J India.*;5(3):115-119.

**Rana AQ, Qureshi DT, Morshed M, Kachhvi ZM, Rana MA, Qureshi ARM (2016):** "Ophthalmological Features of Machado-Joseph Disease." *J Coll Physicians Surg Pak.*;26(11):137-139.

**Raposo M, Bettencourt C, Ramos A (2017):** "Promoter Variation and Expression Levels of Inflammatory Genes IL1A, IL1B, IL6 and TNF in Blood of Spinocerebellar Ataxia Type 3 (SCA3) Patients." *Neuromol Med.*; 19: 41–45.

**Reed GH, Wittwer CT (2004):** "Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis." *Clin Chem.*;50(10):1748-1754.

**Reed GH, Kent J, Wittwer CT (2007):** "High-resolution DNA melting analysis for simple and efficient molecular diagnostics." *Pharmacogenomics*; 8(6): 597–608.

**Riedel BC, Thompson PM, Brinton RD (2016):** "Age, APOE and sex: Triad of risk of Alzheimer's disease." *J Steroid Biochem Mol Biol.*;160:134-147.

**Riess O, Rüb U, Pastore A, Bauer P, Schöls L (2008):** "SCA3: neurological features, pathogenesis and animal models." *Cerebellum.*;7(2):125-37.

**Rubinsztein DC, Orr HT (2016):** "Diminishing return for mechanistic therapeutics with neurodegenerative disease duration: There may be a point in the course of a neurodegenerative condition where therapeutics targeting disease-causing mechanisms are futile." *Bioessays.*;38(10):977-980.

**Rüb U, de Vos RA, Schultz C, Brunt ER, Paulson H, Braak H (2002):** "Spinocerebellar ataxia type 3 (Machado-Joseph disease): severe destruction of the lateral reticular nucleus." *Brain.*;125(Pt 9):2115-2124.

**Saiki RK, Walsh PS, Levenson CH, Erlich HA (1989):** "Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes." *Proc Natl Acad Sci U S A.*;86(16):6230-6234.

**Sanger F, Nicklen S, Coulson AR (1977):** "DNA sequencing with chain-terminating inhibitors." *Proc Natl Acad Sci U S A.*;74(12):5463-5467.

**Schmitt I, Linden M, Khazneh H, Evert BO, Breuer P, Klockgether T, Wuellner U (2007):** "Inactivation of the mouse *Atxn3* (ataxin-3) gene increases protein ubiquitination", *Biochemical and Biophysical Research Communications*;362(3): 734-739,

**Schmitz-Hübsch T, du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, Giunti P, Globas C, Infante J, Kang JS, Kremer B, Mariotti C, Melegh B, Pandolfo M, Rakowicz M, Ribai P, Rola R, Schöls L, Szymanski S, van de Warrenburg BP, Dürr A, Klockgether T, Fancellu R (2006):** "Scale for the assessment and rating of ataxia: development of a new clinical scale." *Neurology.*;66(11):1717-1720.

**Shukla M, Rajgopal Y, Babu PP (2006):** "Activation of calpains, calpastatin and spectrin cleavage in the brain during the pathology of fatal murine cerebral malaria." *Neurochem Int.*;48(2):108-113

**Simko I (2016):** "High-Resolution DNA Melting Analysis in Plant Research." *Trends Plant Sci.*;21(6):528-537

**Simões, AT, Gonçalves N, Koeppen A, Déglon N, Kügler S, Duarte CB, Pereira de Almeida L (2012):** "Calpastatin-mediated inhibition of calpains in the mouse brain prevents mutant ataxin 3 proteolysis, nuclear localization and aggregation, relieving Machado-Joseph disease." *Brain : a journal of neurology*, 135(8), 2428–2439.

**Singh PP, Singh M, Mastana SS (2006):** "APOE distribution in world populations with new data from India and the UK.";33(3):279-308.

**Słomka M, Sobalska-Kwapis M, Wachulec M, Bartosz G, Strapagiel D (2017):** "High Resolution Melting (HRM) for High-Throughput Genotyping-Limitations and Caveats in Practical Case Studies." *Int J Mol Sci*;18(11):2316.

**Stern C (1943):** "THE HARDY-WEINBERG LAW." *Science.*;97(2510):137-138.

**Sullivan R, Yau WY, O'Connor E, Houlden H (2019):** "Spinocerebellar ataxia: an update." *J Neurol.*;266(2):533-544.

**Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, Shimazaki H, Soutome M, Endo K, Ohta S (1993):** "The gene for Machado-Joseph disease maps to human chromosome 14q." *Nature genetics*; 4(3):300–304.

**Takano J, Tomioka, M, Tsubuki S, Higuchi M, Iwata N, Itohara S, Maki M, Saido TC (2005):** "Calpain mediates excitotoxic DNA fragmentation via mitochondrial pathways in adult brains: evidence from calpastatin mutant mice." *The Journal of biological chemistry*;280(16), 16175–16184.

**Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, Forlani S, Rakowicz M, Schöls L, Mariotti C, van de Warrenburg BP, Orsi L, Giunti P, Filla A, Szymanski S, Klockgether T, Berciano J, Pandolfo M, Boesch S, Melegh B, Timmann D, Mandich P, Camuzat A, Clinical Research Consortium for Spinocerebellar Ataxia (CRC-SCA), EUROSCA network, Goto J, Ashizawa T, Cazeneuve C, Tsuji S, Pulst SM, Brusco A, Riess O, Brice A, Stevanin, G (2014):** "Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes." *Brain : a journal of neurology*;137(9): 2444–2455.

**Tindall EA, Petersen DC, Woodbridge P, Schipany K, Hayes VM (2009):** "Assessing high-resolution melt curve analysis for accurate detection of gene variants in complex DNA fragments." *Hum Mutat.*;30(6):876-883.

**Tuite PJ, Rogaeva EA, St George-Hyslop PH, Lang AE (1995):** "Dopa-responsive parkinsonism phenotype of Machado-Joseph disease: confirmation of 14q CAG expansion." *Ann Neurol.*;38(4):684-687.

**van de Warrenburg BP, Hendriks H, Dürr A, van Zuijlen MC, Stevanin G, Camuzat A, Sinke RJ, Brice A, Kremer BP (2005):** "Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch-French cohort." *Ann Neurol.*;57(4):505-512.

**Vergheze J, Holtzer R, Wang C, Katz MJ, Barzilai N, Lipton RB. (2013):** "Role of APOE Genotype in Gait Decline and Disability in Aging" *The Journals of Gerontology*;68(11):1395–1401

**Verpillat P, Camuzat A, Hannequin D, Thomas-Anterion C, Puel M, Belliard S, Dubois B, Didic M, Lacomblez L, Moreaud O, Golfier V, Campion D, Brice A, Clerget-Darpoux F(2002):** "Apolipoprotein E gene in frontotemporal dementia: an association study and meta-analysis." *Eur J Hum Genet.*;10(7):399-405.

**Wahlund S(1928):** "Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet." *Hereditas*; 11(1): 65-106.

**Wang, J, Chuang K, Ahluwalia M, Patel S, Umblas N, Mirel D, Germer S (2005):** "High-throughput SNP genotyping by single-tube PCR with Tm-shift primers." *Biotechniques*; 39(6), 885-893.

**Wang YG, Du J, Wang JL, Chen J, Chen C, Luo YY, Xiao ZQ, Jiang H, Yan XX, Xia K, Pan Q, Tang BS, Shen L (2009):** "Six cases of SCA3/MJD patients that mimic hereditary spastic paraplegia in clinic." *J Neurol Sci.*;285(1-2):121-124.

**Weiß C** (2018): „Lineare Regression.“ *Notfall+ Rettungsmedizin*, Springer, Berlin; 21(5):406-408.

**Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ** (2003): “High-resolution genotyping by amplicon melting analysis using LCGreen.” *Clinical chemistry*, 49(6):853–860

**Wüllner U, Reimold M, Abele M, Bürk K, Minnerop M, Dohmen BM, Machulla HJ, Bares R, Klockgether T** (2005): “Dopamine transporter positron emission tomography in spinocerebellar ataxias type 1, 2, 3, and 6.” *Arch Neurol.*;62(8):1280-1285.

**Yoshizawa T, Watanabe M, Frusho K, Shoji S** (2003): “Magnetic resonance imaging demonstrates differential atrophy of pontine base and tegmentum in Machado-Joseph disease.” *J Neurol Sci.*;215(1-2):45-50.

**Zarepari S, Camicioli R, Sexton G, Bird T, Swanson P, Kaye J, Nutt J, Payami H** (2002): “Age at onset of Parkinson disease and apolipoprotein E genotypes.” *Am J Med Genet.*;107(2):156-161.

**Zhang H, Wu LM, Wu J** (2011): “Cross-talk between apolipoprotein E and cytokines.” *Mediators of Inflammation*; 1–10.

**Zhang L, Ding H, Wang DH, Zhang YL, Baskys A, Chan P, Zhong Y, Cai YN** (2013): “Calpastatin gene (CAST) is not associated with late onset sporadic Parkinson's disease in the Han Chinese population.” *PLoS One.*;8(8):e70935.

**Zhao N, Liu CC, Qiao W, Bu G** (2018): “Apolipoprotein E, Receptors, and Modulation of Alzheimer's Disease.” *Biol Psychiatry.*;83(4):347-357.

**Zhao H, Yang L, Dong Y, Wu ZY** (2020): “Phenotypic variance in monozygotic twins with SCA3.” *Mol Genet Genomic Med.*;8(10):e1438.

**Zhou L, Wang L, Palais R, Pryor R, Wittwer CT** (2005): “High-resolution DNA melting analysis for simultaneous mutation scanning and genotyping in solution.” *Clin Chem.*;51(10):1770-1777.

**Zhou Q, Ni W, Dong Y, Wang N, Gan S-R, Wu ZY** (2014): “The Role of Apolipoprotein E as a Risk Factor for an Earlier Age at Onset for Machado-Joseph Disease Is Doubtful.” *PLoS ONE* 9(11): e111356.

## 8 Erklärung zum Eigenanteil

Die Arbeit wurde in dem Institut für Medizinische Genetik und angewandte Genomik unter Betreuung von Prof. Dr. med. Olaf Rieß und Prof. Dr. rer. nat. Thorsten Schmidt durchgeführt.

Die Konzeption der Studie erfolgte durch Prof. Dr. rer. nat. Thorsten Schmidt.

Sämtliche Versuche wurden nach Einarbeitung durch Labormitglieder von mir eigenständig durchgeführt.

Die statistische Auswertung erfolgte eigenständig durch mich.

Ich versichere, das Manuskript selbständig nach Anleitung durch Prof. Dr. rer. nat. Thorsten Schmidt und mit Unterstützung vom IZKF Promotionskolleg verfasst zu haben und keine weiteren als die von mir angegebenen Quellen verwendet zu haben.

Tübingen, den

---

(Rahel Maria Burger)

## **9 Veröffentlichungen**

**European Human Genetics Virtual Conference 2020:** E-Poster “Potential genetic modifiers of the inherited neurodegenerative disorder SCA3”

**International Congress for Ataxia Research 2022:** Poster “Association of genetic Polymorphism with the Age at Onset in Patients with Spinocerebellar Ataxia Type 3 (SCA3)/ Machado-Joseph-Disease (MJD)”

## **10 Danksagung**

Mein Dank gilt Herr Prof. Dr. med. Olaf Rieß für die Bereitstellung des Themas und der Möglichkeit der Promotion am Institut für medizinische Genetik und angewandte Genomik. Besonders danken möchte ich Thorsten Schmidt für die Konzipierung dieser Studie und die geduldige Unterstützung und Beratung während der Erarbeitung meiner Dissertation. Vielen Dank für die ehrliche und konstruktive Kritik, sowie die Möglichkeit bei jedem Anliegen auf ein offenes Ohr zu stoßen. Die herausragende Betreuung durch Thorsten Schmidt hat maßgeblich zum Abschluss dieser Arbeit beigetragen. Der Arbeitsgruppe der SCA3 danke ich für die herzliche Aufnahme und die tägliche Unterstützung.

Ein besonderer Dank gilt auch dem IZKF Promotionskolleg für die Ermöglichung sich einer anspruchsvollen, experimentellen Doktorarbeit zu widmen. Ich hatte die Möglichkeit Grundlagenforschung kennenzulernen und durch die Teilnahme an Kongressen die Dynamik dieses Forschungsfeldes zu erfahren. Durch den Austausch mit anderen medizinischen Doktoranden im experimentellen Forschungsfeld wurde es mir ermöglicht wichtige Kontakte zu knüpfen, die hoffentlich über die Dauer dieser Arbeit hinaus bestehen werden.