
BACHELORARBEIT

Frau
Toni Kalinsky

Die Rolle von 5'-UTR-Varianten bei hereditärem Mamma- und Ovarialkarzinom

**Untersuchung der 5'-UTRs der 14 Core Gene für
Familiären Brust- und Eierstockkrebs**

**Institut für Humangenetik Universitätsklinikum
Leipzig**

Mittweida, den 23.08.2024

BACHELORARBEIT

Die Rolle von 5'-UTR-Varianten bei hereditärem Mamma- und Ovarialkarzinom

**Untersuchung der 5'-UTRs der 14 Core Gene für
Familiären Brust- und Eierstockkrebs**

**Institut für Humangenetik Universitätsklinikum
Leipzig**

Autorin:

Toni Kalinsky

Studiengang:

Bachelor Biotechnologie

Seminargruppe:

BT21wM-B

Erstprüfer:

Prof. Dr. rer. nat. habil. Röbbbe Wünschiers

Zweitprüfer:

PD Dr. rer. nat. habil. Julia Hentschel

Mittweida, den 23.08.2024

*Science knows no country, because knowledge belongs to humanity, and is the torch
which illuminates the world.*

Louis Pasteur (1822 - 1895)

Bibliografische Angaben

Kalinsky : Die Rolle von 5'-UTR-Varianten bei hereditärem Mamma- und Ovarialkarzinom, Untersuchung der 5'-UTRs der 14 Core Gene für Familiären Brust- und Eierstockkrebs Institut für Humangenetik Universitätsklinikum Leipzig, 58 Seiten, 23 Abbildungen, Hochschule Mittweida, University of Applied Sciences, Fakultät Angewandte Computer- und Biowissenschaften

Englischer Titel: *The role of 5'-UTR variants in hereditary breast and ovarian cancer: Analysis of the 5'-UTRs of the 14 core genes for hereditary breast and ovarian cancer*

Bachelorarbeit, Mittweida, den 23.08.2024

Satz: L^AT_EX

Kurzbeschreibung

Familiärer Brust- und Eierstockkrebs (FBREK) bezeichnet eine genetisch bedingte Prädisposition zur Entwicklung ebendieser Krankheiten, oftmals ausgelöst durch pathogene Veränderungen in bestimmten Genen. In der vorliegenden Bachelorarbeit werden die 5'-UTRs von 14 Genen, welche für FBREK relevant sind, nach genetischen Varianten untersucht. Die Sequenzdaten basieren auf Panel-Sequenzierung sowie auf Sanger-Sequenzierung. Von 70 identifizierten Varianten wurden 44 aufgrund von zu geringen Allelfrequenzen und Abdeckungen von der weiteren Bewertung ausgeschlossen. Es wurden 5 Varianten identifiziert, die benigne sind und 21 Varianten, welche als Varianten unklarer Signifikanz (VUS) eingestuft wurden und womöglich eine negative Auswirkung auf die Genexpression haben.

Abstract

Hereditary Breast and Ovarian Cancer (HBOC) refers to a genetic predisposition to the development of these diseases, often triggered by pathogenic variants in specific genes. In this bachelor thesis, the 5'-UTRs of 14 genes relevant for FBREK are analysed for genetic variants. The sequence data is based on panel sequencing and Sanger sequencing. Of 70 identified variants, 44 were excluded from the further evaluation due to insufficient allele frequencies and coverage. 5 variants were identified that are benign and 21 variants that were categorised as variants of uncertain significance (VUS) and may have a negative effect on gene expression.

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	III
Tabellenverzeichnis	V
Abkürzungsverzeichnis	VI
Danksagung	
1 Einleitung	1
1.1 Genetische Grundlagen und Aufbau der 5'-UTR	4
1.2 Kohorte	7
1.3 Variantenklassifikation nach Richards <i>et al.</i> , 2015	9
1.4 Regulatorische Funktionen der 5'-UTR in der Genexpression	10
1.5 Pathologische Relevanz der 5'-UTR und von 5'-UTR-Varianten	12
2 Zielstellung	14
3 Materialien, Chemikalien und Geräte	15
3.1 Probenmaterial	15
3.2 Molekularbiologische Reagenzien	15
3.3 Chemikalien	17
3.4 Lösungen	17
3.5 Materialien und sonstige Laborausstattung	18
3.6 Geräte	18
3.7 Software und Datenbanken	19
3.8 PCR-Ansätze für die Etablierung der Sanger-Primer	21
3.9 Reaktionsansätze für die Aufreinigung und Sequenzierreaktion	22
4 Methoden	23
4.1 Grundlegende Methoden und Prinzipien	23
4.2 Molekulargenetische Methoden	25
4.2.1 Design der Sanger-Primer	25
4.2.2 Etablierung der Sanger-Primer und Sanger-Sequenzierung	25

4.3	Auswertung der Sangerdaten mittels Software	29
4.4	Vorgehen der Variantenklassifikation	29
5	Ergebnisse und Diskussion	32
5.1	Etablierung der Sanger-Primer und Sanger-Sequenzierung	32
5.2	Genetische Varianten aus Panel-Diagnostik und Sanger-Sequenzierung in den 5'- UTRs der Core Gene für FBREK	36
5.3	Polygenic Risk Score	51
6	Fazit	54
7	Ausblick.....	55
8	Zusammenfassung	57
9	Summary	58
	Literaturverzeichnis	IX
	Anhang	i
A	Ausgewählte Sequenzabschnitte der Sanger-Sequenzierungsdaten	i
B	Kontrolle der PCR-Produkte mittels Gelelektrophorese.....	iii
C	Zusätzliche Daten zu den identifizierten 5'-UTR-Varianten.....	vii

Abbildungsverzeichnis

Abbildung 1.1 Übersicht über die Funktion der Genprodukte der Core Gene für FBREK (erstellt mit BioRender.com).	2
Abbildung 1.2 Aufbau einer eukaryotischen mRNA mit Fokus auf der 5'-UTR (erstellt mit BioRender.com).	5
Abbildung 1.3 5'-UTRs der Core Gene für FBREK (erstellt mit BioRender.com).	6
Abbildung 1.4 Überblick über die FBREK-Kohorte (erstellt mit BioRender.com).	8
Abbildung 4.1 Überblick über das Varianten-Calling und die Klassifikation (erstellt mit Bio-Render.com).	31
Abbildung 5.1 Etablierung der Sanger-Primer mit dem PCR-Programm Touchdown.	32
Abbildung 5.2 Etablierung der Sanger-Primer mit dem PCR-Programm Touchdown HQ.	34
Abbildung 5.3 Etablierung der Sanger-Primer mit dem PCR-Programm Touchdown HQ für <i>ATM</i>	35
Abbildung 5.4 Verteilung der identifizierten Varianten auf die 5'-UTRs der Core Gene für FBREK.	36
Abbildung 5.5 Konservierung einzelnder Sequenzbereiche am Beispiel <i>PTEN</i>	43
Abbildung 5.6 Brustkrebsrisiko für die <i>ATM</i> -Variante c.-111G>A im Verbund eines Haplo- typs nach Yue <i>et al.</i> (2018) (erstellt mit BioRender.com).	46
Abbildung 5.7 Quantitative Verteilung der Varianten im Laufe der Klassifikation.	50
Abbildung 5.8 Boxplot der PRS-Werte der FBREK-Kohorte.	52
Abbildung A.1 Unsauberer Sequenzabschnitt der Sanger-Sequenzierungsdaten.	i
Abbildung A.2 Homozygote Variante c.-205G>A in der 5'-UTR von <i>BRIP1</i> aus der Sanger- Sequenzierung.	ii
Abbildung A.3 Heterozygote Variante c.-205G>A in der 5'-UTR von <i>BRIP1</i> aus der Sanger- Sequenzierung.	ii
Abbildung B.1 Kontrolle der PCR-Produkte der 5'-UTR von <i>ATM</i>	iii
Abbildung B.2 Kontrolle der PCR-Produkte der 5'-UTR von <i>BRIP1</i>	iii
Abbildung B.3 Kontrolle der PCR-Produkte der 5'-UTR von <i>CHEK2</i>	iv

Abbildung B.4 Kontrolle der PCR-Produkte der 5'-UTR von <i>SMARCA4</i>	iv
Abbildung B.5 Kontrolle der PCR-Produkte der 5'-UTR von <i>STK11</i> (Primer A).	v
Abbildung B.6 Kontrolle der PCR-Produkte der 5'-UTR von <i>STK11</i> (Primer B).	vi
Abbildung B.7 Kontrolle der PCR-Produkte der 5'-UTR von <i>STK11</i> (Primer C).	vi

Tabellenverzeichnis

Tabelle 1.1	Transkripte der 14 Core Gene.	7
Tabelle 3.1	Touchdown Reaktionsansatz für die PCR.	21
Tabelle 3.2	Touchdown High Quality (HQ) Reaktionsansatz für die PCR.	21
Tabelle 3.3	Touchdown HQ Reaktionsansatz für die PCR des Primerpaares von <i>ATM</i>	21
Tabelle 3.4	Reaktionsansatz zur Aufreinigung des PCR-Produktes.	22
Tabelle 3.5	Reaktionsansatz für die Sequenzierreaktion des PCR-Produktes.	22
Tabelle 4.1	Touchdown PCR-Programm.	26
Tabelle 4.2	Touchdown HQ PCR-Programm.	27
Tabelle 4.3	Programm zur Aufreinigung des PCR-Produktes.	28
Tabelle 4.4	Sequenzier-Programm für das aufgereinigte PCR-Produkt.	28
Tabelle 5.1	Überblick über die ausgeschlossenen Varianten.	37
Tabelle 5.2	Überblick über die Varianten mit einer benignen Klassifikation Teil I.	38
Tabelle 5.3	Überblick über die Varianten mit einer benignen Klassifikation Teil II.	38
Tabelle 5.4	Überblick über die als VUS eingestufteten Varianten Teil I.	40
Tabelle 5.5	Überblick über die als VUS eingestufteten Varianten Teil II.	40
Tabelle 5.6	Nicht-AUG-Startcodons in den als VUS eingestufteten Varianten.	42
Tabelle C.1	Genomische gnomAD-Daten zu den ausgeschlossenen Varianten.	vii
Tabelle C.2	ClinVar IDs und rs-Nummern der ausgeschlossenen Varianten.	viii
Tabelle C.3	PubMed-IDs der benignen Varianten.	ix
Tabelle C.4	ClinVar IDs und rs-Nummern der benignen Varianten.	ix
Tabelle C.5	PubMed-IDs der als VUS eingestufteten Varianten.	x
Tabelle C.6	ClinVar IDs und rs-Nummern der als VUS eingestufteten Varianten.	xi

Abkürzungsverzeichnis

3'-UTR	<i>three prime untranslated region</i> (3'-untranslatierte Region)
5'-UTR	<i>five prime untranslated region</i> (5'-untranslatierte Region)
A, C, G, T/U	Adenin, Cytosin, Guanin, Thymin/Uracil
AF	Allelfrequenz
ALT	alternative Base(n)
ATM; ATM	<i>Ataxia Telangiectasia Mutated</i> ; Serine Protein Kinase ATM
ATP	Adenosintriphosphat
AUG	Adenin, Uracil, Guanin $\hat{=}$ Startcodon bei Eukaryoten
BA	<i>benign standalone</i> (benigne alleinstehend)
BARD1; BARD1	...	<i>BRCA1-Associated RING Domain Gene 1</i> ; BRCA1-Associated RING Domain Protein 1
bp	<i>base pairs</i> (Basenpaare)
BP	<i>benign supporting</i> (unterstützend benigne)
BRCA1; BRCA1	...	<i>Breast Cancer 1, early onset</i> ; Breast Cancer Type 1 Susceptibility Protein
BRCA2; BRCA2	...	<i>Breast Cancer 2, early onset</i> ; Breast Cancer Type 2 Susceptibility Protein
BRIP1; BRIP1	<i>BRCA1 Interacting Helicase 1</i> ; BRCA1-Associated C-terminal Helicase 1
BS	<i>benign strong</i> (stark benigne)
CADD	<i>Combined Annotation Dependent Depletion</i>
CDH1; CDH1	<i>Cadherin 1</i> ; Cadherin 1
CDS	<i>coding sequence</i> (kodierende Sequenz)
CHEK2; CHEK2	...	<i>Checkpoint Kinase 2</i> ; Serine/Threonine Protein Kinase Chk2
DCIS	duktales Karzinom in situ
ddNTP	Didesoxyribonukleosidtriphosphat
DLG1; DLG1	<i>Discs Large Homolog 1</i> ; Discs Large Homolog 1
DNA	<i>deoxyribonucleic acid</i> (Desoxyribonukleinsäure)
dNTP	Desoxynukleosidtriphosphat
FBREK	Familiärer Brust- und Eierstockkrebs
F + R Primer	<i>forward and reverse primer</i> (Vorwärts- und Rückwärtsprimer)

gnomAD	<i>Genome Aggregation Database</i>
Grpmax AF	<i>Group Maximum Allele Frequency</i> (Gruppe der maximalen Allelfrequenz)
hg38	<i>human genome 38</i> (humanes Genom 38)
HGMD	<i>Human Gene Mutation Database</i>
HGVS	<i>Human Genome Variation Society</i>
HQ	<i>High Quality</i> (hohe Qualität)
IDC	invasives duktales Karzinom
IRES	<i>internal ribosomal entry site</i> (interne ribosomale Eintrittsstelle)
LED	<i>Light Emitting Diode</i> (lichtemittierende Diode)
MANE Select	<i>Matched Annotation from the National Center for Biotechnology Information and European Bioinformatics Institute</i>
miRNA	<i>micro ribonucleic acid</i> (Mikro-Ribonukleinsäure)
mRNA	<i>messenger ribonucleic acid</i> (Boten-Ribonukleinsäure)
N	Adenin, Guanin, Cytosin oder Thymin/Uracil
NGS	<i>Next Generation Sequencing</i>
NST	invasives Mammakarzinom nicht spezifischen Typs
ORF	<i>open reading frame</i> (offener Leserahmen)
<i>PALB2</i> ; PALB2	<i>Partner And Localizer of BRCA2</i> ; Partner And Localizer of BRCA2
PARB-Inhibitoren ..	Poly(Adenosindiphosphat-Ribose)-Polymerase-Inhibitoren
PBS	Proteinbindestelle
PCR	<i>polymerase chain reaction</i> (Polymerase-Kettenreaktion)
pH	pondus hydrogenii
PM	<i>pathogenic moderate</i> (moderat pathogen)
PP	<i>pathogenic supporting</i> (unterstützend pathogen)
prä-mRNA	<i>precursor messenger ribonucleic acid</i> (prä-Boten-Ribonukleinsäure)
PRS	<i>Polygenic Risk Score</i>
PS	<i>pathogenic strong</i> (stark pathogen)
<i>PTEN</i> ; PTEN	<i>Phosphatase and Tensin Homolog</i> ; Phosphatase and Tensin Homolog
PVS	<i>pathogenic very strong</i> (sehr stark pathogen)
R	Purinbase Adenin oder Guanin

RBP	Ribonukleinsäure-bindende Proteine
REF	Referenzbase(n)
<i>RAD51C</i> ; RAD51C	<i>DNA Repair Protein RAD51 Homolog 3</i> ; DNA Repair Protein RAD51 Homolog 3
<i>RAD51D</i> ; RAD51D	<i>DNA Repair Protein RAD51 Homolog 4</i> ; DNA Repair Protein RAD51 Homolog 4
RNA	<i>ribonucleic acid</i> (Ribonukleinsäure)
rpm	<i>revolutions per minute</i> (Umdrehungen pro Minute)
SAP	<i>Shrimp Alkaline Phosphatase</i> (alkalische Phosphatase der Garnele)
S	Svedberg
<i>SMARCA4</i> ; SMARCA4	<i>SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4</i> ; ATP-Dependent Chromatin Remodeler SMARCA4
SNP	<i>single nucleotide polymorphism</i> (Einzelnukleotidpolymorphismus)
SNV	<i>single nucleotide variant</i> (Einzelnukleotidvariante)
<i>STK11</i> ; STK11	<i>Serine/Threonine Kinase 11</i> ; Serine/Threonine Protein Kinase STK11
TBE-Puffer	Tris(hydroxymethyl)aminomethan-Borat-Ethylendiamintetraessigsäure-Puffer
<i>TP53</i> ; TP53	<i>Tumor Protein P53</i> ; Cellular Tumor Antigen p53
UCSC	<i>University of California Santa Cruz</i> (Universität Kalifornien Santa Cruz)
uORF	<i>upstream open reading frame</i> (stromaufwärts gelegener offener Leserahmen)
VUS	<i>variant of uncertain significance</i> (Variante unklarer Signifikanz)

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei all denjenigen bedanken, die mich während der Entstehung meiner Bachelorarbeit unterstützt haben.

Ein besonderer Dank geht an Dr. habil. Julia Hentschel und Dr. Susanna Schubert aus der Humangenetik in Leipzig, unter deren Leitung ich meine Bachelorarbeit anfertigen durfte. Ich wurde fachlich jederzeit kompetent unterstützt und durfte eine sehr angenehme Arbeitsatmosphäre kennenlernen.

Zudem danke ich dem gesamten Team der Molekulargenetik, das mir jederzeit mit Rat und Tat zur Seite stand und mich geduldig in meine Tätigkeiten im Labor eingearbeitet hat. Eure Unterstützung und die gemeinsame Zeit schätze ich sehr.

Ein herzliches Dankeschön geht ebenfalls an Dr. Stephan Drukewitz von der Bioinformatik, der mir bei der Datenbereitstellung geholfen hat.

Ein großes Dankeschön geht an Prof. Dr. Röbbbe Wünschiers für die Betreuung meiner Bachelorarbeit seitens der Hochschule Mittweida. Auch allen Dozenten und Dozentinnen, die mich während meines Studiums begleitet haben, möchte ich danken.

Abschließend möchte ich meiner Familie und meinen Freunden danken, die mir mit ihren unterstützenden Worten und ihrer Ermutigung die nötige Kraft gegeben haben. Es ist schön zu wissen, dass ich mich zu jeder Zeit auf Eure Unterstützung verlassen kann.

1 Einleitung

Bezüglich der Entstehung eines Tumors geht man grundsätzlich von der, nach Knudson (1971) benannten, *two-hit*-Hypothese aus. Laut dieser Hypothese sorgt eine initiale genetische Veränderung (*first hit*) des ersten Allels, etwa eines Tumorsuppressorgens, noch nicht für die Krebsentstehung. Erst die Inaktivierung des zweiten Allels (*second hit*) setzt den Prozess der Tumorentstehung in Gang [Knudson, 1971]. Bei erworbenen Tumoren erfolgen die beiden genetischen Ereignisse im Laufe des Lebens durch den Einfluss von Umweltfaktoren. Bei hereditär bedingten Tumoren hingegen wird der *first hit* bereits von den Eltern vererbt, wodurch die krankheitsrelevante Veränderung in allen somatischen Zellen sowie den Keimzellen vorliegt [Knudson, 1971]. Der erworbene *second hit* im Laufe des Lebens begünstigt schlussendlich die Tumorentstehung. So zeigt sich in der komplexen Vererbung genetischer Veränderungen, wie spezifische pathogene Varianten das Risiko für die Entstehung familiärer Tumorerkrankungen signifikant erhöhen können. Familiärer Brust- und Eierstockkrebs (FBREK) stellt eine erhebliche gesundheitliche Herausforderung für die Betroffenen dar. Dabei begünstigen pathogene Varianten in den sogenannten Core Genen das Risiko für diese Tumorarten mit einem jungen Ersterkrankungsalter. Als Core Gene bezeichnet man die zentralen Gene, welche oft mit FBREK-Erkrankungen im Zusammenhang stehen.

Um die Diagnostik- und Behandlungsmöglichkeiten zu präzisieren, widmet sich das Deutsche Konsortium für FBREK der Erforschung dieser erblichen Krebsformen. Aktuell (Stand 2022) umfasst die Konsensusempfehlung des Konsortiums die Gendiagnostik für 14 Core Gene [URL-1, 2024], deren Genprodukte in Abbildung 1.1 gezeigt sind.

Tumorsuppressoren sind Proteine, welche die Tumorentstehung durch die Regulation des Zellzyklus und damit der Zellproliferation verhindern [Schmidt, 2023]. Unter den Core Genen für FBREK weisen verschiedene Gene die Funktion als Tumorsuppressoren auf. Darunter fällt *Breast Cancer 1, early onset (BRCA1)*, welches für das Breast Cancer Type 1 Susceptibility Protein kodiert. Auch das Gen *Phosphatase And Tensin Homolog (PTEN)* kodiert für den gleichnamigen Tumorsuppressor, welcher zelluläre Signalwege reguliert und Einflüsse auf das Zellwachstum sowie die Differenzierung nimmt [Liu *et al.*, 2022]. Die Serine/Threonine Protein Kinase STK11, die vom Gen *Serine/Threonine Kinase 11 (STK11)* kodiert wird, sowie das Cellular Tumor Antigen p53, welches vom Gen *Tumor Protein P53 (TP53)* kodiert wird, sind ebenfalls als Tumorsuppressoren tätig. STK11 reguliert dabei den Zellstoffwechsel, die Zellteilung sowie die Zelldifferenzierung [S Nakamura, Aoki und Miki, 2021] und TP53 stoppt den Zellzyklus

und steuert die Apoptose bei gravierenden DNA-Schäden [Aubrey, Strasser und Kelly, 2016].

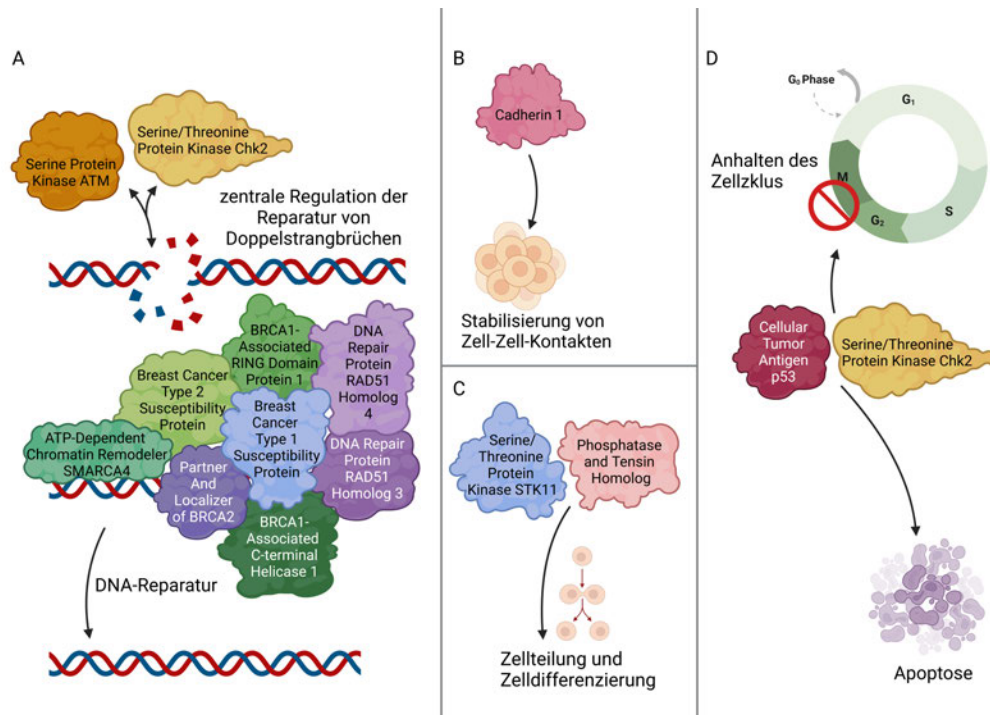


Abbildung 1.1: Übersicht über die Funktion der Genprodukte der Core Gene für FBREK (erstellt mit BioRender.com).

Abbildung 1.1 gibt einen Überblick über das Zusammenwirken der Proteine, welche von den 14 Core Genen codiert werden. Die Hauptfunktion besteht in der homologen Rekombinationsreparatur der DNA [S Nakamura, Aoki und Miki, 2021] (A). Außerdem spielen einige Proteine wichtige Rollen bei der Stabilisierung von Zell-Zell-Kontakten (B), der Zellteilung (C), sowie in der Tumorsuppression zur Verhinderung abnormaler Zellproliferation (D). Auch die Zellzyklusregulation zur Verhinderung der Ausbreitung von geschädigtem Erbmaterail wird übernommen (D).

Durch verschiedenste Ursachen können Doppelstrangbrüche in der DNA auftreten. Zu nennen sind hier beispielsweise die Einflüsse von Strahlung [Clark und Pazdernik, 2009], chemischen Substanzen und mechanischen Kräften sowie enzymatische Einwirkungen [Schmidt, 2023]. Zur Beseitigung von Doppelstrangbrüchen wird in den Zellen die homologe Rekombinationsreparatur eingesetzt. Dafür ist ein intakter Doppelstrang des Schwester-Chromatids notwendig, welcher als Vorlage für die Reparatur dient [Graw, 2020]. Ein Großteil der Genprodukte der, für FBREK relevanten, Core Gene ist bei der homologen Rekombination wichtig. Die Serine Protein Kinase ATM, kodiert durch das Gen *Ataxia Telangiectasia Mutated (ATM)*, dient als Sensor für DNA-Schäden und aktiviert Reparatursignalwege [Clark und Pazdernik, 2009]. Ebenfalls an der Erkennung von DNA-Schäden beteiligt ist die Serine/Threonine Protein Kinase Chk2, welche von

dem Gen *Checkpoint Kinase 2 (CHEK2)* kodiert wird. Dieses Protein stoppt den Zellzyklus für die anstehenden Reparaturprozesse [Bartek, Falck und Lukas, 2001]. Durch die Aktivierung der DNA-Reparatur wird das Protein BRCA1 phosphoryliert [Clark und Pazdernik, 2009]. Ausgehend von dieser Phosphorylierung bindet das Breast Cancer Type 2 Susceptibility Protein, dessen genetische Informationen vom Gen *Breast Cancer 2, early onset (BRCA2)* kodiert werden an BRCA1. BRCA1 interagiert während der DNA-Reparatur außerdem mit den Proteinen DNA Repair Protein RAD51 Homolog 3 sowie Homolog 4, die durch die gleichnamigen Gene (*RAD51C* und *RAD51D*) kodiert werden. Das Protein Partner And Localizer of BRCA2 (*Partner And Localizer of BRCA2 (PALB2)*) dient als Bindeglied zwischen BRCA1 und BRCA2 [Sy, Huen und J Chen, 2009]. Außerdem wird BRCA2 von PALB2 an die Reparaturstelle rekrutiert [Xia *et al.*, 2006]. Mit dem Protein BRCA1 interagieren außerdem die Proteine BRCA1-Associated RING Domain Protein 1 (*BRCA1-Associated RING Domain Gene 1 (BARD1)*) und die BRCA1-Associated C-terminal Helicase 1 (*BRCA1 interacting helicase 1 (BRIP1)*) während der DNA-Reparatur [S Nakamura, Aoki und Miki, 2021]. Auch das Protein Adenosintriphosphat(ATP)-Dependent Chromatin Remodeler SMARCA4, welches durch das Gen *SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4 (SMARCA4)* kodiert wird, kann an BRCA1 binden. SMARCA4 erleichtert den Zugang zu beschädigten DNA-Stellen und rekrutiert viele der Reparaturproteine [Harrod, Lane und Downs, 2020]. Neben den für FBREK relevanten sind viele weitere Proteine an der homologen Rekombinationsreparatur beteiligt.

Das Protein Cadherin 1 (*Cadherin 1 (CDH1)*) vermittelt die Zell-Zell-Adhäsion und dient der Unterdrückung von Tumorwachstum. Demnach kommt es bei Funktionsverlusten dieses Proteins zu verstärkter Proliferation und Metastasierung [Stockinger *et al.*, 2001]. Außerdem an der Unterdrückung des Zellwachstums beteiligt ist das Protein TP53 [Doherty *et al.*, 2019].

Zwischen 5 und 10 % der Brustkrebsfälle lassen sich auf erbliche Faktoren zurückführen [S Nakamura, Aoki und Miki, 2021]. Ein erhöhtes Risiko für ein Mammakarzinom geht mit allen Core Genen einher. Hier bestehen je nach Gen unterschiedliche Lebenszeitriskiken für eine Erkrankung. Beispielsweise für *BRCA1* und *BRCA2* liegt das Risiko jeweils bei etwa 70 % [Kuchenbaecker *et al.*, 2017]. Für *PALB2* liegt das Lebenszeitrisko für Brustkrebs bei 50 % [Yang *et al.*, 2020] und jenes für *PTEN* beläuft sich auf 85 % [Tischkowitz *et al.*, 2020].

Das Kernkrebsrisiko für pathogene *BRCA1*- und *BRCA2*-Varianten konzentriert sich insgesamt auf FBREK. So verursachen pathogene Varianten dieser zwei Gene allein 60 bis 65 % der erblichen Brust- und Eierstockkrebsfälle [Mehaerin, 2023]. Die Core

Gene beherbergen jedoch neben den Risiken für Mamma- und Ovarialkarzinome auch erweiterte Tumorrisiken. Es können etwa auch Pankreas-, Kolon- sowie Prostatakarzinome und Melanome begünstigt werden [Mehaerin, 2023].

1.1 Genetische Grundlagen und Aufbau der 5'-UTR

Während des Evolutionsprozesses der Menschheit haben vor allem die untranslatierten Regionen des Genoms an Größe und Komplexität zugenommen [Hernandez und Jagus, 2016]. Diese Bereiche bieten so einen faszinierenden Schauplatz der genetischen Evolution. Auf molekularer Ebene ist das genetische Material als Desoxyribonukleinsäure (DNA) in Form von Genen organisiert. Gene umfassen die proteinkodierenden Nukleinsäuresequenzabschnitte [Nordheim *et al.*, 2018] sowie die zugehörigen regulatorischen Elemente. Die Transkription erzeugt während der Genexpression aus der DNA eine prä-Boten-Ribonukleinsäure (prä-mRNA) und durch die Prozessierung wird diese in fertige Boten-Ribonukleinsäure (mRNA) überführt. Durch die Translation erfolgt die Übersetzung in ein Polypeptid. Die Genexpressionskontrolle ist entscheidend für den Proteinhaushalt einer Zelle und deren spezialisierte Funktion. Der nicht proteinkodierende Bereich, welcher sich am 5'-Ende einer mRNA, der kodierenden Sequenz (CDS) vorgelagert (upstream) befindet, wird als 5'-untranslatierte Region (5'-UTR) bezeichnet. Diese reicht vom Transkriptionsstart (+1) bis vor das Startcodon der Translation. Die 3'-untranslatierte Region (3'-UTR) schließt sich an das Translationsstopcodon an. Diese Bereiche finden sich auch auf der DNA-Ebene und werden zwar transkribiert, jedoch bei der Translation nicht in eine Aminosäuresequenz übersetzt [Gather, 2020]. Nach der Transkription enthält die 5'-UTR der mRNA vor allem auf die Translation regulatorisch wirkende Bereiche [Rydzek, Lys und Makalowska, 2023], welche in Abbildung 1.2 dargestellt sind.

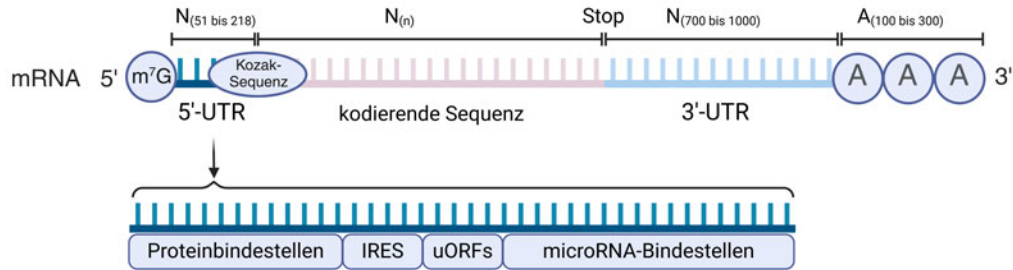


Abbildung 1.2: Aufbau einer eukaryotischen mRNA mit Fokus auf der 5'-UTR (erstellt mit Bio-Render.com).

Gezeigt ist der Aufbau einer eukaryotischen mRNA mit Fokus auf der 5'-UTR. Diese enthält Proteinbindestellen (PBS), interne ribosomale Eintrittsstellen (IRES), stromaufwärts gelegene offene Leserahmen (uORFs) sowie microRNA (miRNA)-Bindestellen. Am 5'-Ende befindet sich die 5'-Kappe. Die Kozaksequenz ist am Übergang zum offenen Leserahmen (ORF) lokalisiert, welcher sich auf die kodierende Sequenz (CDS) bezieht. Anschließend folgt die 3'-UTR mit dem Poly-Adenin-Schwanz. Die Buchstaben beziehen sich auf die jeweiligen Nucleotide (G = Guanin, N = universell, A = Adenin).

Durch die Prozessierung wird eine 5'-Kappe (Cap-Struktur), bestehend aus einem 7-Methylguanidin [Thiemann, Cullen und Klein, 2015], an das 5'-Ende der mRNA angefügt, wodurch die mRNA vor dem Abbau geschützt und der Export aus dem Zellkern unterstützt wird [Nordheim *et al.*, 2018]. In 3'-Richtung folgt die 5'-UTR mit einer mittleren Länge von 51 bis 218 Nucleotiden [Leppek, Das und Barna, 2017]. Für eine ordnungsgemäße Initiation der Translation sind mindestens 25 Nucleotide erforderlich [Abudula, 2001], da das Ribosom das Startcodon AUG (Adenin, Uracil, Guanin) andernfalls nicht mehr erkennen kann [Yoon und Donahue, 1992]. Die regulatorischen Bereiche, welche sich in einer eukaryotischen 5'-UTR befinden, sind teils hoch konserviert (siehe Abbildung 1.2). RNA-bindende Proteine (RBPs) können mit Proteinbindestellen (PBS) in der 5'-UTR interagieren und Einfluss auf die Translation nehmen [Li, X Wang und Jia, 2022]. Außerdem können sich in der 5'-UTR interne ribosomale Eintrittsstellen (IRES) befinden, die ein Sequenzmotiv für eine Sekundärstruktur der mRNA beinhalten, welche als Bindestelle für die 40S (Svedberg) ribosomale Untereinheit dient [Merrick, 2004]. Ausgehend von dieser Sekundärstruktur wird die Translation unabhängig von der Cap-Struktur gestartet, an welcher normalerweise die 40S-Untereinheit des Ribosoms bindet, um den Präinitiationskomplex der Translation zu formieren [Abudula, 2001]. Weist die 5'-UTR einer mRNA ein Startcodon auf, entsteht ein stromaufwärts gelegener offener Leserahmen (uORF). Wird dieser translatiert, kann die Translation der kodierenden Sequenz gehemmt oder die Proteinfunktion gestört werden [Wright *et al.*, 2021]. In der 5'-UTR befinden sich auch Bindestellen, an denen Wechselwirkungen zu micro-RNAs (miRNA) ausgebildet werden können, was die Genexpression reguliert. Diese Binde-

stellen kennzeichnen sich durch einen hohen Grad an Sekundärstrukturen [Ryczek, Lys und Makalowska, 2023]. Die Basen der direkten Umgebung des Startcodons AUG auf einer eukaryotischen mRNA werden als Kozak-Sequenz bezeichnet und sind hochkonserviert [Kozak, 1987]. Diese Sequenzabfolge dient als Ribosomenbindestelle, ist für den Start der Translation entscheidend und beläuft sich bei Vertebraten auf die Basenabfolge GCCRCCAUGG [Kozak, 1987]. Je minimaler die Abweichungen von dieser Konsensussequenz sind, desto höher ist die Wahrscheinlichkeit, dass die Translation an dem AUG in der jeweiligen Kozak-Sequenz eingeleitet wird [Wright *et al.*, 2021].

In den nachfolgenden Ausführungen dieser Arbeit werden die 5'-UTRs der 14 Core Gene betrachtet und auf genetische Varianten geprüft. Da die Untersuchungen auf Panel-Sequenzierungsdaten sowie Sanger-Sequenzierungsdaten beruhen, gibt Abbildung 1.3 einen Überblick darüber, welche Sequenzier-Methode welchen Bereich abdeckt.

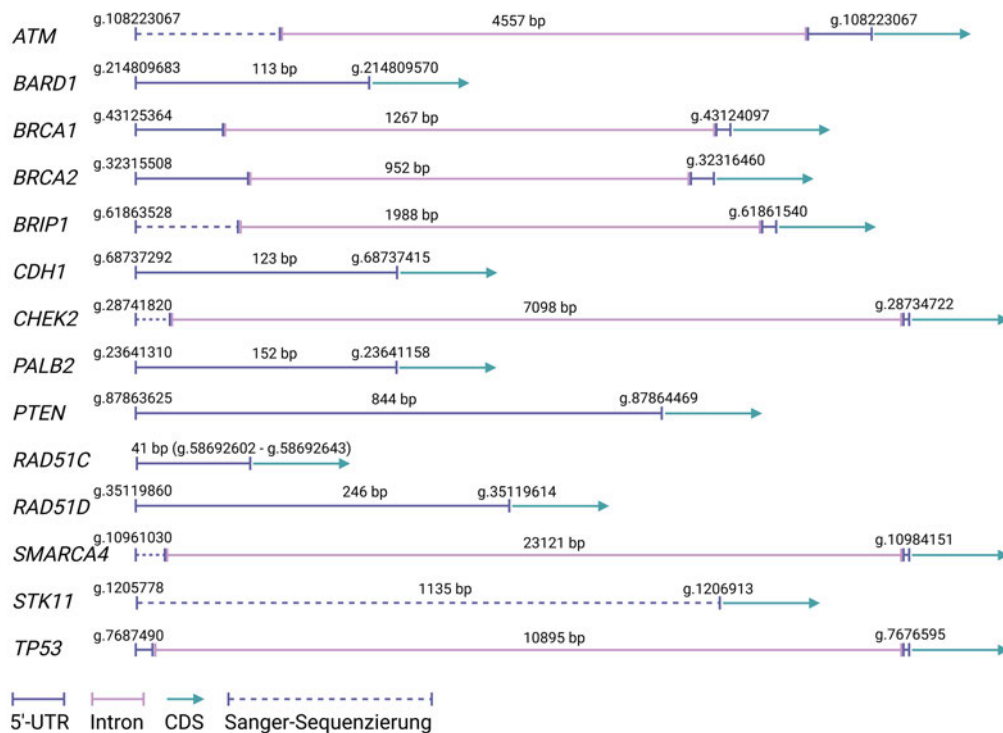


Abbildung 1.3: 5'-UTRs der Core Gene für FBREK (erstellt mit BioRender.com).

Abbildung 1.3 zeigt die 5'-UTRs der 14 Core Gene, die im Zusammenhang mit FBREK stehen. Dabei sind die genomischen Koordinaten angegeben, die die jeweilige 5'-UTR umfassen. Intronische Bereiche, die in den 5'-UTRs liegen, sind ebenfalls mit dargestellt. Für neun der Core Gene ist die Abdeckung der 5'-UTR-Sequenzen über die Panel-Sequenzierung ausreichend. Für Teile der 5'-UTRs von fünf Genen ist eine Sanger-Sequenzierung erforderlich. Diese Bereiche sind mit einer gestrichelten Linie gekennzeichnet. Die Gesamtlängen der 5'-UTRs inklusive intronischer Bereiche sind in Basenpaaren (bp) angegeben. Es ist zu beachten, dass die Abbildung nicht maßstabsgerecht ist.

Aufgrund einer unzureichenden Abdeckung in den Panel-Daten müssen Teile der 5'-UTRs von fünf Genen per Sanger-Sequenzierung separat sequenziert werden. Dabei handelt es sich um Bereiche der 5'-UTRs von *ATM* (g.108223067 bis g.108223186 (119 bp)), *BRIP1* (g.61863284 bis g.61863528 (244 bp)), *CHEK2* (g.28741769 bis g.28741820 (51 bp)), *SMARCA4* (g.10961030 bis g.10961174 (144 bp)) und *STK11* (g.1205778 bis g.1206914 (1136 bp)).

Die verwendeten Transkripte der 14 Core Gene sind in Tabelle 1.1 aufgelistet. Von dem jeweils verwendeten Transkript sind beide Varianten angegeben, sowohl das Transkript aus dem Matched Annotation from the National Center for Biotechnology Information and European Bioinformatics Institute (MANE Select) Transkriptdatensatz als auch jenes vom National Center for Biotechnology Information (NCBI).

Tabelle 1.1: Transkripte der 14 Core Gene.

Gen	MANE Select Transkript	NCBI Transkript
<i>ATM</i>	ENST00000675843.1	NM_000051.4
<i>BARD1</i>	ENST00000260947.9	NM_000465.4
<i>BRCA1</i>	ENST00000357654.9	NM_007294.4
<i>BRCA2</i>	ENST00000380152.8	NM_000059.4
<i>BRIP1</i>	ENST00000259008.7	NM_032043.3
<i>CDH1</i>	ENST00000261769.10	NM_004360.5
<i>CHEK2</i>	ENST00000404276.6	NM_007194.4
<i>PALB2</i>	ENST00000261584.9	NM_024675.4
<i>PTEN</i>	ENST00000371953.8	NM_000314.8
<i>RAD51C</i>	ENST00000337432.9	NM_058216.3
<i>RAD51D</i>	ENST00000345365.11	NM_002878.4
<i>SMARCA4</i>	ENST00000344626.10	NM_003072.5
<i>STK11</i>	ENST00000326873.12	NM_000455.5
<i>TP53</i>	ENST00000269305.9	NM_000546.6

1.2 Kohorte

Die Untersuchungen basieren auf einer Kohorte von 55 FBREK-Patientinnen mit ungeklärten Ursachen für die Tumorerkrankung. Abbildung 1.4 gibt einen Überblick über die Kohorte. Unter bestimmten Bedingungen liegen Indikationen für ein genetisches Screening der Risikogene für FBREK vor. Eine besondere Gewichtung kommt der Familienanamnese zu. So werden sowohl die Patienten selbst als auch deren Geschwister und Kinder sowie die mütterliche und väterliche Linie betrachtet. Gemäß der Checkliste zur Erfassung einer möglichen erblichen Belastung für Brust- und/oder Eierstockkrebs

der Deutschen Krebsgesellschaft e. V. wird ein Score ermittelt. Nimmt dieser mindestens den Wert 3 an, ist eine genetische Ursache wahrscheinlich und eine genetische Beratung wird empfohlen [URL-2, 2023]. Die Patientinnen der Kohorte erfüllen diese Bedingungen. Unter den 55 Patientinnen ist bei 29 Patientinnen vor dem Erreichen des 36. Lebensjahres ein Mammakarzinom diagnostiziert worden. In 24 Fällen trat die Erkrankung ab dem 36. und vor dem 45. Lebensjahr auf und zwei Patientinnen erkrankten mit 45. Ein Ovarialkarzinom ist bis zum derzeitigen Stand (Juni 2024) bei keiner Patientin bekannt. Die Mammakarzinome traten nur in zwei Fällen bilateral (beidseitig) auf. Alle anderen Patientinnen sind von einem unilateralen (einseitig) Mammakarzinom betroffen. Betrachtet man die Brustkrebsarten so zeigt sich bei 45 Fällen ein invasives Mammakarzinom nicht spezifischen Typs (NST). Dies ist die größte Gruppe von Brustkrebs, welche alle invasiven Mammakarzinome zusammenfasst, die keiner spezifischen Art von Brustkrebs angehören [Sinn und Kreipe, 2013]. Als invasiv bezeichnet man einen Tumor, der bereits in umgebendes Gewebe vorgedrungen ist. Bei vier Patientinnen tritt ein duktales Karzinom *in-situ* (DCIS) auf, welches noch nicht in das umgebende Gewebe eingedrungen ist [Sinn und Kreipe, 2013] und sich noch in den Milchgängen befindet. Alle 55 Patientinnen haben bereits die standardmäßig durchgeführte Panel-Diagnostik erhalten, bei welcher die 14 Core Gene auf pathogene Varianten untersucht wurden. Keiner der Fälle konnte über diese Diagnostik gelöst werden. Lediglich bei sieben Patientinnen wurden Varianten unklarer Signifikanz (VUS) gefunden.

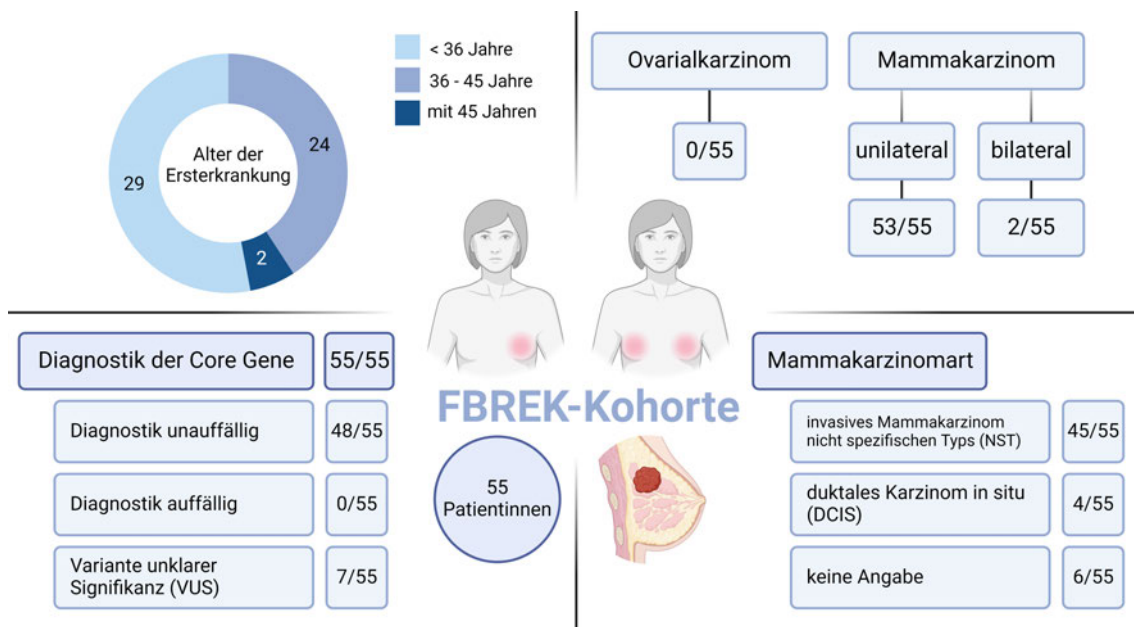


Abbildung 1.4: Überblick über die FBREK-Kohorte (erstellt mit BioRender.com).

Um einen Einblick in die betrachtete Kohorte zu bekommen, sind in Abbildung 1.4 das Ersterkrankungsalter, die Verteilung der Tumorarten sowie die bisherigen Ergebnisse der Gendiagnostik dargestellt. Die Abbildung spiegelt die Kohorte zum Zeitpunkt der Datenerhebung wieder (Stand Juni 2024).

1.3 Variantenklassifikation nach Richards *et al.*, 2015

Genetische Varianten werden grundlegend in fünf Klassen unterteilt. Dabei reichen diese von pathogen und wahrscheinlich pathogen über unklare Signifikanz bis hin zu wahrscheinlich benigne (gutartig) und benigne [Richards *et al.*, 2015]. Zur Einschätzung der Klassifikation einer Variante sind verschiedene Datenbanken hilfreich. Populationsdatenbanken wie die Genome Aggregation Database ([gnomAD](#)) helfen bei der Einordnung der Häufigkeit einer Variante in der Normalbevölkerung. Zusammenhänge zwischen Phänotyp und genetischen Varianten, insbesondere in Bezug auf genetische Erkrankungen finden sich beispielsweise in der Datenbank [ClinVar](#). [VarSome](#) vereint Daten aus verschiedensten Datenbanken und ermöglicht eine umfassende Sicht auf genetische Varianten. Neben Datenbanken sind Scoring-Systeme für *in-silico* Vorhersagen wichtig. Beispielsweise der Combined Annotation Dependent Depletion ([CADD](#))-Score kombiniert viele Einzelinformationen, um eine potenzielle pathogene Wirkung zu bewerten. Enthalten sind im CADD etwa regulatorische Effekte, die durch neuronale Netzwerke vorhergesagt werden, Konservierungsdaten über mehrere Spezies hinweg und Pathogenitätsdaten aus verschiedensten Datenbanken [Schubach *et al.*, 2024]. Über [SpliceAI](#) können präzise Vorhersagen darüber getroffen werden, ob eine Variante das Spleißen der mRNA beeinflusst. Außerdem kann das Heranziehen von aktueller wissenschaftlicher Literatur bei der Variantenklassifikation hilfreich sein.

Durch Richards *et al.* (2015) wurde ein Klassifikationssystem aufgestellt, welches zwischen pathogenen bzw. wahrscheinlich pathogenen und benignen bzw. wahrscheinlich benignen Varianten unterscheidet. Bei einer pathogenen Variante wird ein **P** vorangestellt. Innerhalb dieser Einordnung gibt es verschiedene Gewichtungen, die eine pathogene/wahrscheinlich pathogene Variante näher klassifizieren. Diese reichen von sehr stark (**PVS1**) und stark (**PS1-4**) über moderat (**PM1-6**) zu unterstützend (**PP1-6**). Die Ziffern hinter den jeweiligen Klassifikationen dienen lediglich der Unterscheidung der einzelnen Kriterien. Eine benigne Variante wird mit einem **B** gekennzeichnet. Eine nähere Klassifikation ist über die Einordnungen alleinstehend (**BA1**), stark (**BS1-4**) und unterstützend (**BP1-6**) möglich.

Nicht alle der 14 Core Gene werden nach identischer Vorgehensweise klassifiziert. So gibt es für einige Gene ein erweitertes genspezifisches Klassifizierungssystem. Diese [Variant Curation Expert Panels](#) stammen aus dem Clinical Genome Resource Projekt und gelten beispielsweise für *PTEN*, *BRCA1* sowie *BRCA2*. Durch deren Anwendung werden die allgemeinen Richtlinien [Richards *et al.*, 2015] verfeinert.

Die Klassifizierungsrichtlinien nach Richards *et al.* (2015) gelten außerdem vor allem für

monogene Erkrankungen und Varianten in der CDS, wodurch meist direkte Auswirkungen auf die Aminosäuresequenz entstehen. Für intronische Bereiche und UTRs lassen sich die Kriterien nur eingeschränkt anwenden.

1.4 Regulatorische Funktionen der 5'-UTR in der Genexpression

Die 5'-UTR ist für die Genexpressionsregulation interessant, da sie der CDS vorangestellt ist. Wird die Expressionsregulation aufgrund von genetischen Varianten in der 5'-UTR aus dem Gleichgewicht gebracht, können die Folgen pathologisch relevant sein.

Der Promotorbereich der Transkription kann aus mehreren Kernpromotorelementen bestehen. Bezogen zum Transkriptionsstart (+1) liegen diese zwischen -40 und +40 Nucleotiden lokalisiert. So befinden sie sich auch in der 5'-UTR der mRNA [Shir-Shapira *et al.*, 2019]. Etwa an Position +30 liegt das Downstream Promotor Element [Burke und Kadonaga, 1997]. Kommt es zu einer Veränderung des Abstandes zwischen dem Transkriptionsstart und dem Downstream Promotor Element um drei Nucleotide, führt dies zu einer sieben- bis acht-fachen Verringerung der Transkriptionsrate, da spezifische Transkriptionsfaktoren nicht binden können [Burke und Kadonaga, 1997].

Der direkte Einflussbereich der 5'-UTR ist auf der mRNA-Ebene zu finden und betrifft somit überwiegend die Translation [Leppek, Das und Barna, 2017]. Indirekt beeinflusst die 5'-UTR die Transkription von Genen, wenn die regulatorische Funktion ein, für einen Transkriptionsfaktor kodierendes, Gen betrifft [Wright *et al.*, 2021]. Etwa können Deletionen im Promotorbereich die normale Transkription des Transkriptionsfaktor-Gens unterbinden. Gene, deren Transkription an die Funktionalität dieses Transkriptionsfaktors gekoppelt ist, können so eine beeinträchtigte Transkription erfahren.

Im Anschluss an die Transkription wird die prä-mRNA während der Posttranskription in eine reife mRNA überführt. Das alternative Spleißen erhöht dabei die Vielfalt der mRNA und somit der resultierenden Proteine. Hierbei werden alternative 5'- und 3'-Spleißstellen an den Exons genutzt und beispielsweise Exons übersprungen [Nordheim *et al.*, 2018]. Somit entstehen verschiedene Transkripte auf Grundlage derselben prä-mRNA. Das alternative Spleißen kann auch unterschiedliche 5'-UTRs hervorbringen [Abudula, 2001]. Etwa bei dem Tumorsuppressorgen *Discs Large Homolog 1 (DLG1)* beeinflussen unterschiedlich gespleißte 5'-UTRs die Krebsentstehung [Marziali *et al.*, 2019]. Längere 5'-UTRs enthalten einen kurzen uORF, der die Translation des Haupt-

ORFs hemmt. Werden Transkripte mit kürzerer 5'-UTR verstärkt exprimiert, wird der Haupt-ORF vermehrt abgelesen und der DLG1-Proteinspiegel steigt, wodurch die tumorsuppressive Funktion verstärkt ausgeübt werden kann [Marziali *et al.*, 2019].

Während der Translation wird die reife mRNA an den Ribosomen in eine Polypeptidkette überführt. Hierbei ist der Einfluss der 5'-UTR einer mRNA besonders groß, indem sie die Translationseffizienz der mRNA beeinflusst. Dies geschieht durch verschiedene Faktoren, wie beispielsweise Sekundärstrukturen in der 5'-UTR, Proteinbindestellen, die Cap-Struktur, die Länge der 5'-UTR und uORFs sowie die Kozak-Sequenz [Abudula, 2001]. Normalerweise erfolgt die Initiation der Translation Cap-abhängig in einem sehr komplexen Prozess. Dabei bewegt sich der Präinitiationskomplex ausgehend von der Cap-Struktur entlang der 5'-UTR in 3'-Richtung, bis ein passendes AUG-Startcodon vorliegt [Nordheim *et al.*, 2018]. Über IRES in der 5'-UTR kann jedoch auch die Cap-unabhängige Translationsinitiation erfolgen. Dies ist notwendig, wenn beispielsweise bei starkem zellulären Stress oder in bestimmten Zellzyklusphasen die Cap-abhängige Translation verhindert oder abgeschwächt ist [Gather, 2020]. In diesem Fall kann der Haupt-ORF weiterhin translatiert werden, indem die Ribosomen an IRES binden. IRES befinden sich daher vor allem in mRNAs, die während der Mitose oder im Fall von Apoptose sowie Stress translatiert werden müssen. Varianten in diesen Bereichen können pathogene Auswirkungen haben, da so möglicherweise die Translation trotz Vorhandensein von IRES gehemmt wird [Ryczek, Lys und Makalowska, 2023].

Die von der Kozak-Sequenz bestimmte Umgebung des Startcodons dient der Translation des Haupt-ORFs. Jedoch kann es auch vorkommen, dass während des Scan-Prozesses des Ribosoms auf der Suche nach dem Startcodon uORFs vor dem Haupt-ORF translatiert werden. In diesem Fall wird die Translation des Haupt-ORFs meist gehemmt [Gather, 2020]. Sie kann erst durch Re-Initiation [Abudula, 2001] oder im Falle des Leaky-Scannings [Ryczek, Lys und Makalowska, 2023] erfolgen. Bei der Re-Initiation werden sowohl der uORF als auch der eigentliche Haupt-ORF translatiert [Somers, Pöyry und Willis, 2013]. Beim Leaky-Scanning kommt es beispielsweise in Stresssituationen dazu, dass die kleine ribosomale Untereinheit beim Entlanggleiten an der mRNA einen uORF überspringt und es somit zu einer verstärkten Translation des Haupt-ORFs kommt. Manche Gene werden so nur in Stresssituationen translatiert [H Zhang, Y Wang und Lu, 2019; Young und Wek, 2016] und unter normalen Bedingungen durch die Translation des uORFs gehemmt. Wie effizient eine mRNA translatiert wird, hängt also unter anderem von vorhandenen uORFs ab [Capell, Fellerer und Haass, 2014]. Die Translation des Haupt-ORFs kann jedoch auch durch Sekundärstrukturen wie sehr stabile Haarnadelstrukturen in der 5'-UTR komplett unterdrückt werden, wodurch die Ribosomen gestoppt werden [Capell, Fellerer und Haass, 2014].

1.5 Pathologische Relevanz der 5'-UTR und von 5'-UTR-Varianten

Die Pathologie beschreibt die wissenschaftliche Untersuchung von Krankheiten und Abweichungen von normalen physiologischen Zuständen in Organismen. In der molekularen Pathologie liegt der Fokus auf der Untersuchung von genetischen Veränderungen und Regulationsmechanismen auf zellulärer Ebene. Die Deregulierung der Genexpression durch den Einfluss der 5'-UTR wird mit vielen Krankheiten in Verbindung gebracht. Ein Onkogen, welches hier zu nennen ist, ist *BRCA1* [Ryczek, Lys und Makalowska, 2023]. Sowohl Loss-of-function-Mutationen als auch Deletionen in *BRCA1* erhöhen das Risiko für verschiedene Krebsarten. Ein verringerter mRNA-Spiegel fördert die Krebsentstehung durch fehlende Reparaturmechanismen an der DNA (siehe Abbildung 1.1) [Ryczek, Lys und Makalowska, 2023]. Es existieren zwei verschiedenen lange Transkripte von *BRCA1*, welche durch die Nutzung unterschiedlicher Promotoren bei der Transkription entstehen [Araujo *et al.*, 2012]. Diese unterscheiden sich in der Länge ihrer 5'-UTRs. Während mRNAs mit der kurzen 5'-UTR sowohl in gesundem als auch in Tumor-Brustgewebe gefunden werden, ist das Transkript mit der langen 5'-UTR nur in Gewebeproben von Mammakarzinomen nachgewiesen [Araujo *et al.*, 2012]. Diese lange 5'-UTR-Variante kennzeichnet sich durch mehrere AUGs in der 5'-UTR und eine sehr komplexe Sekundärstruktur [Araujo *et al.*, 2012]. Dadurch werden uORFs bevorzugt translatiert und die Translation der kodierenden Sequenz von *BRCA1* stark reduziert.

Bei verschiedensten Erkrankungen gibt es bereits Hinweise darauf, dass Varianten in der 5'-UTR dafür sorgen, dass neue Leseraster entstehen und somit die eigentlich kodierende Sequenz vermindert abgelesen wird oder ein nicht funktionelles Protein entsteht. Vor allem Varianten, welche neue Startcodons erzeugen, sind wichtige Ursachen für monogene Krankheiten [Wright *et al.*, 2021]. Dabei wird die Translation des Haupt-ORFs durch die vorangehende Translation des uORFs gehemmt. Die Stärke dieser Hemmung hängt unter anderem davon ab, wie stark die Umgebung des neuen Startcodons als Kozaksequenz fungiert. Allerdings haben Varianten, die zu uORFs führen, welche sich mit der kodierenden Sequenz überlappen, einen wesentlich höheren negativen Einfluss auf die Haupt-ORF-Translation [Wright *et al.*, 2021].

Ein Beispiel für pathogene 5'-UTR-Varianten findet sich beim *BRCA1*-Gen. Evans *et al.* (2018) haben eine dominant vererbte, heterozygote Einzelnukleotidvariante (SNV) identifiziert (c.-107A>T). Die Variante sorgt für eine epigenetische Stilllegung des *BRCA1*-Gens über Promotor-Hypermethylierung. Analysen mittels Pyrosequenzierung zeigten eine 50 %-ige Methylierung der DNA, was die heterozygote Variante nachweist [Evans

et al., 2018]. Durch Hypermethylierung des Promotors eines Gens wird die Transkription verhindert. In etwa 10 % der Tumorgewebeproben von somatischem/erworbenem Brustkrebs findet sich solch eine Hypermethylierung bei *BRCA1* [Evans *et al.*, 2018].

In den letzten Jahren sind Varianten in den untranslatierten Bereichen mehr in den Fokus der Forschung gerückt. Beispielsweise wurden von Wilcox *et al.* (2024) basierend auf Genomsequenzierungen seltene 5'- und 3'-UTR Varianten sowie Varianten in den Promotorregionen im Zusammenhang mit Brustkrebs untersucht. Als Datengrundlage dienten 8001 Frauen mit Brustkrebs und über 92000 Frauen ohne Brustkrebs aus der United Kingdom Biobank. Insgesamt wurden 18676 Gene und 35201 mögliche Promotorregionen betrachtet [Wilcox *et al.*, 2024]. Die Studie kommt zu dem Ergebnis, dass Varianten in den kodierenden Regionen nur einen kleinen Anteil der Erbllichkeit ausmachen. Varianten in den untranslatierten Regionen des Genoms sollten zukünftig weiter untersucht werden, da sie eine entscheidende Rolle beim Verständnis der Krankheitsanfälligkeit spielen können [Wilcox *et al.*, 2024].

2 Zielstellung

Um die Relevanz von 5'-UTR-Varianten bei hereditärem Brust- und Eierstockkrebs besser einschätzen zu können, werden in dieser Bachelorarbeit die 5'-UTRs der 14 Core Gene für FBREK [URL-1, 2024] auf auffällige genetische Varianten untersucht. Dazu wird eine Kohorte von 55 Patientinnen herangezogen, deren Ersterkrankungsalter an Brustkrebs unter 36 bzw. 45 Jahren liegt und deren Familienanamnese auffällig ist, was auf eine genetische Ursache hindeutet. Die unauffällige Panel-Diagnostik legt die Vermutung nahe, dass sich die Erkrankungsursache möglicherweise auf Varianten außerhalb der kodierenden Gensequenz bezieht.

Die Untersuchungen erfolgen auf DNA-Ebene. Die Sequenzdaten basieren dabei auf Panel-Diagnostik sowie auf Sanger-Sequenzierung. Mittels des Variantenklassifikationssystems nach Richards *et al.* (2015) werden die gefundenen Varianten hinsichtlich ihrer Pathogenität eingeschätzt. Dazu werden verschiedene Gen-Datenbanken, Scores sowie Vorhersage-Tools angewendet.

Im Fall der Identifizierung relevanter pathogener Varianten können womöglich einige der ungeklärten FBREK-Fälle abgeschlossen werden.

3 Materialien, Chemikalien und Geräte

3.1 Probenmaterial

- DNA-Proben der Patientinnen 1 bis 55 mit ungeklärter FBREK-Diagnose (gelöst in Elutionspuffer; Aufbewahrung bei 4 - 8 °C)

3.2 Molekularbiologische Reagenzien

- 6x DNA Loading Dye (LOT-Nr.: 01136102; Firma: Thermo Fisher Scientific Baltics UAB; Sitz: V. A. Graičiūno 8, LT-02241 Vilnius, Litauen)
- 10x DreamTaq Green Buffer (LOT-Nr.: 2812388; Firma: Thermo Fisher Scientific Baltics UAB; Sitz: V. A. Graičiūno 8, LT-02241 Vilnius, Litauen)
- BigDye™ Terminator v3.1 Cycle Sequencing Kit (LOT-Nr.: 2759910; Firma: Thermo Fisher Scientific Baltics UAB; Sitz: V. A. Graičiūno 8, LT-02241 Vilnius, Litauen)

BigDye™ Terminator v3.1 Cycle Sequencing RR-100 LOT-Nr: 2726534

BigDye™ Terminator v1.1, v3.1 5x Sequencing Buffer LOT-Nr: 2201208

- Desoxynukleosidtriphosphat (dNTP) Set 4 x 25 µmol (LOT-Nr.: 01283322; Firma: Thermo Fisher Scientific Baltics UAB; Sitz: V. A. Graičiūno 8, LT-02241 Vilnius, Litauen)
- DreamTaq DNA Polymerase 5 units/µl (LOT-Nr.: 2879952; Firma: Thermo Fisher Scientific Baltics UAB; Sitz: V. A. Graičiūno 8, LT-02241 Vilnius, Litauen)
- ExoSAP-IT™ Express Reagent (LOT-Nr.: 2875115; Firma: Thermo Fisher Scientific Baltics UAB; Sitz: V. A. Graičiūno 8, LT-02241 Vilnius, Litauen)
- GeneRuler 100 bp Plus DNA Ladder, ready-to-use 0,1 µg/µl (LOT-Nr.: 2699034; Firma: Thermo Fisher Scientific Baltics UAB; Sitz: V. A. Graičiūno 8, LT-02241 Vilnius, Litauen)
- HotStarTaq® Plus DNA Polymerase 1000 units (LOT-Nr.: 175033867; Firma: QIAGEN GmbH; Sitz: QIAGEN Straße 1, 40724 Hilden)

HotStarTaq® DNA Polymerase	5 units/µl
Polymerase-Kettenreaktion (PCR)-Puffer	10x konzentriert
Q-Solution	5x konzentriert

- M13-Primer Forward (F) 10 pmol/µl (LOT-Nr.: 1286460; Sequenz: GTAAAACGACG-GCCAG; Firma: Biolegio; Sitz: Lagelandseweg 56, 6545 CG Nijmegen, Niederlande)

- M13-Primer Reverse (R) 10 pmol/μl (LOT-Nr.: 1286461; Sequenz: CAGGAAA-CAGCTATGAC; Firma: Biolegio; Sitz: Lagelandseweg 56, 6545 CG Nijmegen, Niederlande)
- Midori Green Advance DNA Stain (LOT-Nr.: 451MG06043; Firma: NIPPON Genetics EUROPE GmbH; Sitz: Mariaweilerstraße 28-30, 52349 Düren)
- Sanger-Primer (gelöst in Tris(hydroxymethyl)aminomethan-Ethylendiamintetraessigsäure (TE)-Puffer; inklusive M13-Sequenz; Firma: Biolegio; Sitz: Lagelandseweg 56, 6545 CG Nijmegen, Niederlande)
 - ATM 5'-UTR Primer A Forward (LOT-Nr.: 1404658;
Sequenz: GTAAAACGACGGCCAGCTAAGTCGCTGGCCATTGGT)
 - ATM 5'-UTR Primer A Reverse (LOT-Nr.: 1404659;
Sequenz: CAGGAAACAGCTATGACAGGCAGGGTTCAATTCAGGG)
 - BRIP1 5'-UTR Primer A Forward (LOT-Nr.: 1404660;
Sequenz: GTAAAACGACGGCCAGGACTGGTTGATTCCCTGGCCT)
 - BRIP1 5'-UTR Primer A Reverse (LOT-Nr.: 1404661;
Sequenz: CAGGAAACAGCTATGACGGCTCCCTCCTCAGGTTTTTC)
 - CHEK2 5'-UTR Primer A Forward (LOT-Nr.: 1404664;
Sequenz: GTAAAACGACGGCCAGTTGGCTGAGGGTGGAGTTTG)
 - CHEK2 5'-UTR Primer A Reverse (LOT-Nr.: 1404665;
Sequenz: CAGGAAACAGCTATGACTGTTCAGGCCTGTTTCAGCT)
 - SMARCA4 5'-UTR Primer A Forward (LOT-Nr.: 1404662;
Sequenz: GTAAAACGACGGCCAGCGCGTGTGTGTGAAGGGG)
 - SMARCA4 5'-UTR Primer A Reverse (LOT-Nr.: 1404663;
Sequenz: CAGGAAACAGCTATGACACACTGTGGCCTTCCCCA)
 - STK11 5'-UTR Primer A Forward (LOT-Nr.: 1404666;
Sequenz: GTAAAACGACGGCCAGGGATGGCAGGTTCAACCAAC)
 - STK11 5'-UTR Primer A Reverse (LOT-Nr.: 1404667;
Sequenz: CAGGAAACAGCTATGACCTGTCCCAGGAGCAAACGC)
 - STK11 5'-UTR Primer B Forward (LOT-Nr.: 1404668;
Sequenz: GTAAAACGACGGCCAGGAGGTCCGCTCGGTCGTC)
 - STK11 5'-UTR Primer B Reverse (LOT-Nr.: 1404669;
Sequenz: CAGGAAACAGCTATGACTACCCACCGACATCAGCTCG)
 - STK11 5'-UTR Primer C Forward (LOT-Nr.: 1404670;
Sequenz: GTAAAACGACGGCCAGCAGACGTTTGCGGGGAGG)
 - STK11 5'-UTR Primer C Reverse (LOT-Nr.: 1404671;
Sequenz: CAGGAAACAGCTATGACGGCCCACGGACAAGTATGAA)

3.3 Chemikalien

- Agarose (Firma: Sigma-Aldrich Chemie GmbH; Sitz: Riedstraße 2, 89555 Steinheim)
- Destilliertes Wasser (LOT-Nr.: 2026-01230718; Firma: Henry Schein® Service GmbH; Sitz: Monzastraße 2A, 63225 Langen)
- Ethanol 100 % (LOT-Nr.: L0960; Firma: Honeywell GmbH; Sitz: Strahlenbergerstraße 110-112, 63067 Offenbach)
- Natriumacetat 3 M (LOT-Nr.: 087K8621; Firma: Sigma-Aldrich Chemie GmbH; Sitz: Riedstraße 2, 89555 Steinheim)
- Wasser für Injektionszwecke und andere Anwendungen (LOT-Nr.: 20SEF005; Firma: Fresenius Kabi Deutschland GmbH; Sitz: Siemensstraße 21, 61352 Bad Homburg)
- Hi-DI™ Formamid (LOT-Nr.: 2303888; Firma: Thermo Fisher Scientific; Sitz: 7 Kingsland Grange Woolston, Warrington Cheshire WA1 4SR, England)

3.4 Lösungen

- Agarose-Gel

Agarose	2,8 g
TBE-Puffer 1x	200 ml

- Ethanol 70 %

Ethanol 100 %	70 ml
Wasser	30 ml

- Ethanol-Natriumacetat-Gemisch

Ethanol 100 %	30 µl
Natriumacetat 3 M	1 µl

- Rotiphorese® Tris(hydroxymethyl)aminomethan-Borat-Ethylendiamintetraessigsäure (TBE)-Puffer 10x konzentriert (Chargen-Nr.: 054351288; Firma: Carl Roth GmbH & Co. KG; Sitz: Schoemperlenstraße 3-5, 75185 Karlsruhe)

- TBE-Puffer 1x

TBE-Puffer 10x	1 Volumenanteil
Destilliertes Wasser	9 Volumenanteile

3.5 Materialien und sonstige Laborausstattung

- 8-Cap Strips, weiß, extra robust (Firma: Biozym Biotech Trading GmbH; Sitz: Wehlistraße 27b A-1200 Wien, Österreich)
- Biosphere[®] SafeSeal Tubes 1.5 ml (Firma: SARSTEDT AG & Co. KG; Sitz: Sarstedtstraße 1, 51588 Nümbrecht)
- Combitips advanced[®] 0,1 ml (Firma: Eppendorf SE; Sitz: Barkhausenweg 1, 22331 Hamburg)
- Combitips advanced[®] 1 ml (Firma: Eppendorf SE; Sitz: Barkhausenweg 1, 22331 Hamburg)
- Eppendorf Research[®] plus - mechanische Pipetten 1-Kanal 0,1 - 2,5 µl, 0,5 - 10 µl, 10 - 100 µl, 20 - 200 µl, 100 - 1000 µl (Firma: Eppendorf SE; Sitz: Barkhausenweg 1, 22331 Hamburg)
- Eppendorf Research[®] plus - mechanische Pipetten 8-Kanal 0,5 - 10 µl, 10 - 100 µl (Firma: Eppendorf SE; Sitz: Barkhausenweg 1, 22331 Hamburg)
- Multipette[®] stream (Firma: Eppendorf SE; Sitz: Barkhausenweg 1, 22331 Hamburg)
- Multiply[®] µ-Strip Pro 8er Ketten (Firma: SARSTEDT AG & Co. KG; Sitz: Sarstedtstraße 1, 51588 Nümbrecht)
- PCR 96-Well TW-MT-Platten Universal, farblos (Firma: Biozym Scientific GmbH; Sitz: Steinbrinksweg 27, 31840 Hess. Oldendorf)
- PCR-Platten GK060/H/1E-OA 96-Well mit Halbrahmen für ABI 3100 (Firma: Kisker Biotech GmbH & Co. KG; Sitz: Sonnenschein 8, Postbox 1329, 48565 Steinfurt)
- Septa-Matten G060/S-3100 für ABI 3100 autoklavierbar und wiederverwendbar (Firma: Kisker Biotech GmbH & Co. KG; Sitz: Sonnenschein 8, Postbox 1329, 48565 Steinfurt)

3.6 Geräte

- Applied Biosystems[®] Sanger Sequencing 3500 Series Genetic Analyzer (Firma: Thermo Fisher Scientific; Sitz: Frankfurter Straße 129b, 64293 Darmstadt)
- Applied Biosystems[®] Veriti[™] 96-Well Thermal Cycler (Firma: Thermo Fisher Scientific; Sitz: Frankfurter Straße 129b, 64293 Darmstadt)
- Elektrophorese Netzteil GPS200/400 (Firma: Pharmacia Biotech GmbH, jetzt Phadia GmbH; Sitz: Munzinger Straße 7, 79111 Freiburg i. Breisgau)
- Feinwaage PB602 (Firma: Mettler-Toledo GmbH; Sitz: Ockerweg 3, 35396 Gießen)

- Gelelektrophoresekammer BlueMarine200™ (Firma: SERVA Electrophoresis GmbH; Sitz: Carl-Benz-Straße 7, 69115 Heidelberg)
- Laborschüttler Vibrofix-VF2 (Firma: IKA-Werke GmbH & Co. KG; Sitz: Janke & Kunkel-Straße 10, 79219 Staufen)
- Magnetrührer RH basic (Firma: IKA-Werke GmbH & Co. KG; Sitz: Janke & Kunkel-Straße 10, 79219 Staufen)
- Mikrowelle Micromat 21 GK21MA (Firma: Electrolux Hausgeräte GmbH Markenvertrieb AEG; Sitz: Fürther Str. 246, 90429 Nürnberg)
- Mikrozentrifuge myLab MC-01 (Firma: AHN Biotechnologie GmbH; Sitz: Uthleber Weg 14, 99734 Nordhausen)
- Plattenzentrifuge 5430 (Firma: Eppendorf SE; Sitz: Barkhausenweg 1, 22331 Hamburg)
- Thermocycler Mastercycler® Gradient (Firma: Eppendorf SE; Sitz: Barkhausenweg 1, 22331 Hamburg)
- Transilluminator FastGene FAS-Digi PRO GP-07LED (Firma: NIPPON Genetics EUROPE GmbH; Sitz: Mariaweyerstraße 28-30, 52349 Düren)

3.7 Software und Datenbanken

- Alamut™ Visual Plus Version 1.10 (Firma: SOPHiA GENETICS™; Sitz: Technopole Izarbel 158 allée Fauste d'Elhuyar 64210 Bidart, Frankreich)
- Clinical Genome Resource (National Institutes of Health; <https://clinicalgenome.org/> (zuletzt verwendet am 22.08.2024))
- ClinVar (National Center for Biotechnology Information; <https://www.ncbi.nlm.nih.gov/clinvar/> (zuletzt verwendet am 13.08.2024))
- Catalogue of Somatic Mutations in Cancer (<https://cancer.sanger.ac.uk/cosmic> (zuletzt verwendet am 13.08.2024))
- dbSNP (National Center for Biotechnology Information; <https://www.ncbi.nlm.nih.gov/snp/> (zuletzt verwendet am 12.08.2024))
- Ensembl 112 (European Molecular Biology Laboratory's European Bioinformatics Institute; Mai 2024; https://www.ensembl.org/Homo_sapiens/Info/Index (zuletzt verwendet am 03.06.2024))
- FastGene FAS-DIGI PRO Gel Imaging System (Firma: NIPPON Genetics EUROPE GmbH; Sitz: Mariaweyerstraße 28-30, 52349 Düren)
- GEPADO Xpro Version 4.5.4589, Database Version 4.5.4580 (Firma: GEPADO GmbH; Sitz: Wartburgstraße 46, 01309 Dresden)

- Genome Aggregation Database (gnomAD) v4.1.0 (<https://gnomad.broadinstitute.org/> (zuletzt verwendet am 13.08.2024))
- Human Gene Mutation Database® Professional (HGMD) 2024.2 (QIAGEN 2013 - 2024; <https://my.qiagendigitalinsights.com/bbp/view/hgmd/pro/start.php> (zuletzt verwendet am 13.08.2024))
- IRESbase © 2019 (The Song Lab. All Rights Reserved; <http://reprod.njmu.edu.cn/cgi-bin/iresbase/index.php> (zuletzt verwendet am 17.08.2024))
- OligoPerfect™ (Thermo Fisher Scientific Inc.; 2024; <https://apps.thermofisher.com/apps/oligoperfect/#!/design> (zuletzt verwendet am 03.06.2024))
- Polygenic Score Catalog (European Molecular Biology Laboratory's European Bioinformatics Institute; 31.07.2024; <https://www.pgscatalog.org/> (zuletzt verwendet am 19.08.2024))
- PubMed® (National Center for Biotechnology Information; <https://pubmed.ncbi.nlm.nih.gov/> (zuletzt verwendet am 20.08.2024))
- SEQUENCE PILOT Software For Genetic Analysis (Firma: JSI medical systems GmbH Germany; Sitz: Tullastraße 18, 77955 Ettenheim)
- SNPCheck V3 (Certus Technology Associates Limited; EMQN c/o Manchester Centre for Genomic Medicine; <https://genetools.org/SNPCheck/snpcheck.htm> (zuletzt verwendet am 03.06.2024))
- SpliceAI (Broad Institute; <https://spliceailookup.broadinstitute.org/> (zuletzt verwendet am 09.08.2024))
- University of California Santa Cruz (UCSC) Genome Browser (<https://genome.ucsc.edu/> (zuletzt verwendet am 13.08.2024))
- UCSC In-Silico PCR (University of California Santa Cruz Genome Browser; Jim Kent; <https://genome.ucsc.edu/cgi-bin/hgPcr> (zuletzt verwendet am 03.06.2024))
- Variant Effect Predictor (Ensembl release 112 - May 2024 © EMBL-EBI; <https://www.ensembl.org/Multi/Tools/VEP?db=core> (zuletzt verwendet am 14.08.2024))
- VarSome 12.1.0 (VarSome: the human genomic variant search engine; Christos Kopanos, Vasilis Tsiolkas, Alexandros Kouris, Charles E Chapple, Monica Albarca Aguilera, Richard Meyer, Andreas Massouras; Bioinformatics, Jahrgang 35, Heft 11, 1. Juni 2019, Seite 1978 – 1980; <https://varsome.com/> (zuletzt verwendet am 15.08.2024))

3.8 PCR-Ansätze für die Etablierung der Sanger-Primer

Tabelle 3.1: Touchdown Reaktionsansatz für die PCR.

Reagenz	Volumen [μ l]
10x DreamTaq Green Buffer	2,5
dNTPs (2 mmol/l)	2,5
Q-Solution 5x konzentriert	5
Primer F + R (10 pmol/ μ l)	1
DreamTaq DNA Polymerase (5 U/ μ l)	0,6
Wasser	13,4
DNA (40 ng/ μ l)	1

Tabelle 3.2: Touchdown High Quality (HQ) Reaktionsansatz für die PCR.

Reagenz	Volumen [μ l]
PCR-Puffer 10x konzentriert	2,5
dNTPs (2 mmol/l)	2,5
Q-Solution 5x konzentriert	5
Primer F + R (10 pmol/ μ l)	1
HotStarTaq [®] DNA Polymerase (5 U/ μ l)	0,4
Wasser	11,6
DNA (40 ng/ μ l)	2

Tabelle 3.3: Touchdown HQ Reaktionsansatz für die PCR des Primerpaares von *ATM*.

Reagenz	Volumen Ansatz 1 [μ l]	Volumen Ansatz 2 [μ l]
PCR-Puffer 10x konzentriert	2,5	2,5
dNTPs (2 mmol/l)	2,5	2,5
Q-Solution 5x konzentriert	5	5
Primer F + R (10 pmol/ μ l)	1	1
HotStarTaq [®] DNA Polymerase (5 U/ μ l)	0,4	0,8
Wasser	9,6	11,2
DNA (40 ng/ μ l)	4	2

3.9 Reaktionsansätze für die Aufreinigung und Sequenzierreaktion

Tabelle 3.4: Reaktionsansatz zur Aufreinigung des PCR-Produktes.

Reagenz	Volumen [μ l]
ExoSAP-IT™ Express Reagent	0,45
Wasser	3,55
PCR-Produkt	1

Tabelle 3.5: Reaktionsansatz für die Sequenzierreaktion des PCR-Produktes.

Reagenz	Volumen [μ l]
BigDye® Terminator v1.1, v3.1 5x Sequencing Buffer	2
BigDye® Terminator v3.1 Cycle Sequencing RR-100	0,7
Wasser	3,3
M13 Primer Forward bzw. Reverse	0,5
gereinigtes PCR-Produkt	5

4 Methoden

Als Referenzgenom aller betrachteten Daten diente das human genome 38 (hg38). Die Transkripte der 14 Gene beziehen sich auf den Matched Annotation from the National Center for Biotechnology Information and European Bioinformatics Institute (MANE Select) Transkriptdatensatz (siehe Tabelle 1.1).

4.1 Grundlegende Methoden und Prinzipien

Die 5'-UTR-Bereiche der relevanten Core Gene wurden auf Abdeckung in der routinemäßig durchgeführten Panel-Diagnostik geprüft. Die Panel-Sequenzierung umfasst eine definierte Auswahl an Genen, welche über Next Generation Sequencing (NGS) sequenziert werden. Bei ausreichender Abdeckung konnten die Daten direkt bioinformatisch untersucht und nach genetischen Varianten gefiltert werden.

Die fünf Gene, deren 5'-UTRs keine ausreichende Abdeckung in der Panel-Diagnostik aufweisen, wurden mittels Sanger-Sequenzierung untersucht. Diese setzt das Primerdesign für die zu amplifizierenden Zielbereiche voraus. Dabei beinhalten die Primer sowohl einen Abschnitt der zu amplifizierenden genomischen Zielsequenz sowie eine M13-spezifische Sequenz [Arneemann *et al.*, 2019]. In Vorbereitung auf die Sequenzierung wird mit den Primern eine Touchdown-PCR durchgeführt. Diese dient der Verhinderung des Entstehens unspezifischer PCR-Produkte [Müller und Prange, 2016], welche sich sonst mit der zunehmenden Anzahl an PCR-Zyklen anreichern und mit der Zielsequenz konkurrieren. Bei der Touchdown-PCR wird daher die Annealing-Temperatur während der ersten PCR-Zyklen schrittweise herabgesetzt, wodurch das korrekte PCR-Produkt spezifischer amplifiziert wird [Don *et al.*, 1991].

Anschließend wird das Vorhandensein des korrekten PCR-Produkts mittels Agarose-Gelelektrophorese überprüft. Bei dieser Methode wird die negative Ladung des Phosphat-Rückgrats der DNA ausgenutzt. Über ein elektrisches Feld wird die amplifizierte DNA bei einer Spannung von 150 V nach Größe aufgetrennt. Kleinere Fragmente wandern schneller zur positiv geladenen Anode. Durch das Mitführen von Größenmarkern bekannter Länge kann die amplifizierte DNA einer Fragmentgröße zugeordnet werden. Die Bildgebung der DNA-Banden erfolgt dann in einem Transilluminator. Auch unspezifisch amplifizierte Produkte können über die Gelelektrophorese sichtbar gemacht werden. Nachfolgend muss das PCR-Produkt aus dem Reaktionsansatz aufgereinigt werden. Dies ist notwendig, um überschüssige Primer und einzelsträngige DNA zu entfernen

sowie die übrigen dNTPs zu inaktivieren. Die Reinigung erfolgt enzymatisch [URL-3, 2024]. Die Exonuklease I baut dabei die einzelsträngigen Oligonukleotide ab und das Protein Shrimp Alkaline Phosphatase (SAP) ist dafür zuständig, das Phosphat am fünften Kohlenstoffatom der dNTPs zu entfernen [URL-4, 2024].

Das aufgereinigte PCR-Produkt wird im Anschluss in einer Sequenzier-PCR mit der Kettenabbruchmethode nach Frederick Sanger [Sanger, Nicklen und Coulson, 1977] bearbeitet. Eingesetzt wird hierbei nur ein Sequenzierprimer (Forward oder Reverse) je Reaktion. Diese sogenannten M13-Primer sind universell einsetzbar. Sie binden an die M13-spezifische Sequenz, welche durch die Sanger-Primer an die Sequenzierprodukte angehängt wurde [Arnemann *et al.*, 2019]. Somit muss nicht für jede Zielsequenz ein individueller Sequenzierprimer designiert werden. Neben den dNTPs werden auch Didesoxyribonukleosidtriphosphate (ddNTPs) in wesentlich geringerer Konzentration der Sequenzierreaktion zugegeben. Diesen Nukleotiden fehlt die 3'-Hydroxylgruppe, mit welcher normalerweise das nachfolgende Nukleotid verknüpft wird [Clark und Pazdernik, 2009]. Demnach kommt es nach Einbau eines ddNTPs zum Kettenabbruch. Die vier ddNTPs sind unterschiedlich fluoreszenzmarkiert [Brown, 2007], was in der späteren Kapillar-Sequenzierung die Zuordnung der Basen zulässt.

Nach der Sequenzier-PCR muss erneut eine Reinigung von den nicht umgesetzten ddNTPs und Primern sowie von störenden Salzen erfolgen, um später das Hintergrundrauschen in den Sequenzdaten zu verhindern. Über Waschschriffe und eine Ethanol-fällung kann dies realisiert werden. Bei der Ethanolpräzipitation mit Natriumacetat wird die DNA bei niedrigem *pondus hydrogenii* (pH)-Wert ausgefällt. Das Natriumacetat liefert monovalente Kationen, die zur Bildung eines unlöslichen Niederschlags der DNA beitragen [Jurk, 2022], und verringert den pH-Wert der Lösung, während das Ethanol die Löslichkeit der DNA durch Wasserentzug verringert. Durch Zentrifugation kann der DNA-Niederschlag isoliert werden. Überschüssiges Salz wird durch eine Reinigung des DNA-Pellets mit 70 %-igem Ethanol entfernt [Jurk, 2022]. Eine anschließende Zugabe von Formamid dient der Resuspendierung der Proben in Vorbereitung auf die Kapillarelektrophorese [URL-5, 2024].

Die Kapillar-Sequenzierung trennt die durch den Kettenabbruch unterschiedlich langen DNA-Fragmente nach Größe. Dabei werden Kapillaren mit einem Innendurchmesser von maximal 100 µm verwendet [Westermeier, 2019]. Durch Fluoreszenzdetektion der Farbstoffe an den ddNTPs kann die DNA-Sequenz ausgelesen werden [Schaaf und Zschocke, 2018]. Dies erfolgt jeweils aus beiden Richtungen der Sequenz (Forward und Reverse).

4.2 Molekulargenetische Methoden

Die molekulargenetischen Methoden werden im Nachfolgenden einmal ausführlich erläutert. Bei erneuter Durchführung wird nur auf geänderte Parameter eingegangen.

4.2.1 Design der Sanger-Primer

Für die fünf Gene *ATM*, *BRIP1*, *CHEK2*, *SMARCA4* und *STK11*, deren 5'-UTR-Bereiche nicht über die Panel-Diagnostik abgedeckt sind, wurden Sanger-Primer designt. Dafür wurde die jeweilige Sequenzinformation der 5'-UTR aus der aktuellen Ensembl-Version 112 genutzt und das Transkript mit Alamut™ Visual Plus abgeglichen. Mittels OligoPerfect™ erfolgte das Primerdesign. Die Länge der Targetsequenz in Basenpaaren und die zugelassene Region zur Primerbindung wurden so variiert, dass die 5'-UTR von dem Primerpaar mit ausreichend Abstand umgeben wird und die Primer möglichst keine Haarnadelstrukturen bilden. Sonstige Parameter, wie der GC-Gehalt wurden gemäß der Standardvorgaben belassen. Die ausgewählten Primerpaare wurden mit SNPCheck V3 auf relevante Einzelnukleotidpolymorphismen (SNPs) im Primerbereich geprüft. Bei vorhandenen SNPs, welche die Primerbindung beeinträchtigen können, wurde ein anderes Primerpaar gewählt. Außerdem wurde anhand der Software UCSC *In-Silico* PCR sichergestellt, dass die Primer nur die Zielsequenz einschließen und an keine weitere Stelle im Genom binden. Für *STK11* wurde aufgrund der Länge der Target-Region auf drei Primerpaare (A bis C) zurückgegriffen. Die erstellten Primer sind dem Abschnitt 3.2 zu entnehmen.

4.2.2 Etablierung der Sanger-Primer und Sanger-Sequenzierung

Um die designten Primer auf Funktionalität und Richtigkeit zu testen, wurde eine Etablierung der sieben Primerpaare an einer DNA-Probe durchgeführt. Dafür wurde von den Primerpaaren jeweils eine Primerverdünnung aus Forward und Reverse Primer hergestellt. Diese bestand aus je 10 µl Forward und Reverse Primer sowie 80 µl Wasser. Die DNA-Probe wurde auf 40 ng/µl verdünnt.

Für die Touchdown-PCR wurde ein Mastermix aus PCR-Puffer, dNTPs, Q-Solution, DNA-Polymerase und Wasser gemäß Tabelle 3.1 angesetzt. Die Q-Solution unterstützt

dabei die Stabilität der DNA. Anschließend wurden davon für jedes Primerpaar zweimal 24 µl in beschriftete Reaktionsgefäße pipettiert. Jede der Primerverdünnungen für die 5'-UTR von *ATM*, *BRIP1*, *CHEK2*, *SMARCA4* sowie *STK11* A, B und C wurde zu 1 µl in die zwei Reaktionsgefäße hinzugegeben. Jeweils ein Reaktionsgefäß wurde als Leerwert belassen. In das jeweils andere Reaktionsgefäß wurde 1 µl der verdünnten DNA pipettiert. Nach einem kurzen Anschneiden der Reaktionsgefäße wurden die Ansätze herunterzentrifugiert und in einen Thermocycler gestellt. Das PCR-Programm aus Tabelle 4.1 lief nun für etwa eine Stunde.

Tabelle 4.1: Touchdown PCR-Programm.

Temperatur [°C]	Zeit [min]	Wiederholungen
95	5	1x
95 60 72	0,5 0,5 1	3x
95 58 72	0,5 0,5 1	3x
95 55 72	0,5 0,5 1	29x
72	7	1x
4	∞	

Für die Gelelektrophorese als Kontrolle der PCR-Produkte wurde ein 1,4 %-iges Agarose-Gel aus 2,8 g Agarose und 200 µl 1x TBE-Puffer angesetzt. Der 10x aufkonzentrierte TBE-Puffer wurde vorher auf die gewünschte Konzentration verdünnt. Die Agarose-TBE-Puffer-Lösung wurde mehrfach aufgekocht bis keine Schlieren mehr vorhanden waren. Das Gel wurde anschließend für 20 Minuten mit einem Magnetrührer gerührt und nach dem Abkühlen mit 10 µl Midori Green versetzt. Danach wurde es auf den Gel-schlitten mit eingesetzten Kämmen gegossen. Nach Erstarren des Agarose-Gels wurden die Kämmen entfernt und die PCR-Produkte sowie die GeneRuler 100 bp Plus DNA Ladder in die Geltaschen pipettiert. Für 30 Minuten bei 150 V erfolgte nun die Auftrennung der DNA-Fragmente nach Größe. Im Anschluss konnte der genutzte Bereich des Gels mit einem Skalpell vorsichtig vom restlichen Gel getrennt und unter einem Transilluminator betrachtet werden. Verwendet wurde hierbei eine schonende Technologie mit blau/grüner Light Emitting Diode (LED).

Mit den Primerpaaren für die 5'-UTR von *ATM*, *SMARCA4* sowie *STK11* (Primer A, B und C) wurde erneut eine PCR angesetzt, da bei vorheriger PCR für diese Primerpaare

keine DNA-Banden auf dem Gel sichtbar waren. Der Reaktionsansatz ist Tabelle 3.2 zu entnehmen. Erneut wurde ein Leerwert pro Primerpaar pipettiert. Dem anderen Reaktionsansatz wurden je 2 μl DNA (40 ng/ μl) hinzugegeben. Es erfolgte eine PCR für etwa zwei Stunden nach dem Programm aus Tabelle 4.2.

Tabelle 4.2: Touchdown HQ PCR-Programm.

Temperatur [°C]	Zeit [min]	Wiederholungen
95	15	1x
95	0,5	3x
60	1	
72	0,5	
95	0,5	3x
58	1	
72	0,5	
95	0,5	29x
55	1	
72	0,5	
72	10	1x
4	∞	

Mit den PCR-Produkten erfolgte erneut eine Gelelektrophorese bei 150 V für 30 Minuten. Anschließend konnte das Gel im Transilluminator betrachtet werden.

Einzig für das Primerpaar, welches die 5'-UTR von *ATM* abdeckt, wurde keine ausreichend intensive DNA-Bande detektiert. Aus diesem Grund wurde ein erneuter PCR-Ansatz nach Tabelle 3.3 gestartet. Dabei wurde in Reaktionsansatz 1 die eingesetzte DNA-Menge verdoppelt und bei Ansatz 2 erfolgte eine Verdopplung des Volumens der eingesetzten DNA-Polymerase. Die PCR erfolgte ebenfalls nach dem Programm aus Tabelle 4.2. Die Gelelektrophorese wurde im Anschluss bei den bekannten Bedingungen durchgeführt.

Wurde die amplifizierte DNA als eindeutige Bande mit der entsprechenden Länge in dem Agarose-Gel identifiziert, konnten die PCR-Produkte aus dem Reaktionsansatz aufgereinigt werden. Dafür wurde ein Reinigungsmix gemäß Tabelle 3.4 pipettiert und im Thermocycler das Programm aus Tabelle 4.3 gestartet.

Tabelle 4.3: Programm zur Aufreinigung des PCR-Produktes.

Temperatur [°C]	Zeit [min]	Wiederholungen
37	4	1x
80	1	1x
4	∞	

Für die Sequenzierreaktion des gereinigten PCR-Produkts erfolgte das Pipettieren der Reaktionsansätze nach Tabelle 3.5. Das anschließende PCR-Programm aus Tabelle 4.4 lief etwa 75 Minuten.

Tabelle 4.4: Sequenzier-Programm für das aufgereinigte PCR-Produkt.

96 °C	2 min	1x
96 °C	10 sek	25x
50 °C	5 sek	
60 °C	2 min	
4 °C	∞	

Im Anschluss folgte die Ethanolfällung und Reinigung. Hierfür wurden zu jedem Reaktionsansatz 30 µl Ethanol (100 %) und 1 µl Natriumacetat hinzugegeben. Nach einem kurzen Invertieren der Proben wurden diese für zehn Minuten bei Raumtemperatur inkubiert und anschließend für 50 Minuten bei 4680 Umdrehungen pro Minute (rpm) zentrifugiert. Die Proben wurden anschließend mit etwas Schwung über dem Abfluss invertiert, um den Überstand zu entfernen. 100 µl Ethanol (70 %) wurden nun auf die pelletierten Proben gegeben. Es erfolgte eine erneute Zentrifugation bei 4680 rpm für zehn Minuten und der Überstand wurde verworfen. Für die Sicherstellung, dass das Ethanol vollständig entfernt wurde, wurden die Proben drei bis vier Minuten bei 95 °C mit offenem Deckel im Cycler getrocknet. Das Sequenzierprodukt wurde anschließend unter einem Abzug in 30 µl Hi-Di™ Formamid resuspendiert und bei Raumtemperatur für 5 Minuten inkubiert. Die Proben wurden nachfolgend komplett in eine 96-Well-Sequenzierplatte überführt und diese mit einer Silikon-Septe verschlossen. Leere Wells wurden mit mindestens 10 µl Hi-Di™ Formamid befüllt. Luftblasen wurden durch kurzes Zentrifugieren eliminiert. Nun konnte die Platte in den Applied Biosystems® Sanger Sequencing 3500 Series Genetic Analyzer geladen und die Kapillarsequenzierung gestartet werden. Pro 8er-Reihe der 96-Well-Platte wurden etwa 40 Minuten zur Sequenzierung benötigt. Im Anschluss können die Daten zur Auswertung in die SEQUENCE PILOT Software geladen werden.

Nach erfolgreicher Primer-Etablierung erfolgte die Sanger-Sequenzierung der Patientendaten Primerweise. Für die PCR-Reaktionen der Primerpaare für *BRIP1* und *CHEK2* wurde der Reaktionsansatz gemäß Tabelle 3.1 mit dem Touchdown-Programm aus Tabelle 4.1 angesetzt. Für Die 5'-UTRs der Primerpaare *SMARCA4* und *STK11* (Primer A, B und C) wurden jeweils der Reaktionsansatz aus Tabelle 3.2 und das PCR-Programm aus Tabelle 4.2 angewendet. Für *ATM* wurde sich ebenfalls für das Programm aus Tabelle 4.2 entschieden. Hierbei wurde jedoch gemäß Tabelle 3.3 (Ansatz 1) die eingesetzte DNA-Menge verdoppelt. Alle sieben Primerpaare wurden bei den 55 Patientinnen der Kohorte eingesetzt. Lediglich bei einer Patientin reichte die vorhandene DNA nur noch für die Sequenzierung der 5'-UTRs von *SMARCA4*, *BRIP1* und *CHEK2*.

4.3 Auswertung der Sangerdaten mittels Software

Die Auswertung der Sanger-Sequenzierungsdaten erfolgte manuell in der SEQUENCE PILOT Software For Genetic Analysis. Nach dem Einladen der Sequenzen in die Software werden unsaubere Sequenzenden so abgeschnitten, dass der Zielbereich noch ausreichend abgedeckt ist. Abweichungen der Sequenz zum Referenztranskript werden von der Software markiert. Diese werden dann manuell begutachtet. Es erfolgt eine Einschätzung, ob es sich um eine tatsächliche genetische Variante oder um eine unsaubere Sequenz handelt (siehe Anhang Abbildung A.1, A.2 und A.3). Hilfreich ist es dabei, dass die Sequenz aus beiden Richtungen sequenziert wurde. Somit sollte eine echte Variante immer auf beiden Strängen vorhanden sein. Nützlich ist außerdem der phred quality score, welcher die Qualität des base callings misst und Werte zwischen 0 und 99 annehmen kann. Dabei deutet ein Wert von unter 20 auf eine Fehleranfälligkeit des base callings von über 1 % hin [Liao, Satten und Hu, 2017]. Je höher der Wert ist, desto besser ist die Qualität des base callings. Sind alle Varianten technisch validiert, können sie der Variantenklassifikation unterzogen werden und medizinisch validiert werden.

4.4 Vorgehen der Variantenklassifikation

Für die Varianten, die aus den NGS-Daten stammen, wurde durch die Bioinformatik lediglich ein Filter auf die 5'-UTRs der 14 Core Gene gesetzt. Eine weitere Bedingung für die Filterung der Varianten war, dass die Variante in mindestens einem der 55 ungelösten FBREK-Fälle auftaucht.

Zu Beginn wurden die Varianten aus der Panel-Diagnostik hinsichtlich der Allelfrequenz (AF) und der Sequenziertiefe beurteilt. Die Sequenziertiefe oder auch Coverage gibt dabei an, wie viele Reads der betrachtete DNA-Abschnitt aufweist. Eine höhere Coverage erhöht die Genauigkeit der Sequenzierdaten. Anhand der AF konnte geschlussfolgert werden, ob es sich bei der Variante um eine homozygote ($AF \approx 1,0$) oder heterozygote ($AF \approx 0,5$) Variante handelt. Außerdem konnten Varianten mit einer sehr geringen Allelfrequenz (z.B. $AF \approx 0,03$) ausgeschlossen werden, da es sich hierbei vermutlich um Sequenzierartefakte handelt. Von Sequenzierartefakten spricht man, wenn während des Sequenzierprozesses Fehler durch beispielsweise technische Limitationen auftreten. Artefakte können auf Primerrückstände sowie Sequenzierfehler, etwa an repetitiven Stellen zurückzuführen sein [Graw, 2020]. Auch können Allelfrequenzen, welche deutlich unter 50 % liegen, Hinweise auf genetische Mosaikose sein. Somatischer Mosaizismus tritt dann auf, wenn eine genetische Variante nur in einem Teil der Körperzellen vorhanden ist. Er entsteht, wenn genetische Veränderungen nach der Befruchtung während der Zellteilung auftreten [Aluri und Cooper, 2023]. Für die hier betrachtete Fragestellung wurden vordergründig Varianten näher klassifiziert, die eindeutig homozygot oder heterozygot vorliegen.

Kommt eine Variante mit einem Prozentsatz von mehr als 10 % in der Kohorte bzw. mehr als 5 % in den genomischen Daten von [gnomAD](#) vor, so kann das Klassifikationskriterium BA1 vergeben werden. Es dient als eigenständiger Nachweis dafür, dass eine Variante benigne ist [Richards *et al.*, 2015]. $PM2_{\text{Supporting}}$ wird vergeben, wenn die Variante nicht in [gnomAD](#) vorhanden ist. Weisen jedoch andere Parameter darauf hin, dass die Variante unbedenklich ist, so wird die Variante in Kombination mit $PM2_{\text{Supporting}}$ als VUS eingestuft.

Der Δ -Score gibt die Wahrscheinlichkeit für die Beeinflussung des Spleißens durch die Variante an und reicht von 0 bis 1 [de Sainte Agathe *et al.*, 2023]. Die vorhergesagten Auswirkungen der Variante auf das Spleißen sind bei einem Δ -Score $\leq 0,1$ sehr gering [Richards *et al.*, 2015]. Daher kann BP4 vergeben werden. Bei einem Δ -Score $> 0,2$ kann PP3 vergeben werden. Das Kriterium $BP4_{\text{Supporting}}$ kann außerdem bei einem CADD zwischen 17,3 und 22,7 angewendet werden. Ein CADD zwischen 0,15 und 17,7 deutet auf BP4 (Moderat) hin [Richards *et al.*, 2015].

Wenn die Variante in [ClinVar](#) oder der [Human Gene Mutation Database](#) bereits beschrieben wurde, wurden vorhandene Literatur und funktionelle Studien in die Klassifikation einbezogen. Belegen funktionelle Studien die krankheitsverursachende Auswirkung der Variante auf das Gen oder das Genprodukt, so kann das Klassifikationskriterium PS3 vergeben werden [Richards *et al.*, 2015]. Wenn durch funktionelle Studien nachweislich keine Pathogenität von der Variante ausgeht, so wird BS3 angewendet.

Über Alamut™ Visual Plus wurde beurteilt, ob die Variante in der Kozak-Sequenz (von c.-6 bis c.3) [Kozak, 1987] liegt oder ein neues Startcodon gebildet wird. Außerdem werden die Varianten mit dem UTR-Annotator des [Ensembl Variant Effect Predictors](#) analysiert. Das Vorgehen der Variantenklassifikation ist zusammenfassend in Abbildung 4.1 dargestellt.

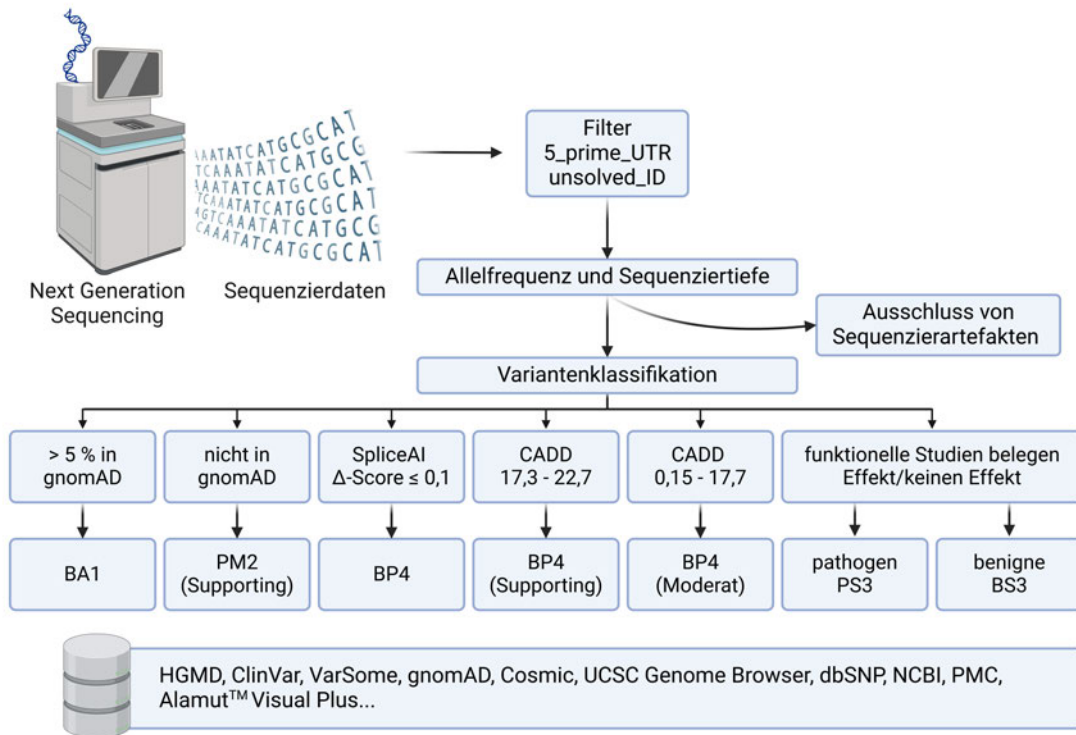


Abbildung 4.1: Überblick über das Varianten-Calling und die Klassifikation (erstellt mit BioRender.com).

In Abbildung 4.1 wird das Vorgehen der Variantenfilterung und der Klassifikation dargestellt. Die NGS-Daten werden nach 5'-UTR-Varianten gefiltert, welche in der Kohorte von Abbildung 1.4 vorkommen. Anschließend erfolgt ein Vorsortieren auf Grundlage der Allelfrequenz und der Abdeckung der Varianten. Die Klassifikation mithilfe von verschiedenen Scores und Datenbanken erfolgt im Anschluss. Ebendiese Klassifikation wird auch auf die Varianten angewendet, welche mit Sanger-Sequenzierung identifiziert wurden.

5 Ergebnisse und Diskussion

5.1 Etablierung der Sanger-Primer und Sanger-Sequenzierung

Die Abbildungen 5.1, 5.2 und 5.3 zeigen die Ergebnisse der Agarose-Gelelektrophorese bei der Etablierung der Sanger-Primer. Durch das Vorhandensein der erwarteten DNA-Bande kann von der korrekten Amplifikation der Zielsequenz mittels der Sanger-Primer ausgegangen werden. Für die Etablierung der Primerpaare wurden die Reaktionsbedingungen und PCR-Programme so lange angepasst, bis eine aussagekräftige Bande auf der richtigen Höhe im Gel und keine Nebenprodukte detektiert wurden.

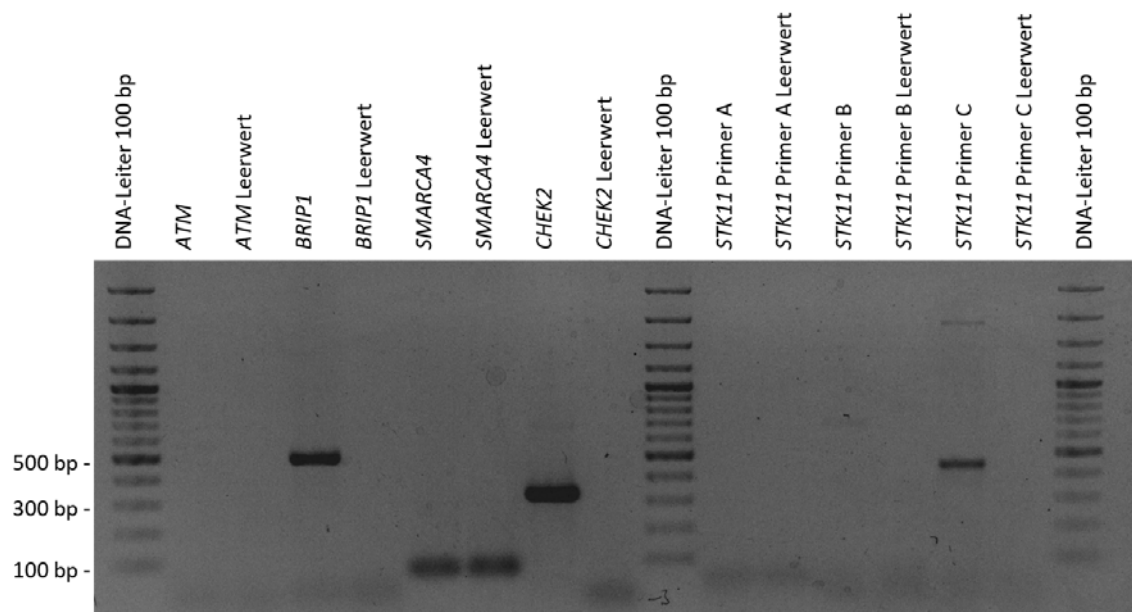


Abbildung 5.1: Etablierung der Sanger-Primer mit dem PCR-Programm Touchdown.

In Abbildung 5.1 sind alle sieben Primerpaare mit dem Reaktionsansatz aus Tabelle 3.1 und dem PCR-Programm aus Tabelle 4.1 bearbeitet worden. Deutliche DNA-Banden der korrekt amplifizierten Zielregionen zeigen sich bei den 5'-UTRs der Gene *BRIP1* (472 bp), *CHEK2* (300 bp) und bei Primer C von *STK11* (410 bp). Bei dem Ansatz von Primer C wurde außerdem ein unspezifisches Nebenprodukt mit einer Länge von über 1000 bp detektiert. Bei jedem Primer ist neben dem normalen Reaktionsansatz ein Leerwert im Gel aufgetragen. Als Längenreferenz der DNA-Banden dient eine 100 bp DNA-Leiter.

Das PCR-Programm Touchdown aus Tabelle 4.1 erzeugt bei *BRIP1*, *CHEK2* sowie bei Primer C der 5'-UTR von *STK11* eindeutige DNA-Banden auf der erwarteten Höhe im

Gel (siehe Abbildung 5.1). Bei Primer C von *STK11* ist jedoch zusätzlich ein unerwartetes Nebenprodukt von über 1000 bp Länge im Gel aufgetrennt. Dieses ist deutlich schwächer als die eigentliche DNA-Bande. Das der Primer an einer anderen Stelle im Genom bindet, wurde zuvor beim Primer-Design ausgeschlossen. Ein Grund für die zusätzliche Bande im Gel könnten Haarnadelstrukturen sein, die die Migration im Gel erschweren und somit zu einer scheinbar längeren DNA-Bande führen [Jernigan *et al.*, 1987]. Die Leerwerte der sieben Primerpaare weisen jeweils keine Verunreinigungen durch genomische DNA auf. Unterhalb des Längenmarkers von 100 bp sind schwach die Primer zu erkennen. Einzig bei *SMARCA4* ist sowohl in dem Ansatz mit als auch in dem Ansatz ohne Template-DNA ein unspezifisches Produkt mit einer Länge von etwas unter 100 bp zu sehen. Zurückzuführen ist dies wahrscheinlich auf Primerdimere, die durch einen kleinen komplementären Abschnitt zwischen den beiden Primern entstanden sind. Somit dient ein Primer als Template und das gewünschte PCR-Produkt wird nicht bzw. kaum gebildet [Mülhardt, 2013].

Das Fehlen eines korrekt amplifizierten PCR-Produktes bei den Primerpaaren von *ATM*, *SMARCA4* sowie *STK11* Primer A und B kann mit dem PCR-Programm bzw. der verwendeten Polymerase zusammenhängen. Um die Funktionalität der Primerpaare erneut zu überprüfen, wurde ein ebenfalls etabliertes Touchdown-PCR-Programm mit zugehörigem Reaktionsansatz eingesetzt (siehe Tabelle 3.2 und Tabelle 4.2). Die beiden PCR-Programme unterscheiden sich nur in den Zeiten der jeweiligen PCR-Phasen. Die Temperaturen sind bei beiden Touchdown-Programmen identisch. Zum einen weist das Programm aus Tabelle 4.2 eine initiale Denaturierung von 15 Minuten auf, während diese bei dem Programm aus Tabelle 4.1 nur fünf Minuten andauert. Durch eine längere Denaturierung wird sichergestellt, dass wirklich die gesamte DNA denaturiert vorliegt, was speziell bei genomischer DNA als Ausgangsmaterial wichtig ist. Liegt die DNA nicht komplett denaturiert vor, so kann dies negative Auswirkungen auf das Primer-Annealing haben [Müller und Prange, 2016].

Im Vergleich zu dem PCR-Programm aus Tabelle 4.1 ist die Annealingzeit bei jenem aus Tabelle 4.2 von 30 Sekunden auf 60 Sekunden verdoppelt. In der Regel schwankt die Annealingzeit zwischen 10 Sekunden und 60 Sekunden [Lorenz, 2012]. Eine längere Annealingzeit kann die Wahrscheinlichkeit der korrekten Primer-Bindung erhöhen und somit gleichzeitig die Ausbeute steigern. Außerdem benötigen GC-reiche Regionen länger für das Primer-Annealing.

Die Elongationszeit richtet sich nach der Länge der PCR-Produkte sowie der eingesetzten Polymerase [Müller und Prange, 2016]. In beiden Reaktionsansätzen (siehe Tabelle 3.1 und Tabelle 3.2) werden *Thermus aquaticus*-Polymerasen eingesetzt. Diese benötigen für 1000 bp eine Elongationszeit von 30 bis 60 Sekunden [Mülhardt, 2013]. Bei

beiden PCR-Programmen trifft dies zu.

Das Ergebnis der Primer-Etablierung mit dem zweiten PCR-Programm aus Tabelle 4.2 ist in Abbildung 5.2 zu sehen und zeigt bei *SMARCA4* sowie allen drei Primern für *STK11* DNA-Banden der erwarteten Länge. Erneut zeigt sich bei *ATM* keine DNA-Bande. Der Primer C von *STK11* wurde aufgrund der unspezifischen zusätzlichen DNA-Bande in Abbildung 5.1 ebenfalls erneut etabliert.

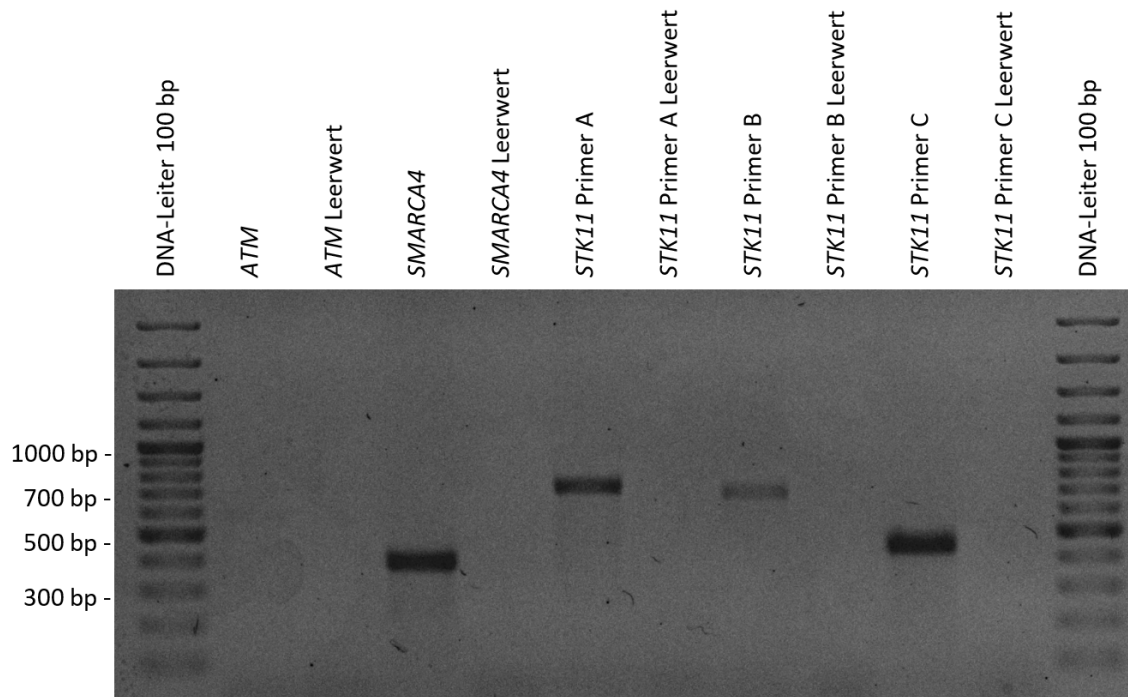


Abbildung 5.2: Etablierung der Sanger-Primer mit dem PCR-Programm Touchdown HQ.

Die zweite Etablierung (Abbildung 5.2) erfolgte mit dem Reaktionsansatz von Tabelle 3.2 und dem PCR-Programm aus Tabelle 4.2. Angewendet wurde dieses Programm auf die Primer für die 5'-UTRs der Gene *ATM*, *SMARCA4* und *STK11*. Die DNA-Bande des PCR-Produktes fehlt lediglich bei *ATM*. Das PCR-Produkt von *SMARCA4* weist eine Länge von 363 bp auf. Jenes von Primer A (*STK11*) ist 695 bp lang und das PCR-Produkt von Primer B (*STK11*) ist 651 bp lang. Das PCR-Produkt von Primer C der 5'-UTR von *STK11* ist 410 bp lang.

Das fehlende PCR-Produkt bei der 5'-UTR von *ATM* lässt beispielsweise auf nicht funktionierende Primer schlussfolgern. Außerdem kann es sein, dass sich die PCR aufgrund des vergleichsweise hohen GC-Gehaltes des PCR-Produktes als schwieriger erweist [Frey *et al.*, 2008]. Dem fehlenden Amplifikat kann etwa das Erhöhen der Menge an Template-DNA entgegenwirken [Müller und Prange, 2016].

Es wurden mit dem PCR-Programm aus Tabelle 4.2 für die 5'-UTR von *ATM* zwei Reaktionsansätze bearbeitet. Ersterer enthielt die doppelte Menge an eingesetzter DNA

und im zweiten Reaktionsansatz wurde die Menge der DNA-Polymerase verdoppelt. Das Ergebnis ist in Abbildung 5.3 dargestellt und weist für beide Reaktionsansätze eine schwache DNA-Bande der korrekten Länge von 541 bp auf.

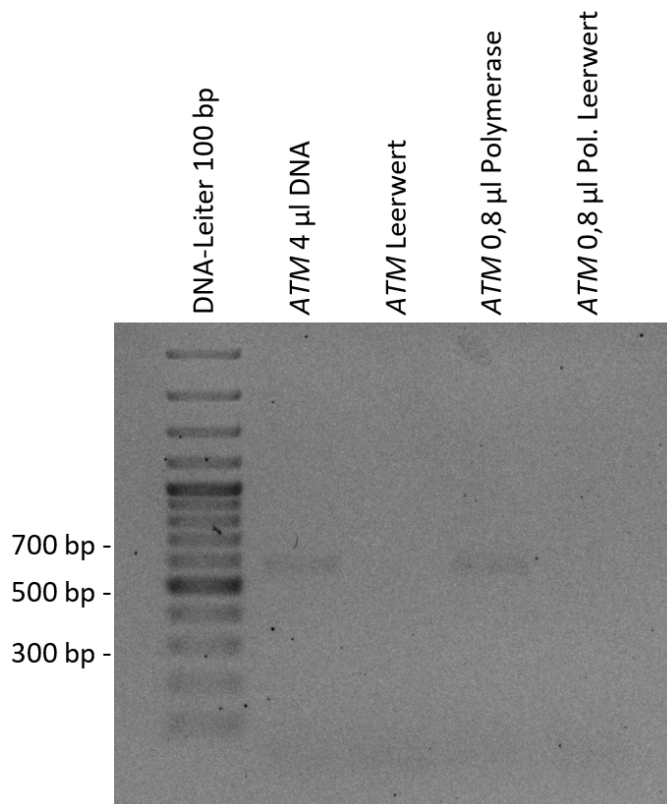


Abbildung 5.3: Etablierung der Sanger-Primer mit dem PCR-Programm Touchdown HQ für *ATM*.

Für die 5'-UTR von *ATM* wurden zwei Reaktionsansätze nach Tabelle 3.3 mit dem PCR-Programm aus Tabelle 4.2 angesetzt. Es zeigt sich in beiden Fällen eine schwache Bande des PCR-Produktes mit einer Länge von 541 bp.

Für die Überprüfungen der PCR-Produkte der 55 Proben wurden für die sieben Primerpaare jeweils stichprobenartig bis zu zehn Proben in der Gelelektrophorese aufgetrennt. Diese Gel-Kontrollen sind im Anhang in den Abbildungen B.1, B.2, B.3, B.4, B.5, B.6 und B.7 dargestellt.

5.2 Genetische Varianten aus Panel-Diagnostik und Sanger-Sequenzierung in den 5'-UTRs der Core Gene für FBREK

Im Nachfolgenden werden die identifizierten Varianten aus der Sanger-Sequenzierung sowie der Panel-Diagnostik beurteilt und unter anderem nach den aktuellen Richtlinien für Variantenklassifikationen der Association for Clinical Genomic Science klassifiziert [Stand Februar 2024; Originalleitlinien: Richards *et al.*, 2015]. Das Vorgehen der Klassifizierung ist dem Abschnitt 4.4 auf Seite 29 zu entnehmen. Anhand der Klassifizierung kann eine Einschätzung über die Relevanz der Varianten zur Erkrankung gegeben werden.

Für die identifizierten Varianten wurden verschiedene Informationen betrachtet, um eine Beurteilung vorzunehmen. Insgesamt wurden über die Panel-Diagnostik 60 Varianten identifiziert. Über die Sanger-Sequenzierung konnten 10 weitere genetische Varianten gefunden werden. Die Verteilung der Varianten auf die 5'-UTRs der 14 Core Gene ist in dem Diagramm aus Abbildung 5.4 dargestellt.

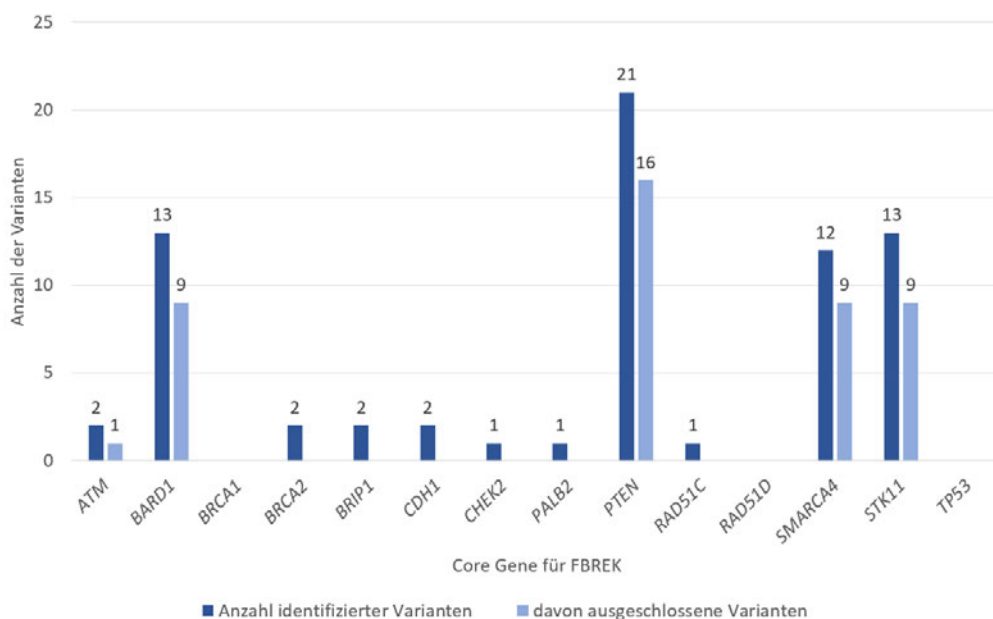


Abbildung 5.4: Verteilung der identifizierten Varianten auf die 5'-UTRs der Core Gene für FBREK.

Das Diagramm aus Abbildung 5.4 zeigt die Verteilung der Varianten auf die 5'-UTRs der 14 Core Gene. Wurden Varianten unter anderem aufgrund von technischen Limitationen von der Klassifikation ausgeschlossen, sind diese in einem helleren Farbton gekennzeichnet. In den 5'-UTRs der Gene *BRCA1*, *RAD51D* sowie *TP53* konnten in der betrachteten Kohorte keine Varianten gefunden werden. Insgesamt wurden von 70 identifizierten Varianten 44 Varianten ausgeschlossen.

Die 44 eliminierten Varianten sind in der Tabelle 5.1 dargestellt. Für alle Varianten ist neben dem Gen, in dessen 5'-UTR die Variante vorkommt, auch die Nomenklatur gemäß der Human Genome Variation Society (HGVS) angegeben. Es folgt das Chromosom sowie die genomische Position. Außerdem sind die Referenzbase(n) (REF) und die alternative(n) (ALT) Base(n) der Variante sowie die Häufigkeit in der Kohorte angegeben. Zusätzlich ist der CADD aufgelistet.

Tabelle 5.1: Überblick über die ausgeschlossenen Varianten.

Gen	HGVS	Chromosom	Position	REF	ALT	in Kohorte	CADD
<i>ATM</i>	c.-105G>A	11	g.108223112	G	A	1	8,08
<i>BARD1</i>	c.-64T>A	2	g.214809633	A	T	1	1,45
	c.-80T>G	2	g.214809649	A	C	1	5,04
	c.-105A>C	2	g.214809674	T	G	1	10,07
	c.-114A>C	2	g.214809683	T	G	1	7,38
	c.-75T>A	2	g.214809644	A	T	1	1,71
	c.-44T>G	2	g.214809613	A	C	2	4,08
	c.-91T>G	2	g.214809660	A	C	4	10,87
	c.-64T>G	2	g.214809633	A	C	22	1,54
	c.-99T>G	2	g.214809668	A	C	24	7,31
<i>PTEN</i>	c.-588A>G	10	g.87863882	A	G	1	20,60
	c.-273A>C	10	g.87864197	A	C	1	17,91
	c.-503A>C	10	g.87863967	A	C	1	19,44
	c.-607T>G	10	g.87863863	T	G	1	19,44
	c.-575T>G	10	g.87863895	T	G	1	16,96
	c.-623C>A	10	g.87863847	C	A	1	18,55
	c.-383A>C	10	g.87864087	A	C	1	18,90
	c.-829A>C	10	g.87863641	A	C	1	21,30
	c.-327T>G	10	g.87864143	T	G	1	18,43
	c.-305T>G	10	g.87864165	T	G	1	16,30
	c.-341T>G	10	g.87864129	T	G	5	17,99
	c.-596T>G	10	g.87863874	T	G	5	12,19
	c.-332T>G	10	g.87864138	T	G	11	18,91
	c.-385A>C	10	g.87864085	A	C	22	20,50
	c.-335A>C	10	g.87864135	A	C	52	17,15
c.-356A>C	10	g.87864114	A	C	54	20,10	
<i>SMARCA4</i>	c.-90C>G	19	g.10961116	C	G	1	15,31
	c.-80C>G	19	g.10961126	C	G	1	18,73
	c.-50A>C	19	g.10961156	A	C	1	16,87
	c.-147G>C	19	g.10961059	G	C	1	16,95
	c.-120T>G	19	g.10961086	T	G	1	18,17
	c.-111G>T	19	g.10961095	G	T	1	14,69
	c.-100A>C	19	g.10961106	A	C	1	17,19
	c.-72C>G	19	g.10961134	C	G	1	22,40
c.-55A>C	19	g.10961151	A	C	1	19,34	
<i>STK11</i>	c.-302G>C	19	g.1206612	G	C	1	8,05
	c.-296T>C	19	g.1206618	T	C	1	5,15
	c.-329C>T	19	g.1206585	C	T	1	8,50
	c.-841G>T	19	g.1206073	G	T	1	8,09
	c.-524T>G	19	g.1206390	T	G	1	10,65
	c.-456G>T	19	g.1206458	G	T	1	11,17
	c.-134T>G	19	g.1206780	T	G	1	8,30
	c.-290T>G	19	g.1206624	T	G	2	9,12
c.-107dup	19	g.1206797	C	CT	32	2,72	

Weitere genomische Daten aus gnomAD sind im Anhang in Tabelle C.1 gezeigt. Gibt es zu der Variante keine genomischen Daten bzw. ist sie nicht in gnomAD verzeichnet, ist dies vermerkt. Die Group Maximum Allele Frequency (Grpmax AF) bezieht sich auf die Population mit dem häufigsten Vorkommen der Variante in gnomAD. Es ist auch die Anzahl homozygoter Träger der Variante im Anhang in Tabelle C.1 verzeichnet. Wenn vorhanden, sind die ClinVar IDs sowie die rs-Nummern zu den Varianten im Anhang in Tabelle C.2 angegeben.

Alle 44 ausgeschlossenen Varianten haben keine PubMed-Referenz. Es handelt sich mit Ausnahme von c.-107dup (Duplikation) in *STK11* um Einzelnukleotidvarianten. Von den 44 Varianten wurden 13 aufgrund einer zu geringen Coverage von ≤ 11 Reads an der Position der Variante ausgeschlossen und 31 Varianten wiesen eine zu geringe Allelfrequenz ($AF \leq 0,14$) auf. Alle Varianten wurden trotzdem mit SpliceAI untersucht. Dabei trat bei keiner Variante ein Δ -Score $> 0,2$ auf. Lediglich die *ATM*-Variante c.-105G>A wies einen Δ -Score von 0,13 auf.

Bei 5 Varianten ergaben die Kriterien eine benigne Klassifikation. Diese Varianten sind in den Tabellen 5.2 und 5.3 dargestellt. Finden sich zu den Varianten PubMed-Einträge, so sind diese im Anhang in Tabelle C.3 zu finden. Die zugehörigen ClinVar IDs und rs-Nummern sind ebenfalls im Anhang in Tabelle C.4 aufgelistet.

Tabelle 5.2: Überblick über die Varianten mit einer benignen Klassifikation Teil I.

Gen	HGVS	Chromosom	Position	REF	ALT	in Kohorte	CADD
<i>BRCA2</i>	c.-52A>G	13	g.32315655	A	G	19	11,61
<i>BRIP1</i>	c.-31+12G>A	17	g.61863272	C	T	26	11,18
<i>CHEK2</i>	c.-7+62del	22	g.28741706	CA	C	2	6,41
<i>PALB2</i>	c.-47G>A	16	g.23641204	C	T	1	12,94
<i>STK11</i>	c.-311C>T	19	g.1206603	C	T	3	8,82

Tabelle 5.3: Überblick über die Varianten mit einer benignen Klassifikation Teil II.

Gen	HGVS	gnomAD Grpmax AF	gnomAD homozygot	Art	Kriterien	Klasse
<i>BRCA2</i>	c.-52A>G	0,1742	1995	SNV	BA1, BP4, BS3	benign
<i>BRIP1</i>	c.-31+12G>A	0,4909	5518	SNV	BA1, BP4	benign
<i>CHEK2</i>	c.-7+62del	0,01253	27	Deletion	BP4	likely benign
<i>PALB2</i>	c.-47G>A	0,09557	308	SNV	BA1, BP4	benign
<i>STK11</i>	c.-311C>T	0,02011	18	SNV	BP4	likely benign

Zu den benignen Varianten aus den vorangehenden Tabellen wurden sowohl *in-silico*-Daten wie der CADD und die Vorhersage von SpliceAI als auch vorhandene Literatur und funktionelle Studien betrachtet. Der CADD bewegt sich hier zwischen Werten von 6,41 bis 12,94 und lässt somit das Kriterium BP4 zu. Auch die SpliceAI-Vorhersagen

sind unauffällig. Drei Varianten kommen mit einer Häufigkeit von mehr als 5 % in gnomAD vor, weshalb hier BA1 vergeben werden kann. Zusätzlich wurde für eine Variante in funktionellen Studien nachgewiesen, dass diese keinen nennenswerten Effekt auf die Genexpression hat. Somit ist das Kriterium BS3 relevant. Wurde für eine Variante lediglich BP4 vergeben, reicht dies nicht für eine klare Zuordnung zu benigne aus und die Variante wird als wahrscheinlich benigne eingestuft.

Die *BRCA2*-Variante c.-52A>G wurde von Fraile-Bethencourt *et al.* (2018) mittels Luciferase-Assays in ihrer Wirkung untersucht. Dabei zeigte sich keine wesentliche Veränderung der Promotoraktivität. Diese wies mit der Variante noch 97 % auf. Sowohl in der Kohorte als auch in gnomAD kommt die Variante mit 34,5 % und 17,4 % recht häufig vor, was ebenfalls gegen eine krankheitsverursachende Wirkung spricht.

Die intronische *CHEK2*-Variante c.-7+62del ist durch den Spleißvorgang nicht mehr in der fertigen mRNA vorhanden und somit unauffällig.

Zu der *PALB2*-Variante c.-47G>A gibt es bisher keine funktionellen Studien. In einer Studie zu *PALB2*-Varianten bei familiärem Brustkrebs wurde diese Variante zwar detektiert, jedoch nicht als pathogen eingestuft [Hellebrand *et al.*, 2011]. Ebenso wurde die Variante in Studien zu spanischen und italienischen Brustkrebsfällen berichtet, jedoch nicht näher betrachtet [Balía *et al.*, 2010; Garcia *et al.*, 2009]. In einer ebenfalls spanischen Studie von Blanco *et al.* (2013) wurde für diese *PALB2*-Variante vorhergesagt, dass sie eine kryptische Spleißstelle zerstört. Kryptische Spleißstellen sind Spleißstellen, welche im Normalfall keine Verwendung finden. Findet dort aus ungeklärten Umständen doch ein Spleißvorgang statt, so führt dies zu einem defekten Genprodukt [Brown, 2007]. Dies bestätigt unter anderem die benigne Einstufung der Variante. In einer Assoziationsstudie zu Brustkrebs von Wu *et al.* (2018) wurde für diese Variante keine Signifikanz in Bezug zu Brustkrebs detektiert.

Letztendlich wurden 21 Varianten als VUS eingestuft. Informationen hierzu sind den Tabellen 5.4 und 5.5 zu entnehmen. Bei vorhandenen PubMed-Einträgen sind diese im Anhang in Tabelle C.5 angegeben. Tabelle C.6 im Anhang zeigt die ClinVar IDs und rs-Nummern der als VUS eingestuftten Varianten. Werden widersprüchliche Kriterien für eine Variante vergeben, wird diese als VUS eingestuft. Der CADD bei den 21 Varianten unklarer Signifikanz bewegt sich zwischen 1,61 und 20,5 und deutet so auf das Kriterium BP4 hin. Für einige der Varianten belegen funktionelle Studien, dass sie an der Pathogenität beteiligt sind (PS3). Zwei der 21 Varianten kommen nicht in den genomischen Daten von gnomAD bzw. gar nicht in gnomAD vor, wodurch PM2_{Supporting} vergeben wird.

Tabelle 5.4: Überblick über die als VUS eingestuftten Varianten Teil I.

Gen	HGVS	Chromosom	Position	REF	ALT	in Kohorte	CADD
<i>ATM</i>	c.-111G>A	11	g.108223106	G	A	50	10,46
<i>BARD1</i>	c.-48T>C	2	g.214809617	A	G	1	3,11
	c.-83C>T	2	g.214809652	G	A	2	10,49
	c.-78G>A	2	g.214809647	C	T	2	9,70
	c.-30G>C	2	g.214809599	C	G	52	8,10
<i>BRCA2</i>	c.-26G>A	13	g.32316435	G	A	29	9,46
<i>BRIP1</i>	c.-205G>A	17	g.61863458	C	T	35	5,78
<i>CDH1</i>	c.-54G>C	16	g.68737362	G	C	1	9,68
	c.-71C>G	16	g.68737345	C	G	1	13,53
<i>PTEN</i>	c.-734G>A	10	g.87863736	G	A	1	20,50
	c.-534_-532dup	10	g.87863921	C	CGCG	1	18,80
	c.-511G>A	10	g.87863959	G	A	10	20,50
	c.-488_-486dup	10	g.87863967	A	AGCG	2	18,50
	c.-326G>C	10	g.87864144	G	C	54	16,70
<i>RAD51C</i>	c.-26C>T	17	g.58692618	C	T	22	1,61
<i>SMARCA4</i>	c.-147G>C	19	g.10961059	G	C	1	16,95
	c.-72C>T	19	g.10961134	C	T	1	19,37
	c.-43C>T	19	g.10961163	C	T	1	15,02
<i>STK11</i>	c.-590C>T	19	g.1206324	C	T	1	6,83
	c.-554G>A	19	g.1206360	G	A	1	6,33
	c.-973C>G	19	g.1205941	C	G	1	15,46

Tabelle 5.5: Überblick über die als VUS eingestuftten Varianten Teil II.

Gen	HGVS	gnomAD Grpmax AF	gnomAD homozygot	Art	Kriterien	Klasse
<i>ATM</i>	c.-111G>A	0,5998	20089	SNV	BA1, BP4, PS3	VUS*
<i>BARD1</i>	c.-48T>C	0,7833	41405	SNV	BA1, BP4	VUS*
	c.-83C>T	0,03193	49	SNV	BP4, PS3	VUS
	c.-78G>A	0,7819	31168	SNV	BA1, BP4	VUS*
	c.-30G>C	0,7801	41389	SNV	BA1, BP4	VUS*
<i>BRCA2</i>	c.-26G>A	0,3727	3886	SNV	BA1, BP4, PS3	VUS
<i>BRIP1</i>	c.-205G>A	0,7777	15228	SNV	BA1, BP4	VUS*
<i>CDH1</i>	c.-54G>C	0,003597	0	SNV	BP4, PS3	VUS
	c.-71C>G	0,013	8	SNV	BP4	VUS*
<i>PTEN</i>	c.-734G>A	0,0000802	0	SNV	BP4	VUS*
	c.-534_-532dup	0,016	22	Insertion	BP4	VUS*
	c.-511G>A	0,3351	1493	SNV	BA1, BP4, PS3	VUS
	c.-488_-486dup	0,003062	3	Insertion	BP4	VUS*
	c.-326G>C	0,9936	75955	SNV	BA1	VUS*
<i>RAD51C</i>	c.-26C>T	0,2087	2768	SNV	BA1, BP4, PS3	VUS
<i>SMARCA4</i>	c.-147G>C	0,001128	0	SNV	BP4	VUS*
	c.-72C>T	0,002368	4	SNV	BP4	VUS*
	c.-43C>T	0,003184	2	SNV	BP4	VUS*
<i>STK11</i>	c.-590C>T	0,0006556	0	SNV	BP4	VUS*
	c.-554G>A	keine Daten	-	SNV	BP4, PM2 ^{Supporting}	VUS
	c.-973C>G	n. in gnomAD	-	SNV	BP4, PM2 ^{Supporting}	VUS

VUS* = Variante mit konträren Studien oder Gewinn/Verlust eines alternativen Startcodons; kein Kriterium für diesen Sachverhalt vergebbar; Einstufung als VUS abweichend von den Klassifikationskriterien

Bei einigen Varianten ergeben die Kriterien keine Klassifikation als VUS, da lediglich BP4 bzw. BA1 vergeben wurden. Diese Varianten wurden trotzdem als VUS eingestuft, da entweder konträre Studien zu den Varianten existieren oder ein neues alternatives Startcodon gebildet wird. Für diesen Fall lässt sich kein Kriterium eindeutig vergeben (in Tabelle gekennzeichnet mit VUS*). Tatsächlich wurde mittlerweile nachgewiesen, dass auch bei Menschen [Cao und Slavoff, 2020] die Translation nicht nur an der Basenabfolge AUG initiiert werden kann. So ist vor allem in den UTRs die Translationsinitiation an einer Vielzahl von Startcodons möglich [Andreev *et al.*, 2022]. Daraus entspringt eine vielseitige Zusammensetzung des Proteoms [Andreev *et al.*, 2022]. Zum einen kann ausgehend von einer alternativen Translationsinitiationsstelle ein neuer ORF entstehen oder es kommt im Fall eines *in-frame*-Startcodons zu einer N-terminalen Verlängerung des Proteins. Nicht-AUG-Startcodons sind sowohl an der Expressionsregulation als auch an der Entstehung von Krankheiten beteiligt [Diaz de Arce, Noderer und C Wang, 2018]. Auch mehrere Gene, welche bei Krebs eine Rolle spielen, werden nicht-AUG initiiert [Andreev *et al.*, 2022]. Es können auch kleine ORFs entstehen, deren mRNA eine regulatorische Funktion hat. Das nicht-AUG-Startcodon erfordert einen spezifischen Sequenzkontext [Diaz de Arce, Noderer und C Wang, 2018], wobei vor allem die Positionen -4 bis +4 [Andreev *et al.*, 2022] relevant sind und sich der Nukleotidkontext auf jedes Startcodon anders auswirkt. Meist handelt es sich um palindromische Sequenzen. Im 'Idealfall' kann die Expressionseffizienz der eines AUG-Startcodons ähneln [Diaz de Arce, Noderer und C Wang, 2018]. Durch nachgeschaltete Sekundärstrukturen kann die Initiationseffizienz an der Stelle noch verstärkt werden [Andreev *et al.*, 2022]. Diaz de Arce, Noderer und C Wang (2018) führten Expressionsanalysen mit verschiedenen Startcodons in Zellkulturen durch. Dabei zeigten sich vor allem bei CUG, GUG und ACG Initiationseffizienzen zwischen 18 und 50 % in Bezug zu der von AUG. CUG ist meist zwei bis dreimal effizienter als andere Nicht-AUG-Codons [Andreev *et al.*, 2022]. Auch AUU kann als alternatives Startcodon genutzt werden [Tzani *et al.*, 2016]. Von Cao und Slavoff (2020) wurden zusätzlich zu den bereits erwähnten noch UUG und AUC identifiziert. Beispielsweise eine *BARD1*-Variante sorgt laut Literatur für einen vorzeitigen Startcodon-Gewinn [Da Costa E Silva Carvalho *et al.*, 2020]. Dabei handelt es sich um die Variante c.-83C>T. In der Publikation ist nicht weiter beschrieben, um welches Startcodon es sich handelt. Es entstehen die drei folgenden neuen Codon-Möglichkeiten: GTT, TTG und TGC. Auf mRNA-Ebene entsteht also UUG, welches von Cao und Slavoff (2020) als mögliches Startcodon identifiziert wurde. Alle Varianten wurden auf alternative Startcodons untersucht. In Tabelle 5.6 sind die Varianten aufgelistet, bei denen es sich in der Referenzsequenz um eines der eben erwähnten nicht-AUG-Startcodons handelt oder bei denen durch die Variante ein neues Startcodon entsteht.

Tabelle 5.6: Nicht-AUG-Startcodons in den als VUS eingestuftten Varianten.

Gen	Variante	alternatives Startcodon in Referenz	alternatives Startcodon durch Variante
<i>ATM</i>	c.-111G>A	nein	ACG
<i>BARD1</i>	c.-83C>T	nein	UUG
	c.-30G>C	CUG	nein
<i>BRIP1</i>	c.-205G>A	nein	AUC
<i>CDH1</i>	c.-71C>G	nein	CUG
<i>PTEN</i>	c.-326G>C	GUG	nein
<i>SMARCA4</i>	c.-147G>C	UUG	nein
	c.-72C>T	nein	GUG
	c.-43C>T	nein	CUG
<i>STK11</i>	c.-590C>T	nein	CUG
	c.-554G>A	UUG	nein

Bei sieben der Varianten aus Tabelle 5.6 entsteht ein neues alternatives Startcodon. Bei drei dieser sieben Varianten entsteht das effiziente CUG. Bei vier Varianten aus Tabelle 5.6 besteht in der Referenzsequenz eine mögliche alternative Startstelle, welche durch die Variante verloren geht. Zwar kann keine eindeutige Aussage darüber getroffen werden, ob genau diese Startcodons auch genutzt werden, aber die Konservierung der Sequenz um und nach dem Startcodon kann einen Hinweis geben, ob eine Translationsinitiation an der Startstelle wahrscheinlich ist. Dafür wird der PhyloP-Score herangezogen. Dieser dient der Beurteilung der evolutionären Konservierung einzelner Nucleotide über verschiedene Spezies hinweg [URL-6, 2024]. Hohe positive Werte deuten auf evolutionäre Konservierung hin und somit auf eine mögliche funktionelle Relevanz der Basenposition. Negative Werte belegen, dass die Stellen einer schnellen evolutionären Veränderung unterliegen [URL-6, 2024]. Bei fast allen Varianten mit alternativem Startcodon ist die Konservierung vergleichsweise gering. Da jedoch nichts Gegenteiliges für die Varianten belegt ist, erfolgt die Einstufung als VUS. Bei der *PTEN*-Variante c.-326G>C jedoch findet sich eine flächendeckend höhere Konservierung, als es normalerweise für nicht kodierende Bereiche üblich ist. Im Fall dieser Variante wird ein mögliches alternatives Startcodon (GUG) eliminiert. Dies ist in Abbildung 5.5 gezeigt.

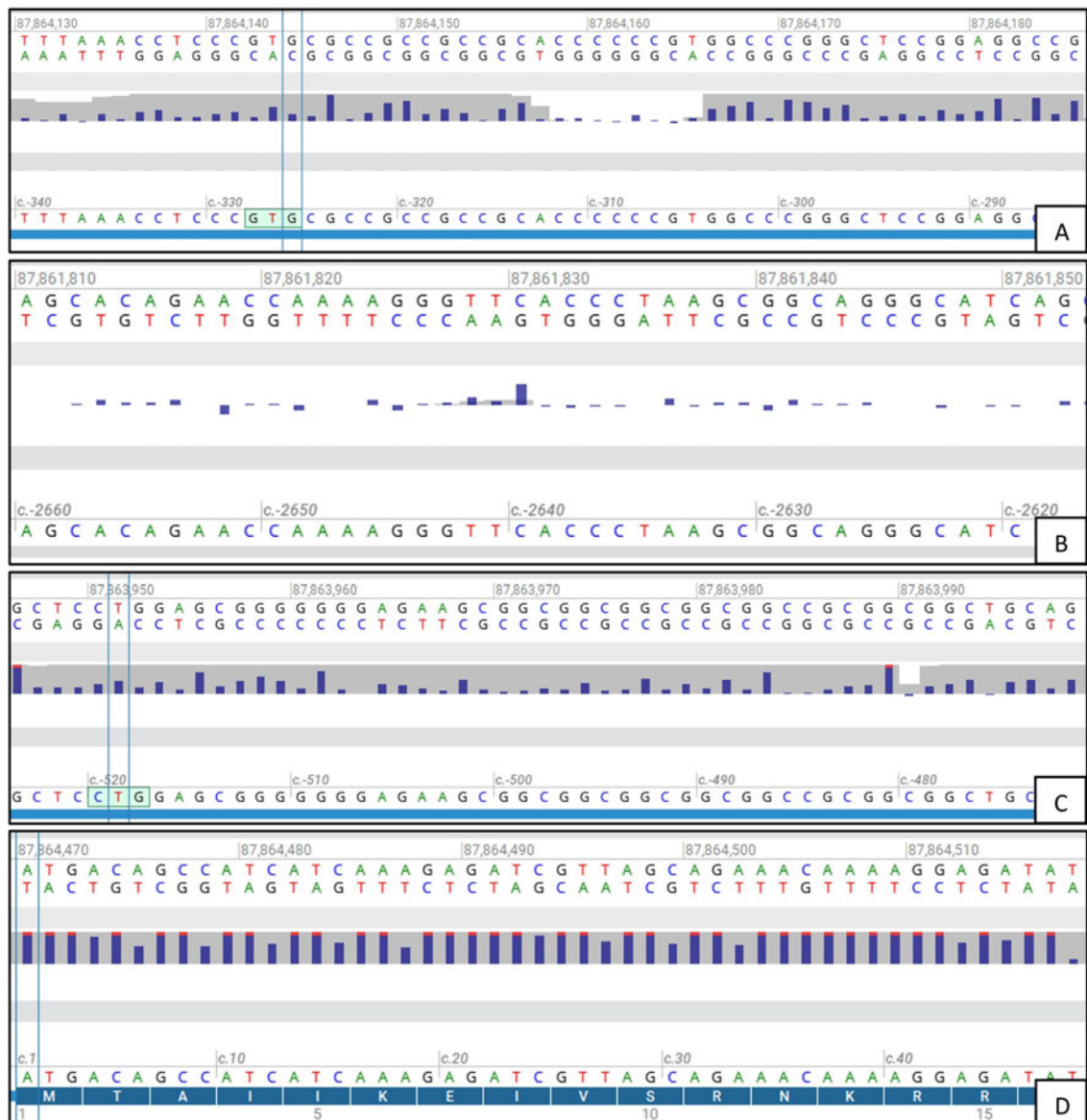


Abbildung 5.5: Konservierung einzelner Sequenzbereiche am Beispiel *PTEN*.

Abbildung 5.5 gibt einen Einblick in die Konservierung verschiedener Sequenzbereiche upstream von *PTEN*, aus der 5'-UTR von *PTEN* sowie in der kodierenden Region von *PTEN*. Die Konservierungsdaten beziehen sich auf den PhyloP-Score.

(A) Gezeigt ist die Referenzsequenz um und nach dem alternativen Startcodon GUG (GTG), an Position c.-326, wobei die Konservierung der Region mit den blauen Balken gekennzeichnet ist. Diese sind überwiegend deutlich im positiven Bereich und somit die Sequenz vergleichsweise konserviert.

(B) Die Sequenz aus diesem intronischen Bereich upstream der 5'-UTR von *PTEN* ist kaum konserviert, was an den Konservierungsbalken zu erkennen ist. Sie dient dem Vergleich der Konservierungswerte.

(C) Das markierte Startcodon CUG (CTG) an Stelle c.-519 kennzeichnet die Initiationsstelle von *PTEN* α . Die Konservierungen an und nach dem Startcodon sind ähnlich zu jenen aus Abbildung 5.5 (A).

(D) Als Referenz ist die Konservierung zu Beginn der kodierenden Sequenz von *PTEN* gezeigt. Die Konservierungswerte sind hier deutlich höher als im Bereich upstream des AUG-Startcodons.

In Bezug auf *PTEN* gibt es mehrere Isoformen mit einer N-terminalen Verlängerung des normalen *PTEN*-Proteins. Vor allem *PTEN*-Long [Hopkins, Fine *et al.*, 2013] (auch *PTEN* α genannt [Liang, He *et al.*, 2014]), *PTEN* β und *PTEN* ϵ sind hier zu nennen [Liang, X Chen *et al.*, 2017]. Das klassische Protein *PTEN* reguliert das Zellwachstum und dient als Tumorsuppressor und Phosphatase. Durch die Hemmung von zellulären Signalwegen ist *PTEN* in der Lage, die Apoptose von entarteten Zellen zu fördern. Außerdem interagieren viele Proteine mit *PTEN*, die bei der tumorsuppressiven Funktion relevant sind [Hopkins, Hodakoski *et al.*, 2014].

Das N-terminal verlängerte *PTEN* α wurde von Hopkins, Fine *et al.* (2013) entdeckt und entspringt dem alternativen Startcodon CUG 519 bp vor dem Startcodon von *PTEN*. Die zusätzliche Domäne ist 173 Aminosäuren lang und evolutionär über mehrere Spezies hinweg konserviert [Hopkins, Fine *et al.*, 2013]. Die Funktion von *PTEN* α ähnelt größtenteils jener von *PTEN*. *PTEN* α kann jedoch im Gegensatz zu *PTEN* aus der Zelle sezerniert werden [Hopkins, Hodakoski *et al.*, 2014]. Außerdem kann es in andere Zellen eindringen, da es einen Polyargininabschnitt besitzt, durch den die Zellwand durchdrungen werden kann [Hopkins, Hodakoski *et al.*, 2014]. Somit ist *PTEN* α essenziell für die Tumorsuppression auch in angrenzenden Zellen und Geweben. Wichtig ist *PTEN* α vor allem im Mitochondrien-Metabolismus, da es an der Induktion der ATP-Produktion beteiligt ist. Dabei interagiert *PTEN* α mit *PTEN* [Liang, He *et al.*, 2014].

Die Translation von *PTEN* β wird 438 bp upstream des Startcodons von *PTEN* initiiert. Der Initiation dient eine nahezu palindromische Sequenz um das Startcodon AUU herum [Liang, X Chen *et al.*, 2017]. Das Protein *PTEN* β verlängert *PTEN* N-terminal um 146 Aminosäuren. Diese Isoform ist vor allem im Nukleolus lokalisiert und übt dort Interaktionen zu Nucleolin aus, indem es als Phosphatase fungiert. Somit spielt *PTEN* β eine wichtige Rolle bei der Ribosomen-Synthese [Liang, X Chen *et al.*, 2017].

PTEN ϵ wird ausgehend von einem CUG-Startcodon in der 5'-UTR von *PTEN* translatiert, welches 216 bp vom AUG des *PTEN*s entfernt liegt und *PTEN* somit um 72 Aminosäuren verlängert. Es befindet sich hauptsächlich in der Zellmembran und kann somit Tumorzellen destabilisieren [Q Zhang *et al.*, 2021]. Es zeigt sich also für alle drei Isoformen des Proteins *PTEN*, dass dadurch wichtige Funktionen übernommen werden. Missense-Varianten in den Isoformen, welche in Bezug auf *PTEN* in der 5'-UTR liegen, stören also mit hoher Wahrscheinlichkeit die Proteinfunktion. Nachgewiesen ist für einige Varianten, dass die Phosphatase-Aktivität und die Bindungsaffinität zu anderen Proteinen gestört werden [Liang, X Chen *et al.*, 2017].

Da die Konservierung im Bereich der *PTEN*-Variante c.-326G>C jener im Bereich des Startcodons von *PTEN* α ähnelt (siehe Abbildung 5.5), ist es denkbar, dass durch den Verlust des alternativen Startcodons eine Störung der Proteinfunktion vorliegt, sollte je-

nes Startcodon für eine weitere Isoform oder regulatorische mRNA genutzt werden. Außerdem liegt diese Variante sowohl in *PTEN α* als auch in *PTEN β* . Betrachtet man die dabei entstehende Polypeptidkette, so wirkt c.-326G>C in beiden Fällen als Missense-Variante. Da diese *PTEN*-Variante jedoch sowohl in der Kohorte als auch in gnomAD mit mehr als 98 % vorkommt, ist der eigentliche Wildtyp wesentlich seltener vorhanden. Betrachtet man diesen stattdessen als 'Variante', würde also ein alternatives Startcodon entstehen.

Die *PTEN*-Variante c.-511G>A liegt in Bezug auf *PTEN α* an der Position c.10G>A. Damit bewirkt diese Variante einen Aminosäure-Austausch von Glycin zu Arginin [Ferri *et al.*, 2019]. Für diesen Austausch ist im Zusammenhang mit chronischer myeloischer Leukämie ein erhöhtes Erkrankungsrisiko beschrieben [Ferri *et al.*, 2019]. Der Aminosäure-Austausch von einer kleinen, unpolaren Aminosäure ohne Seitenkette (Glycin) zu einer komplexen, polaren und positiv geladenen Aminosäure (Arginin) in der Polypeptidkette erhöht die Wahrscheinlichkeit, dass negative Auswirkungen auf die Struktur bzw. Funktion des Proteins folgen.

Die zwei *PTEN*-Varianten c.-734G>A und c.-534_-532dup befinden sich beide in keiner der eben beschriebenen Isoformen. Allerdings liegen die Varianten in einem Bereich, in dem mehrere Varianten mit Brustkrebs in Verbindung gebracht werden [Ren *et al.*, 2021]. In der HGMD werden diese Varianten als eventuell krankheitsverursachend eingestuft, wobei die Pathogenität noch ungeklärt ist. Die *PTEN*-Variante c.-734G>A wird außerdem in ClinVar als VUS geführt. Die Duplikation c.-488_-486dup liegt in-frame in *PTEN α* und ebenfalls in der Region mit den Varianten unklarer Signifikanz in Bezug auf Brustkrebs [Ren *et al.*, 2021].

Die *ATM*-Variante c.-111G>A ist vor allem bei Patienten mit asiatischer Herkunft mit verschiedenen Krebsarten verbunden. Die Variante befindet sich in der Promotorregion von *ATM* und verändert die Sekundärstruktur [Zhao *et al.*, 2019]. In einer Studie von Yue *et al.* (2018) wurde diese Variante bei 524 Brustkrebsfällen und 518 Kontrollen der weltgrößten Volksgruppe (Han-Chinesen) genotypisiert. Bei der Studie handelt es sich um eine Assoziationsstudie zwischen den untersuchten Varianten und der Kontroll- bzw. Erkrankengruppe. Die Varianten wurden nicht funktionell untersucht. Dabei waren zwei Varianten, darunter c.-111G>A, deutlich mit dem Brustkrebsrisiko assoziiert. Die Variante zeigte sich bei erkrankten Personen wesentlich öfter als bei gesunden Kontrollen. Der homozygote Genotyp AA deutet auf eine zweifach höhere Erkrankungswahrscheinlichkeit hin [Yue *et al.*, 2018]. Auch ein Haplotyp aus vier SNPs, der die *ATM*-Variante einschließt, ist relevant. Als Haplotyp bezeichnet man einen Sequenzbereich der DNA, welcher durch spezifische Polymorphismen geprägt ist [Arnemann *et al.*, 2019]. Bestimmte Haplotypen können mit der Zugehörigkeit zu einer Ethnie oder mit bestimmten Krankheitsrisiken korrelieren [Arnemann *et al.*, 2019]. Für den Haplotyp aus *ATM*

wurden verschiedene Zusammensetzungen der vier SNPs betrachtet. Das A-Allel der Variante c.-111G>A war im Zusammenhang mit dem Haplotyp einmal mit einem 1,98-fachen und einmal mit einem 6,04-fachen Brustkrebsrisiko verbunden [Yue *et al.*, 2018]. Dieser Zusammenhang ist in Abbildung 5.6 gezeigt.

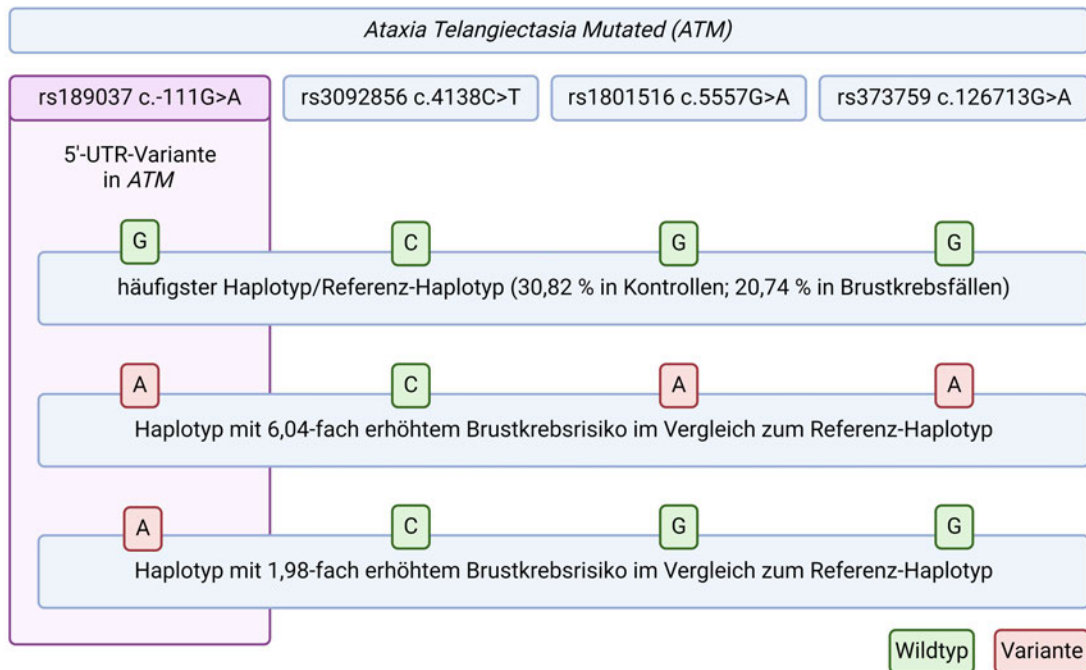


Abbildung 5.6: Brustkrebsrisiko für die *ATM*-Variante c.-111G>A im Verbund eines Haplotyps nach Yue *et al.* (2018) (erstellt mit BioRender.com).

In Abbildung 5.6 sind verschiedene Konstellationen des von Yue *et al.* (2018) untersuchten Haplotyps des *ATM*-Gens dargestellt. Der häufigste Haplotyp beinhaltet sowohl bei den Kontrollen als auch bei den Brustkrebsfällen die jeweiligen Wildtyp-Allele. Für die Zusammensetzung A-C-A-A wurde ein 6,04-faches Brustkrebsrisiko im Vergleich zum Referenzhaplotyp beobachtet und für die Kombination A-C-G-G lag das beobachtete Risiko bei einer 1,98-fachen Erhöhung. Andere Haplotyp-Konstellationen wurden nicht mit einem erhöhten Brustkrebsrisiko in Verbindung gebracht [Yue *et al.*, 2018].

Auch mit anderen Tumorrisiken ist die *ATM*-Variante verbunden. So wird etwa ein moderates Risiko im Zusammenhang mit einem Bronchialkarzinom angegeben [de Alencar, Formiga und de Lima, 2020]. Auch Untersuchungen von H Wang *et al.* (2010) deuten darauf hin, dass diese Variante mit Brustkrebs assoziiert ist. Insgesamt stammen die meisten Veröffentlichungen zu dieser *ATM*-Variante aus dem asiatischen Raum. Von Zhao *et al.* (2019) wurde eine Metaanalyse über 15 Studien zu ebendieser Variante veröffentlicht, die einen Überblick über Tumorrisiken der Variante gibt. Es besteht ein eindeutiger Zusammenhang zwischen dem A-Allel und Bronchialkarzinomen, Mammakarzinomen sowie Mundhöhlenkarzinomen. In dieser Studie wurde die Stärke der As-

soziation zwischen der Variante und dem Tumorrisiko mittels Quotenverhältnis (Odds Ratio) beurteilt. Als Odds Ratio bezeichnet man ein statistisches Maß, welches behilflich ist, die Stärke des Zusammenhangs zwischen einer Exposition und einem bestimmten Ergebnis zu quantifizieren [Arneemann *et al.*, 2019]. Dabei werden die Quoten verglichen, dass beispielsweise eine Krankheit bei der Gruppe, die einem Risikofaktor (hier der Variante) ausgesetzt (exponiert) ist, im Vergleich zu einer nicht exponierten Gruppe auftritt. Die Berechnung erfolgt nach folgender Formel:

$$\text{Odds Ratio} = \frac{\text{Odds in der exponierten Gruppe}}{\text{Odds in der nicht exponierten Gruppe}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \cdot d}{b \cdot c}$$

a = Anzahl der Personen in der exponierten Gruppe, die erkrankt sind

b = Anzahl der Personen in der exponierten Gruppe, die nicht erkrankt sind

c = Anzahl der Personen in der nicht exponierten Gruppe, die erkrankt sind

d = Anzahl der Personen in der nicht exponierten Gruppe, die nicht erkrankt sind

Ein Odds Ratio von 1 spricht dafür, dass keine Assoziation vorliegt [Arneemann *et al.*, 2019]. Liegt der Wert > 1 , so ist die Erkrankung in der Kohorte mit der Variante wahrscheinlicher. Bei einem Wert < 1 ist die Erkrankung in der exponierten Gruppe weniger wahrscheinlich. In der Metaanalyse von Zhao *et al.* (2019) ergab sich ein Odds Ratio von 1,207 (95 % Konfidenzintervall) zwischen der ATM-Variante c.-111G>A und den zuvor erwähnten Tumorerkrankungen. Für Europäer gibt es zu dieser Variante erst eine Publikation aus Belarus [Damiola *et al.*, 2014], welche ebenfalls in der Metaanalyse inkludiert ist. Der hierbei vergleichsweise geringe Probenumfang lässt keine sichere Aussage über die Bedeutung der Variante bei Europäern zu. Daher kann die Pathogenität der Variante bei Europäern nicht eigenständig beurteilt werden und es können nur die asiatischen Ergebnisse herangezogen werden. Allgemein kann die Variante als Biomarker für ein erhöhtes Krebsrisiko angesehen werden [Zhao *et al.*, 2019].

Die BARD1-Variante c.-48T>C verändert eine Transkriptionsbindestelle und hat außerdem ein mittleres Risiko für den Einfluss auf die Regulierung von Spleißstellen [Alshatwi *et al.*, 2012]. Insgesamt wird dadurch womöglich die Promotorregion von BARD1 beeinflusst [Alshatwi *et al.*, 2012].

Für die drei BARD1-Varianten c.-48T>C, c.-78G>A und c.-30G>C existieren in Bezug auf die funktionellen Auswirkungen konträre Betrachtungen [Alshatwi *et al.*, 2012; Shi *et al.*, 2019], was zusätzlich für eine Einstufung als VUS spricht.

Die BRCA2-Variante c.-26G>A wurde in einer Studie von Maia *et al.* (2012) betrachtet,

deren Ziel es war, cis-wirkende SNPs in *BRCA2* und deren Einfluss auf die Expression zu untersuchen. Cis-wirkende Elemente dienen als regulatorische Faktoren auf DNA-Abschnitte in unmittelbarer Nähe [Arneemann *et al.*, 2019]. Die Variante c.-26G>A liegt innerhalb einer Bindestelle für die RNA-Polymerase II in der Promotorregion [Maia *et al.*, 2012]. Die Studie liefert Beweise dafür, dass die Variante im Verbund eines Haplotyps die Bindung von Transkriptionsfaktoren verändert. Sapkota *et al.* (2014) untersuchten verschiedene SNPs, die vermutlich mit Brustkrebs assoziiert sind. Unter anderem die *BRCA2*-Variante c.-26G>A zeigte eine starke statistische Signifikanz und trägt zur Anfälligkeit für Brustkrebs bei [Sapkota *et al.*, 2014].

Die *BRIP1*-Variante c.-205G>A wurde in mehreren Studien auf einen Zusammenhang mit erhöhtem Krebsrisiko untersucht. Song *et al.* (2007) haben unter anderem diese Variante bei über 2000 Brust- und bei über 1500 Eierstockkrebsfällen genotypisiert. Außerdem wurde die Variante bei über 2000 nicht erkrankten Kontrollen untersucht. Im Kontext mit fünf anderen SNPs wurde die Variante in einem Haplotyp zusammengefasst. In der Studie zu Brust- und Eierstockkrebs konnte kein signifikanter Unterschied in der Haplotyp-Häufigkeit zwischen erkrankten Personen und den gesunden Kontrollen festgestellt werden [Song *et al.*, 2007]. In weiteren Studien von Ma *et al.* (2013) und Martinez-Nava *et al.* (2016) konnte ebenfalls kein erhöhtes Krebsrisiko in Zusammenhang mit der Variante festgestellt werden. In der asiatischen Bevölkerung jedoch konnten Assoziationen zwischen dem SNP und einem erhöhten Krebsrisiko festgestellt werden [Di Liu *et al.*, 2018]. Da zusätzlich durch den SNP eine neue AUC-Startcodonstelle entsteht (siehe Tabelle 5.6), kann nicht eindeutig von der Gutartigkeit der Variante ausgegangen werden, was eine Einstufung als VUS zur Folge hat.

Zu der *CDH1*-Variante c.-54G>C existieren mehrere Studien mit funktionellen und statistischen Analysen. Mucaki *et al.* (2016) haben verschiedene Mutationen in FBREK-Risiko-Genen in Laboranalysen untersucht. Dabei wurde festgestellt, dass die Variante c.-54G>C die Bindung des Transkriptionsfaktors E2F verändert. Dieser Transkriptionsfaktor ist an der Aktivierung der Transkription vieler Gene beteiligt, die unter anderem bei der DNA-Synthese und im Zellzyklus relevant sind [Löffler, 2008]. Von A Nakamura *et al.* (2002) wurde die Promotoraktivität an einem Luciferase-Reportergen mit und ohne die Variante c.-54G>C gemessen. Die Promotoraktivität mit dem C-Allel der Variante war eindeutig geringer als jene mit dem Wildtyp-Allel. Bezogen zur 100 %-igen Promotoraktivität beim Wildtyp zeigte sich nur eine Aktivität von $65,4 \pm 12,1$ % [A Nakamura *et al.*, 2002].

Mit der *CDH1*-Variante c.-71C>G wurde noch keine Expressionsanalyse durchgeführt. Sie wurde zwar in zwei unabhängigen Studien als benigne eingestuft [Avizienyte *et al.*, 2001; Garcia-Pelaez *et al.*, 2023], jedoch entsteht bei dieser Variante das effiziente CUG als alternatives Startcodon. Daher muss der Variante eine unklare Signifikanz zugewie-

sen werden.

Zu der *RAD51C*-Variante c.-26C>T wurde von Gresner, Gromadzinska, Twardowska *et al.* (2014) eine Assoziationsstudie durchgeführt. Zwar handelt es sich hier um Patienten mit Kopf-Hals-Tumoren, jedoch spielt das Gen *RAD51C* auch eine wichtige Rolle bei FBREK und anderen Tumorerkrankungen. Es wurde ein signifikanter Zusammenhang zwischen dem SNP und dem Erkrankungsrisiko gefunden (Odds Ratio 2,0) [Gresner, Gromadzinska, Twardowska *et al.*, 2014]. Ebenfalls von Gresner, Gromadzinska, Jablonska *et al.* (2014) wurden 132 Frauen mit Brustkrebs im Vergleich zu einer krebsfreien Kontrollgruppe untersucht. Im Mittelpunkt der Untersuchungen stand hierbei der Zusammenhang von 5'-UTR-SNPs in *RAD51C* und dem Brustkrebsrisiko in Bezug auf oxidativen Stress. Die Variante c.-26C>T konnte dabei mit signifikant erhöhten Werten von Thiobarbitursäure-reaktiven Substanzen im Blutplasma der Tumorpatientinnen in Verbindung gebracht werden [Gresner, Gromadzinska, Jablonska *et al.*, 2014]. Bei diesen Substanzen handelt es sich um Reaktionsprodukte der Reaktion von Thiobarbitursäure mit Malondialdehyd. Letzteres ist ein Abbauprodukt der Lipidperoxidation [Arnemann *et al.*, 2019], einem Prozess, bei dem freie Radikale Zellen oxidativ schädigen. Thiobarbitursäure-reaktive Substanzen können als Marker für die Lipidperoxidation und somit ebenfalls als Marker für den oxidativen Stress dienen [Arnemann *et al.*, 2019]. Eine Erhöhung des oxidativen Stresses kann als Risikofaktor für die Tumorentstehung angesehen werden. Sowohl DNA-Schäden als auch die Deregulation von Signalwegen können dadurch ausgelöst werden.

Die *SMARCA4*-Variante c.-72C>T wurde bisher lediglich in einer Publikation erwähnt und ebenfalls als wahrscheinlich benigne klassifiziert [Laraqui *et al.*, 2021]. Allerdings entsteht auch hier ein alternatives Startcodon (GUG), weshalb die Klassifikation VUS gewählt wurde.

Mittels des Variant Effect Predictors konnten keine zuverlässigen Aussagen über die 5'-UTR-Varianten gewonnen werden. Die Zuverlässigkeit dieses Tools ist vor allem bei kodierenden Varianten hoch, während sie bei UTR-Varianten stark schwanken kann. Teils wurden zu gut vorbeschriebenen Varianten durch das Tool keine Auswirkungen vorhergesagt. Aus diesem Grund wurde der Fokus in der Auswertung der Varianten eher auf vorbeschriebene Daten und funktionelle Studien gelegt.

Außerdem wurden verschiedene Datenbanken zu uORFs sowie IRES nach den Core Genen durchsucht. Bezüglich uORFs existiert die Datenbank [uORFdb](#) der Universität Münster. Diese enthält für die Core Gene keine uORFs. Zu vorhandenen IRES sind über die Datenbank [IRESite](#) ebenfalls keine Daten zu den Core Genen enthalten. Das letzte Update dieser Datenbank erfolgte außerdem im Jahr 2019, sodass man davon ausgehen kann, dass die Datenbank nicht mehr aktuell ist. Des Weiteren ist die Da-

tenbank [Human IRES Atlas](#) nicht mehr erreichbar. Einzig in der Datenbank [IRESbase](#) finden sich vorhandene IRES des Menschen. Für die Core Gene ist nur für *TP53* eine IRES-Sequenz verzeichnet. Diese weist eine Länge von 174 bp auf. Betrachtet man die Position dieser IRES in Alamut™ Visual Plus, so fällt auf, dass diese IRES nicht in der 5'-UTR von *TP53* sondern intronisch liegt und mit dem fünften Exon von *TP53* überlappt.

Ein zusammenfassender Überblick über die Anzahl an identifizierten und klassifizierten Varianten findet sich in Abbildung 5.7.

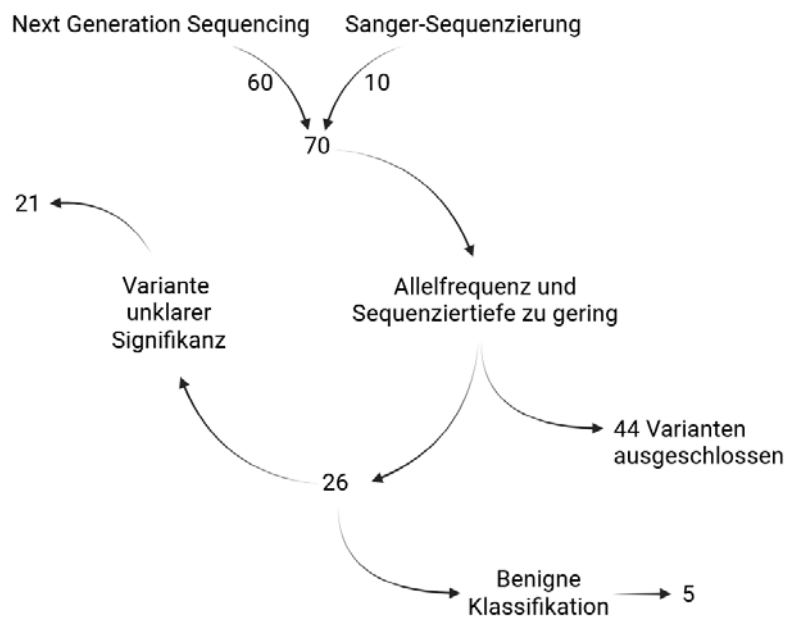


Abbildung 5.7: Quantitative Verteilung der Varianten im Laufe der Klassifikation.

Abbildung 5.7 zeigt abschließend zur Variantenklassifikation, wie viele Varianten in die jeweiligen Klassen eingeteilt bzw. ausgeschlossen wurden. Schlussendlich konnten 21 Varianten als VUS und 5 Varianten als benigne eingestuft werden.

5.3 Polygenic Risk Score

Der Polygenic Risk Score (PRS) ist ein, auf polygenetischen Faktoren beruhendes, Maß, welches das Risiko abschätzt, an einer genetisch bedingten Krankheit zu erkranken. Im Gegensatz zu monogenetischen Faktoren, wie pathogenen Varianten in einzelnen Genen, werden beim PRS viele kleine Sequenzvarianten bzw. SNPs berücksichtigt. Dabei ist nicht ein SNP allein krankheitsverursachend, sondern die Menge und Zusammensetzung der SNPs in verschiedenen Genen beeinflusst das Erkrankungsrisiko. Die statistischen Berechnungen zum PRS beruhen auf dem Vorhandensein oder Fehlen verschiedener genetischer Varianten [Khera *et al.*, 2018]. Über genomweite Assoziationsstudien können für verschiedene Krankheiten SNPs identifiziert werden, die das Erkrankungsrisiko erhöhen [Sueta *et al.*, 2012]. Der PRS beruht einzig auf genetischen Faktoren und berücksichtigt keine Umweltfaktoren [Khera *et al.*, 2018]. Er ist außerdem krankheitsspezifisch sowie jeweils auf eine Ethnie begrenzt.

Im Polygenic Score Catalog des European Bioinformatics Institute finden sich über 4800 PRS-Kombinationen für verschiedene Erkrankungen [URL-7, 2024]. Allein für Brustkrebs listet die Datenbank bis zu 159 PRS-IDs auf, welche sich in der Anzahl an inkludierten Varianten stark unterscheiden. Die hier betrachteten PRS-Werte richten sich nach dem *PRS313_BC*, welcher 313 SNPs beinhaltet [Mavaddat *et al.*, 2019]. Der von Mavaddat *et al.* (2019) erstellte PRS basiert auf Erkrankten und Kontrollen mit europäischer Abstammung und zählt derzeit zu den zuverlässigsten Polygenic Risk Scores für Brustkrebs. Die SNPs, welche in der vorliegenden Bachelorarbeit identifiziert wurden, doppeln sich nicht mit den SNPs aus dem PRS nach Mavaddat *et al.* (2019).

Der PRS kann auf verschiedene Weisen angegeben werden. Etwa als absoluter Wert als Summe der Effekte der risikobehafteten Varianten oder als relativer Wert in Bezug auf eine Referenzpopulation. Auch in Perzentilen kann der PRS dargestellt werden. In der Statistik ist ein Perzentil eine Größe, die angibt, wie ein Wert im Vergleich zu einer Verteilung der Werte in einer Referenzpopulation steht [Mittag und Schüller, 2023]. Dafür wird der Datensatz nach Größe geordnet und in 100 gleiche Teile zerlegt. Das 50. Perzentil beispielsweise gibt an, dass 50 % der Kohorte einen höheren und 50 % einen niedrigeren oder gleich hohen PRS aufweisen. Liegt der PRS einer Person über dem 97. Perzentil, so ist er höher als der PRS von 97 % der Referenzpopulation. Diese Person hat also ein erhöhtes Erkrankungsrisiko. Im Gegensatz dazu hat eine Person, die beispielsweise im 10. Perzentil liegt, einen PRS, der niedriger ist als jener von 90 % der Referenzpopulation. Das genetische Erkrankungsrisiko ist in diesem Fall sehr gering. In der medizinischen Diagnostik ist die 97. Perzentile oft der Grenzwert, um patho-

logische Auffälligkeiten zu diagnostizieren. Personen mit einem PRS über der 97. Perzentile werden oft als Hochrisikopatienten in Bezug auf die Krankheit eingestuft. Werte über der 85. Perzentile deuten ebenfalls auf ein erhöhtes Risiko hin, werden jedoch noch nicht als stark pathologisch eingestuft.

Basierend auf der Panel-Diagnostik, welche die Patientinnen der Kohorte erhalten haben, wurde für jede Patientin der PRS aus den 313 SNPs [Mavaddat *et al.*, 2019] berechnet. Über einen Boxplot kann die Verteilung der PRS-Werte in der Kohorte abgebildet werden (siehe Abbildung 5.8).

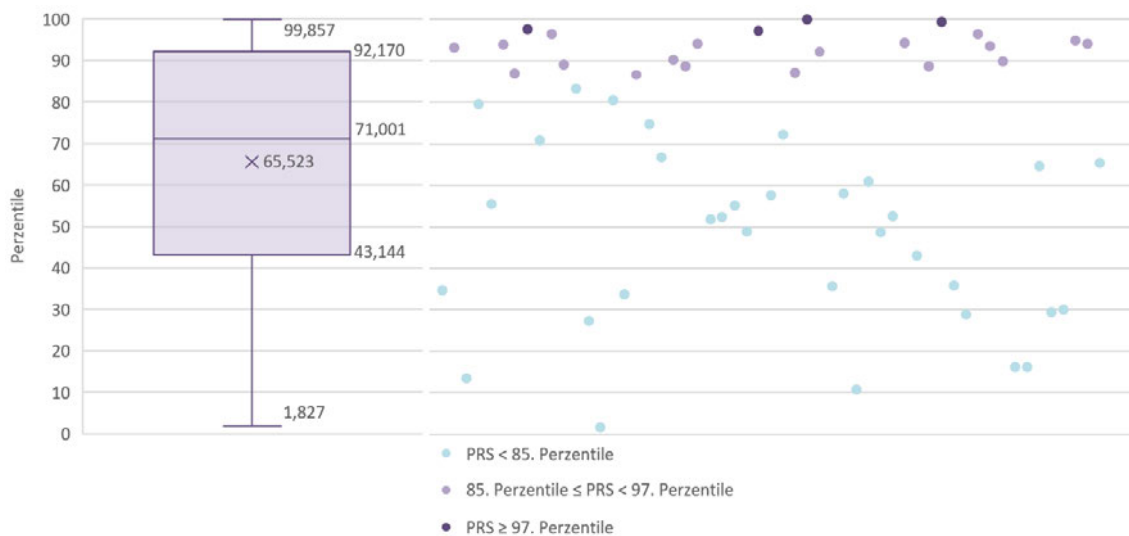


Abbildung 5.8: Boxplot der PRS-Werte der FBREK-Kohorte.

Die Abbildung 5.8 zeigt die PRS-Werte der FBREK-Kohorte als Boxplot mit den zusätzlichen Einzelwerten auf der rechten Seite. Es fällt auf, dass vier Patientinnen einen PRS über der 97. Perzentile zeigen (dunkles Violett). Über der 85. aber noch unter der 97. Perzentile liegt der PRS von 18 Patientinnen (helles Violett). Die restlichen PRS-Werte sind in einem hellblauen Farbton gekennzeichnet und bewegen sich zwischen Werten von der 83. bis zur 1. Perzentile. Der Boxplot zeigt, dass der Mittelwert etwa bei der 66. Perzentile und der Median bei der 71. Perzentile liegt. Das erste Quantil liegt bei der 43. Perzentile und das dritte Quantil bei der 92. Perzentile. Es gibt keine Ausreißer.

Aus den PRS-Daten wird ersichtlich, dass die Hälfte der Patientinnen einen PRS über der 71. Perzentile aufweisen. 50 % der Werte liegen zwischen dem ersten und dritten Quantil also zwischen der 43. und der 92. Perzentile. Allgemein zeigen die PRS-Werte einen Trend, der sich im oberen Drittel der Perzentilen bewegt. Von 22 PRS-Werten über der 85. Perzentile liegen vier über der kritischen 97. Perzentile. Daraus lässt sich ableiten, dass das Brustkrebsrisiko der erkrankten Frauen durchaus auch mit dem PRS zu erklären ist. Standardmäßig ist der PRS bisher jedoch noch nicht in den diagnosti-

schen Leitlinien enthalten.

Die Bedeutung des PRS im Hinblick auf die individuelle Risikoeinschätzung und auf Präventionsempfehlungen, vor allem für Brustkrebs-Patientinnen ohne pathogene Varianten, zeigte in Untersuchungen von Stiller *et al.* (2023) einen hohen Stellenwert. Hierbei wurden das 10-Jahresrisiko sowie die Präventionsempfehlungen unter Einbezug des PRS für Patientinnen mit und ohne pathogene Varianten betrachtet. Es zeigte sich, dass das 10-Jahresrisiko durch Hinzunahme des PRS in 13,6 % der Fälle verändert wurde und die empfohlenen Präventionsmaßnahmen in 12 % der Fälle anders ausfielen als bei einer Diagnostik ohne den PRS [Stiller *et al.*, 2023].

Für die 55 FBREK-Patientinnen aus dieser Arbeit konnten keine pathogenen Varianten durch die standardmäßige Diagnostik identifiziert werden. Demnach ist möglicherweise der PRS für die Diagnostik dieser Patientinnen ein sinnvolles unterstützendes Instrument. Von Stiller *et al.* (2023) konnte nachgewiesen werden, dass bei 11,1 % der Patientinnen ohne pathogene Varianten das 10-Jahresrisiko durch den Einbezug des PRS über den Schwellenwert von 5 % hinweg gestiegen ist.

6 Fazit

Zusammenfassend haben die Ergebnisse der Analyse der 5'-UTRs der 14 Core Gene für FBREK gezeigt, dass durchaus relevante genetische Varianten in den 5'-UTRs zu finden sind. Einerseits können neben dem AUG auch andere Startcodons gebildet werden, welche zu kleinen uORFs mit regulatorischen Funktionen oder zu N-terminalen Verlängerungen der Genprodukte führen. Vor allem Varianten in den Promotorregionen und in Transkriptionsfaktorbindestellen können die Expressionseffizienz beeinflussen. Im Verbund von Haplotypen sind viele SNPs in Assoziationsstudien auffällig, indem sie deutlich häufiger bei erkrankten Personen als bei Personen ohne Tumorerkrankung vorkommen.

Außerdem wurde durch die Untersuchungen und die Literaturrecherche zu spezifischen Varianten deutlich, dass sich unterschiedliche Tumorerkrankungen oft nicht getrennt voneinander betrachten lassen. Varianten in Genen, die bisher nur im Zusammenhang mit anderen Tumorerkrankungen betrachtet wurden, können ebenfalls relevant für Brustkrebs sein, da diese Gene in grundlegenden zellulären Prozessen involviert sind, welche in verschiedenen Tumorentitäten eine Relevanz haben.

Die Klassifikation nach Richards *et al.* (2015) ließ sich außerdem nur eingeschränkt anwenden, da diese Richtlinien auf monogene Krankheiten und kodierende Regionen von Genen spezialisiert sind. Aus diesem Grund wurden Varianten teilweise abweichend von den Kriterien als VUS* klassifiziert.

Die identifizierten Varianten in der vorliegenden Arbeit sind mit hoher Wahrscheinlichkeit nicht allein krankheitsverursachend. Durch die Akkumulation verschiedener risikobehafteter SNPs jedoch kann das Brustkrebsrisiko deutlich erhöht werden. Wie der PRS zeigt, ist es wichtig, nicht nur einzelne Gene auf pathogene Varianten zu untersuchen, sondern auch das polygenetische Netzwerk von Sequenzvarianten in die Diagnostik einzubeziehen. Dabei spielen sowohl Varianten in kodierenden Bereichen als auch Varianten in nicht kodierenden Regionen eine wichtige Rolle.

Neben den genetischen Faktoren gibt es jedoch auch weitere Bedingungen, unter denen das Risiko einer Tumorerkrankung steigt. Hierbei spielt vor allem der Lebensstil eine entscheidende Rolle: Rauchen, Alkoholkonsum, übermäßige Strahlenbelastung sowie eine ungesunde Ernährung und Bewegungsmangel können das Krebsrisiko erhöhen.

7 Ausblick

Um die Untersuchungen zu den 5'-UTRs bei FBREK zu vertiefen, könnten verschiedene Ansätze hilfreich sein. Zum einen könnte die Kohorte auf ungeklärte Ovarialkarzinomfälle ausgeweitet werden, um eine direkte Gegenüberstellung zwischen 5'-UTR-Varianten bei Brust- und Eierstockkrebs zu erreichen. In der hier untersuchten Kohorte sind zwar zum Zeitpunkt der Datenerhebung keine Ovarialkarzinomfälle bekannt, jedoch weisen diese beiden Tumorerkrankungen eine ähnliche genetische Prädisposition auf. Auch die Behandlungsmöglichkeiten überschneiden sich teilweise. So können bei beiden Tumorerkrankungen Poly(Adenosindiphosphat-Ribose)-Polymerase (PARB)-Inhibitoren eingesetzt werden [Kast, Rhiem und Schmutzler, 2020]. In Tumorzellen, in denen die DNA-Reparatur aufgrund von pathogenen Varianten nicht mehr ordnungsgemäß ablaufen kann, werden durch den Einsatz von PARB-Inhibitoren weitere DNA-Schäden in Kauf genommen, damit die Tumorzellen absterben. Außerdem ist es ratsam, bei weiteren Untersuchungen, wie etwa Assoziationsstudien, auch eine vergleichbare aber kreisfreie Kontroll-Kohorte einzubeziehen. Für aussagekräftige Assoziationsstudien ist zudem ein größerer Kohortenumfang notwendig.

Die bisherige Datenlage in Bezug auf funktionelle Studien ist im Bereich der 5'-UTR-Varianten teilweise noch lückenhaft. So ist es denkbar, in weiteren Analysen einige der identifizierten Varianten funktionell zu untersuchen, zu denen es noch keine weiteren Studien gibt. Somit wird das Verständnis für 5'-UTR-Varianten und deren Funktion möglicherweise ausgebaut. Die Umsetzung kann etwa in einem *in-vitro*-Reporter-Gen-Assay erfolgen, wodurch anhand der Genexpression beurteilt werden kann, ob die Varianten Einfluss auf die mRNA-Stabilität oder die Initiation der Translation haben.

Mit anschließenden Western-Blot-Analysen oder Massenspektrometrie können die resultierenden Proteine untersucht werden. Damit kann auf die Menge und Struktur der entstehenden Proteine geschlossen werden und es können quantitative sowie qualitative Unterschiede in der Proteinexpression identifiziert werden.

Wenn Varianten in Transkriptionsfaktorbindestellen vorliegen, kann die Interaktion des Transkriptionsfaktors mit der Bindestelle über Chromatin-Immunopräzipitation untersucht werden. Über eine Formaldehyd-Fixierung wird die Transkriptionsfaktor-DNA-Bindung festgehalten. Im Anschluss wird die DNA fragmentiert und über einen Antikörper, der spezifisch für den Transkriptionsfaktor ist, werden die Fragmente isoliert, welche an der Bindung beteiligt sind [Mülhardt, 2013]. Über die Menge der Interaktionen zwischen Transkriptionsfaktor und DNA mit der Variante sowie ohne die Variante, können Rück-

schlüsse auf den Einfluss der Variante gezogen werden.

Neben den 14 Core Genen können auch andere Gene durch ähnliche 5'-UTR-Varianten beeinflusst werden. Zu nennen sind hier etwa Gene, die in den selben Signalwegen involviert sind wie die Core Gene. Gene, die mit den Core Genen funktionell interagieren können hierfür ebenfalls wichtig sein. So interagiert beispielsweise das Genprodukt von *Fanconi Anemia Complementation Group D2* mit *BRCA1* [Garcia-Higuera *et al.*, 2001]. Auch pathogene Varianten in anderen Genen können Brustkrebs begünstigen. Zu nennen sind hier etwa Veränderungen in den DNA-Reparatur-Genen *MutL Homolog 1* und *MutS Homolog 2* [Malik *et al.*, 2019]. Die Ausweitung der Analyse der 5'-UTRs auf weitere Gene kann also sinnvoll sein.

Darüber hinaus lässt sich das Beispiel der *BRCA1*-Variante c.-107A>T aus der Einleitung erneut aufgreifen. Durch diese Variante wird der Promotor hypermethyliert, was in einer epigenetischen Stilllegung mündet [Evans *et al.*, 2018]. Dies ist ein Indiz dafür, dass auch die Untersuchung auf genomweite Methylierungsveränderungen nützlich sein kann.

In weiteren Ausführungen und Untersuchungen kann der Zusammenhang zwischen den Patientinnen mit risikobehaftetem PRS und den identifizierten Varianten unklarer Signifikanz näher betrachtet werden. Wenn man die 22 Patientinnen mit einem PRS über der 85. Perzentile betrachtet, ist bereits jetzt auffällig, dass diese durchweg ein sehr junges Ersterkrankungsalter aufweisen. So liegt der Altersdurchschnitt hier bei 34,9 Jahren. Allein bei den vier Patientinnen mit einem PRS über der 97. Perzentile liegt das durchschnittliche Ersterkrankungsalter bei 32,8 Jahren.

Die Patientinnen der Kohorte können allein aufgrund der Varianten unklarer Signifikanz keinen Korrekturbericht erhalten. Betrachtet man jedoch jene Patientinnen mit einem risikobehafteten PRS, kann dies zu einer Neubewertung führen und die Patientinnen profitieren möglicherweise von einem angepassten Behandlungsprogramm.

8 Zusammenfassung

Im Fokus dieser Bachelorarbeit stand die Untersuchung der 5'-UTRs der 14 Core Gene für familiären Brust- und Eierstockkrebs. Dazu diente eine Kohorte von 55 Mammakarzinom-Patientinnen mit jungem Erkrankungsalter, bei denen die Analysen der kodierenden Regionen der Core Gene keine Auffälligkeiten ergaben. Durch die regulatorische Rolle der 5'-UTRs lag die Vermutung nahe, dass sich bei diesen Patientinnen mögliche krankheitsrelevante Varianten in den untranslatierten Bereichen upstream der CDS befinden. Molekulargenetisch dienten Sanger-Sequenzierungsdaten und NGS-Daten als Grundlage für weitere Untersuchungen. Die Einordnung der Varianten erfolgte unter anderem nach dem Klassifikationssystem von Richards *et al.* (2015). Es wurden sowohl Varianten-Datenbanken und Literatur-Datenbanken als auch *in-silico*-Vorhersagen für die Klassifikation angewendet. Insgesamt wurden 70 Varianten identifiziert, von denen 44 Varianten aufgrund von technischen Limitierungen ausgeschlossen werden mussten. Fünf der übrig gebliebenen Varianten wurden als benigne eingestuft und 21 Varianten wiesen unklare Signifikanzen hinsichtlich ihrer Pathogenität auf. Oft wurde für die Varianten in funktionellen Studien ein Beitrag zur Pathogenität vorhergesagt. Sowohl im Zusammenhang eines Haplotyps als auch als einzeln wirkende Variante sind negative Einflüsse in Studien berichtet worden. Beispielsweise liegen die Varianten in Transkriptionsfaktorbindestellen oder in N-terminal verlängerten Isoformen der Proteine. Da jedoch keine der Varianten allein krankheitsverursachend ist und als pathogen eingestuft werden konnte, wurde der PRS für die 55 Patientinnen mit einbezogen. 22 der Patientinnen wiesen einen PRS über der 85. Perzentile auf und für vier dieser Patientinnen lag der PRS über der kritischen 97. Perzentile.

Die Ergebnisse deuten darauf hin, dass durch die Ansammlung von kritischen SNPs das Erkrankungsrisiko für Brustkrebs erhöht wird. Sowohl Varianten in den kodierenden Bereichen als auch solche in untranslatierten Regionen können eine Relevanz für die polygenetischen Ursachen für Brustkrebs haben. Um das Verständnis für 5'-UTR-Varianten künftig auszubauen, ist es wichtig, funktionelle Studien sowie Assoziationsstudien durchzuführen.

9 Summary

The focus of this bachelor thesis was the investigation of the 5'-UTRs of the 14 core genes for hereditary breast and ovarian cancer. A cohort of 55 breast cancer patients with a young age of onset was used for this purpose. The analyses of the coding regions of the core genes showed no abnormalities. Due to the regulatory role of the 5'-UTRs, it was suspected to assume that these patients might have disease-relevant variants in the untranslated regions upstream of the CDS. Molecular genetic Sanger sequencing data and NGS data served as the basis for further investigations. The variants were classified according to the Richards *et al.* (2015) classification system. Variant databases and literature databases as well as *in-silico* predictions were used for the classification. A total of 70 variants were identified, of which 44 variants had to be excluded due to technical limitations. Five of the remaining variants were classified as benign and 21 variants showed uncertain significance of pathogenicity. Damage was often predicted for the variants in functional studies. Negative influences have been reported in studies both in the context of a haplotype and as a single variant. For example, the variants are located in transcription factor binding sites or in N-terminally elongated isoforms of the proteins. However, since none of the variants alone is disease-causing and could be classified as pathogenic, the PRS was included for the 55 patients. 22 of the patients had a PRS above the 85th percentile and for four of these patients the PRS was above the critical 97th percentile.

The results indicate that the accumulation of several critical SNPs increases the risk of developing breast cancer. Both variants in the coding regions and those in untranslated regions may be relevant to the polygenetic causes of breast cancer. It is important to conduct functional studies and association studies to increase the understanding of 5'-UTR variants in the future.

Literaturverzeichnis

- Abudula, A (2001). Identifizierung und funktionelle Analyse von 5´ UTR-Spleissvarianten der mRNA des humanen Cathepsin L. Diss. Jena: Friedrich-Schiller-Universität Jena.
- Alshatwi, A, T Hasan, N Syed, G Shafi und B Grace (2012) Identification of functional SNPs in BARD1 gene and in silico analysis of damaging SNPs: based on data procured from dbSNP database. *PLoS one* 7(10):e43939.
- Aluri, J und M Cooper (2023) Somatic mosaicism in inborn errors of immunity: Current knowledge, challenges, and future perspectives. *Seminars in immunology* 67:S.101761.
- Andreev, D *et al.* (2022) Non-AUG translation initiation in mammals. *Genome biology* 23(1):S.111.
- Araujo, P *et al.* (2012) Before It Gets Started: Regulating Translation at the 5 UTR. *International Journal of Genomics* 2012:475731–475739.
- Arnemann, J *et al.* (2019). Haplotyp; MLPA; Odds Ratio; Thiobarbitursäure-reaktive Substanzen. In: *Lexikon der Medizinischen Laboratoriumsdiagnostik*. Hrsg. von A Gressner und T Arndt. Berlin und Heidelberg: Springer-Verlag GmbH Deutschland, 1066, 1570, 1669–1670, 1777, 1960, 2282. ISBN: 978-3-662-48985-7. DOI: [10.1007/978-3-662-48986-4](https://doi.org/10.1007/978-3-662-48986-4).
- Aubrey, B, A Strasser und G Kelly (2016) Tumor-Suppressor Functions of the TP53 Pathway. *Cold Spring Harbor perspectives in medicine* 6(5).
- Avizienyte, E, V Launonen, R Salovaara, T Kiviluoto und L Aaltonen (2001) E-cadherin is not frequently mutated in hereditary gastric cancer. *Journal of medical genetics* 38(1):49–52.
- Balia, C *et al.* (2010) PALB2: a novel inactivating mutation in a Italian breast cancer family. *Familial cancer* 9(4):531–536.
- Bartek, J, J Falck und J Lukas (2001) CHK2 kinase - a busy messenger. *Nature reviews. Molecular cell biology* 2(12):877–886.
- Blanco, A *et al.* (2013) Analysis of PALB2 gene in BRCA1/BRCA2 negative Spanish hereditary breast/ovarian cancer families with pancreatic cancer cases. *PLoS one* 8(7):e67538.
- Brown, T (2007). Genomes, Transcriptomes and Proteomes. In: *Genomes 3*. New York: Garland Science Pub, 105, 357. ISBN: 978-0-8153-4138-3.
- Burke, T und J Kadonaga (1997) The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Cold Spring Harbor Laboratory Press - Genes Dev.* 11(22):3020–3031.
- Cao, X und S Slavoff (2020) Non-AUG start codons: Expanding and regulating the small and alternative ORFeome. *Experimental cell research* 391(1):S.111973.
- Capell, A, K Fellerer und C Haass (2014) Progranulin transcripts with short and long 5´ untranslated regions (UTRs) are differentially expressed via posttranscriptional and translational repression. *Journal of Biological Chemistry* 289(37):25879–25889.

- Clark, D und N Pazdernik (2009). DNA-Synthese in vitro. In: *Molekulare Biotechnologie: Grundlagen und Anwendungen*. SpringerLink Bücher. Heidelberg: Spektrum Akademischer Verlag, 102, 103, 492, 493. ISBN: 9783827421890. DOI: [10.1007/978-3-8274-2189-0](https://doi.org/10.1007/978-3-8274-2189-0).
- Da Costa E Silva Carvalho, S *et al.* (2020) Germline variants in DNA repair genes associated with hereditary breast and ovarian cancer syndrome: analysis of a 21 gene panel in the Brazilian population. *BMC medical genomics* 13(1):S.21.
- Damiola, F *et al.* (2014) Contribution of ATM and FOXE1 (TTF2) to risk of papillary thyroid carcinoma in Belarusian children exposed to radiation. *International journal of cancer* 134(7):1659–1668.
- de Alencar, V, M Formiga und V de Lima (2020) Inherited lung cancer: a review. *Ecan-cermedicalscience* 14:S.1008.
- de Sainte Agathe, J *et al.* (2023) SpliceAI-visual: a free online tool to improve SpliceAI splicing variant interpretation. *Human genomics* 17(1):S.7.
- Di Liu, L *et al.* (2018) Four common polymorphisms of BRIP1 (rs2048718, rs4988344, rs4986764, and rs6504074) and cancer risk: evidence from 13,716 cancer patients and 15,590 cancer-free controls. *Aging* 10(2):266–277.
- Diaz de Arce, A, W Noderer und C Wang (2018) Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. *Nucleic acids research* 46(2):985–994.
- Don, R, P Cox, B Wainwright, K Baker und J Mattick (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic acids research* 19(14):S.4008.
- Donehower, L *et al.* (2019) Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. *Cell reports* 28(5):1370–1384.
- Evans, D *et al.* (2018) A Dominantly Inherited 5' UTR Variant Causing Methylation-Associated Silencing of BRCA1 as a Cause of Breast and Ovarian Cancer. *Hum Genet* (2):213–220.
- Ferri, C *et al.* (2019) Single nucleotide polymorphism in PTEN-Long gene: A risk factor in chronic myeloid leukemia. *Gene* 694:71–75.
- Fraille-Bethencourt, E *et al.* (2018) Genetic dissection of the BRCA2 promoter and transcriptional impact of DNA variants. *Breast cancer research and treatment* 171(1):53–63.
- Frey, U, H Bachmann, J Peters und W Siffert (2008) PCR-amplification of GC-rich regions: 'slowdown PCR'. *Nature protocols* 3(8):1312–1317.
- Garcia, M *et al.* (2009) Analysis of FANCB and FANCN/PALB2 fanconi anemia genes in BRCA1/2-negative Spanish breast cancer families. *Breast cancer research and treatment* 113(3):545–551.
- Garcia-Higuera, I *et al.* (2001) Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Molecular cell* 7(2):249–262.
- Garcia-Pelaez, J *et al.* (2023) Genotype-first approach to identify associations between CDH1 germline variants and cancer phenotypes: a multicentre study by the European Reference Network on Genetic Tumour Risk Syndromes. *The Lancet. Oncology* 24(1):91–106.

- Gather, F (2020). Analyse der Expression der humanen induzierbaren NO-Synthase (iNOS) : Einfluss der 5'-UTR auf die Expression der humanen iNOS und Expression der humanen iNOS in Modellen der neuronalen Differenzierung. Diss. Mainz: Johannes Gutenberg-Universität Mainz. DOI: [10.25358/OPENSOURCE-3100](https://doi.org/10.25358/OPENSOURCE-3100).
- Graw, J (2020). Veränderungen im Genom: Mutationen. In: *Genetik*. 7. aktualisierte und überarbeitete Auflage. Berlin und Heidelberg: Springer Spektrum, 126, 549. ISBN: 9783662609088.
- Gresner, P, J Gromadzinska, E Jablonska *et al.* (2014) Single nucleotide polymorphisms in noncoding regions of Rad51C do not change the risk of unselected breast cancer but they modulate the level of oxidative stress and the DNA damage characteristics: a case-control study. *PLoS one* 9(10):e110696.
- Gresner, P, J Gromadzinska, E Twardowska, K Rydzynski und W Wasowicz (2014) Rad51C: a novel suppressor gene modulates the risk of head and neck cancer. *Mutation research* 762:47–54.
- Harrod, A, KA Lane und JA Downs (2020) The role of the SWI/SNF chromatin remodeling complex in the response to DNA double strand breaks. *DNA repair* 93:S.102919.
- Hellebrand, H *et al.* (2011) Germline mutations in the PALB2 gene are population specific and occur with low frequencies in familial breast cancer. *Human mutation* 32(6):2176–2188.
- Hernandez, G und R Jagus (2016). Evolution of Translational Initiation: From Archaea to Eukarya; On the Origin and Early Evolution of Translation in Eukaryotes. In: *Evolution of the Protein Synthesis Machinery and Its Regulation*. Cham: Springer International Publishing, 62–87.
- Hopkins, B, B Fine *et al.* (2013) A secreted PTEN phosphatase that enters cells to alter signaling and survival. *Science (New York, N.Y.)* 341(6144):399–402.
- Hopkins, B, C Hodakoski, D Barrows, S Mense und R Parsons (2014) PTEN function: the long and the short of it. *Trends in biochemical sciences* 39(4):183–190.
- Jernigan, R, A Sarai, B Shapiro und R Nussinov (1987) Relationship between curved DNA conformations and slow gel migration. *Journal of biomolecular structure & dynamics* 4(4):561–567.
- Jurk, M (2022). Isolierung und Reinigung von Nukleinsäuren. In: *Bioanalytik*. Hrsg. von J Kurreck, J Engels und F Lottspeich. Berlin und Heidelberg: Springer Spektrum, 749–753. ISBN: 978-3-662-61706-9.
- Kast, K, K Rhiem und R Schmutzler (2020) Familiärer Brust- und Eierstockkrebs (einschließlich Hormone). *Der Gynäkologe* 53(11):739–747.
- Khera, A *et al.* (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics* 50(9):1219–1224.
- Knudson, A (1971) Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America* 68(4):820–823.
- Kozak, M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research* 15(20):8125–8148.

- Kuchenbaecker, K *et al.* (2017) Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* 317(23):2402–2416.
- Laraqui, A *et al.* (2021) Identification of a novel pathogenic variant in PALB2 and BARD1 genes by a multigene sequencing panel in triple negative breast cancer in Morocco. *Journal of genomics* 9:43–54.
- Leppek, K, R Das und M Barna (2017) Functional 5' UTR mRNA Structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology* 19(3):158–174.
- Li, Z, X Wang und R Jia (2022) Poly(rC) Binding Protein 1 Represses the Translation of STAT3 through 5' UTR. *Current Gene Therapy* 22(5):397–405.
- Liang, H, X Chen *et al.* (2017) PTEN-beta is an alternatively translated isoform of PTEN that regulates rDNA transcription. *Nature communications* 8:S.14771.
- Liang, H, S He *et al.* (2014) PTEN-alpha, a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism. *Cell metabolism* 19(5):836–848.
- Liao, P, G Satten und Y Hu (2017) PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. *Genetic epidemiology* 41(5):375–387.
- Liu, A, Y Zhu, W Chen, G Merlino und Y Yu (2022) PTEN Dual Lipid- and Protein-Phosphatase Function in Tumor Progression. *Cancers* 14(15):S.3666.
- Löffler, G (2008). RNA und Genexpression. In: *Basiswissen Biochemie. Mit Pathobiochemie*. Heidelberg: Springer Medizin Verlag, 243–263, 307. ISBN: 978-3-540-76512-7.
- Lorenz, T (2012) Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *Journal of visualized experiments : JoVE* (63):e3998.
- Ma, X *et al.* (2013) BRIP1 variations analysis reveals their relative importance as genetic susceptibility factor for cervical cancer. *Biochemical and biophysical research communications* 433(2):232–236.
- Maia, A *et al.* (2012) Effects of BRCA2 cis-regulation in normal breast and cancer risk amongst BRCA2 mutation carriers. *Breast cancer research : BCR* 14(2):S.63.
- Malik, S *et al.* (2019) Expressional analysis of MLH1 and MSH2 in breast cancer. *Current problems in cancer* 43(2):97–105.
- Martinez-Nava, G, J Fernandez-Nino, V Madrid-Marina und K Torres-Poveda (2016) Cervical Cancer Genetic Susceptibility: A Systematic Review and Meta-Analyses of Recent Evidence. *PloS one* 11(7):e0157344.
- Marzialì, F, M Dizanzo, A Cavatorta und D Gardiol (2019) Differential expression of DLG1 as a common trait in different human diseases: an encouraging issue in molecular pathology. *Journal of Biological Chemistry* 400(6):699–710.
- Mavaddat, N *et al.* (2019) Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *American journal of human genetics* 104(1):21–34.
- Mehaerin, Y (2023). *Vortrag: Familiärer Brust- und Eierstockkrebs. FBREK*. 4. Wintersymposium des GCCC. Zentrum für Familiären Brust- und Eierstockkrebs der Univer-

- sitätsmedizin Göttingen. URL: <https://www.youtube.com/watch?v=Rh2gu5s2maY> (aufgerufen am 31.05.2024).
- Merrick, W (2004) Cap-dependent and cap-independent translation in eukaryotic systems. *Gene* 332:1–11.
- Mittag, H und K Schüller (2023). Statistik: Eine interdisziplinäre Einführung mit interaktiven Elementen. In: 7., vollständig überarbeitete und aktualisierte Auflage. Lehrbuch. Berlin und Heidelberg: Springer Spektrum, 99–104. ISBN: 9783662682234.
- Mucaki, E *et al.* (2016) A unified analytic framework for prioritization of non-coding variants of uncertain significance in heritable breast and ovarian cancer. *BMC medical genomics* 9:S.19.
- Mülhardt, C (2013). Das Werkzeug; Die Polymerasekettenreaktion. In: *Der Experimentator Molekularbiologie Genomics*. Berlin und Heidelberg: Springer Spektrum, 59, 85, 88, 223–224. ISBN: 978-3-642-34636-1.
- Müller, H und D Prange (2016). PCR-Optimierung. In: *PCR - Polymerase-Kettenreaktion*. Berlin und Heidelberg: Springer Spektrum, 2, 3, 11, 100. ISBN: 978-3-662-48235-3.
- Nakamura, A *et al.* (2002) Characterization of DNA polymorphisms in the E-cadherin gene (CDH1) promoter region. *Mutation research* 502(1-2):19–24.
- Nakamura, S, D Aoki und Y Miki (2021). Molecular Basis of BRCA1 and BRCA2: Homologous Recombination Deficiency and Tissue-Specific Carcinogenesis. In: *Hereditary Breast and Ovarian Cancer. Molecular Mechanism and Clinical Practice*. Singapore: Springer, 15–30.
- Nordheim, A *et al.* (2018). Gene und Genprodukte; Schlüsseltechnologien. In: *Molekulare Genetik*. Stuttgart und New York: Georg Thieme Verlag KG, 295–298, 358–371, 390–398, 409, 543. ISBN: 978-3-13-242637-5. DOI: 10.1055/b-006-149922.
- Ren, M *et al.* (2021) Germline variants in hereditary breast cancer genes are associated with early age at diagnosis and family history in Guatemalan breast cancer. *Breast cancer research and treatment* 189(2):533–539.
- Richards, S *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine: official journal of the American College of Medical Genetics* 17(5):405–424.
- Ryczek, N, A Lys und I Makalowska (2023) The Functional Meaning of 5UTR in Protein-Coding Genes. *International Journal of Molecular Sciences* 24(3).
- Sanger, F, S Nicklen und A Coulson (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 74(12):5463–5467.
- Sapkota, Y *et al.* (2014) Assessing SNP-SNP interactions among DNA repair, modification and metabolism related pathway genes in breast cancer susceptibility. *PloS one* 8(6):e64896.
- Schaaf, C und J Zschocke (2018). Genetische Labordiagnostik. In: *Basiswissen Human-genetik*. Bd. 3. Berlin und Heidelberg: Springer-Verlag GmbH Deutschland, 200–202. ISBN: 978-3-662-56146-1.

- Schmidt, O (2023). Genetik und Molekularbiologie. In: Berlin, Heidelberg: Springer Berlin Heidelberg, 51, 220–223. ISBN: 978-3-662-66946-4. DOI: [10.1007/978-3-662-66947-1](https://doi.org/10.1007/978-3-662-66947-1).
- Schubach, M, T Maass, L Nazaretyan, S Roener und M Kircher (2024) CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. *Nucleic acids research* 52(D1):D1143–D1154.
- Shi, J *et al.* (2019) Functional Polymorphisms in BARD1 Association with Neuroblastoma in a regional Han Chinese Population. *Journal of Cancer* 10(10):2153–2160.
- Shir-Shapira, H *et al.* (2019) Identification of evolutionarily conserved downstream core promoter elements required for the transcriptional regulation of Fushi tarazu target genes. *PLOS ONE* 14(4).
- Sinn, H und H Kreipe (2013) A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast care (Basel, Switzerland)* 8(2):149–154.
- Somers, J, T Pöyry und A Willis (2013) A perspective on mammalian upstream open reading frame function. *The International Journal of Biochemistry Cell Biology* 45(8):1690–1700.
- Song, H *et al.* (2007) Tagging single nucleotide polymorphisms in the BRIP1 gene and susceptibility to breast and ovarian cancer. *PloS one* 2(3):e268.
- Stiller, S, S Drukewitz, K Lehmann, J Hentschel und V Strehlow (2023) Clinical Impact of Polygenic Risk Score for Breast Cancer Risk Prediction in 382 Individuals with Hereditary Breast and Ovarian Cancer Syndrome. *Cancers* 15(15).
- Stockinger, A, A Eger, J Wolf, H Beug und R Foisner (2001) E-cadherin regulates cell growth by modulating proliferation-dependent beta-catenin transcriptional activity. *The Journal of cell biology* 154(6):1185–1196.
- Sueta, A *et al.* (2012) A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. *Breast cancer research and treatment* 132(2):711–721.
- Sy, S, M Huen und J Chen (2009) PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proceedings of the National Academy of Sciences of the United States of America* 106(17):7155–7160.
- Thiemann, F, P Cullen und H Klein (2015). Grundlagen der molekularen Diagnostik. In: *Molekulare Diagnostik. Grundlagen der Molekularbiologie, Genetik und Analytik*. 2. Aufl. Weinheim und Berlin: Wiley-VHC Verlag Co. KGaA, S.19. ISBN: 978-3-527-33502-2.
- Tischkowitz, M, C Colas, S Pouwels und N Hoogerbrugge (2020) Cancer Surveillance Guideline for individuals with PTEN hamartoma tumour syndrome. *European journal of human genetics : EJHG* 28(10):1387–1393.
- Tzani, I *et al.* (2016) Systematic analysis of the PTEN 5' leader identifies a major AUU initiated proteoform. *Open biology* 6(5).

- URL-1. (2024) *Deutsches Konsortium Familiärer Brust- und Eierstockkrebs*. URL: <https://www.konsortium-familiaerer-brustkrebs.de/das-konsortium/> (aufgerufen am 30. 05. 2024).
- URL-2. (2023) *Deutsche Krebsgesellschaft: Zertifizierung: Erhebungsbögen: Zertifizierung der DKG: Dokumente*. URL: <https://www.krebsgesellschaft.de/zertdokumente.html> (aufgerufen am 25. 07. 2024).
- URL-3. (2024) *Thermo Fisher Scientific: ExoSAP-IT™ Reinigungsreagenz für PCR-Produkte*. URL: <https://www.thermofisher.com/order/catalog/product/de/de/78205.10.ML> (aufgerufen am 12. 07. 2024).
- URL-4. (2024) *Thermo Fisher Scientific: Shrimp Alkaline Phosphatase (SAP)*. URL: <https://www.thermofisher.com/order/catalog/product/783901000UN> (aufgerufen am 12. 07. 2024).
- URL-5. (2024) *Thermo Fisher Scientific: Hi-Di™ Formamid*. URL: <https://www.thermofisher.com/order/catalog/product/de/de/4311320> (aufgerufen am 22. 07. 2024).
- URL-6. (2024) *Thermo Fisher Scientific: iontorrent: Annotation sources: phyloP score*. URL: <https://ionreporter.thermofisher.com/ionreporter/help/GUID-03D1F68A-E646-4B49-AD59-AF2F51874BD2.html> (aufgerufen am 06. 08. 2024).
- URL-7. (2024) *EMBL-EBI: Polygenic Scores (PGS)*. URL: <https://www.pgscatalog.org/browse/scores/> (aufgerufen am 18. 08. 2024).
- Wang, H *et al.* (2010) Association between ataxia telangiectasia mutated gene polymorphisms and breast cancer in Taiwanese females. *Anticancer research* 30(12):5217–5221.
- Westermeier, R (2019). Kapillarelektrophorese. In: *Lexikon der Medizinischen Laboratoriumsdiagnostik*. Hrsg. von A Gressner und T Arndt. Berlin und Heidelberg: Springer-Verlag GmbH Deutschland, S.1313. ISBN: 978-3-662-48985-7. DOI: 10.1007/978-3-662-48986-4.
- Wilcox, N *et al.* (2024) Analysis of variants in untranslated and promoter regions and breast cancer risk using whole genome sequencing data. *medRxiv*.
- Wright, C, N Quaife, L Ramos-Hernandez und P Danecek (2021) Non-coding region variants upstream of MEF2C cause severe developmental disorder through three distinct loss-of-function mechanisms. *The American Journal of Human Genetics* 108(6):1083–1094.
- Wu, Y *et al.* (2018) Association between rs120963, rs152451, rs249935, rs447529, rs8053188, and rs16940342 Polymorphisms in the PALB2 Gene and Breast Cancer Susceptibility: A Meta-Analysis. *Oncology research and treatment* 41(12):780–786.
- Xia, B *et al.* (2006) Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Molecular cell* 22(6):719–729.
- Yang, X *et al.* (2020) Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 38(7):674–685.
- Yoon, H und T Donahue (1992) Control of translation initiation in *Saccharomyces cerevisiae*. *Molecular Microbiology* 6(11):1413–1419.

- Young, S und C Wek (2016) Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. *Journal of Biological Chemistry* 291(33):16927–16935.
- Yue, L *et al.* (2018) Association of ATM and BMI-1 genetic variation with breast cancer risk in Han Chinese. *Journal of cellular and molecular medicine* 22(7):3671–3678.
- Zhang, H, Y Wang und J Lu (2019) Function and Evolution of Upstream ORFs in Eukaryotes. *Trends in Biochemical Sciences* 44(9):782–794.
- Zhang, Q *et al.* (2021) PTEN-epsilon suppresses tumor metastasis through regulation of filopodia formation. *The EMBO journal* 40(10):e105806.
- Zhao, Z, L Xia, C Zhao und J Yao (2019) ATM rs189037 (G A) polymorphism increased the risk of cancer: an updated meta-analysis. *BMC medical genetics* 20(1):S.28.

Anhang A: Ausgewählte Sequenzabschnitte der Sanger-Sequenzierungsdaten

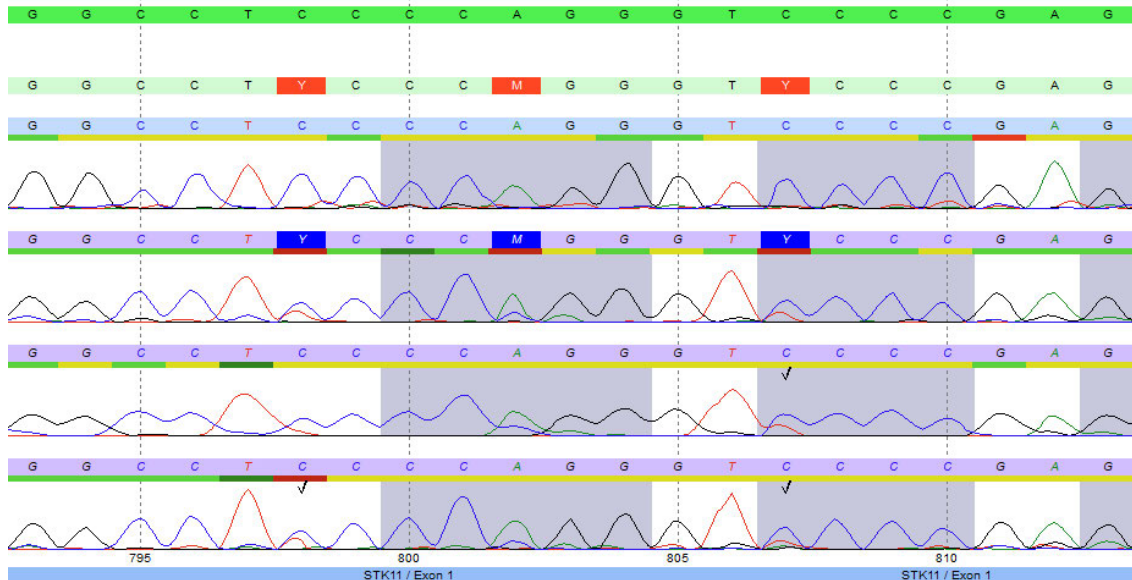


Abbildung A.1: Unsauberer Sequenzabschnitt der Sanger-Sequenzierungsdaten.

Die Abbildung A.1 zeigt einen Sequenzabschnitt aus der Software SEQUENCE PILOT. In diesem Fall ist der Sequenzabschnitt vierfach abgedeckt. Die Sequenzdaten wurden mittels Sanger-Sequenzierung ermittelt. Die Konsensussequenz ist im oberen Bereich in dem intensiveren Grünton dargestellt. Die Sequenzabfolge darunter spiegelt die sequenzierte Basenabfolge wieder. In rot sind Abweichungen der sequenzierten Sequenz von der Konsensussequenz gekennzeichnet. Die Unstimmigkeiten treten nie in allen vier Strängen auf. Es handelt sich um eine unsaubere Sequenzierung und somit nicht um echte Varianten. Solche Sequenzierfehler werden manuell beurteilt und können aus der Sequenz entfernt werden.

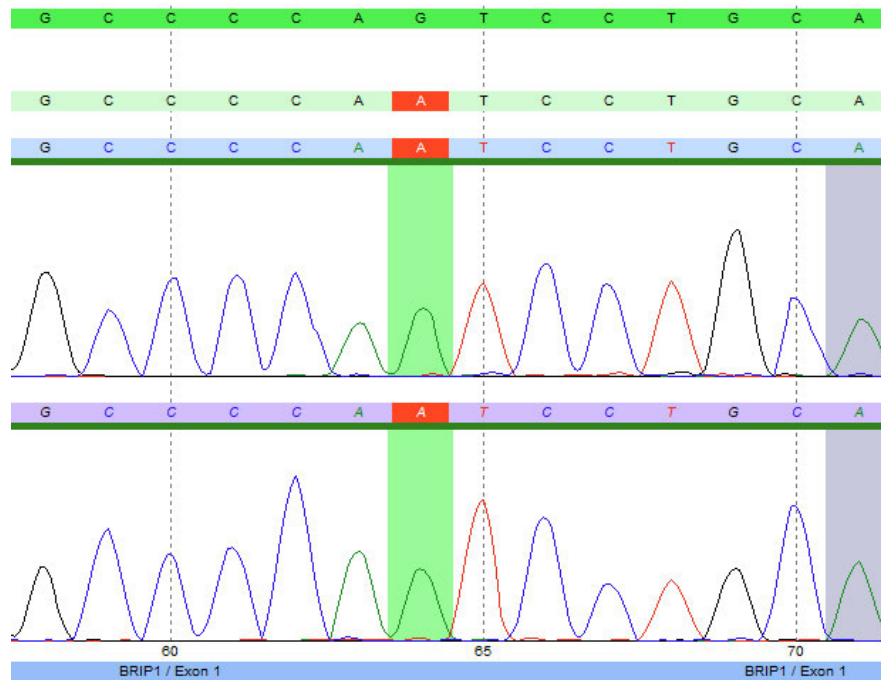


Abbildung A.2: Homozygote Variante c.-205G>A in der 5'-UTR von *BRIP1* aus der Sanger-Sequenzierung.

Bei der Variante aus Abbildung A.2 handelt es sich um die homozygote Variante c.-205G>A aus der 5'-UTR von *BRIP1*. Sequenziert wurde Adenin, obwohl die Referenz an dieser Stelle Guanin ist.

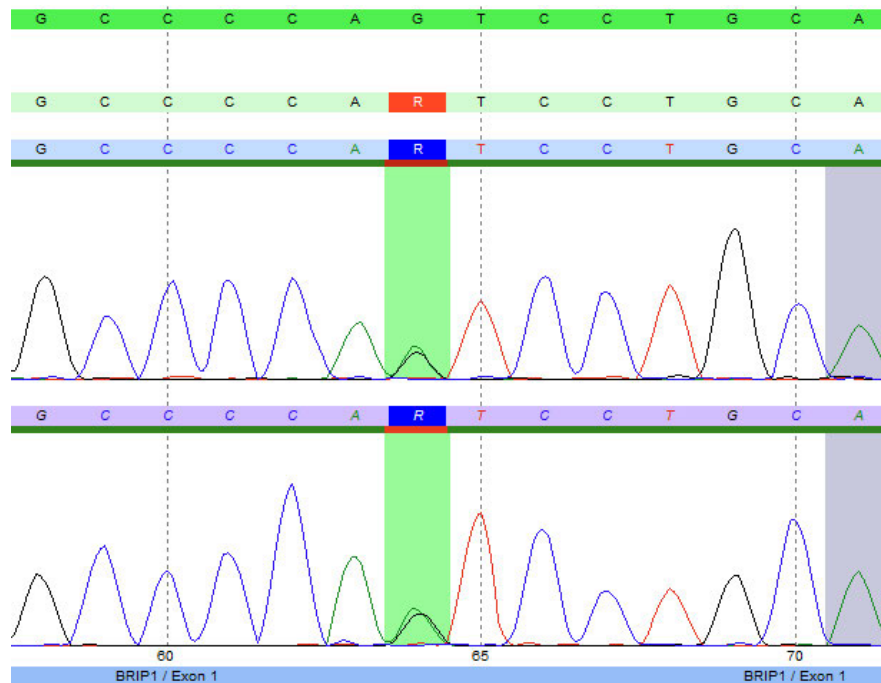


Abbildung A.3: Heterozygote Variante c.-205G>A in der 5'-UTR von *BRIP1* aus der Sanger-Sequenzierung.

In Abbildung A.3 ist die heterozygote Variante c.-205G>A in der 5'-UTR von *BRIP1* gezeigt. Sequenziert wurden in gleichen Teilen die Basen Adenin (Variante) und Guanin (Referenz).

Anhang B: Kontrolle der PCR-Produkte mittels Gelelektrophorese

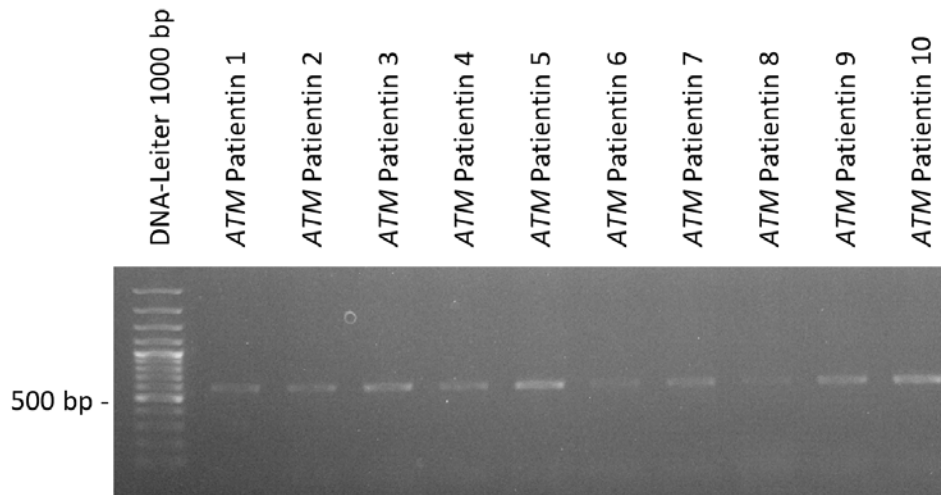


Abbildung B.1: Kontrolle der PCR-Produkte der 5'-UTR von *ATM*.

Abbildung B.1 zeigt das Agarose-Gel, in welchem die PCR-Produkte der 5'-UTR von *ATM* anhand der Patientinnen 1 bis 10 überprüft wurden. Es sind schwache, jedoch deutliche DNA-Banden der korrekten Länge von 541 bp zu erkennen.

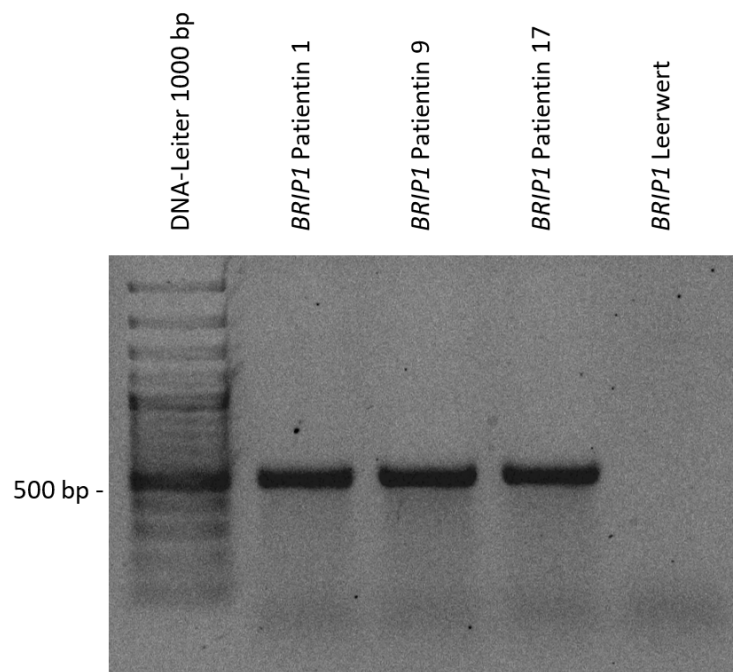


Abbildung B.2: Kontrolle der PCR-Produkte der 5'-UTR von *BRIP1*.

Anhand von drei Patientinnen wurde das PCR-Produkt der 5'-UTR von *BRIP1* überprüft. Gezeigt wird dies in Abbildung B.2. Das PCR-Produkt hat eine Länge von 472 bp.

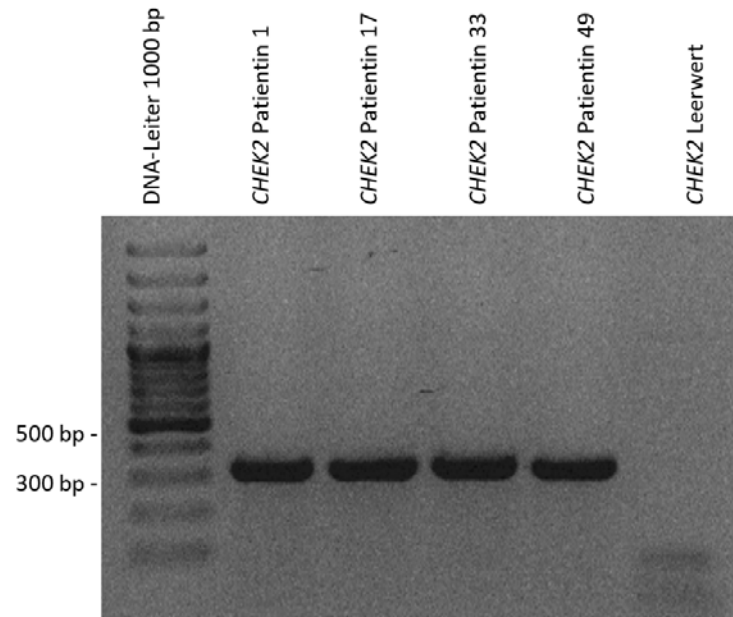


Abbildung B.3: Kontrolle der PCR-Produkte der 5'-UTR von *CHEK2*.

Abbildung B.3 zeigt die PCR-Produkte von drei Patientenproben für die 5'-UTR von *CHEK2* mit einer Länge von 300 bp.

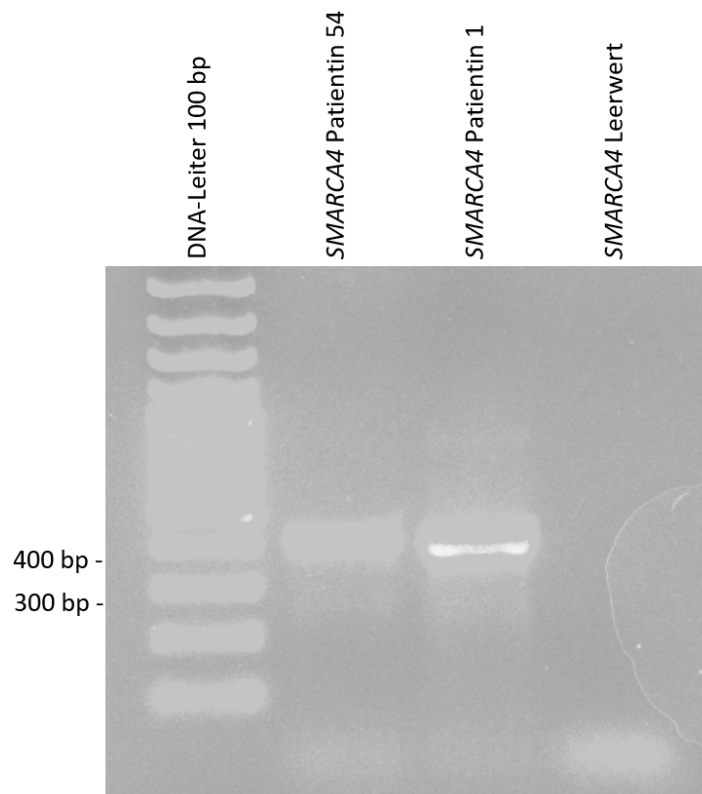


Abbildung B.4: Kontrolle der PCR-Produkte der 5'-UTR von *SMARCA4*.

In Abbildung B.4 sind die PCR-Produkte von zwei Patientenproben für die 5'-UTR von *SMARCA4* mit einer Länge von 363 bp gezeigt.

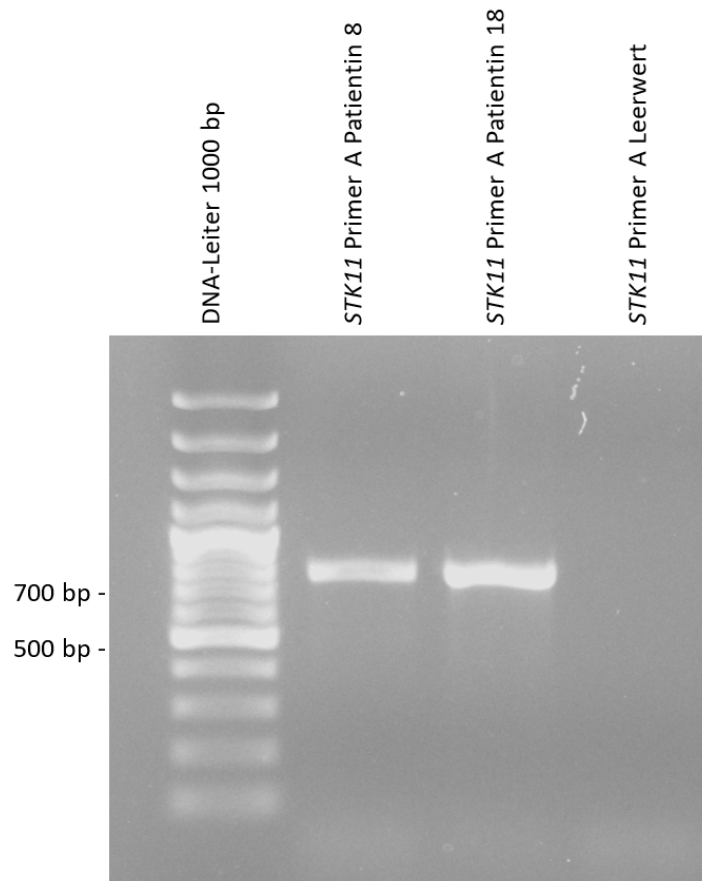


Abbildung B.5: Kontrolle der PCR-Produkte der 5'-UTR von *STK11* (Primer A).

Die Kontrolle der PCR-Produkte des Primerpaars A für *STK11* erfolgte in Abbildung B.5 anhand von zwei Patientinnen. Das PCR-Produkt hat eine Länge von 695 bp.

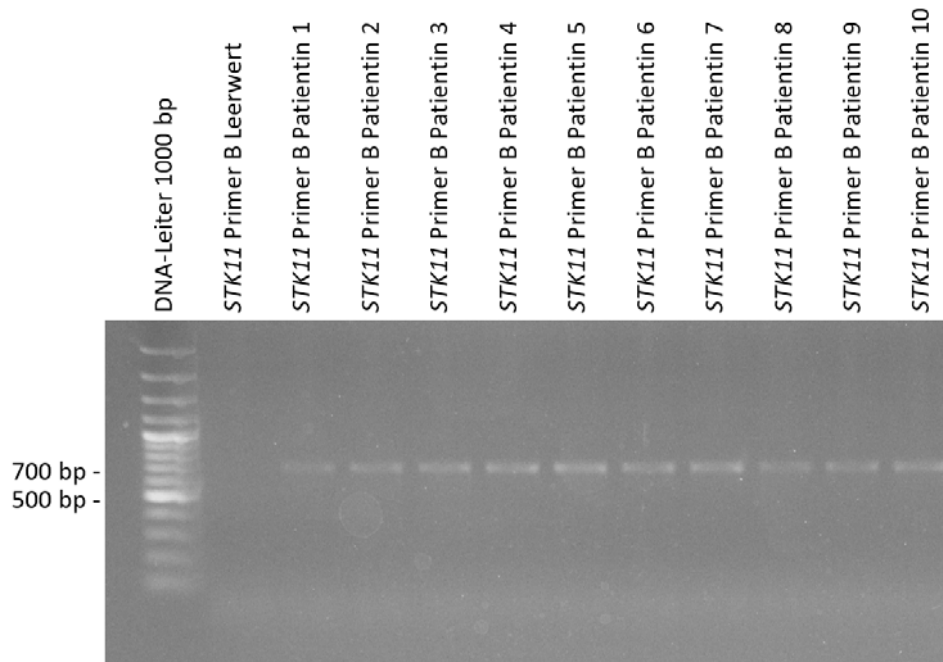


Abbildung B.6: Kontrolle der PCR-Produkte der 5'-UTR von *STK11* (Primer B).

Anhand der Patientenproben 1 bis 10 wurde das Vorhandensein der PCR-Produkte von Primerpaar B für *STK11* überprüft. Das Ergebnis ist in Abbildung B.6 dargestellt. Das PCR-Produkt hat eine Länge von 651 bp.

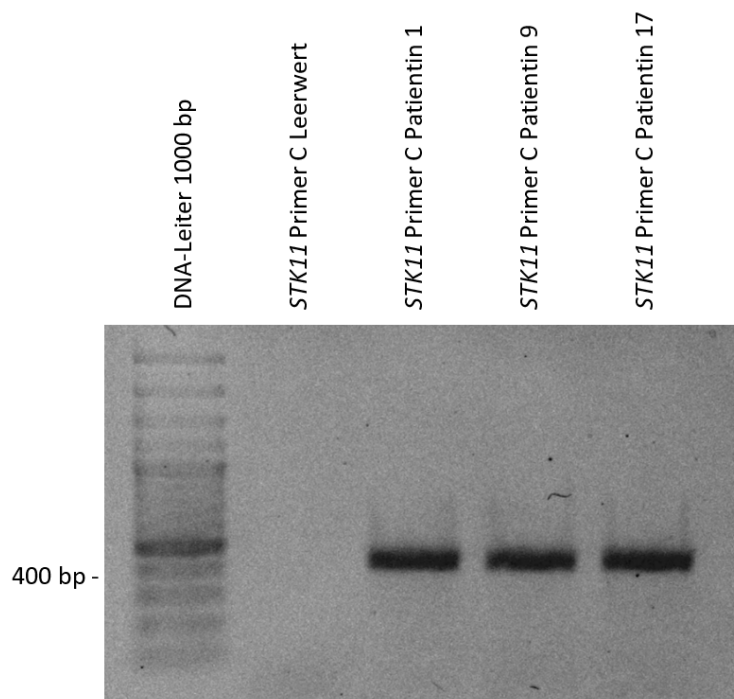


Abbildung B.7: Kontrolle der PCR-Produkte der 5'-UTR von *STK11* (Primer C).

Abbildung B.7 zeigt die PCR-Produkte von drei Patientinnen für das Primerpaar C der 5'-UTR von *STK11* mit einer Länge von 410 bp.

Anhang C: Zusätzliche Daten zu den identifizierten 5'-UTR-Varianten

Tabelle C.1: Genomische gnomAD-Daten zu den ausgeschlossenen Varianten.

Gen	Variante	gnomAD Grpmax AF	gnomAD homozygot
<i>ATM</i>	c.-105G>A	keine Daten	keine Daten
<i>BARD1</i>	c.-64T>A	keine Daten	keine Daten
	c.-80T>G	nicht in gnomAD	nicht in gnomAD
	c.-105A>C	keine Daten	keine Daten
	c.-114A>C	keine Daten	keine Daten
	c.-75T>A	keine Daten	keine Daten
	c.-44T>G	nicht in gnomAD	nicht in gnomAD
	c.-91T>G	keine Daten	keine Daten
	c.-64T>G	nicht in gnomAD	nicht in gnomAD
	c.-99T>G	nicht in gnomAD	nicht in gnomAD
<i>PTEN</i>	c.-588A>G	keine Daten	keine Daten
	c.-273A>C	nicht in gnomAD	nicht in gnomAD
	c.-503A>C	nicht in gnomAD	nicht in gnomAD
	c.-607T>G	nicht in gnomAD	nicht in gnomAD
	c.-575T>G	nicht in gnomAD	nicht in gnomAD
	c.-623C>A	keine Daten	keine Daten
	c.-383A>C	0,001346	0
	c.-829A>C	keine Daten	keine Daten
	c.-327T>G	keine Daten	keine Daten
	c.-305T>G	keine Daten	keine Daten
	c.-341T>G	keine Daten	keine Daten
	c.-596T>G	nicht in gnomAD	nicht in gnomAD
	c.-332T>G	keine Daten	keine Daten
	c.-385A>C	keine Daten	keine Daten
	c.-335A>C	keine Daten	keine Daten
c.-356A>C	keine Daten	keine Daten	
<i>SMARCA4</i>	c.-90C>G	nicht in gnomAD	nicht in gnomAD
	c.-80C>G	nicht in gnomAD	nicht in gnomAD
	c.-50A>C	nicht in gnomAD	nicht in gnomAD
	c.-147G>C	0,001128	0
	c.-120T>G	nicht in gnomAD	nicht in gnomAD
	c.-111G>T	keine Daten	keine Daten
	c.-100A>C	nicht in gnomAD	nicht in gnomAD
	c.-72C>G	nicht in gnomAD	nicht in gnomAD
	c.-55A>C	0,00001209	0
<i>STK11</i>	c.-302G>C	nicht in gnomAD	nicht in gnomAD
	c.-296T>C	nicht in gnomAD	nicht in gnomAD
	c.-329C>T	keine Daten	keine Daten
	c.-841G>T	keine Daten	keine Daten
	c.-524T>G	nicht in gnomAD	nicht in gnomAD
	c.-456G>T	keine Daten	keine Daten
	c.-134T>G	keine Daten	keine Daten
	c.-290T>G	keine Daten	keine Daten
	c.-107dup	0,005613	1

Tabelle C.2: ClinVar IDs und rs-Nummern der ausgeschlossenen Varianten.

Gen	Variante	ClinVar ID	rs-Nummer
<i>BARD1</i>	c.-75T>A		rs1696486045
	c.-64T>G		rs2106172512
<i>PTEN</i>	c.-588A>G		rs1858356492
	c.-383A>C	1879136	rs2132143026
	c.-305T>G		rs1858375675
	c.-341T>G		rs1215644762
	c.-385A>C		rs1858369934
	c.-335A>C		rs1589595433
	c.-356A>C		rs2132143232
<i>SMARCA4</i>	c.-147G>C		rs923273826
	c.-111G>T		rs2083757463
<i>STK11</i>	c.-329C>T		rs889606970
	c.-107dup	328224	rs927999961

Tabelle C.3: PubMed-IDs der benignen Varianten.

Gen	Variante	PubMed-IDs
<i>BRCA2</i>	c.-52A>G	34375487, 34249459, 17428325, 27067391, 21161372, 25243787, 29766361, 29860059, 35899106, 25380576, 37204320
<i>BRIP1</i>	c.-31+12G>A	19584272, 19935797, 32888493, 25592580
<i>PALB2</i>	c.-47G>A	25741868, 29879995, 32894086, 21356067, 30362319, 30458447, 18302019, 18446436, 20180015, 21618343, 22052327, 22310028, 23448497, 23824750, 23935836, 19921424, 20852946, 22692731, 19635604, 24556926, 23587053, 30309218

Tabelle C.4: ClinVar IDs und rs-Nummern der benignen Varianten.

Gen	Variante	ClinVar ID	rs-Nummer
<i>BRCA2</i>	c.-52A>G	209600	rs206118
<i>BRIP1</i>	c.-31+12G>A	136585	rs4988340
<i>CHEK2</i>	c.-7+62del	1232772	rs199585274
<i>PALB2</i>	c.-47G>A	126578	rs8053188
<i>STK11</i>	c.-311C>T	328217	rs149756065

Tabelle C.5: PubMed-IDs der als VUS eingestuftten Varianten.

Gen	Variante	PubMed-IDs
<i>ATM</i>	c.-111G>A	28698656, 26872812, 28827732, 24756757, 26870349, 32104210, 31548641, 27258253, 31247975, 34950210, 23663450, 25010664, 26985396, 21739480, 25541996, 28874875, 29928473, 19193393, 35255575, 26139146, 29691986, 22529920, 25196645, 25759145, 34995174, 21050672, 17428325, 20004998, 20816691, 21108427, 21937043, 22960875, 23154078, 23925578, 23993922, 24819391, 26504743, 28806901, 29246212, 29369221, 29714915, 29756124, 30709340, 31201228, 32171823, 32329754, 34124196, 34105905, 22438227, 22229509, 36895830, 29218269, 24364026, 32201398, 29854078, 35740268, 22420423, 25151201, 36316110, 21908578, 28642860, 22126766, 30197789, 25771414, 26629166, 26019482, 25553725, 29928112, 34522244, 25258552, 28825373, 36976737, 30531820, 36308385, 33407290, 24023351, 33513745, 21187516, 31973052, 25386189, 20171797, 25966232, 37425424
<i>BARD1</i>	c.-48T>C	23056176, 30132831, 31123171, 31258718, 32708251, 32021290, 35650591, 30793172, 35768791
	c.-83C>T	32039725
	c.-78G>A	31258718
	c.-30G>C	19252927, 31258718, 32708251, 18059179
<i>BRCA2</i>	c.-26G>A	25741868, 19276285, 23663450, 22513257, 12466288, 21161372, 19644020, 26979245, 33499154, 16760289, 24337145, 21702907, 21356067, 18724707, 21939546, 23767878, 18694767, 33643918, 34295994, 28944232, 34287479, 19012493, 22809218, 32284662, 23755158, 17945002, 28076423, 24938600, 29460995
<i>BRIP1</i>	c.-205G>A	27415837, 28415781, 15113441, 34050209, 19138047, 17342202, 23473757, 23644138, 29466248, 19012493, 24301948, 29138235, 17504528
<i>CDH1</i>	c.-54G>C	27067391
	c.-71C>G	27067391, 21356067, 29295527, 11332401, 11948460, 16801346, 24204729, 27146957
<i>PTEN</i>	c.-734G>A	25741868
	c.-511G>A	22146979, 21633361, 17033968, 30738963, 34050721, 32547095, 33240400, 36010895
	c.-326G>C	36292575, 33240400
<i>RAD51C</i>	c.-26C>T	29879995, 31905201, 25343521, 21980511, 26406419, 22843497
<i>SMARCA4</i>	c.-72C>T	34646395

Tabelle C.6: ClinVar IDs und rs-Nummern der als VUS eingestuftten Varianten.

Gen	HGVS	ClinVar ID	rs-Nummer
<i>ATM</i>	c.-111G>A	302240	rs189037
<i>BARD1</i>	c.-48T>C	334200	rs17489363
	c.-83C>T	334202	rs71579840
	c.-78G>A	334201	rs34732883
	c.-30G>C	334199	rs1129804
<i>BRCA2</i>	c.-26G>A	125965	rs1799943
<i>BRIP1</i>	c.-205G>A	324354	rs2048718
<i>CDH1</i>	c.-54G>C	221181	rs5030874
	c.-71C>G	221174	rs34033771
<i>PTEN</i>	c.-734G>A	189524	rs786205432
	c.-534_-532dup	492718	rs34413673
	c.-511G>A	810888	rs12573787
	c.-488_-486dup	548871	rs1237307954
	c.-326G>C	440217	rs2943772
<i>RAD51C</i>	c.-26C>T	324173	rs12946397
<i>SMARCA4</i>	c.-147G>C		rs923273826
	c.-72C>T	328015	rs559144002
	c.-43C>T	328016	rs574982898
<i>STK11</i>	c.-590C>T	888658	rs1050439702
	c.-554G>A		
	c.-973C>G		

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich diese Arbeit selbstständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die Arbeit noch nicht anderweitig für Prüfungszwecke vorgelegt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.



Toni Kalinsky

Mittweida, 23.08.2024