# GEZIELTE ANREICHERUNGS- UND NEUE DNA-SEQUENZIERUNGSSTRATEGIEN FÜR DIE MOLEKULARE ANALYSE VON FANCONI-ANÄMIE-GENEN



DISSERTATION ZUR ERLANGUNG DES NATURWISSENSCHAFTLICHEN DOKTORGRADES DER JULIUS-MAXIMILIANS-UNIVERSITÄT WÜRZBURG

VORGELEGT VON

ISABELL ROST

GEBOREN IN ROTHENBURG O. D. T.

WÜRZBURG, MAI 2017

 $\odot$   $\odot$ 

Eingereicht am: \_\_\_\_\_

Mitglieder der Promotionskommission:

Vorsitzender:

Erstgutachter: Prof. Dr. Detlev Schindler (Institut für Humangenetik) Zweitgutachter: Prof. Dr. Thomas Dandekar (Lehrstuhl für Bioinformatik)

Tag des Promotionskolloquiums:

Doktorurkunde ausgehändigt am: \_\_\_\_\_

# Erklärungen nach §4 Abs. 3 Satz 3, 5, 8 der Promotionsordnung der Fakultät für Biologie

# Affidavit

I hereby declare that my thesis entitled: "Gezielte Anreicherungs- und neue DNA-Sequenzierungsstrategien für die molekulare Analyse von Fanconi-Anämie-Genen" is the result of my own work.

I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore I verify that the thesis has not been submitted as part of another examination process neither in identical nor in similar form.

#### Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation: "Gezielte Anreicherungs- und neue DNA-Sequenzierungsstrategien für die molekulare Analyse von Fanconi-Anämie-Genen", eigenständig, d. h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen, als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

\_\_\_\_\_, den\_\_\_\_\_

Unterschrift

# Inhaltsverzeichnis

INHALTSVERZEICHNIS IV					
KUR	KURZFASSUNG1				
SUM	MAR	Υ			
1.	EINL	EITUNG	5		
1.1	DN	A-Reparatursysteme	5		
	1.1.1	Basenexzisionsreparatur			
	1.1.2	Nukleotidexzisionsreparatur	7		
	1.1.3	Basenfehlpaarungs-Reparatur	9		
	1.1.4	Reparatur von DNA-Doppelstrangläsionen			
1.2	Fan	ICONI-ANÄMIE			
	1.2.1	Erstbeschreibung und Inzidenz			
	1.2.2	Zellulärer und klinischer Phänotyp			
	1.2.3	Genetischer Hintergrund			
	1.2.4	Der FA/BRCA-Signalweg			
	1.2.5	Therapiemöglichkeiten			
1.3	Ide	ntifizierung von FA-Genen und Mutationsanalyse			
	1.3.1	Methoden zur Komplementationsgruppen-Bestimmung und Identifizierung von FA-Genen			
	1.3.2	Mutationsanalyse in der FA-Routinediagnostik			
	1.3.3	Möglichkeiten durch Hochdurchsatz-Sequenzierungsmethoden			
1.4	Fra	GESTELLUNG			
2.	MAT	ERIALIEN	33		
21	An	TKÖRPER	33		
2.1	211	Primärantikärtser	33		
	2.1.2	Sekundärantikörper	33		
2.2	ENZ	YME	33		
2.3	GRO	ĎBENSTANDARDS			
2.4	NG	S-Panels zur Anreicherung	34		
2.5	PRI	MER	34		
2.6	/ I RIVER				
2.7	KO	MMERZIELLE ZELLLINIEN			
2.8	KO	MPETENTE BAKTERIENSTÄMME			
2.9	2.9 PLASMIDE				

# Inhaltsverzeichnis

	2.10	Kulturmedien, Puffer und Reagenzien	36
	2.11	Sonstige Puffer und Lösungen	38
	2.12	Sonstige Verbrauchsmaterialien	38
2.13 Geräte		Geräte	39
	2.14	SOFTWAREPROGRAMME UND DATENBANKEN	41
3.	ME	THODEN	42
	3.1 A	LLGEMEINE METHODEN DER ZELLKULTUR	42
	3.1.1	Kultivierung von adhärenten und Suspensionszellen	42
	3.1.2	Rryokonservierung und Auftauen von Zellen	42
	3.1.3	Bestimmung der Zellzahl	43
	3.2 0	Gentransfer	43
	3.2.1	Klonierung	43
	3.2.2	? Transformation	44
	3.2.3	3 Transduktion	44
	3.3 I	MMUNFLUORESZENZ	44
	3.4 Z	ELLZYKLUS-ANALYSE AM DURCHFLUSSZYTOMETER	45
	3.5 Z	ELLVIABILITÄTS-TEST	45
	3.5.1	MMC-Sensitivität	45
	3.5.2	? UV-Sensitivität	46
	3.6 A	ILGEMEINE MOLEKULARGENETISCHE METHODEN	47
	3.6.1	Isolation von Nukleinsäuren	47
	3.6.2	2 Aufreinigung von DNA	48
	3.6.3	9 Quantifizierung von DNA	49
	3.6.4	cDNA-Synthese	50
	3.6.5	PCR-Amplifikation	50
	3.6.0	Größenauftrennung von DNA-Fragmenten	52
	3.6.7	Methoden zur Sequenzierung	53
	3.7 S	PEZIELLE METHODEN: ANREICHERUNG MIT FA-GENPANELS UND NGS	55
	3.7.1	Sondendesign zur Erstellung eines FA-Genpanels	55
	3.7.2	Probenvorbereitung zur Anreicherung	55
	3.7.3	B Klonale Amplifikation durch emPCR	58
	3.7.4	Sequenzierung am GS Junior	60
	3.7.5	Auswertung von NGS-Daten	60
	3.8 P	ROTEINANALYTISCHE METHODEN	61
	3.8.1	Proteinextraktion	61
	3.8.2	2 Quantifizierung von Proteinextrakten	62
	3.8.3	3 SDS-Gelelektrophorese	62
	3.8.4	Proteintransfer	62

# Inhaltsverzeichnis

		3.8.5	Immunblot-Analyse	62
		3.8.6	Entfernen von Antikörperbindungen an Membranen	63
4.		ERGE	BNISSE	64
	4.1	Eta	BLIERUNG EINES FA-GENPANELS	64
		4.1.1	Abdeckung der ROI durch Sonden	64
		4.1.2	Anreicherung und Sequenzierung von Patienten bekannter Komplementationsgruppe	67
	4.2	Ana	LYSE VON PATIENTEN UNBEKANNTER KOMPLEMENTATIONSGRUPPE MIT HILFE VON SEQCAP EZ-	
	An	REICHE	RUNGSPANELS	74
		4.2.1	Anreicherung mit verschiedenen SeqCap EZ-Libraries und Sequenzierung am GS Junior	74
		4.2.2	Alignment mit Hilfe der NextGENe-Software	77
		4.2.3	Zuordnung der Patienten zu Komplementationsgruppen	83
	4.3	MUI	ATIONSANALYSE VON FA-PATIENTEN MIT HILFE EINES KOMMERZIELLEN KREBSPANELS	109
		4.3.1	Alignment mit Hilfe der NextGENe-Software	109
		4.3.2	Bestätigung bereits zuvor bekannter Mutationen	112
		4.3.3	Nachweis von Mutationen, die mittels SeqCap EZ-Panelsequenzierung nicht detektiert wurden	113
		4.3.4	MLPA- und Immunblot-Analyse zur Eingrenzung der Komplementationsgruppe	115
		4.3.5	Erfolgreiche Komplementationsgruppen-Zuordnung	117
		4.3.6	Nicht zugeordnete Patienten trotz Mutationsanalyse	119
	4.4	MUI	'ATIONSANALYSE MIT HILFE VON WES	120
		4.4.1	Alignment mit Hilfe der NextGENe-Software	121
4.4.2		4.4.2	Erfolgreiche Komplementationsgruppen-Zuordnung durch WES: Identifizierung eines dritten FA-Q-Patienten	123
		4.4.3	Bestätigung bereits bekannter Mutationen bei FA-07, FA-10 somie FA-24	128
		4.4.4	Nachweis pathogener Mutationen in anderen Genen der DNA-Reparatur	129
	4.5	SEQ	uenzanalysen bei FANCD2	134
		4.5.1	Pseudogen-Analyse	134
		4.5.2	Bestätigung hypomorpher Mutationen	136
5.		DISKU	USSION	. 140
	5.1	Eta	BLIERUNG EINES ANREICHERUNGSPANELS ZUR NGS-ANALYSE VON FA-PATIENTEN	140
	5.2	VER	SCHIEDENE NGS-STRATEGIEN ZUR MUTATIONSANALYSE VON FA-GENEN	142
		5.2.1	Anreicherung ausgewählter Zielregionen	142
		5.2.2	Abdeckung der angereicherten Zielregionen durch unterschiedliche Anreicherungs- und Sequenziermethoden	145
	5.3	Kon	IPLEMENTATIONSGRUPPEN-ZUORDNUNG DURCH DIE GEZIELTE ANREICHERUNG VON FA-GENEN	147
		5.3.1	Priorisierung, Klassifizierung und Validierung von Varianten	147
		5.3.2	Zuordnung der Patienten und Identifizierung neuer Mutationen	150
		5.3.3	Einige Patienten blieben trotz gezielter Anreicherung ohne Zuordnung	152
	5.4	Сна	RAKTERISIERUNG EINZELNER KOMPLEMENTATIONSGRUPPEN	155
		5.4.1	Erweiterung des Mutationsspektrums in selteneren Komplementationsgruppen	155
		5.4.2	Identifizierung eines vierten UBE2T-defizienten Patienten	158

5.4.3	Fanconi-Anämie bei einer Patientin mit XP-assoziierter Missense-Mutation	
5.4.4	Mutationen im zentralen Protein des FA-Signalweges: Sequenzanalyse bei FA-D2-Patienten	
5.5 NACH	IWEIS VON MUTATIONEN IN WEITEREN GENEN DER DNA-REPARATUR	
LITERATU	RVERZEICHNIS	
ABKÜRZUI	NGSVERZEICHNIS	VIII
DANKSAGU	J <b>NG</b>	XI
ANHANG.		XII
Primerseq	UENZEN	XII
Mutagene	se-Primer	XII
InFusion-	Primer	XII
Primer zu	r Bestätigung von Mutationen durch Sanger-Sequenzierung	XII
VEKTORKA	RTEN	XVII
GENLISTE I	DER SEQCAP EZ-ANREICHERUNGSPANELS	XVIII
GENLISTE I	DES TRUSIGHT CANCER-ANREICHERUNGSPANELS	XXI
EIGENE PU	BLIKATIONEN, DIE IM RAHMEN DER DISSERTATION ENTSTANDEN SIND	XXIV
WEITERE P	UBLIKATIONEN	XXIV
Beiträge 2	zu Fachkonferenzen	XXIV

# Kurzfassung

Fanconi-Anämie (FA) ist, mit Ausnahme von Mutationen in FANCR/RAD51, eine autosomal-rezessive oder X-chromosomal vererbte Krankheit, die sich durch eine ausgesprochene klinische als auch genetische Heterogenität auszeichnet. Neben einem fortschreitenden Knochenmarksversagen zählen zu den typischen Merkmalen eine Vielzahl an angeborenen Fehlbildungen, wie beispielsweise Radialstrahlanomalien, Minderwuchs oder Pigmentierungsstörungen. Zudem besteht für FA-Patienten ein überdurchschnittlich hohes Risiko bereits in jungen Jahren an akuter myeloischer Leukämie oder soliden Tumoren zu erkranken. Bislang konnten in 21 FA-Genen (FANCA, -B, -C, -D1, -D2, -E, -F, -G, -I, -J, -L, -M, -N, -O, -P, -Q, -R, -S, -T, -U oder -V) krankheitsverursachende Mutationen identifiziert werden, deren Proteinprodukte maßgeblich an der Aufrechterhaltung der Genomstabilität beteiligt sind und Komponenten des FA/BRCA-DNA-Reparaturweges darstellen. In der klassischen FA-Mutationsanalyse kommen meist Sanger-Sequenzierungen sowie MLPA- und Immunblot-Analysen zum Einsatz. Da im Wesentlichen keine Genotyp-Phänotyp-Korrelation besteht, gestaltet sich, gerade bei seltenen FA-Komplementationsgruppen, der Nachweis von krankheitsverursachenden Mutationen oftmals sehr zeit- und kostenintensiv. Während der letzten Jahre wurden verschiedene Strategien zur Anreicherung und Sequenzierung entwickelt, welche die parallele Sequenzanalyse einzelner ausgewählter Gene, ganzer Exome oder sogar des gesamten Genoms und somit eine kosten- und zeiteffiziente Mutationsanalyse ermöglichen. In der vorliegenden Arbeit wurden unterschiedliche Anreicherungsmethoden mit anschließender Hochdurchsatzsequenzierung auf ihre Anwendbarkeit in der molekulargenetischen FA-Diagnostik getestet, um klassische Mutationsanalyse-Methoden zu ergänzen oder möglicherweise sogar ganz ersetzen zu können.

Der erste Teil der Arbeit befasste sich mit der Etablierung eines FA-spezifischen Genpanels zur Genotypisierung von FA-Patienten. Nachdem die Methode zunächst anhand von FA-Patienten mit bekannten Mutationen optimiert werden musste, erwies sie sich als effizienter Ansatz zum Nachweis krankheitsverursachender Mutationen bei FA-Patienten unbekannter Komplementationsgruppe. Durch die FA-Panelanalyse konnten 37 von 47 unklassifizierten Patienten einer FA-Komplementationsgruppe zugeordnet werden, indem deren kausalen Mutationen bestimmt wurden. In einem weiteren Ansatz sollte die Anwendbarkeit eines kommerziellen Anreicherungspanels zur FA-Diagnostik untersucht werden. Auch hier konnte ein Großteil der krankheitsverursachenden Mutationen von fünf bekannten wie auch 13 nicht zugeordneten FA-Patienten detektiert und somit eine molekulargenetische Diagnose bei neun weiteren, zuvor unklassifizierten FA-Patienten, gestellt werden. Ferner wurden sechs ausgewählte Patienten, zusätzlich zur Panelanreicherung, per Exomanalyse untersucht. Zum einen konnten Mutationen in bekannten FA-Genen bestätigt oder neu identifiziert werden. Zum anderen wurden auch potentiell pathogene Mutationen in DNAaußerhalb des FA/BRCA-Signalweges bei zwei Patienten mit Reparaturgenen unbestätigter Verdachtsdiagnose FA verifiziert. So wurde bei mehreren Mitgliedern einer Familie mit unterschiedlichen Tumorerkrankungen eine zuvor unbeschriebene homozygote Nonsense-Mutation in der BER-Glykosylase NTHL1 nachgewiesen, für welche bislang erst zwei pathogene Mutationen als Auslöser eines neuen Krebssyndroms bekannt sind. Bei einem weiteren Patienten wurden compound-heterozygote Mutationen in *RPA1* detektiert, ein Gen für das bislang noch kein Krankheitsbild bekannt ist. Mit Hilfe der drei verschiedenen Anreicherungsstrategien konnten insgesamt 47 von 60 unklassifizierten FA-Patienten 13 verschiedenen Komplementationsgruppen eindeutig zugeordnet werden. Es zeigte sich dabei ein breites Spektrum an neuen, bislang unbeschriebenen FA-Mutationen. Den größten Anteil an der Gesamtzahl der nachgewiesenen Mutationen hatten Spleißmutationen, die auf eine Auswirkung auf das kanonische Spleißmuster untersucht wurden, um einen pathogenen Effekt nachweisen zu können.

Weiterhin schloss die Arbeit die Charakterisierung einzelner FA-Patienten bzw. Komplementationsgruppen mit ein. Dazu zählen die seltenen Untergruppen FA-T und FA-Q, für die jeweils ein neuer Patient identifiziert werden konnte. Durch die funktionelle Charakterisierung der dritten jemals beschriebenen FA-Q-Patientin konnten Einblicke in das Zusammenspiel der Reparatur von DNA-Quervernetzungen und der Nukleotidexzisionsreparatur gewonnen und die phänotypische Variabilität von FA durch die subjektive als auch zelluläre UV-Sensitivität der Patientin ergänzt werden. Darüber hinaus konnte das Mutationsspektrum in FA-I sowie FA-D2 erweitert werden. Eine genauere Untersuchung der Pseudogenregionen von *FANCD2* ermöglichte dabei die gezielte Mutationsanalyse des Gens.

Insgesamt konnten die Ergebnisse dieser Arbeit dazu beitragen, das Mutationsspektrum in FA zu erweitern und durch die Identifizierung und Charakterisierung einzelner Patienten neue Einblicke in verschiedene Komponenten des FA/BRCA-Signalweges zu erhalten. Es zeigte sich, dass neue DNA-Sequenzierungsstrategien in der FA-Diagnostik eingesetzt werden können, um eine effiziente Mutationsanalyse zu gewährleisten und klassische Methoden in Teilbereichen zu ersetzen.

# Summary

Fanconi anemia (FA) is, with the exception of mutations in FANCR/RAD51, an autosomal recessive or X-linked inherited disease that is characterized by a remarkable clinical and genetic heterogeneity. In addition to progressive bone marrow failure, typical features include a multitude of developmental malformations, such as radial ray anomalies, growth retardation or cutaneous pigment displacement. Additionally, FA patients have a higher risk for developing acute myelogenous leukemia or solid tumors early in life. To date, pathogenic mutations have been identified in 21 FA genes (*FANCA*, -*B*, -*C*, - *D1*, -*D2*, -*E*, -*F*, -*G*, -*I*, -*J*, -*L*, -*M*, -*N*, -*O*, -*P*, -*Q*, -*R*, -*S*, -*T*, -*U* or -*V*) whose protein products are responsible for maintaining genomic integrity and constitute components of the FA/BRCA DNA repair pathway. Typical methods for FA mutation analysis comprise Sanger sequencing as well as MLPA and immunoblot analyses. As no definite genotype-phenotype correlation exists, pathogenic mutation detection in rare subgroups is often quite time-consuming and cost-intensive. Within the last few years, distinct strategies for both enrichment and sequencing of a subset of genes, whole exomes or even the whole genome have been developed that facilitate a cost-effective and time-saving mutation analysis. In the present work different target-enrichment strategies followed by high-throughput sequencing were tested for their applicability in molecular genetic diagnostics of FA in order to complement or even replace classic strategies for mutation analysis.

The first part of this work addressed the establishment of an FA-specific gene panel for genotyping FA patients. After optimizing this method by means of FA patients with known mutations, this proved to be an efficient approach for detecting pathogenic mutations in FA patients of unknown complementation groups. Due to FA gene panel analysis, 37 of 47 unclassified FA patients were assigned to a complementation group based on the identification of their causative mutations. In another approach, a commercial enrichment panel was tested for its application in FA diagnostics. Again, most pathogenic mutations of five classified and 13 unclassified FA patients were detected, enabling a molecular diagnosis for nine previously unclassified FA patients. Moreover, six selected patients were studied by exome analysis in addition to panel enrichment. This allowed for mutations in known FA complementation groups to be confirmed or newly identified. Additionally, potentially pathogenic variants in DNA-repair genes outside the FA/BRCA pathway were verified in two patients with an unconfirmed suspected diagnosis of FA. One previously undescribed homozygous nonsense mutation in the BER glycosylase NTHL1 was detected in several members of one family with various tumors. For this gene, only two distinct pathogenic mutations were previously described to cause a novel cancer syndrome. In another patient, compound heterozygous mutations in RPA1 were detected, a gene for which no disease pattern is yet known. By means of the three different enrichment strategies a total of 47 of 60 unclassified FA patients were definitely assigned to 13 diverse complementation groups. In this context, a broad spectrum of previously undescribed mutations was identified. The majority of all verified mutations were splice mutations that were examined for an effect on the canonical splicing pattern in order to verify a pathogenic effect.

Additionally, this work also included the characterization of individual FA patients and complementation groups, respectively. These include the rare subgroups FA-T and FA-Q, for each of which one new patient was identified. Functional characterization of the third ever described FA-Q patient allowed new insights into the interplay of DNA interstrand-crosslink and nucleotide excision repair and broadened the spectrum of phenotypic variability of FA by the subjective and cellular UV sensitivity of this patient. Furthermore, the mutation spectrum in both FA-I and FA-D2 was expanded. Here, a closer investigation of the pseudogene regions of *FANCD2* facilitated a precise mutation screening of the gene.

Overall, the results of this work broadened the mutation spectrum of FA and allowed new insights into diverse components of the FA/BRCA pathway by identifying and characterizing individual patients. It became apparent that novel strategies for DNA sequencing can be applied in FA diagnostics to ensure an efficient mutation analysis, as well as to replace some parts of classical approaches.

# 1. Einleitung

#### 1.1 DNA-Reparatursysteme

Jeden Tag entstehen in unserem Genom neue Mutationen, die sowohl durch exogene Faktoren, wie UV-Licht, chemische Substanzen oder natürliche radioaktive Strahlung, als auch durch den Einfluss endogener Faktoren, wie reaktive Stoffwechselprodukte oder Fehler bei der Replikation, hervorgerufen werden können. Um die Mutationsrate dennoch gering zu halten, existieren in unserem Organismus verschiedene, spezifische DNA-Reparatursysteme, die bestimmte Mutationen erkennen und beheben können. Die Auswirkungen von Veränderungen im Genom können vielfältig sein. Zum einen sind Mutationen ein wesentlicher Bestandteil der Evolution. So unterscheidet sich das Humangenom vom Genom des Schimpansen um weniger als 1,3 % [Ebersberger et al., 2002; Chimpanzee Sequencing Analysis Consortium, 2005]. Genetische Variabilität existiert aber nicht nur zwischen verschiedenen Spezies sondern auch beim Vergleich verschiedener Individuen innerhalb einer Art. Beispielsweise beträgt die Varianz bei verschiedenen Menschen nur etwa 0,5 % und ist für die Unterschiede im Phänotyp verantwortlich. Allerdings können Mutationen innerhalb eines Individuums auch pathogene Auswirkungen auf einen Organismus sowie die Entstehung schwerer Erkrankungen zur Folge haben. Treten diese in somatischen Zellen auf, so wirken sie sich nur auf die jeweils betroffene Person aus. Entsteht eine pathogene Variante hingegen bereits in der Keimbahn, so kann diese auch auf Nachkommen übertragen werden.

Zur Behebung der schätzungsweise rund 70 000 DNA-Läsionen, die jeden Tag in einer humanen Zelle entstehen [Tubbs und Nussenzweig, 2017], stehen spezialisierte DNA-Reparaturmechanismen zu Verfügung, um den fehlerfreien Ablauf der DNA-Transkription und -Replikation zu gewährleisten. Dabei sind enzymatische Signalkaskaden für die Regulation dieser Abläufe von entscheidender Bedeutung. Beispielsweise kann man während des Zellzyklus den periodischen Auf- und Abbau von Cyclinen beobachten [Evans et al., 1983]. Zusammen mit Cyclin-abhängigen Serin/Threonin-Proteinkinasen (*cyclin-dependent kinases*; CDKs) bilden sie funktionell aktive Holoenzyme, welche zellzyklusabhängig verschiedene Substrate phosphorylieren und dadurch das Voranschreiten des Zellzyklus regulieren. In dessen Verlauf werden verschiedene Kontrollpunkte durchlaufen, um eine fehlerfreie Weitergabe der Erbinformation an die Tochterzelle zu gewährleisten. Tritt ein DNA-Schaden auf, wird der Zellzyklus am entsprechenden Kontrollpunkt angehalten, um die Reparatur des Schadens zu ermöglichen. Dabei unterscheidet man den G1/S-Phase-, Intra-S-Phase-, G2/M-Phase- sowie den Mitose-Kontrollpunkt. Falls eine Behebung des Schadens nicht möglich ist, tritt die Zelle in die prämature Seneszenz ein, geht in die p53-vermittelte Apoptose über oder entartet. Ein Überblick über wichtige Bestandteile des komplexen Netzwerkes an DNA-Reparatursystemen soll in den folgenden Abschnitten gegeben werden.

#### 1.1.1 Basenexzisionsreparatur

Einer der häufigsten endogenen DNA-Schäden im menschlichen Genom ist die Modifikation einzelner Basen. Diese Veränderungen entstehen z. B. durch Alkylierung, Desaminierung oder oxidative Reaktionen, die durch reaktive Sauerstoffspezies (ROS) ausgelöst werden können, welche als Nebenprodukt der Zellatmung in den Mitochondrien entstehen. Ein häufig genutzter Reparaturweg zur Behebung dieser DNA-Schädigungen ist die Basenexzisionsreparatur (BER) (Abbildung 1), wozu im Wesentlichen vier Enzyme benötigt werden [Kubota et al., 1996]. Dazu zählen eine Glykosylase, eine Endonuklease, eine Polymerase sowie eine Ligase. Zunächst wird die geschädigte Base von einer auf eine DNA-Läsion spezialisierten DNA-Glykosylase erkannt. Diese spaltet die N<sup>c</sup>-glykosidische Bindung zwischen der Base und der Desoxyribose, so dass eine Apurin/Apyrimidin (AP)-Stelle entsteht. Bifunktionale Glykosylasen besitzen zudem eine intrinsische AP-Lyaseaktivität, wodurch die Phosphodiester-Bindung des DNA-Rückgrates gespalten werden kann. Diese Aufgabe übernimmt bei monofunktionalen Glykosylasen die AP-Endonuklease APE1 [Demple et al., 1991]. Durch die Spaltung der Phosphodiester-Bindung 5<sup>c</sup> der AP-Stelle entsteht jeweils ein 3<sup>c</sup>-OH- sowie ein 5<sup>c</sup>-Desoxyribose-5<sup>c</sup>-Phosphat (5<sup>c</sup>-dRP)-Ende.



Abbildung 1: Schematische Übersicht der Basenexzisionsreparatur. Eine modifizierte Base (Stern) wird von einer spezifischen Glykosylase als Schaden erkannt und entfernt, wobei die Zucker-Phosphat-Kette intakt bleibt. Dadurch entsteht zunächst eine abasische Stelle (AP-Stelle). Durch Spaltung der Phosphodiester-Bindung entsteht ein Einzelstrangbruch, der durch die Synthese eines (*short patch*) oder mehrerer Nukleotide (*long patch*) und anschließender Ligation behoben wird.

Ab diesem Punkt gibt es zwei Möglichkeiten der weiteren Reparatur. Entweder wird durch die DNA-Polymerase  $\beta$  ein Nukleotid eingebaut sowie der 5'-dRP-Rest entfernt, so dass der LigIII-XRCC1-Komplex den Einzelstrangbruch verschließen kann (*short patch*) [Caldecott et al., 1994; Matsumoto und Kim, 1995]. Alternativ dazu besteht die Möglichkeit, dass durch die Polymerasen  $\delta$  und  $\varepsilon$  PCNA-abhängig mehrere Nukleotide synthetisiert werden, wodurch einige Basen 3' der Schnittstelle verdrängt werden. Diese können durch die Flap-Endonuklease 1 (FEN1) entfernt und die Lücke durch eine Ligase verschlossen werden (*long patch*) [Frosina et al., 1996; Klungland und Lindahl, 1997; Kim et al., 1998].

Schädigungen des BER-Reparaturweges durch Mutationen in essenziellen Genen haben schwerwiegende Folgen für den Organismus. So werden verschiedene Formen der adenomatösen Polyposis durch Keimbahn-Mutationen in *MUTYH* oder *NTHL1*, die für jeweils eine der bislang 11 bekannten BER-Glykosylasen codieren, hervorgerufen [Al-Tassan et al., 2002; Weren et al., 2015].

#### 1.1.2 Nukleotidexzisionsreparatur

Zur Entfernung Helix-verformender Läsionen, wie etwa Thymin-Dimere, die durch UV-induzierte Vernetzungen entstehen können, wird die Nukleotidexzisionsreparatur (NER) genutzt. Dem Reparaturweg kommt eine wichtige Bedeutung in der G1-Phase des Zellzyklus zu. Die NER lässt sich in zwei Reparaturwege unterteilen, die sich in ihrer initialen Schadenserkennung unterscheiden (Abbildung 2A+B). Die Globale Genom-NER (GG-NER) ist für die Schadenserkennung des gesamten Genoms verantwortlich, wohingegen die Transkriptions-gekoppelte NER (*transcription-coupled;* TC-NER) Blockaden während der Transkription behebt, welche die RNA Polymerase II an der Elongation des Transkriptes hindern. Defekte in Genen, die an der NER beteiligt sind, haben schwere klinische Krankheitsbilder zur Folge, wie Xeroderma pigmentosum (XP), Cockayne-Syndrom (CS) oder Trichothiodystrophie [Cleaver et al., 2009]. Zu den Symptomen zählen unter anderem Überempfindlichkeit gegenüber UV-Licht, Hautpathologien, Entwicklungsverzögerungen oder ein erhöhtes Krebsrisiko.

Zur initialen Schadenserkennung während der GG-NER ist ein Komplex aus XPC, RAD23B sowie CETN2 essenziell, welcher die DNA fortwährend nach Helix-verformenden Läsionen absucht [Masutani et al., 1994; Sugasawa et al., 1998; Nishi et al., 2005]. Cyclobutan-Pyrimidin-Dimere (CPD), welche durch die Verknüpfung von zwei benachbarten Pyrimidinen unter Ausbildung eines Butanringes entstehen und den häufigsten durch UV-Strahlung induzierten Schaden darstellen, sind allerdings nur ein schwaches Substrat für den XPC-Komplex [Sugasawa et al., 2001]. Bei einer solchen Schädigung bindet zunächst das DDB1-DDB2-Heterodimer an die Schadensstelle, das sich mit weiteren Proteinen (Cullin 4A, Roc 1) zu einem größeren Komplex zusammenlagert und als Ubiquitin-E3-Ligase fungiert [Groisman et al., 2003]. Durch die aktive Ligase wird XPC an den Ort des Schadens rekrutiert und zusammen mit Cullin 4A sowie DDB2 ubiquitiniert, woraufhin der E3-Ligase-Komplex von der DNA entlassen wird [Sugasawa et al., 2005]. Danach kann die Entfernung des Schadens beginnen, welche mit dem TC-NER-Signalweg identisch ist.



Abbildung 2: DNA-Reparatur durch Nukleotid-Exzision. A: Während der Transkription wird die RNA-Polymerase II (RNAPII) an der Elongation durch einen Schaden (gelbes Symbol) gehindert. Die initiale Erkennung der TC-NER (unten links) erfolgt durch CSA und CSB. Bei der GG-NER (unten rechts) ist ein stabiler Komplex aus XPC, RAD23B (23B) sowie CETN2 beteiligt. Diesem kann noch das DDB1-DDB2-Heterodimer voran geschaltet sein. B: Die Entfernung des Schadens folgt einem gemeinsamen Ablauf. Es lagert sich, unter Beteiligung von XPB, ein Multiprotein-Komplex (TFIIH) an, der eine XPD-vermittelte Helikasefunktion besitzt. Weitere Faktoren kommen hinzu, wie RPA, XPA sowie die Endonukleasen XPF-ERCC1 und XPG. Letztere sind für Einschnitte an der Schadensstelle zuständig. Anschließend können die Neusynthese der entfernten Nukleotide sowie die Ligation des DNA-Stranges erfolgen. Verändert nach [Cleaver et al., 2009].

Der initiale Schritt der TC-NER wird zunächst durch die Blockade der RNA-Polymerase II ausgelöst, wodurch die Bildung eines Komplexes aus CSA und CSB initiiert wird. Die RNA-Polymerase wird entfernt, so dass die Läsion für eine Reparatur zugänglich wird. Nach dieser initialen Schadenserkennung wird bei beiden NER-Signalwegen der sogenannte TFIIH-Komplex an den Ort des Schadens rekrutiert. Weiterhin binden RPA, XPA sowie XPG und die DNA kann im Bereich der Läsion entwunden werden. Durch XPA wird zunächst die XPF-ERCC1-Endonuklease rekrutiert, welche 5' vom Schaden schneidet [Volker et al., 2001]. Dadurch kommt es zur Aktivierung von XPG, so dass ein weiterer Einschnitt 3' der Läsion erfolgen kann. Nachdem der Schaden entfernt wurde, kann der Bereich der entstandenen Lücke durch Bindung von DNA-Polymerasen ( $\delta$  und  $\varepsilon$ ) sowie weiterer Faktoren neu synthetisiert und die Lücke durch DNA-Ligase I verschlossen werden.

#### 1.1.3 Basenfehlpaarungs-Reparatur

Bei der Reparatur von Basenfehlpaarungen, die auch als Mismatch-Reparatur (MMR) bezeichnet wird, können sowohl einzelne, falsch gepaarte Basen, als auch kleinere Insertionen sowie Deletionen, wie sie oft bei repetitiven Sequenzen von Mikrosatelliten vorkommen, behoben werden. Eine der häufigsten Fehlpaarungen entsteht zwischen Guanin und Thymin, ausgelöst durch die Methylierung eines Cytosins zu 5'-Methylcytosin. Durch hydrolytische Desaminierung entsteht daraus wiederum Thymin, welches bei der nächsten Replikation mit einem Adenin paart, sofern die Fehlpaarung zuvor nicht erkannt und aufgehoben wurde.

Ein wesentlicher Bestandteil in der Detektion von Basenfehlpaarungen ist MSH2, das zunächst mit weiteren Proteinen Heterodimere bildet. So erkennt der als MutS $\alpha$  bezeichnete Komplex aus MSH2 und MSH6 Fehlpaarungen einzelner Basen sowie bis zu zwei Nukleotide umfassende Insertionen oder Deletionen [Drummond et al., 1995]. Durch das MSH2-MSH3-Heterodimer (MutS $\beta$ ) können auch längere Insertionen und Deletionen von bis zu 14 Nukleotiden detektiert werden [Palombo et al., 1996]. MutS-Heterodimere bilden zusammen mit MUTL-Endonukleasen einen Komplex, der für die Strang-Diskriminierung wichtig ist. Die gemeinsame Komponente der verschiedenen humanen MUTL-Heterodimere ist MLH1. Der am besten charakterisierte Zusammenschluss besteht dabei aus MLH1 und PMS2 und wird als MutL $\alpha$  bezeichnet. Zusammen mit PCNA wird die Bindung der Exonuklease 1 vermittelt, welche die Fehlpaarung entfernt. Danach können die Neusynthese der entfernten Nukleotide durch Polymerase  $\delta$  und die anschließende Verknüpfung des DNA-Stranges durch eine Ligase stattfinden. Eine Inaktivierung dieses Reparaturweges führt zu einer erhöhten Mutationsrate sowie einem gesteigerten Krebsrisiko. Keimbahnmutationen in MMR-Genen sind z. B. mit dem Hereditären Nicht-Polypösen Kolonkarzinom (*bereditary non-polyposis colon cancer;* HNPCC) assoziiert.

#### 1.1.4 Reparatur von DNA-Doppelstrangläsionen

DNA-Doppelstrangbrüche (DSBs) können durch Zytostatika, ionisierende Strahlung oder freie Radikale hervorgerufen werden. Sie induzieren in der Regel einen Replikationsstopp und können u.a. zu Translokationen und Deletionen führen. Es existieren verschiedene Mechanismen, mit welchen solche Schäden behoben werden können, wie das *Non-homologous end joining* (NHEJ) oder die Homologe Rekombination (HR). Ein wichtiger Bestandteil der Doppelstrangbruch-Reparatur ist der MRN-Komplex, bestehend aus MRE11, RAD50 sowie NBS1. Der Komplex fungiert zum einen als früher Sensor von Doppelstrangbrüchen, zum anderen werden ihm aber auch noch weitere Funktionen zugesprochen [Lamarche et al., 2010]. So ist er an der Aktivierung von Zellzyklus-Kontrollpunkten sowie an der Aufrechterhaltung von Telomeren beteiligt und spielt eine wichtige Rolle während der Meiose und Replikation. Mutationen in einer Komponente des MRN-Komplexes führen entweder zur AT-ähnlichen Krankheit (*Ataxia telangietasia-like disease*; ATLD) [Stewart et al., 1999], zum Nijmegen-Breakage-Syndrom (NBS) [Varon et al., 1998] oder sind mit der NBS-ähnlichen Krankheit [Waltes et al., 2009] assoziiert. Eine Übersicht über verschiedene Möglichkeiten der DSB-Reparatur ist in den folgenden Abschnitten sowie in Abbildung 3 und Abbildung 4 wiedergegeben.

# a. Non-homologous end joining

Eine Reparatur von Doppelstrangbrüchen durch die Ligation von DNA-Enden (*end joining*) ist während jedes Abschnittes der Interphase möglich und weist eine hohe Leistungsfähigkeit auf, da keine Reparaturvorlage, wie ein Schwesterchromatid oder homologes Chromosom, nötig ist. Während der G1- sowie frühen S-Phase wird ein Doppelstrangbruch zunächst vom MRN-Komplex gebunden, der die Proteinkinase ATM rekrutiert (Abbildung 3). Die Kinase phosphoryliert im Anschluss verschiedene Substrate, beispielsweise das Histon H2AX [Burma et al., 2001]. Durch die Modifikation an Serin 139 von H2AX, das nun als γH2AX bezeichnet wird, kann MDC1 an die Schadensstelle binden, wo es als Plattform zur Anlagerung weiterer DNA-Schadensproteine dient, wie etwa die Ubiquitinligasen RNF8 und RNF168 [Stewart et al., 2003; Mailand et al., 2007; Schwertman et al., 2016]. Die Funktion dieser Ligasen fördert wiederum die Bindung von 53BP1, das nur an Nukleosomen bindet, welche sowohl an H2AK15 ubiquitiniert als auch an H4K20 methyliert sind.

Im Anschluss wird 53BP1 ebenfalls von ATM phosphoryliert, was die Bindung von RIF1 und PTIP ermöglicht. Durch diesen Zusammenschluss wird die Anlagerung wichtiger Faktoren der Homologen Rekombination, wie beispielsweise CtIP und BRCA1, und damit die Resektion der DNA am Doppelstrangbruch verhindert. Dies wird zusätzlich durch die Bindung der DNA-Helikase B (HELB) an RPA unterstützt, da durch die 5'-3'-Translokaseaktivität von HELB die Nukleaseaktivität von EXO1, BLM sowie DNA2 unterdrückt wird [Tkac et al., 2016].

An die offenen Enden des Doppelstrangbruchs kann anschließend das Ku70-Ku80-Heterodimer binden, das die katalytische Untereinheit der Protein-Kinase DNA-PK, DNA-PKcs (*protein kinase catalytic subunit*), rekrutiert und mit ihr zusammen den aktiven DNA-PK-Komplex bildet [Gottlieb und Jackson, 1993]. Weitere Effektorproteine werden zur Schadensstelle rekrutiert und durch die Kinase-Aktivität aktivierend phosphoryliert. Dazu gehört die Nuklease Artemis, welche sowohl eine 5'-Exonuklease-Aktivität besitzt, als auch im Komplex mit DNA-PK als Endonuklease wirkt. Durch diesen Zusammenschluss werden die Enden des Doppelstrangbruchs prozessiert, so dass diese anschließend ligiert werden können. Die Ligation wird durch den XRCC4-LIG4-Komplex sowie weitere Faktoren, wie XLF und PAXX, vermittelt [Nick McElhinny et al., 2000; Riballo et al., 2009; Ochi et al., 2015].



Abbildung 3: Übersicht über die Doppelstrangbruch-Reparatur durch klassisches nicht-homologes (*non-homologous*) sowie alternatives *end joining*. Der MRN-Komplex bindet an die Schadensstelle und rekrutiert ATM. Durch Phosphorylierung (P) von H2AX (dunkelgrau: Histone) wird die Bindung von MDC1 vermittelt, wodurch weitere Signalkaskaden ausgelöst werden. So werden die Nukleosomen, die den DNA-Bruch flankieren, weiter modifiziert. Dabei wird H4K20 methyliert, während Histon 2A (H2A) durch RNF168 an K15 ubiquitiniert wird (oben links). Durch ATM phosphoryliertes 53BP1 bindet an Nukleosomen, die beide Modifikationen aufweisen und fungiert so als Bindestelle für RIF1 und PTIP. Dadurch wird die Resektion (rote Pfeile) der DNA-Enden an der Schadensstelle, welche durch Proteine der Homologen Rekombination (braun) gefördert wird, verhindert (rechts oben). Denselben Effekt besitzt die Bindung von HELB an RPA (Mitte). Zur Verbindung der DNA-Enden stehen zwei Möglichkeiten zur Verfügung. Entweder wird der Doppelstrangbruch durch das Ku70-Ku80-Heterodimer gebunden und durch Ligase IV ligiert (unten links). Bei Vorhandensein von Mikrohomologien kann ein alternativer Weg genutzt werden, an welchem PARP1, POLQ sowie Ligase I oder III beteiligt sind (unten rechts). Verändert nach [Hustedt und Durocher, 2016].

Durch Studien mit Ku70-, Ku80- oder Ligase IV-defizienten Hefemutanten konnte gezeigt werden, dass neben dem klassischen NHEJ weitere alternative Varianten (*alternative end-joining*; alt-EJ) existieren [Boulton und Jackson, 1996; Wilson et al., 1997]. Eine gemeinsame Voraussetzung stellt das Vorhandensein kurzer homologer Abschnitte am Doppelstrangbruch dar, weshalb alternatives *end joining* auch als Mikrohomologievermitteltes *end joining (microhomology-mediated end joining;* MMEJ) bezeichnet wird. In Abbildung 3 ist beispielsweise der häufig genutzte Polymerase Theta-vermittelte alt-EJ-Weg dargestellt. Anstelle des Ku70-Ku80-Heterodimers bindet hierbei PARP1 an die Schadensstelle. Nach der Prozessierung des Doppelstrangbruchs von beiden Seiten durch die Exonuklease CtIP erfolgen Basenpaarungen zwischen den homologen Bereichen. POLQ scheint dabei sowohl für die Suche nach Mikrohomologien, die Entfernung von nicht-homologen Überhängen sowie für die Basenpaarungen zwischen den homologen Bereichen zuständig zu sein [Wyatt et al., 2016]. Für das Verbinden der Enden wird anschließend entweder DNA Ligase I oder III benötigt.

#### b. Homologe Rekombination

Die Homologe Rekombination (HR) ermöglicht eine fehlerfreie Reparatur von DSBs, sofern ein homologes Schwesterchromatid als Reparaturvorlage zur Verfügung steht. Dieser Vorgang beschränkt sich daher auf die S- sowie G2-Phase des Zellzyklus und ist genauestens kontrolliert, um eine illegitime Rekombination mit partiell homologen DNA-Abschnitten zu vermeiden. Eine Übersicht über diesen Reparaturweg ist in Abbildung 4 zusammengestellt. Zunächst wird der DSB vom MRN-Komplex gebunden, welcher die Aktivierung der ATM-Kinase vermittelt. Analog zur Reparatur durch NHEJ werden verschiedene Substrate durch ATM phosphoryliert (Abbildung 4). Durch die erhöhte Aktivität der CDK-Kinase während der S-Phase werden nun allerdings auch NBS1 sowie CtIP phosphoryliert, wodurch zunächst die Endonuklease-Aktivität des MRN-Komplexes stimuliert wird. Zudem wird die Bindung von BRCA1 an CtIP vermittelt, welches wiederum inhibierend auf 53BP1 wirkt. Ein weiteres Ziel der CDK-vermittelten Phosphorylierung stellt HELB dar, welches daraufhin aus dem Zellkern transportiert wird. Damit werden sowohl die 53BP1- als auch die HELB-vermittelte Unterdrückung der DNA-Resektion am Doppelstrangbruch aufgehoben. Die Prozessierung der Enden wird zudem durch EXO1 und DNA2-BLM unterstützt [Bolderson et al., 2010; You und Bailis, 2010]. Als Resultat entstehen freie 3'-Enden, die von RPA-Proteinen gebunden werden.

Im nächsten Schritt vermitteln BRCA1, BRCA2 sowie PALB2, unter Verdrängung von RPA, die Anlagerung von RAD51-Proteinen an die einzelsträngige DNA. Durch Polymerisation entstehen RAD51-Filamente, die sich an komplementäre Regionen des homologen Stranges anlagern und so die Strang-Invasion einleiten. Zunächst bildet sich durch die Verdrängung des Schwesterchromatids eine Schlaufe aus, die als *D-Loop*-Struktur (*displacement loop*) bezeichnet wird. Sie kann direkt unter Beteiligung der MUS81-EME1-Exonuklease aufgelöst werden. Klassischerweise bildet sich allerdings aus dem D-Loop eine doppelte *Holliday-Junction* (HJ)-Struktur, die sich auf verschiedene Weise auflösen lässt. Unter Beteiligung des BLM-TOPOIIIA-RMI1/2 (BTR)-Komplexes entstehen *Non-Crossover*-Produkte, wohingegen durch SLX1-SLX4 oder MUS81-EME1 sowohl *Crossover*-Produkte gebildet werden können.



Abbildung 4: Übersicht über die Doppelstrangbruch-Reparatur durch Homologe Rekombination. Durch die erhöhte Aktivität der CDK-Kinase in der S-Phase werden NBS1, CtIP sowie HELB phosphoryliert (P). Dies führt zur verstärkten Endonuklease-Aktivität des MRN-Komplexes sowie zum Export von HELB aus dem Zellkern. An phosphoryliertes CtIP kann BRCA1 binden (oben links), das reprimierend auf 53BP1 wirkt. Dadurch wird die Anlagerung von DNA2-BLM sowie EXO1 ermöglicht, wodurch die Resektion der DNA (rote Pfeile) weiter gefördert wird. Einzelsträngige DNA wird zunächst von RPA gebunden. BRCA1, BRCA2 sowie PALB2 fördern den Austausch von RPA durch RAD51, so dass sich RAD51-Filamente ausbilden können, welche die Strang-Invasion vermitteln (Mitte). Dabei können unterschiedliche Strukturen entstehen, die über verschiedene Proteinkomplexe gelöst werden können (unten). Eine Alternative in Form des RAD52- und XPF-ERCC1-vermittelten Einzelstrang-Annealings ist rechts oben dargestellt. Verändert nach [Hustedt und Durocher, 2016].

Eine Variante der Homologen Rekombination besteht in der Auffüllung und Verknüpfung der RPAgebundenen Einzelstränge im Bereich von repetitiven Sequenzen (*single strand annealing*; SSA). Daran beteiligt ist neben RAD52, welches für die Ausrichtung und Anlagerung der repetitiven Sequenzabschnitte zuständig ist, auch die ERCC1-XPF-Endonuklease [Ciccia und Elledge, 2010]. Eine weitere Möglichkeit der HR-Reparatur besteht außerdem im SDSA (*synthesis-dependent strand-annealing*)-Signalweg, der an die Strang-Invasion anschließt. Dabei kommt es zur Hybridisierung der neu synthetisierten Stränge, wodurch nur Produkte ohne *Crossover* entstehen. Hierbei spielt die Helikase RTEL1 eine wichtige Rolle [Barber et al., 2008].

#### c. Reparatur von Interstrang-Quervernetzungen

Die kovalente Verknüpfung eines DNA-Doppelstranges, auch als *interstrand crosslink* (ICL) bezeichnet, stellt eine große Herausforderung für das DNA-Reparatursystem dar. ICL-induzierte Schäden werden normalerweise während der G2-Phase des Zellzyklus repariert, wenn die Replikationsgabel am Voranschreiten gehindert wird und an der Schadensstelle stagniert. Bei der Entfernung des ICLs entsteht in der Regel ein DNA-Doppelstrangbruch der, falls er nicht kontrolliert induziert wird, zu einem freien DNA-Ende führen kann. Der genaue Ablauf des FA/BRCA-Reparaturweges wird unter 1.2.4 näher beschrieben.

Kürzlich wurde zudem ein alternativer Weg zur Entfernung eines ICLs an stagnierten Replikationsgabeln aufgeklärt. Semlow et al. 2016 [Semlow et al., 2016] konnten anhand von Psoralen-induzierten Quervernetzungen zeigen, dass diese auch ohne Einschnitte in den DNA-Strang gelöst werden können (Abbildung 5). Psoralen verbindet sich unter Einwirkung von UVA-Licht mit gegenüber liegenden Thyminen zu einem Addukt. Durch eine Glykosylase kann die N-glykosidische Bindung, welche die Desoxyribose mit der Base verknüpft, an einer Seite der DNA-Quervernetzung gespalten werden. Dadurch entsteht eine AP-Stelle auf der einen Seite sowie ein Thymin-Psoralen-Monoaddukt auf der gegenüberliegenden Seite, dargestellt in Abbildung 5A. Unter Beteiligung der Transläsions-DNA-Polymerase REV1 kann zunächst über die AP-Stelle hinweg synthetisiert werden. Der CMG-Komplex (CDC45-MCM2-7-GINS), welcher für die Entwindung des DNA-Doppelstranges zuständig ist, wird daher nicht an der Schadensstelle blockiert und fällt ab, sondern bleibt an der DNA gebunden. Letztlich müssen sowohl die AP-Stelle als auch das Monoaddukt entfernt werden, was vermutlich durch Endonukleasen vermittelt wird. Als verantwortliche Glykosylase konnte in der beschriebenen Studie die BER-Glykosylase NEIL3 identifiziert werden. Bislang wurde dieser Reparaturweg allerdings nur mit Hilfe von Plasmidkonstrukten in zellfreien Xenopus-Eiextrakten gezeigt. Eine Bedeutung dieses Reparaturweges *in vivo* bleibt damit unbestimmt.



Abbildung 5: Übersicht zur Entfernung eines Psoralen-induzierten ICLs durch die Gykosylase NEIL3. A: Ein DNA-Ausschnitt zeigt die Quervernetzung zweier Thymine (I) durch Psoralen (oben). Durch die Spaltung einer N-glykosidischen Bindung (rote Pfeilspitze) entstehen eine abasische Stelle (AP-Stelle) sowie ein Thymin-Psoralen-Monoaddukt (unten). B: Durch Psoralen sind zwei DNA-Stränge verknüpft (blau). Durch das Voranschreiten der Replikationsgabeln nähert sich auch die CMG-Helikase dem Psoralen-induzierten ICL an. Die Glykosylase NEIL3 spaltet auf einer Seite des ICLs die N-glykosidische Bindung zwischen der Desoxyribose und der Base. Dadurch entsteht eine AP-Stelle (Kreis) sowie ein Monoaddukt (blau, nach außen geklappt). Nach Transläsionssynthese (TLS) und Entfernung der AP-Stelle sowie des Addukts ist der Schaden behoben. "G": Guanin. "A": Adenin. "C": Cytosin. Verändert nach [Semlow et al., 2016].

# 1.2 Fanconi-Anämie

## 1.2.1 Erstbeschreibung und Inzidenz

Der Schweizer Kinderarzt Guido Fanconi (Abbildung 6) berichtete erstmals vor 90 Jahren über den Fall dreier Brüder, die im Alter von fünf bis sieben Jahren eine progressiv fortschreitende Anämie entwickelten, welche schließlich in allen drei Fällen zum Tode führte. Gemeinsame klinische Merkmale dieser Kinder waren, neben der schweren Anämie, auch diverse Fehlbildungen wie Mikrozephalie, Hypogonadismus sowie Hyperpigmentierungen der Haut. Dieses Krankheitsbild veröffentlichte Fanconi im Jahre 1927 als "Familiäre infantile perniziosaartige Anämie (perniziöses Blutbild und Konstitution)" [Fanconi, 1927], welches später nach ihm benannt wurde und uns heute als Fanconi-Anämie (FA) bekannt ist.



Abbildung 6: Der Schweizer Kinderarzt Guido Fanconi auf einer Aufnahme um 1959 (Quelle: Ze'ev Aleksandrowicz).

Mit einer Prävalenz von 1-5 pro 1 000 000 zählt FA zu den seltenen Erkrankungen, welche in allen ethnischen Gruppen auftritt [Joenje und Patel, 2001]. Die Heterozygoten-Frequenz liegt bei rund 1:200 [Rosenberg et al., 2011], allerdings ist diese Rate in bestimmten Populationen weit höher. So beträgt die Trägerschaft bei Aschkenasim 1:89, bei der Afrikaans-sprechenden Bevölkerung Südafrikas 1:77 und ist unter den Gitanos, einer Gruppe spanischer Roma, sogar mit 1:64–1:70 angegeben [Rosendorff et al., 1987; Verlander et al., 1995; Callen et al., 2005].

#### 1.2.2 Zellulärer und klinischer Phänotyp

Die Ursache der Erkrankung war lange Zeit unbekannt. 1964 konnte erstmals gezeigt werden, dass FA mit chromosomaler Instabilität einhergeht, welche sowohl spontan auftritt als auch durch die Zugabe von alkylierenden, DNA-quervernetzenden Substanzen wie Mitomycin C (MMC) oder Diepoxybutan (DEB) induziert und verstärkt werden kann [Schroeder et al., 1964]. Diese zelluläre Hypersensitivität lässt sich zur Differentialdiagnostik nutzen, indem man die Chromosomenbruchrate und die Anzahl von Rearrangements von behandelten und unbehandelten Zellen am Mikroskop bestimmt. Aufgrund der erhöhten Instabilität lassen sich dabei charakteristische Chromatidbrüche und Chromosomenaberrationen, sogenannte Radialfiguren, beobachten, welche durch die Zusammenlagerung nicht-homologer Chromatiden entstehen. Des Weiteren lässt sich der klastogene und zytotoxische Effekt von MMC und DEB auch via Durchflusszytometrie nachweisen, da Zellen von FA-Patienten während der späten S-Phase/frühen G2-Phase des Zellzyklus arretieren [Seyschab et al., 1993]. Dieser Zellzyklusarrest grenzt FA deutlich von anderen Syndromen ab, welche ebenfalls eine erhöhte Chromosomenbrüchigkeit, nicht aber den FA-typischen Zellzyklusarrest aufweisen. Dazu zählen beispielsweise NBS [New et al., 2005], Ataxia telangiectasia (AT) [Taniguchi et al., 2002], das Warsaw-Breakage- oder auch das Roberts-Syndrom [van der Lelij et al., 2010].

Klinisch erscheint FA sehr heterogen [Shimamura und Alter, 2010], wobei selbst innerhalb von Familien Unterschiede bezüglich des Phänotyps auftreten können [Koc et al., 1999; Neveling et al., 2009]. Dabei ist eine Beteiligung von Umwelteinflüssen, epigenetischer Faktoren oder die Wirkung von *Modifier*-Genen für die unterschiedliche Ausprägung desselben Genotyps möglich. Die Bedeutung des ethnischen Hintergrundes konnte beispielsweise durch Untersuchungen des *FANCC*-Gens gezeigt werden. In westlichen Ländern sind biallelische Mutationen in diesem Gen bei etwa 10–15 % der FA-Patienten ursächlich. Diese Gruppe zählt nach FA-A (~60 %) und noch vor FA-G (~10 %) zu den häufigsten FA-Komplementationsgruppen. Aufgrund einer *Founder*-Mutation liegt der Anteil von FA-C-Patienten bei den Aschkenasim allerdings bei etwa 80 % [Verlander et al., 1995]. Die häufigste Mutation ist hier die als IVS4+4A>T veröffentlichte Spleiß-Mutation c.456+4A>T, die lange nur innerhalb dieser Gruppe bekannt war und dort mit einem schweren Phänotyp assoziiert ist [Verlander et al., 1994; Verlander et al., 1995]. Hingegen konnte bei einer Untersuchung von japanischen FA-Patienten diese Mutation bei acht Patienten im homozygoten Zustand nachgewiesen werden, wobei diese einen deutlich milderen Phänotyp zeigten [Futaki et al., 2000].

häufigsten phänotypischen Auffälligkeiten von FA-Patienten Zu zählen den das progressive Knochenmarksversagen sowie eine Vielzahl verschiedener kongenitaler Fehlbildungen, die in Tabelle 1 zusammengefasst sind. Dazu gehören Radialstrahlanomalien, Minderwuchs, Organfehlbildungen sowie Pigmentierungsstörungen, die häufig in Form von Café-au-lait-Flecken, aber auch als hypopigmentierte Vitiligo-Flecken auftreten können. Allerdings ist hierbei ist zu beachten, dass rund 25-30 % der FA-Patienten keine angeborenen Fehlbildungen zeigen. Typischerweise manifestiert sich das progressive Knochenmarksversagen bereits im Alter von sieben bis acht Jahren, welches bis hin zur Panzytopenie führen kann [Kutler et al., 2003]. Zur Wiederherstellung des hämatopoetischen Systems ist deshalb in vielen Fällen eine Stammzelltransplantation nötig. Weiterhin besteht für FA-Patienten ein stark erhöhtes Risiko, schon früh an einer akuten myeloischen Leukämie (AML) oder soliden Tumoren, insbesondere Plattenepithelkarzinome des Mund- und Rachenraumes, zu erkranken.

Einigen seltenen FA-Untergruppen wird allerdings nur ein FA-ähnlicher Phänotyp zugeschrieben, da bei den wenigen zugeordneten Patienten bislang das für FA charakteristische Knochenmarksversagen ausblieb. Dazu zählt beispielsweise der einzige bekannte FA-O-Patient, bei dem weder hämatologische Auffälligkeiten noch eine Tumorentstehung beobachtet wurden [Vaz et al., 2010]. Dies gilt ebenso für beide FA-R-Patienten [Ameziane et al., 2015; Wang et al., 2015] wie auch für den einzigen bisher identifizierte FA-U-Patienten [Park et al., 2016]. Bei der kürzlich beschriebenen FA-S-Patientin wurde mit 23 Jahren Brustkrebs beschrieben, allerdings blieb auch hier ein Versagen des Knochenmarks bislang aus [Sawyer et al., 2015].

Fehlbildungen	Häufigkeit [%]
Kleinwuchs	40
Pigmentierungsstörungen der Haut	40
Obere Extremitäten (Daumen, Hand, Radius, Ulna )	35
Skelett (Kopf, Gesicht, Hals, Wirbelsäule)	20
Nieren	20
Augen	20
Ohren	10
Herz	6
Genitaltrakt und Keimdrüsen	
Männer	25
Frauen	2
Entwicklungsverzögerungen	10
Untere Extremitäten (Fuß, Bein)	5
Gastrointestinal	5
Zentrales Nervensystem	3

Tabelle 1: Übersicht über häufige kongenitale Fehlbildungen bei FA-Patienten. Verändert nach [Shimamura und Alter, 2010].

#### 1.2.3 Genetischer Hintergrund

Wie beim klinischen Phänotyp zeigt sich auch genetisch eine hohe Heterogenität. FA wird, mit Ausnahme von *FANCB* und *FANCR*, autosomal-rezessiv vererbt (*FANC-A*, -*C*, -*D1/BRCA2*, -*D2*, -*E*, -*F*, -*G*, -*I*, -*J/BRIP1*, -*L*, -*M*, -*N/PALB2*, -*O/RAD51C*, -*P/SLX4*, -*Q/ERCC4*, -*S/BRCA1*, -*T/UBE2T*, -*U/XRCC2*, -*V/REV7*). Die Komplementationsgruppe FA-M muss allerdings kritisch betrachtet werden, da die Gruppe um Singh et al. 2009 [Singh et al., 2009] zeigen konnte, dass der zuvor beschriebene FA-M-Indexpatient [Meetei et al., 2005] auch Träger biallelischer *FANCA*-Mutationen ist. Zudem konnte die defiziente FANCD2-Monoubiquitinierug der untersuchten Zelllinie durch ektopische Expression von *FANCA*, nicht aber von *FANCM*, behoben werden. Allerdings blieb die Hypersensitivität der Patientenzellen gegenüber DNAquervernetzenden Substanzen auch nach der Korrektur der Zellen mit *FANCA* erhalten.

Eine Ausnahme zur autosomal-rezessiven Vererbung von FA bildete bislang nur *FANCB*, welches einem Xchromosomalen Erbgang folgt. Dagegen wurde eine autosomal-dominante Vererbung *per se* bislang nicht beschrieben. Jedoch wurden kürzlich zwei Arbeiten über *de novo*-Mutationen in *FANCR/RAD51* bei FA-Patienten veröffentlicht, wobei für die jeweiligen Mutationen ein dominant-negativer Effekt auf die Proteinfunktion nachgewiesen werden konnte [Ameziane et al., 2015; Wang et al., 2015]. Somit sind aktuell 21 Gene bekannt, die mit FA in Verbindung gebracht werden können (Tabelle 2). Spezifische Mutationen in *FANCQ* sind zudem nicht nur mit der Entstehung von FA, sondern auch mit Xeroderma pigmentosum sowie dem XFE-Progeroid-Syndrom assoziiert. Dies unterstreicht die schwierige phänotypische Abgrenzung und genetische Komplexität, die FA zugrunde liegt.Da nach wie vor einige Patienten keiner bekannten Komplementationsgruppe zugeordnet werden konnten, scheint die Existenz weiterer FA-assoziierter Gene wahrscheinlich.

Eine häufig diskutierte Frage stellt die erhöhte Krebsprädisposition für heterozygote Anlagenträger von FA-Mutationen dar. In einer Studie des Internationalen Fanconi-Anämie-Registers (IFAR) konnte für die meisten heterozygoten Träger kein erhöhtes Krebsrisiko festgestellt werden [Berwick et al., 2007]. Bei bestimmten FA-Genen besteht allerdings auch für heterozygote Träger, wie Eltern oder Geschwister, eine Prädisposition für bestimmte Krebserkrankungen. So sind monoallelische Mutationen in den Genen *FANCD1/BRCA2*, *FANCJ/BRIP1*, *FANCN/PALB2*, *FANCO/RAD51C* sowie *FANCS/BRCA1* mit einem erhöhten Risiko für Brust- und Ovarialkarzinome assoziiert [Ford et al., 1998; Wong et al., 2011], während heterozygote Funktionsverlustmutationen in *FANCM* mit dreifach-negativen Brustkrebs in Verbindung gebracht wurden [Neidhardt et al., 2016]. Eine schwache Assoziation zwischen monoallelischen FA-Mutationen und einer erhöhten Prädisposition für Brustkrebs besteht weiterhin für die Gene *FANCC* und *FANCP/SLX4*. Hier sind allerdings noch weitere Studien nötig, um diese zu bestätigen [Berwick et al., 2007; Landwehr et al., 2011; Bakker et al., 2013].

	Waitore			Ungefährer Anteil
eindeutig [Mehta und	der entsprechenden Kor d Tolar, 1993-2017] sowie	nplementationsgrupp e den Angaben der of	pe zugeordnet werden konnten, aufgeführt. Moo ffiziellen FA-Mutationsdatenbank (http://www.roo	lifiziert und erweitert nach ckefeller.edu/fanconi/).
Lokalisatio	on sowie die Funktion w	ährend der ICL-Rep	paratur. Weiterhin ist der jeweilige ungefähre Ante	eil an FA-Patienten, welche
			0	

Takalla 2. Ühernicht der hielene identifizierten EA Cone. Aufgelistet eind die Beneichnungen der Cone. ihre

UCII	Bezeichnungen	Genlocus	Funktion im FA/BRCA-Signalweg	in FA [%]
FANCA	-	16q24.3	Kernkomplex	60-70
FANCB	FAAP90; FAAP95	Xp22.2	Kernkomplex	~2
FANCC	-	9q22.3	Kernkomplex	~14
FANCD1	BRCA2	13q12.3	Homologe Rekombination	~3
FANCD2	-	3p25.3	ID-Komplex	~3
FANCE	-	6p21.31	Kernkomplex	~3
FANCF	-	11p14.3	Kernkomplex	~2
FANCG	XRCC9	9p13.3	Kernkomplex	~10
FANCI	KIAA1794	15q26.1	ID-Komplex	~1
FANCJ	BRIP1; BACH1	17q23.2	Homologe Rekombination; 5'-3'-Helikase	~2
FANCL	PHF9; FAAP43	2p16.1	Kernkomplex; E3-Ubiquitin-Ligase	~0,3
FANCM	FAAP250	14q21.3	Kernkomplex; DNA-Translokase; wichtig für ATR-Aktivierung	~0,1
FANCN	PALB2	16p12.2	Homologe Rekombination	~0,6
FANCO	RAD51C	17q22	Homologe Rekombination	~0,1
FANCP	SLX4; BTBD12	16p13.3	Wichtiges Gerüstprotein für Nukleasen XPF- ERCC1, MUS81-EME1, SLX1; Lösen des ICLs	~0,5
FANCQ	ERCC4; XPF	16p13.12	Struktur-spezifische Endonuklease im Enzymkomplex mit ERCC1	~0,1
FANCR	R <i>AD51</i>	15q15.1	Homologe Rekombination	~0,1
FANCS	BRCA1	17q21.31	Homologe Rekombination	~0,1
FANCT	UBE2T; HSPC150	1q32.1	E2-Ubiquitin-konjugierendes Enzym	~0,2
FANCU	XRCC2	7q36.1	Homologe Rekombination	~0,1
FANCV	REV7/MAD2L2	1p36.22	Transläsions-Synthese	~0,1

## 1.2.4 Der FA/BRCA-Signalweg

Die Proteinprodukte aller bislang identifizierten FA-Gene sind Teil des sogenannten FA/BRCA-Doppelstrangbruch-Reparaturweges und für die Aufrechterhaltung der genomischen Stabilität entscheidend. DNA-quervernetzende Substanzen wie MMC oder Cisplatin verknüpfen beide Stränge kovalent und irreversibel. Dadurch wird die Replikationsgabel am Voranschreiten entlang des DNA-Stranges gehindert und stagniert an der Schadensstelle. Essenzielle Prozesse, wie Replikation und Transkription, können nicht erfolgen, wodurch es zur Apoptose der Zelle kommt, insofern der Schaden nicht behoben wird. Als Folge von blockierten Replikationsgabeln wird der Replikations-Kontrollpunkt des Zellzyklus durch die ATR-Kinase aktiviert, welche spezifisch auf Replikationsblockaden wie auch auf Replikationsstress reagiert. Das wiederum hat die Aktivierung des FA/BRCA-Signalweges zur Folge, wodurch der Schaden repariert werden kann [Andreassen et al., 2004; Wang et al., 2007; Collis et al., 2008; Tomida et al., 2013]. Zur Reparatur von DNA-Interstrang-Quervernetzungen ist ein komplexes Netzwerk an DNA-Reparaturfaktoren nötig, dessen exakter Ablauf sowie alle daran beteiligten Faktoren bis heute nicht vollständig geklärt sind. Obwohl für einige FA-Proteine Orthologe in anderen Organismen beschrieben wurden, ist der ICL-DNA-Reparaturweg nur in Vertebraten hoch konserviert, wie aus Abbildung 7 ersichtlich wird.



Abbildung 7: Konservierung der FA-Gene bei Eukaryoten. Die Proteine des Kernkomplexes (links), des ID-Komplex (Mitte) sowie *downstream* davon agierenden Effektoren (rechts) sind jeweils in Gruppen zusammengefasst. "?": Die Existenz bzw. Identität eines Orthologs ist unklar. "O\*": *rfs-1* ist das einzige RAD51-Paralog in *C. elegans* [Ward et al., 2007]. "J\*\*": CHL1 besitzt Ähnlichkeit zu humanem FANCJ und DDX11 [Amann et al., 1997; Skibbens, 2004].

Im Zentrum dieses DNA-Reparaturweges steht die Monoubiquitinierung des so genannten ID-Komplexes, eines Heterodimers bestehend aus FANCI und FANCD2. Diese Modifikation wird durch den Kernkomplex vermittelt, der aus acht FA-Proteinen (FANC-A, -B, -C, -E, -F, G, -L, -M) sowie mehreren FA-assoziierten Proteinen (FAAPs) zusammengesetzt ist, wie FAAP20, FAAP24 und FAAP100. Die FA-Proteine des Kernkomplexes werden häufig auch als *Upstream*-Gene bezeichnet, da ihre Funktion vor der Monoubiquitinierung des ID-Komplexes liegt, was sich an der fehlenden FANCD2-Monoubiquitinierung bei der Analyse von *Upstream*-Patienten per FANCD2-Immunblot zeigt. Die fehlende Modifikation von FANCD2 kann zudem auch bei FANCI-Patienten beobachtet werden[Levitus et al., 2004; Smogorzewska et al., 2007]. Dagegen werden FA-Effektor-Proteine, die nach der FANCD2-Monoubiquitinierung aktiviert werden, auch als *Donnstream*-Gene bezeichnet werden. Eine Übersicht über die verschiedenen FA-Proteine ist in Abbildung 8 dargestellt.



Abbildung 8: Schematische Darstellung des FA/BRCA-Signalweges. FA-Proteine sind farbig, assoziierte Faktoren in grau abgebildet. Eine DNA-Quervernetzung (rotes Symbol) führt zu einer Aktivierung des FA/BRCA-Signalweges. 1: Ein Komplex um FANCM (gelb) erkennt stagnierte Replikationsgabeln. Es folgt die Rekrutierung der Kernkomplex-Proteine (Orange- und Rottöne) an die Schadensstelle. 2: In seiner aktiven Form überträgt der Kernkomplex durch die E3-Ligase-Aktivität von FANCL Ubiquitin (U) vom Ubiquitin-konjugierenden Enzym UBE2T (gelb) auf den ID-Komplex (grün), welcher bereits an der Schadensstelle an die DNA gebunden ist. Der ubiquitinierte ID-Komplex rekrutiert Effektorproteine zur Schadensaufhebung (3. –5.). 3: Zunächst wird die Quervernetzung durch Endonukleasen aus dem Doppelstrang gelöst (Pfeilspitzen). 4: Die entstandene Lücke des einen Stranges wird durch Transläsions-Polymerasen wie REV1 und Pol  $\zeta$  (FANCV+REV3) aufgefüllt. 5: Die Reparatur des verbleibenden Bruches wird schließlich durch Homologe Rekombination vermittelt, an welcher verschiedene FA-Proteine (violett) beteiligt sind. Verändert nach [Schindler et al., 2011].

Für die initiale Erkennung einer DNA-Quervernetzung während der S-Phase ist FANCM im Signalweg essenziell. Es besitzt eine N-terminale Helikase-Domäne, die eine ATP-abhängige DNA-Translokase-Aktivität enthält [Meetei et al., 2005; Mosedale et al., 2005]. Anders als alle anderen Kernkomplex-Proteine ist FANCM, zusammen mit seinem Bindungspartner FAAP24, konstitutiv ans Chromatin gebunden [Kim et al., 2008]. Gemeinsam mit dem MHF1-MHF2-Heterodimer bilden FANCM-FAAP24 einen stabilen Komplex, welcher eine hohe Affinität für verzweigte DNA-Strukturen, wie *Holliday Junctions* oder Replikationsgabeln aufweist und diese stabilisiert [Ciccia et al., 2007; Singh et al., 2010; Yan et al., 2010]. Zudem ist der Komplex zur Aktivierung des ATR-vermittelten S-Phase-Kontrollpunktes nötig [Collis et al., 2008; Huang et al., 2010], wodurch es folglich zur Phosphorylierung von FANCA, FANCE, FANCG, FANCM sowie FANCD2 und FANCI kommt [Andreassen et al., 2004; Qiao et al., 2004; Wang et al., 2007; Collis et al., 2008; Ishiai et al., 2008; Collins et al., 2009; Singh et al., 2013]. Über diesen positiven *Feedback-Loop* zwischen FANCM und ATR wie auch durch die Bindung des FANCM-FAAP24-Komplexes an die Schadensstelle wird die Anlagerung der

übrigen Kernkomplex-Proteine vermittelt. Der genaue Prozess der Rekrutierung des Kernkomplexes ist jedoch nicht in allen Details geklärt. Allerdings ist die Bindung des Kernkomplexes sowohl von der Zellzyklus-Phase als auch von FANCM abhängig [Mi und Kupfer, 2005; Kim et al., 2008].

Die Funktion des Kernkomplexes besteht in der Übertragung von Ubiquitin auf den ID-Komplex. Dieser Prozess folgt einer Enzymkaskade, bestehend aus E1-, E2- und E3-Enzymen. Im Einzelnen wird dabei zunächst in einem ATP-abhängigen Schritt freies Ubiquitin über eine Thioesterbrücke an das Ubiquitinaktivierende Enzym (E1) gebunden. Das aktivierte Ubiquitin wird im Anschluss, ebenfalls unter Ausbildung einer Thioesterbrücke, auf das Ubiquitin-konjugierende Enzym (E2) übertragen. Im FA-Signalweg konnte hierfür UBE2T, welches mit FANCT identisch ist, als entscheidendes E2-Enzym identifiziert werden [Machida et al., 2006; Alpi et al., 2007; Alpi et al., 2008; Rajendra et al., 2014]. Letztlich wird über eine E3-Protein-Ligase das Substrat erkannt und Ubiquitin auf einen Lysinrest des spezifischen Zielproteins transferiert. Die katalytische E3-Ligase-Aktivität wird hierbei von FANCL vermittelt.

Durch verschiedene Co-Immunpräzipitations (Co-IP)-Studien und *Pulldown*-Assays sowie unter Zuhilfenahme von Tiermodellen konnte eine Aufteilung des Kernkomplexes in drei Subkomplexe beobachtet werden [Thomashevski et al., 2004; Medhurst et al., 2006; Huang et al., 2014; Swuec et al., 2017; van Twest et al., 2017]. Es besteht eine direkte Interaktion zwischen FANCB und FANCL sowie FAAP100, die zusammen einen Subkomplex innerhalb des Kernkomplexes bilden [Medhurst et al., 2006; Ling et al., 2007; Swuec et al., 2017; van Twest et al., 2017]. Dieser wird als BL100-Komplex bezeichnet und scheint die einzelnen Komponenten vor proteolytischer Degradierung zu schützen. *In vitro* ist die Bildung des BL100-Komplexes nicht für die E3-Ligase-Aktivität von FANCL essenziell, wie Studien mit rekombinanten Proteinen belegen [Alpi et al., 2008; Sato et al., 2012]. Allerdings wird die Aktivität der E3-Ligase-Funktion von FANCL um das fünf- bis sechsfache gesteigert, wenn eine Interaktion mit FANCB und FAAP100 stattfinden kann [Rajendra et al., 2014]. Strukturelle Analysen zeigten, dass der BL100-Komplex als Homodimer (BL100)<sub>2</sub> vorliegt, mit je zwei FAAP100-Proteinen im Zentrum [Swuec et al., 2017; van Twest et al., 2017]. Diese bilden zusammen mit zwei FAAP100-Proteinen eine Art Gerüst, um zwei FAANCL-Proteine zur Monoubiquitinierung des ID-Komplexes entsprechend auszurichten (Abbildung 8).

Zur Bildung bzw. Lokalisation des BL100-Komplexes im Zellkern ist neben FANCM auch FANCA wichtig, das im Komplex mit FANCG und FAAP20 vorliegt (AG20-Komplex) [Garcia-Higuera et al., 1999; Garcia-Higuera et al., 2000; Ali et al., 2012; Huang et al., 2014]. Die Bindung zwischen FANCA und FANCG, die sich gegenseitig stabilisieren, ist vermutlich konstitutiv und ein frühes Ereignis im FA-Signalweg. Ein stimulatorischer Effekt des AG20-Komplexes auf den BL100-Komplex scheint hingegen nicht zu bestehen [van Twest et al., 2017]. Stattdessen wird über FANCG die Verbindung zum dritten Subkomplex, bestehend aus FANCC, FANCE und FANCF (CEF-Komplex), vermittelt [Garcia-Higuera et al., 1999; de Winter et al., 2000c; Gordon und Buchwald, 2003]. Bei FANCC handelt es sich um ein vorwiegend cytoplasmatisches Protein, das, abhängig von FANCE, in den Zellkern transloziert [Yamashita et al., 1994; Youssoufian, 1994;

Pace et al., 2002; Taniguchi und D'Andrea, 2002]. Dort interagieren beide Proteine mit FANCF, welches wiederum über seine N-terminale Domäne an FANCM binden kann und so die Verbindung des übrigen Kernkomplexes mit FANCM herstellt [Deans und West, 2009]. Im Gegensatz zum AG20-Komplex stimuliert der CEF-Komplex die katalytische Aktivität des BL100-Komplexes deutlich. Bei einem gemeinsamen Vorliegen der BL100- und CEF-Komplexe ist die Aktivität der Ligase-Funktion um das 250-fache erhöht, im Vergleich zur isolierten Aktivität von FANCL [van Twest et al., 2017]. Zudem stellt der CEF-Komplex eine Verbindung zwischen BL100 und dem ID-Komplex her. Dabei können sowohl FANCC als auch FANCE an den ID-Komplex binden, wohingegen FANCF eine stabilisierende Rolle dabei innezuhaben scheint [van Twest et al., 2017].

Die Proteine des ID-Komplexes stellen Paraloge dar und ähneln sich in Größe und Struktur. Im Gegensatz zu FANCD2 scheint die Ubiquitinierung von FANCI für die DNA-Reparatur nicht essenziell zu sein. Vielmehr wird angenommen, dass Phosphorylierungen regulierend auf die Funktion und Lokalisation von FANCI wirken und auch dessen unmodifizierte Form wichtige Aufgaben während der Rekrutierung des Kernkomplexes innehat [Ishiai et al., 2008]. Lange war unklar, ob der ID-Komplex erst nach der Ubiquitinierung an die Schadensstelle rekrutiert wird oder schon vorher dorthin transloziert. In einer kürzlich erschienenen Studie konnte durch Kryo-Elektronenmikroskopie eine Struktur des humanen ID-Komplexes erstellt werden, anhand dessen eine neue, funktionell entscheidende Domäne des FANCD2-Proteins identifiziert werden konnte [Liang et al., 2016]. Diese als Tower-Domäne bezeichnete C-terminale Struktur wird für die Bindung des ID-Komplexes an ICLs benötigt, was sowohl in vitro als auch in vitro nachgewiesen werden konnte. Weiterhin konnte gezeigt werden, dass der Komplex aus FANCD2 und FANCI schon vor der Ubiquitinierung durch blockierte Replikationsgabeln an die DNA bindet, was schließlich den Auslöser für die Monoubiquitinierung des ID-Komplexes durch den Kernkomplex darstellt. Der modifizierte ID-Komplex rekrutiert anschließend Downstream-Effektor-Proteine an die Schadensstelle, mit welchen er in DNA-Reparatur-Foci kolokalisiert. Bei diesen Proteinen handelt es sich um Bestandteile der Transläsionssynthese (TLS) und der Homologen Rekombination, welche zur Schadensauflösung beitragen.

Um die Quervernetzung der DNA-Stränge zu lösen und den entstandenen Schaden zu reparieren, ist ein Zusammenspiel von TLS, HR sowie NER nötig (Abbildung 8). Zunächst wird SLX4/FANCP durch ubiquitiniertes FANCD2 an die Schadensstelle rekrutiert. Dort rekrutiert sowie aktiviert das Gerüstprotein verschiedene Struktur-spezifische Endonukleasen, wie die Heterodimere ERCC4/FANCQ-ERCC1, MUS81-EME1 und SLX1. Nachdem jeweils auf beiden Seiten des ICLs geschnitten wurde, kann dieser von einem der kovalent verbundenen Stränge gelöst werden.

Im Gegensatz zu den an der Replikation beteiligten Polymerasen können Transläsions-Polymerasen, wie REV1 und Pol ζ, deren katalytische Untereinheit aus REV3 und FANCV/REV7 besteht [Gan et al., 2008; Kim et al., 2012], auch über Schäden hinweg replizieren. So wird gewährleistet, dass der intakte Strang, an dem

der ICL noch gebunden ist, dennoch repliziert werden kann. Der neu synthetisierte Strang dient anschließend als Vorlage für die Homologe Rekombination, um den Doppelstrangbruch am komplementären Strang zu beheben. Eine wichtige Rolle spielt dabei die Rekombinase RAD51 (FANCR), indem sie zunächst RPA von den einzelsträngigen DNA (ssDNA)-Enden des Doppelstrangbruches verdrängt und anschließend den Strangaustausch der HR vermittelt. Hierfür sind die RAD51-Paralogen FANCO und FANCU sowie weitere RAD51-assoziierte Faktoren nötig. Dazu zählen FANCD1 sowie dessen Bindungspartner FANCN, welcher für die Stabilisierung und Lokalisation von FANCD1 notwendig ist und die Bindung von FANCD1 mit FANCS vermittelt [Zhang et al., 2009]. Aufgrund dieser Proteininteraktionen können in Zelllinien dieser Subgruppen nach ICL-Schadensinduktion keine oder nur sehr stark vermindert RAD51-Foci detektiert werden [Godthelp et al., 2002; Reid et al., 2007; Vaz et al., 2010; Sawyer et al., 2015; Park et al., 2016], im Gegensatz zu FANCJ-defizienten Zellen [Litman et al., 2005]. Damit scheint die Helikase FANCJ, welche über BRCT-Domänen mit FANCS interagiert [Cantor et al., 2001], eine Funktion nach der RAD51-Filamentbildung aufzuweisen. Die exakte Funktion dieser zweiten Helikase im FA-Reparaturweg ist nach wie vor nicht geklärt. Allerdings konnte anhand von FANCI-knockouts in DT40-Zellen gezeigt werden, dass die Interaktion mit FANCS nicht für die Reparatur des DNA-Schadens durch Quervernetzungen nötig ist, wohl aber die funktionelle Helikase-Domäne [Bridge et al., 2005]. Die Entfernung des noch gebundenen ICLs erfolgt im Anschluss über den NER-Signalweg [Mouw und D'Andrea, 2014]. Nachdem die DNA-Reparatur abgeschlossen ist, vermittelt USP1-UAF1 die Deubiquitinierung des ID-Komplexes [Nijman et al., 2005; Smogorzewska et al., 2007] und die beteiligten Faktoren dissoziieren vom Chromatin.

#### 1.2.5 Therapiemöglichkeiten

Die derzeitige Standardbehandlung von FA-Patienten besteht vor allem aus einer symptomatischen Therapie. Durch chirurgische Maßnahmen können angeborene Fehlbildungen, wie beispielsweise der Hand oder von inneren Organen, korrigiert bzw. operativ behandelt werden. Zudem spielen sie eine wichtige Rolle bei der Behandlung von Tumorerkrankungen, v. a. bei Plattenepithelkarzinomen des Kopf- und Halsbereiches (*head and neck squamous cell caricnomas*; HNSCCs), da Chemo- und Radiotherapie bei FA-Patienten mit erheblichen Nebenwirkungen verbunden sind und nur sehr schlecht toleriert werden.

Androgenderivate, wie Oxymetholon oder Danazol und/oder Zytokine, wie G-CSF (granulocyte-colony stimulating factor), werden häufig bei beginnendem Knochenmarksversagen eingesetzt [Rackoff et al., 1996; Scheckenbach et al., 2012]. Der genaue Wirkmechanismus von Androgenen ist noch immer nicht vollständig geklärt, doch die Therapie bewirkt in der Regel einen Anstieg bzw. eine Stabilisierung des Hämoglobin-Levels sowie der Thrombozytenzahl. Allerdings sprechen viele Patienten nach einer gewissen Behandlungszeit nicht mehr und manche Patienten überhaupt nicht auf diese Therapiemaßnahme an. Zu den Nebenwirkungen zählt ein erhöhtes Risiko für die Entstehung von Lebertumoren, weshalb eine regelmäßige Kontrolle beachtet werden sollte. Weibliche Patienten reagieren zudem häufig mit einer Virilisierung auf die Androgentherapie.

Die einzige dauerhafte Behandlungsmöglichkeit zur Wiederherstellung des blutbildenden Systems besteht zurzeit in der allogenen hämatopoetischen Stammzelltransplantation (HSZT). Damit können jedoch nicht die Schäden in den übrigen Körperzellen behoben werden. Das erhöhte Tumorrisiko, z. B. für HNSCCs, bleibt somit weiterhin bestehen. Als Spender kommen neben Geschwistern auch unverwandte Spender zunehmend in Frage. Entscheidend für die Fortschritte der HSZT bei FA-Patienten war u. a. eine entsprechende Anpassung der Protokolle bei der Konditionierung. Dabei wird das Knochenmark des Empfängers zerstört sowie dessen Immunsystem unterdrückt, wodurch allerdings auch die Gefahr von Infektionen steigt. Neben der Abstoßung des Transplantats durch den Empfänger ist zudem eine *Graft-versus-Host-Disease* (GvHD) möglich. Dabei handelt es sich um eine immunologische Reaktion, ausgelöst durch die T-Lymphozyten des Spenders ("*Graft*<sup>e</sup>), die das Gewebe des Empfängers ("*Host*<sup>e</sup>) als fremd erkennen und angreifen. Durch eine T-Zell-Depletion der Spenderzellen vor der Transplantation wird versucht, die GvHD möglichst zu vermeiden [Zecca et al., 2014].

Diese Probleme ließen sich durch eine Lentivirus-vermittelte Gentherapie umgehen, da hierbei die hämatopoetischen Stammzellen (*hematopoetic stem cells;* HSCs) des Patienten genutzt werden. Momentan laufen verschiedene Gentherapie-Studien [Gonzalez-Murillo et al., 2010; Rio et al., 2014], beispielsweise für die Korrektur von *FANCA*, die sich aber noch in der Versuchsphase befinden. Das Ziel besteht darin, über einen retroviralen Gentransfer HSCs des Patienten *ex vivo* zu korrigieren und anschließend zu reimplantieren. Aufgrund eines Wachstumsvorteils der korrigierten Zellen gegenüber der FA-Zellen, wie er auch bei FA-Mosaikpatienten nach der Reversion einer Mutation beobachtet werden kann, sollten sich diese in der Zellpopulation allmählich durchsetzen. Des Weiteren bestehen Ansätze zur Genom-Editierung bei FA durch das CRISPR/Cas9-System. Anders als beim Einsatz von Viren lässt sich hierbei der Einbau zielgerichtet leiten. Bei ersten Versuchen mit Patientenzellen konnte *in vitro* die Korrektur des betroffenen Gens in induzierten pluripotenten Stammzellen gezeigt werden [Osborn et al., 2015; Osborn et al., 2016].

#### 1.3 Identifizierung von FA-Genen und Mutationsanalyse

## 1.3.1 Methoden zur Komplementationsgruppen-Bestimmung und Identifizierung von FA-Genen

#### a. Komplementationsgruppen-Bestimmung

Erstmals konnte im Jahr 1980 durch somatische Zellhybridisierung gezeigt werden, dass es sich bei FA um eine genetisch heterogene Erkrankung handelt [Zakrzewski und Sperling, 1980]. Das Prinzip der Komplementationsgruppen-Bestimmung untersucht die Aufhebung des Defektes von FA-Zellen nach der Fusion mit nicht-verwandten Patientenzellen durch Testung FA-typischer Eigenschaften. Ist die erzeugte Hybridzelllinie weiterhin sensitiv gegenüber DNA-quervernetzenden Substanzen, gehören beide Patienten derselben Komplementationsgruppe an. Handelt es sich hingegen um unterschiedliche Komplementationsgruppen, so kann durch Zellfusion der zugrunde liegende Defekt korrigiert und die Sensitivität der Hybridzelllinie aufgehoben werden. Durch diese Strategie konnten bis zum Jahr 2004 elf verschiedene Komplementationsgruppen identifiziert werden, welchen bis dahin nur neun Gene zugeordnet werden konnten [Levitus et al., 2004].

Eine moderne Variante der oben beschriebenen Methode stellt die retrovirale Komplementation dar. Dabei werden Patientenzellen unbekannter Komplementationsgruppe mit retroviralen Vektoren transduziert. Diese Vektoren enthalten jeweils die Wildtypsequenz eines einzelnen FA-Gens, die in die Zellen eingebracht und dort exprimiert wird [Hanenberg et al., 2002]. Die Zellen werden anschließend ebenfalls auf die typische Hypersensitivität gegenüber DNA-quervernetzende Substanzen untersucht [Chandra et al., 2005]. Ist diese aufgehoben, so gehört der Patient der durch das lentivirale Konstrukt bestimmten Komplementationsgruppe an. Anschließend kann die Mutationsanalyse, beispielsweise durch Sanger-Sequenzierung des entsprechenden Gens, beginnen.

#### b. Strategien zur Identifizierung von FA-Genen

Die Kartierung und Klonierung von FA-Genen begann 1992 mit der Identifizierung von FANCC [Strathdee et al., 1992a]. Dazu wurde eine FA-C-Linie mit einer cDNA-Expressionsbibliothek transfiziert, um mögliche Kandidaten durch das Testen der MMC- bzw. DEB-Hypersensitivität zu testen. Diese Strategie führte in den folgenden acht Jahren zur Identifizierung von FANCA, FANCG, FANCF und FANCE [Lo Ten Foe et al., 1996; de Winter et al., 1998; de Winter et al., 2000b; de Winter et al., 2000a]. Mit einer alternativen Strategie konnte FANCA, zeitgleich mit Erscheinen der Arbeit über die erfolgreiche Expressionsklonierung, auch durch Positionsklonierung identifiziert werden [Apostolou et al., 1996]. Die Entdeckung, dass FANCG mit dem Reparaturgen XRCC9 identisch ist, das zuvor in chinesischen Hamsterzellen untersucht und beschrieben worden war, war ein erster konkreter Hinweise für die Beteiligung von FA-Genen an der DNA-Reparatur [de Winter et al., 1998]. Kurze Zeit später konnte gezeigt werden, dass FANCA mit FANCC und FANCG einen Protein-Komplex im Zellkern bildet [Garcia-Higuera et al., 1999], ohne jedoch dessen genaue Funktion zu kennen.

Positionsklonierungen spielten ferner bei der Identifizierung des hochkonservierten Gens FANCD2 eine wichtige Rolle. Zunächst waren zwei Patientenzelllinien zur Komplementationsgruppe FA-D zusammengefasst worden, charakterisiert durch die Referenzzelllinie HSC62 sowie die Zelllinie PD20 [Strathdee et al., 1992b]. Durch die Fusion von PD20-Zellen mit Mikrozellen, welche jeweils einzelne Chromosomen enthielten, konnte zunächst die Aufhebung der MMC-Hypersensitivität durch Chromosom 3 gezeigt werden [Whitney et al., 1995]. Die chromosomale Position von FANCD konnte mit Hilfe von Positionsklonierungen in PD20-Zellen auf einen Bereich von 200 kb auf 3p25.3 eingegrenzt werden [Hejna et al., 2000]. Die Mutationsanalyse von Kandidatengenen innerhalb dieser Region durch die Arbeitsgruppe um Timmers et al. führte 2001 schließlich zum Erfolg [Timmers et al., 2001]. In der gleichen Arbeit konnte gezeigt werden, dass beide bis dato bekannten FA-D-Zelllinien unterschiedlichen Komplementationsgruppe FA-D1, PD20

hingegen der Komplementationsgruppe FA-D2 zugeordnet. Zeitgleich erschien eine Studie von Garcia-Higuera et al. [Garcia-Higuera et al., 2001], in der gezeigt wurde, dass ein intakter Kernkomplex zur Modifizierung von FANCD2 nötig ist, FANCD1 hingegen nicht. Weiterhin wurde darin die Interaktion von FANCD2 mit BRCA1 beschrieben. Aus den Ergebnissen konnte man schließen, dass FA-Proteine in einem linearen Signalweg interagieren. Zudem konnte erstmals eine direkte Verbindung zwischen dem Kernkomplex und der DNA-Reparaturmaschinerie hergestellt werden. Aufgrund dieser Verknüpfung wurden fortan verschiedene Kandidatengene getestet, was kurz darauf zur Identifizierung von *FANCD1/BRCA2* führte [Howlett et al., 2002].

Durch eine Studie von Meetei et al. im Jahr 2003, die eigentlich die Mechanismen zur Entstehung des Bloom-Syndroms untersuchen wollten, konnte ein großer Proteinkomplex aus HeLa-Zellextrakten isoliert werden [Meetei et al., 2003a]. Dieser als BRAFT bezeichnete Komplex bestand aus den bis dahin fünf bekannten FA-Kernkomplex-Proteinen (FANCA, -C, -E, -F, -G), der BLM-Helikase, Topoisomerase IIIa, RPA und weiteren BLM- und FA-assoziierten Proteinen (FAAPs). Nach Zugabe von Salz dissoziierte der Bloom-Syndrom- vom FA-Kernkomplex, der so erstmals mit allen assoziierten Faktoren aufgereinigt und analysiert werden konnte. Die verschiedenen FAAPs (FAAP250, FAAP100, FAAP43) wurden später durch Massenspektrometrie und Co-IP näher charakterisiert und die entsprechenden Gene folglich in noch nicht auf Mutationen gescreent. Dies führte zur zugeordneten Patienten Identifizierung von FANCL/FAAP43/PHF9, FANCB/FAAP95/FAAP90 sowie FANCM/FAAP250/KLAA1596 [Meetei et al., 2003b; Meetei et al., 2004; Meetei et al., 2005]. Nachdem 2006 ebenfalls über eine Co-IP-Analyse PALB2 als BRCA2-Bindungspartner bestätigt wurde, konnte auch dieses Gen nur Jahr später von zwei Arbeitsgruppen einer weiteren Komplementationsgruppe (FA-N) zugeordnet werden [Reid et al., 2007; Xia et al., 2007]. Auch SLX4 wurde aufgrund seiner beschriebenen Funktion über einen Kandidatengen-Screen als FA-Gen identifiziert [Kim et al., 2011; Stoepker et al., 2011].

In drei unabhängigen Untersuchungen konnte *FANCI* im Jahr 2007 als weiteres FA-Gen bestimmt werden. Dabei führten sowohl Kopplungsanalysen [Dorsman et al., 2007], ein Substrat-Screen für die Kinasen ATM und ATR [Smogorzewska et al., 2007] sowie eine bioinformatische Suche nach Sequenzhomologien zu *FANCD2* [Sims et al., 2007] zum Erfolg. Über die Strategie der Homozygotiekartierung konnten ferner *FANCJ*, das zeitgleich auch über eine klassische Positionsklonierung zugeordnet werden konnte, sowie *FANCO* identifiziert werden [Levitus et al., 2005; Levran et al., 2005b; Vaz et al., 2010].

Bei den zuletzt beschriebenen FA-Genen FANCQ, -R, -S, -T, -U und -V wurden erstmals Hochdurchsatz-Sequenzierungen (*next-generation sequencing*; NGS) zur Identifizierung neuer FA-Gene genutzt. So wurden die Komplementationsgruppen FA-Q, -S, -U und -V durch Gesamt-Exom-Sequenzierung (*whole exome sequencing*; WES) identifiziert [Bogliolo et al., 2013; Sawyer et al., 2015; Bluteau et al., 2016; Park et al., 2016]. Zudem konnte bei FA-R ein Patient über WES, ein anderer über eine Gesamt-Genom-Sequenzierung (*whole genome sequencing*, WGS) [Ameziane et al., 2015; Wang et al., 2015] zugeordnet werden. Des Weiteren führte WES zur Identifizierung eines der beiden beschriebenen FA-T-Patienten [Hira et al., 2015], wohingegen die biallelischen Mutationen des zweiten FA-T-Patienten durch zwei unterschiedliche Arbeitsgruppen mit verschiedenen Methoden bestimmt werden konnten. So führte 2015 bei Rickman et al. [Rickman et al., 2015] eine Sequenzierung des gesamten Transkriptoms (RNA-Seq), bei Virts et al. [Virts et al., 2015] eine retrovirale Komplementation zum Ziel. Die Entdeckung von *UBE2T* als neues FA-Gen unterstreicht nochmals, wie vielfältig die Methoden zur Identifizierung neuer FA-Gene sein können.

#### 1.3.2 Mutationsanalyse in der FA-Routinediagnostik

#### a. Bestätigung der Verdachtsdiagnose

Um die Verdachtsdiagnose FA bestätigen oder ausschließen zu können, nutzt man in der klinischen Diagnostik die Hypersensitivität von FA-Zellen gegenüber DNA-quervernetzenden Substanzen. Man kultiviert dazu zur Zellteilung stimulierte periphere Blutlymphozyten oder immortalisierte Zelllinien in Medium, welchem verschiedene Konzentrationen an DEB oder MMC zugegeben wurden. Anschließend werden die Zellen entweder am Mikroskop auf eine erhöhte Chromosomenbrüchigkeitsrate untersucht [Oostra et al., 2012] oder hinsichtlich des Zellzyklusprofils mittels Durchflusszytometrie analysiert [Schindler et al., 1987; Seyschab et al., 1995]. FA-Patienten zeigen schon bei geringen Konzentrationen an DEB oder MMC eine erhöhte Brüchigkeit, die mit zunehmender ICL-Reagenz-Konzentration weiter ansteigt. Durch beide Methoden lassen sich auch Mosaik-Patienten von gesunden Kontrollen und FA-Patienten unterscheiden, sofern sich die Reversion im untersuchten Material nicht vollständig durchgesetzt hat. Bei der Chromosomenbruchanalyse grenzen sich dabei mit steigender Konzentration an DEB oder MMC immer deutlicher zwei Zellpopulationen ab. Dabei zeigt eine Population kaum Bruchereignisse, ähnlich wie es bei Kontrollen der Fall ist. Die zweite Population hingegen weist eine stark erhöhte Chromosomenbruchrate auf, wie sie bei FA-Patienten beobachtet wird. Fällt die Untersuchung einer Blutkultur in der Chromosomenbruchanalyse oder Durchflusszytometrie negativ aus, sollte bei bestehendem Verdacht auf FA eine Fibroblastenkultur untersucht werden. Zeigt diese nach Behandlung mit DEB oder MMC eine erhöhte Rate an Zellen, die in der G2-Phase arretiert sind (typischerweise 40 % oder mehr), kann die Diagnose FA dennoch bestätigt und auf einen Mosaikstatus des Patienten geschlossen werden.

#### b. Molekulargenetische Mutationsanalyse

Sofern die Diagnose FA beim Patienten bestätigt wurde, schließt sich üblicherweise eine Mutationsanalyse an. Obwohl im Allgemeinen keine Genotyp-Phänotyp-Korrelation bei FA vorliegt [Neveling et al., 2009], hat die Identifizierung der Mutationen, die der Erkrankung zugrunde liegen, sowohl für den Patienten, als auch für dessen Angehörige einigen Nutzen. Für die genetische Beratung kann bei einem Kinderwunsch das Wissen über die ursächlichen Mutationen für die Risikoabschätzung genutzt werden, sowohl für den Patienten, dessen Geschwister oder andere Verwandte. Dies ist insbesondere bei konsanguinen Verbindungen von Bedeutung. Bei einer erneuten Schwangerschaft haben Eltern eines FA-Patienten zudem die Möglichkeit, eine mutationsspezifische pränatale Untersuchung durchführen zu lassen. Weiterhin kann man das Vorliegen der Erkrankung in bislang nicht diagnostizierten Geschwisterkindern untersuchen, bei welchen sich die Symptome noch nicht manifestiert haben oder die generell phänotypisch unauffällig sind. Da bestimmte FA-Gene mit einer Prädisposition für Brust- und Ovarialkarzinome assoziiert sind, ist das Wissen über einen möglichen Heterozygotenstatus für alle Angehörigen von Patienten dieser Komplementationsgruppen von Bedeutung. Ferner kann bei Verlaufskontrollen ein möglicher Mosaikstatus auf DNA-Ebene abgeklärt werden.

Da Mutationen in FANCA die häufigste Ursache bei FA sind, wird in der Regel mit der Mutationsanalyse durch Sanger-Sequenzierung dieses Gens begonnen. Dabei ist das Mutationsspektrum sehr heterogen und die Mutationen finden sich über alle 43 Exons von FANCA verteilt [Levran et al., 2005a; Castella et al., 2011]. Weiterhin handelt es sich bei den meisten pathogenen Varianten um private Mutationen, d. h. diese sind sehr selten und treten meist nur innerhalb einer Familie oder einer bestimmten Population auf. Bei rund einem Drittel aller FA-A-Patienten finden sich große Deletionen, die ein bis mehrere Exons betreffen können oder sogar das ganze FANCA-Gen einschließen. Eine mögliche Ursache dafür liegt in den zahlreichen Alu-Repeat-Elementen, die sich in der FANCA-Sequenz finden. Es konnte gezeigt werden, dass rund 75 % aller FANCA-Deletionen über Alu-vermittelte Rekombination entstehen [Flynn et al., 2014]. Eine sehr sensitive und effiziente Methode zur Detektion solcher Deletionen stellt die Multiplex Ligation-dependent Probe Amplification (MLPA) dar. Diese wird in der Regel der konventionellen Sanger-Sequenzierung oder NGS vorgeschalten [Ameziane et al., 2008; Ameziane et al., 2012; Gille et al., 2012]. Eine weitere Möglichkeit zur Detektion von großen Deletionen ist die Durchführung einer Microarray-basierten komparativen genomischen Hybridisierung (Array-CGH). Der Vorteil dieser Methode gegenüber einer MLPA besteht darin, dass gleichzeitig mehrere Gene auf strukturelle Veränderungen hin untersucht werden können [Chandrasekharappa et al., 2013]. Dazu sind allerdings spezielle, meist selbst erstellte Arrays nötig, um die entsprechenden Gene analysieren zu können.

Sofern der Patient nicht zur Komplementationsgruppe FA-A gehört, sollten sukzessive weitere FA-Gene sequenziert werden. Nach *FANCA* sind Mutationen in *FANCC* (~14%) und *FANCG* (~10%) die häufigste Ursache für FA. Durch die Abklärung dieser Gene sowie von *FANCE*, *FANCF* und bei männlichen FA-Patienten *FANCB*, können rund 85% aller untersuchten FA-Patienten einer Komplementationsgruppe zugeordnet werden [Ameziane et al., 2008; Gille et al., 2012].

Sollte dies weiterhin zu keinem Ergebnis führen, empfiehlt sich eine funktionelle Testung auf Proteinebene, sofern eine Zelllinie des Patienten verfügbar ist. Dafür macht man sich die Monoubiquitinierung von FANCD2 zunutze, durch die sich das Molekulargewicht des Proteins von 155 auf 162 kDa erhöht. Diese Modifizierung lässt sich mithilfe eines Immunblots durch das Vorhandensein einer Doppelbande nachweisen. Ist diese detektierbar, so muss ein Protein betroffen sein, das *downstream* des ID-Komplexes lokalisiert ist. Fehlt die zweite Bande, konnte FANCD2 nicht monoubiquitiniert werden und die Mutationen befinden sich
demnach in einem der Upstream-Gene oder FANCI. Ist keine der Banden detektierbar, so ist FANCD2 selbst mutiert und sollte direkt durch Sanger-Sequenzierung überprüft werden.

Eine Kenntnis über den ethnischen Hintergrund kann zur Detektion bestimmter *Founder*-Mutationen ebenso in die Mutationsanalyse mit einfließen, wie auch vereinzelt bestimmte phänotypische Auffälligkeiten. So ist bei Patienten der Komplementationsgruppen FA-D1 und FA-N bereits in jungen Jahren die Entstehung von Leukämie oder soliden Tumoren, wie Medulloblastome oder eines Wilms-Tumors, zu beobachten [Alter et al., 2007; Reid et al., 2007].

#### 1.3.3 Möglichkeiten durch Hochdurchsatz-Sequenzierungsmethoden

Durch die Entwicklung von Methoden zur Hochdurchsatz-Sequenzierung ist es möglich, parallel ein bestimmtes Genset, das gesamte Exom oder sogar Genom in einem einzigen Sequenzierlauf zu analysieren. Prinzipiell lassen sich alle Verfahren in drei Schritte unterteilen. Zunächst muss eine DNA-Bibliothek erstellt werden, bevor diese durch klonale Amplifikation vermehrt und schließlich parallel sequenziert werden kann. Die erste Plattform wurde im Jahr 2005 von 454 Life Sciences eingeführt [Margulies et al., 2005]. Das System, das später von Roche übernommen wurde und in den GS FLX- sowie GS Junior-Sequenziergeräten verwendet wird, nutzt das Prinzip der Pyrosequenzierung [Hyman, 1988]. Diese Methode wird bereits seit Mitte der 1990er-Jahren als Hochdurchsatzverfahren zur Sequenzierung genutzt, insbesondere zur SNP (*single nucleotide polymorphism*)-basierten Genotypisierung. Aufgrund der kurzen Leselängen und der hohen Fehlerrate bei Homopolymeren konnte sie jedoch nicht zur *de novo*-Sequenzierung eingesetzt werden [Ronaghi et al., 1996; Ronaghi et al., 1998; Ronaghi, 2001].

Da die Bestimmung der Basenabfolge einer Matrize während der Strangsynthese stattfindet, wird dieses Prinzip auch als *Sequencing by synthesis* bezeichnet. Im Falle einer Pyrosequenzierung wird durch den Einbau eines Nukleotids Pyrophosphat freigesetzt, was zur Aktivierung einer enzymatischen Kaskade führt. Letztlich wird dabei ein Lichtsignal über eine Luziferase erzeugt, das proportional zur Anzahl der eingebauten Nukleotide ist. 454 Life Sciences entwickelte ein Verfahren, bei welchem die Sequenzierreaktion nicht mehr in Lösung, sondern auf einem Trägermaterial stattfindet. Dabei handelt es sich um eine sogenannte Pikotiterplatte (PTP), die aus Millionen einzelner Vertiefungen (Wells) besteht. Diese Wells dienen als lokale und voneinander abgegrenzte Reaktionsräume während der Sequenzierung. Die zu analysierenden DNA-Fragmente sind zudem, ebenso wie alle benötigten Enzyme, an *Beads* gekoppelt und auf der PTP immobilisiert. Das System wurde schnell für verschiedene Anwendungen genutzt, wie die Sequenzierung von Bakteriengenomen oder des humanen Genoms [Andries et al., 2005; Wheeler et al., 2008]. Mit Hilfe der neuen Technologie dauerte die Sequenzierung des Humangenoms nur noch wenige Monate und konnte, nur kurz nach Veröffentlichung der ersten Humangenomsequenz durch die Gruppe um Craig Venter [Levy et al., 2007], die noch die klassische Sanger-Sequenzierung genutzt hatten, fertiggestellt werden.

Seit der ersten Einführung durch 454 Life Sciences haben sich innerhalb kurzer Zeit verschiedene Systeme entwickelt und auf dem Markt etabliert. Hauptsächlich haben sich dabei, neben Roche 454, das SOLiD

(Sequencing by Oligonucleotide ligation and detection)-System von Applied Biosystems sowie die Plattform von Illumina durchgesetzt, welche ebenfalls auf der Sequencing by synthesis-Methode basiert. Mittlerweile scheinen jedoch die Sequenziersysteme von Illumina die Konkurrenz zu verdrängen. Mit Hilfe der NGS-Systeme werden Reads einer durchschnittlichen Länge von etwa 100–500 bp (Basenpaare) sowie mehrere Gigabasen (Gb) an Daten generiert. Nach dem Alignment können die Daten bioinformatisch ausgewertet und die Varianten bewertet werden. Durch die Entwicklung, Optimierung sowie Automatisierung verschiedener Methoden zur Anreicherung und Sequenzierung konnten die Kosten zusehends gesenkt werden.

Die Einführung der NGS-Technologie in Forschungslabors hat in kurzer Zeit zur Identifizierung zahlreicher neuer Krankheits-assoziierter Gene sowie zur molekularen Aufklärung seltener monogener Erkrankungen, wie dem Miller- oder Sensenbrenner-Syndrom, beigetragen [Gilissen et al., 2010; Ng et al., 2010]. Auch für heterogene Erkrankungen konnten Kandidatengene als neue krankheitsverursachende Gene bestätigt werden. Ein Beispiel hierfür ist das *ERCC4*-Gen, das 2013 als erstes FA-Gen durch WES in zwei zuvor nicht zugeordneten Patienten identifiziert werden konnte [Bogliolo et al., 2013].

Durch die Entwicklung verschiedener Anreicherungsstrategien und die Weiterentwicklung der Sequenziertechnologien, z. B. durch die Einführung von Benchtop-Sequenzierern, konnte NGS auch in Diagnostiklabors etabliert werden. Dies erlaubt eine kosteneffizientere und zeitsparendere Analyse krankheitsrelevanter Gene, insbesondere bei heterogenen Erkrankungen wie mentaler Retardierung, bei welcher die Analyse einer Vielzahl von möglichen Kandidatengenen nötig ist [Brett et al., 2014]. Für diagnostische Zwecke werden hauptsächlich bestimmte Gensets oder ganze Exome sequenziert, wohingegen die Genomsequenzierung bislang eher für Forschungszwecke eingesetzt wird. Mittlerweile gibt es zahlreiche kommerzielle Anreicherungspanels, die auf bestimmte Anwendungen zugeschnitten sind, beispielsweise die simultane Untersuchung bestimmter krebsassoziierter Gene. Zudem besteht bei der Panelsequenzierung die Möglichkeit, die Anzahl und Art der Gene bzw. Regionen von Interesse (region of interest; ROI), die angereichert und sequenziert werden sollen, selbst zu bestimmen. Weiterhin ist die Problematik von Nebenbefunden nicht gegeben bzw. deutlich reduziert, da zur Abklärung einer bestimmten Fragestellung nur die entsprechend relevanten Gene sequenziert und analysiert werden [Dorschner et al., 2013]. Dagegen steht, dass keine weiteren Kandidatengene untersucht werden können, sondern bei einem negativen Ergebnis weitere Sequenzier- oder Analysemethoden einbezogen werden müssen. Die Wahl der Anreicherungsmethode sollte demnach auf die jeweilige Anwendung und Zielsetzung abgestimmt werden.

# 1.4 Fragestellung

FA erscheint sowohl klinisch als auch genetisch sehr heterogen und es ist in der Regel kaum möglich, über eine Genotyp-Phänotyp-Korrelation auf die genetische Ursache der Erkrankung zu schließen. Aus diesem Grund sind klassische molekulargenetische Methoden sowohl sehr zeitaufwändig als auch kostenintensiv, insbesondere bei sehr seltenen Komplementationsgruppen. Zudem können noch immer nicht alle Patienten einer bekannten Komplementationsgruppe zugeordnet werden, was darauf schließen lässt, dass noch weitere, bislang nicht identifizierte Gene eine Rolle im FA/BRCA-Signalweg spielen und bei Defekten zu FA führen. Aus diesem Grund soll in der vorliegenden Arbeit die Anwendung verschiedener Anreicherungs- und Hochdurchsatz-Sequenzierungsmethoden zur Mutationsanalyse bei FA-Patienten untersucht werden, die sowohl zur Diagnostik als auch zu Forschungszwecken dienen. Im Vordergrund steht dabei zunächst die Möglichkeit der gezielten Anreicherung von FA-Genen als Ergänzung oder Ersatz zu klassischen Analysemethoden. Durch die Etablierung und Validierung eines selbst erstellten spezifischen Genpanels sollen neben der Analyse aller bislang identifizierten FA-Gene auch vielversprechende Kandidatengene analysiert werden, die sich durch ihre Funktion oder Interaktion im FA/BRCA-Signalweg auszeichnen. Dadurch soll es möglich sein, bei einem Ausschluss bekannter FA-Gene parallel aussichtsreiche neue FA-Gene untersuchen zu können.

In einem zweiten Ansatz soll ein kommerzielles Anreicherungspanel auf einen möglichen Einsatz in der Diagnostik getestet werden, welches einen Großteil der FA-Gene enthält. Dieses wird mit einem anderen *Benchtop*-Sequenziergerät analysiert, so dass sowohl der Vergleich verschiedener Anreicherungs- als auch Sequenziermethoden gegeben ist.

Nachdem manche Patienten trotz einer Panelanalyse keiner Komplementationsgruppe zugeordnet werden können und die molekulare Ursache damit ungeklärt bleibt, soll für einzelne Patienten die Mutationsanalyse auf das gesamte Exom ausgeweitet werden. Hierbei liegt das Augenmerk auf Patienten mit phänotypisch auffälligen Merkmalen sowie der Identifizierung potentieller Kandidatengene.

In der Arbeit sollen zudem potentiell krankheitsverursachende Varianten durch verschiedene Methoden, wie *in silico*-Analysen oder funktionelle Untersuchungen an Patientenmaterial, auf ihre Pathogenität untersucht werden. Dies gilt insbesondere für Mutationen, die bislang nicht beschrieben wurden oder bei Varianten in seltenen Komplementationsgruppen. Damit sollen auch Besonderheiten im Mutationsspektrum einzelner Gene bestimmt werden können, um diese gegebenenfalls in der Routinediagnostik berücksichtigen zu können. Da im Vorfeld während der Routinediagnostik sowie durch Kooperation mit einer französischen Arbeitsgruppe mehrere neue FA-D2-Patienten bestimmt werden konnten, soll dieses Gen ebenfalls einer näheren Betrachtung unterzogen werden. Eine Analyse der *FANCD2*-Pseudogene soll dabei zu einer effektiven Mutationsanalyse in diesem Gen beitragen und das Mutationsspektrum in dieser Komplementationsgruppe erweitert werden.

# 2. Materialien

# 2.1 Antikörper

# 2.1.1 Primärantikörper

Spezifität	Bestellnummer	Hersteller	Wirt	Verdünnung
FANCA	A301-980A-T	Bethyl	Kaninchen	1:1 000
FANCB	ab186729	Abcam	Kaninchen	1:1 000
FANCD2	sc-20022	Santa Cruz	Maus	1:200 (Foci) 1:500 (Immunblot)
	NB100-182	Novus	Kaninchen	1:800 (Foci) 1:10 000 (Immunblot)
FANCI	AB589	A. Smogorzewska		1:1 000
FANCJ	NB100-416	Novus	Kaninchen	1:1 000
Histon-H3	ab1791	Abcam	Kaninchen	1:1 000
NTHL1	ab191413	Abcam	Kaninchen	1:7 500
p84	ab487	Abcam	Maus	1:2 000
RAD50	GTX70228	GeneTex	Maus	1:2 000
RAD51	ab63801	Abcam	Kaninchen	1:800
Tubulin	ab44928	Abcam	Maus	1:1 000
UBE2T	ab140611	Abcam	Kaninchen	1:1 000
Vinculin	ab129002	Abcam	Kaninchen	1:10 000
XPF	ab17790	Abcam	Maus	1:200
YY1	ab199998	Abcam	Kaninchen	1:5 000

# 2.1.2 Sekundärantikörper

HRP-gekoppelte sekundäre Antikörper stammten von Abcam bzw. Santa Cruz. Alexa Fluor®-gekoppelte Sekundärantikörper wurden von Molecular Probes bezogen.

# 2.2 Enzyme

Bezeichnung	Hersteller
BIO-X-ACT <sup>TM</sup> Short DNA Polymerase	Bioline
DNase I, recombinant, RNase-free	Roche
ExoSAP-IT	Affymetrix
Halt <sup>TM</sup> Protease & Phosphatase Inhibitor Cocktail	Thermo Scientific
Micrococcal nuclease	Thermo Scientific

Agilent Technologies
NEB
NEB
NEB
Invitrogen
NEB
NEB

# 2.3 Größenstandards

Bezeichnung	Hersteller
Precision Plus Protein <sup>TM</sup> All Blue Prestained Protein	BioRad
Standards	Dioitad
100 bp DNA ladder	NEB
1 kb DNA ladder	NEB

## 2.4 NGS-Panels zur Anreicherung

SeqCap EZ Libraries wurden bei Roche NimleGen bestellt. Es wurden sowohl SeqCap EZ Choice als auch SeqCap EZ Developer Libraries erstellt und geordert. Zudem fanden Sequenzierungen unter Verwendung des TruSight Cancer Panels (Illumina) statt. Dieses diente zur spezifischen Anreicherung von 94 Genen sowie 284 SNPs, die mit einer Prädisposition für eine erbliche Tumorerkrankung assoziiert sind. Eine Übersicht über die in den SeqCap EZ-Panels enthaltenen Gene ist im Anhang aufgeführt. Die durch das TruSight Cancer-Panel angereicherten Gene sowie die rs-Nummern der enthaltenen SNPs sind ebenfalls im Anhang aufgeführt. Die entsprechenden Koordinaten der angereicherten Regionen des Krebspanels sind zudem online hinterlegt (http://support.illumina.com/downloads/trusight\_cancer\_product\_files.html).

Die Anreicherung sowie Sequenzierung ganzer Exome wurde von kommerziellen Dienstleistern (BGI, Shenzhen/China; Genome Diagnostics Nijmegen, Nijmegen/Niederlande) durchgeführt, nicht aber die Datenanalyse, die selbst vorgenommen wurde. Es wurden dabei *Agilent SureSelect Human All Exon Kits* sowie das *NimbleGen SeqCap EZ Human Exome Library-Kit* zur Anreicherung verwendet.

# 2.5 Primer

Zum Entwerfen von Primern wurde das Programm Primer-BLAST von NCBI genutzt. Es greift auf Primer3 und BLAST zurück, um spezifische Primer zu generieren [Ye et al., 2012]. NGS-Primer wurden mit zusätzlicher HPLC-Aufreinigung geordert. Alle Oligonukleotide wurden von den Firmen Biolegio und Sigma-Aldrich bezogen. Sie wurden in lyophilisierter Form verschickt, so dass sie zur Herstellung einer Stammlösung von 100 bzw. 1000 µM zunächst mit demineralisiertem Wasser (ddH<sub>2</sub>O) gelöst wurden. Diese Stammlösungen wurden bei -20 °C gelagert und zur Herstellung von Arbeitslösungen (10 µM) verwendet. Die Primer zur Anreicherung sind in den entsprechenden Protokollen angegeben ("*NimbleGen SeqCap EZ Library LR User's*  *Guide*", Roche NimbleGen; "*Multiplex Identifier (MID) Adaptors for Rapid Library Preparations*"; Roche NimbleGen). Primer zur Nachsequenzierung von Varianten sind im Anhang angegeben.

# 2.6 Kit-Systeme

Bezeichnung	Hersteller
Agencourt CleanSEQ System	Beckman Coulter
Agilent DNA 7500 Kit	Agilent
Agilent High Sensitivity DNA Kit	Agilent
BigDye® Terminator v1.1 Cycle Sequencing Kit	Thermo Scientific
DNA Clean & Concentrator <sup>TM</sup> Kit	Zymo Research
FastStart™ High Fidelity PCR System, dNTPack	Roche
GeneJET Gel Extraction Kit	Thermo Scientific
GeneJET Genomic DNA Purification Kit	Thermo Scientific
GS FLX Titanium Rapid Library MID Adaptors Kit	Roche
GS FLX Titanium Rapid Library Preparation Kit	Roche
GS Junior Titanium emPCR Kit (Lib-L)	Roche
GS Junior Titanium PicoTiterPlate Kit	Roche
GS Junior Titanium Sequencing Kit	Roche
HighPure RNA Isolation Kit	Roche
illustra GenomiPhi V2 DNA Amplification Kit	GE Healthcare
In-Fusion® HD Cloning Kit	Clontech Laboratories
Lenti-X <sup>TM</sup> Expression System (EF1alpha Version)	Clontech Laboratories
MinElute PCR Purification Kit	Qiagen
NE-PER® Nuclear and Cytoplasmic Extraction Kit	Thermo Scientific
NucleoSpin® Gel and PCR Clean-up Kit	Machery Nagel
Pierce <sup>™</sup> Crosslink Magnetic IP/Co-IP Kit	Thermo Scientific
PureLink® Quick Gel Extraction Kit	Invitrogen
QIAGEN Plasmid Plus Maxi Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
QuantiFluor® dsDNA System	Promega
Qubit® dsDNA BR Assay Kit	Thermo Scientific
Qubit® Protein Assay Kit	Thermo Scientific
Quick-gDNA™ MiniPrep Kit	Zymo Research
SALSA® MLPA® probemix P002 BRCA1 (CE-IVD)	MRC-Holland
SALSA MLPA P045 BRCA2/CHEK2 probemix	MRC-Holland
SALSA MLPA probemix P031-B2/P032-B2 FANCA	MRC-Holland
SeqCap EZ Hybridization and Wash Kit	Roche
Subcellular Protein Fractionation Kit	Thermo Scientific
Zyppy <sup>™</sup> Plasmid Miniprep Kit	Zymo Research

# 2.7 Kommerzielle Zelllinien

Bezeichnung	Hersteller
Lenti-X <sup>TM</sup> 293T Cell Line	Clontech Laboratories
HAP1 control cells (C665)	Horizon Discovery
Human NTHL1 5bp deletion knockout cell line	Horizon Discovery

# 2.8 Kompetente Bakterienstämme

Bezeichnung	Hersteller
One Shot® TOP10 Chemically Competent E. coli	Thermo Scientific
XL10-Gold Ultracompetent Cells	Stratagene
MAX Efficiency® DH5a <sup>TM</sup> Competent Cells	Thermo Scientific
Stellar <sup>TM</sup> Competent Cells	Clontech Laboratories

# 2.9 Plasmide

Die Vektorkarten der angegebenen Plasmide sind im Anhang abgebildet.

Bezeichnung	Hersteller
pBlueScript II SK (+)	Agilent Technologies
pLVX-EF1a-IRES-Puro	Clontech Laboratories
pWPXL:XPFwt916-HA	J. Surallés

# 2.10 Kulturmedien, Puffer und Reagenzien

Gängige Chemikalien, die im Folgenden nicht aufgeführt sind, wurden von den Firmen Amersham Biosciences, AppliChem, Carl Roth, Life Technologies, Merck sowie Sigma Aldrich im analytischen Reinheitsgrad bezogen.

Bezeichnung	Hersteller
Agarose	Serva
Agencourt AMPure XP	Beckman Coulter
Amniopan	PAN-Biotech
Ampicillin (Stammlösung 10 mg/ml)	Carl Roth
Ampuwa <sup>®</sup> (ddH <sub>2</sub> O)	Fresenius Kabi
β-Mercaptoethanol	AppliChem
COT Human DNA, Fluorometric Grade	Roche
Cycloheximid (100 mg/ml)	Sigma-Aldrich
DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)	Life Technologies
Dulbecco's Modified Eagle Medium (DMEM)	Sigma-Aldrich
DMSO (Dimethylsulfoxid)	AppliChem
dNTP Set (100 mM)	Bioline
DPBS (Dulbecco's Phosphate-Buffered Saline)	PAN-Biotech

Dynabeads® M-270 Streptavidin Ethidiumbromid Fötales Kälberserum (FCS) HDGreen® Plus Safe DNA Dye HEPES (2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure) HiDi<sup>TM</sup> Formamid Hoechst 33342 Hydroxyurea (HU) Iscove's Modified Dulbecco's Medium (IMDM) Immersionsöl Immobilon Western Chemiluminescent HRP Substrat IP Lysis/Wash Buffer Kanamycin (Stammlösung 50 mg/ml) Minimum Essential Media (MEM) + GlutaMAX<sup>TM</sup>-I Mitomycin C (MMC) NuPAGE® LDS Sample Buffer (4x) NuPAGE® MOPS SDS Running Buffer (20x) NuPAGE® Sample Reducing Agent (10x) NuPAGE® Transfer Buffer (20x) NuPAGE® Tris-Acetate SDS Running Buffer (20x) Oligo(dT)<sub>12-18</sub> Primer Orange G PenStrep Propidiumiodid FluoroPure<sup>™</sup> Grade ProLong® Gold Antifade Mountant with DAPI Protease Inhibitor Cocktail Tabletten (100x) Proteinase K solution Puromycin (10 mg/ml) QuantiFluor® dsDNA System Restore<sup>™</sup> Immunblot Stripping Buffer RPMI 1640 + GlutaMAX<sup>TM</sup>-I SDS solution (20%) Solution  $6 - VB-48^{TM} - PI$ PerfeCTa® SYBR® Green FastMix® for iQ<sup>TM</sup> Tet System Approved Fetal Bovine Serum (FBS) TrypLE<sup>TM</sup> Express Tween® 20

Invitrogen AppliChem Sigma-Aldrich Intas Gibco Applied Biosystems Life Technologies Sigma-Aldrich Gibco Carl Roth Merck Millipore Thermo Scientific Carl Roth Gibco Medac Thermo Scientific Thermo Scientific Thermo Scientific Thermo Scientific Thermo Scientific Thermo Scientific Carl Roth Gibco Thermo Scientific Thermo Scientific Roche AppliChem Life Technologies Promega Thermo Scientific Gibco AppliChem ChemoMetec A/S Quanta Biosciences? Clontech Laboratories Gibco AppliChem

Reagenz	Zusammensetzung
DAPI-Färbepuffer Durchflusszytometrie	100 mM Tris (pH 7,4); 154 mM NaCl; 1 mM CaCl <sub>2</sub> ; 0,5 mM MgCl <sub>2</sub> ; 0,2 % BSA; 0,1 % NP40
DMEM + FBS	DMEM + GlutaMAX <sup>TM</sup> -I; 10 % FBS
DNA-Probenpuffer (6x)	40 % (w/v) Sucrose; 0,25 % Bromphenolblau; 0,25 % Xylencyanol
DNA-Probenpuffer Orange G (10x)	100 mg Orange G; 15 ml Glycerin; 10 ml 0,5 M EDTA; 5 ml 10 % SDS; ad ddH <sub>2</sub> O 50 ml
Einfriermedium	RPMI-Medium; 10 % FCS; 10 % DMSO
Einfriermedium HAP1-Zellen	IMDM; 20 % FCS; 10 % DMSO
2x HEPES Buffered Saline (HeBs) (pH 7,0)	1,636 g NaCl; 1,19 g HEPES; 0,0213 g Na <sub>2</sub> HPO <sub>4</sub> ; 80 ml ddH <sub>2</sub> O
IMDM + FCS + PenStrep	IMDM; 10 % FCS; 100 U/ml Penicilin; 100 mg/ml Streptomycin
LB-Medium	$25 \text{ g LB-Broth}$ ; ad $1 \text{ l } ddH_2O$
Lysis-Puffer (pH 7,4)	155 mM NH4Cl; 10 mM KHCO3; 0,1 mM EDTA
MEM + FBS	MEM + GlutaMAX <sup>TM</sup> -I; 10 % FBS
MEM + FBS + Puro	MEM + GlutaMAX <sup>TM</sup> -I; 10 % FBS; 0,01 % Puromycin
MEM + FCS	MEM + GlutaMAX <sup>TM</sup> -I; 10 % FCS
PBS-T	1x PBS; 0,1 % Tween <sup>®</sup> 20
RPMI + FCS	RPMI 1640 + GlutaMAX <sup>TM</sup> -I; 15 % FCS
Selektionsplatten	25 g LB-Broth; 15 g Agar; ad 1 l ddH <sub>2</sub> O; 1 ml Antibiotikum
SE-Puffer (pH 8,0)	75 mM NaCl; 25 mM EDTA
TBE-Puffer (5x)	450 mM Tris; 450 mM Borsäure; 10 mM EDTA
TBS (10x) (pH 7,6)	24 g Tris base, 88 g NaCl, ad 1 l dd $\mathrm{H_{2}O}$
TBS-T	1x TBS; 0,1 % Tween <sup>®</sup> 20
TE-Puffer	10 mM Tris-HCl; 1 mM EDTA

# 2.11 Sonstige Puffer und Lösungen

# 2.12Sonstige Verbrauchsmaterialien

Bezeichnung	Hersteller
Deckgläser	Carl Roth
Einmalspritzen (1ml; 5 ml)	Brand
Erlenmeyerkolben (DURAN®)	Duran Group
FACS-Röhrchen (unsteril)	BD Biosciences
Falcons (15 ml; 50 ml)	Greiner Bio-One
Fuchs-Rosenthal-Zählkammer	Brand
Handschuhe (Nitril, Latex)	Carl Roth
iBlot® 2 PVDF mini Transfer Stacks	Thermo Scientific
Kanüle (21 G)	B. Braun

17 11 1	
Kryoröhrchen	Nunc
Multiwell-Kulturplatten	Greiner Bio-One, Nunc
NC-Slide A2 <sup>™</sup>	ChemoMetec A/S
NC-Slide A8 <sup>TM</sup>	ChemoMetec A/S
NuPAGE <sup>TM</sup> Novex <sup>TM</sup> Bis-Tris Protein-Gele; 1,0 mm; 10-well	Thermo Scientific
NuPAGE <sup>TM</sup> Novex <sup>TM</sup> Tris-Acetat Protein-Gele; 1,0 mm; 10-well	Thermo Scientific
Parafilm	Pechiney
Pasteurpipetten	Brand
PCR-Reaktionsgefäße (0,2 ml; 8er-Streifen)	Eppendorf
Petrischalen	Greiner bio-one
Pipettenspitzen(10 µl; 20 µl; 100 µl; 200 µl; 1250 µl)	Sarstedt
Roti®-PVDF, Porengröße 0,45 µm	Carl Roth
Reaktionsgefäße (200 µl - 2000 µl)	Eppendorf
Serologische Pipetten (Glas; 2 ml; 5 ml; 10 ml; 20 ml)	Brand
Serologische Pipetten (Kunststoff; 2 ml; 5 ml; 10 ml; 25 ml)	Greiner Bio-One
Slide flasks (Zellkulturkammern auf Glasobjektträgern)	Sarstedt, Thermo Scientific
Spritzenvorsatzfilter Minisart®	Sartorius AG
Zellkulturflaschen (50 ml, 200 ml)	Corning, Nunc, Sarstedt
Zubehör für ABI PRISM 310 Genetic Analyser	PE Biosystems
Zubehör für GS Junior	Roche
Zubehör für QIAvac 24 Plus	Qiagen

# 2.13Geräte

Bezeichnung	Modell	Hersteller
Analysenwaage	MC1	Sartorius
Autoklav	Autoclave Systec V-150	Systec GmbH
Bioanalyzer	2100 Bioanalyzer	Agilent
Blot-Systeme	BlueBlot Wet 100 Tank	SERVA Electrophoresis
	iBlot® 2 Gel-Transfersystem	Thermo Fisher
CO <sub>2</sub> -Inkubator	NuAire NU-5810E	Integra Biosciences
Drehrad	Reax2	Heidolph
Durchflusszytometer	BD™ LSR II	BD Biosciences
Elektrophoresekammern	Compact M	Biometra
	SE260	AA Hoefer
Fluorometer	QuantiFluor®-ST	Promega
	Qubit®	Life Technologies
Homogenisiersystem	ULTRA-TURRAX® Tube Drive	IKA
Imaging-Systeme	FluorChem® HD2	Alpha Innotech

	MicroChemi	DNR Bio-Imaging Systems
Kühlsystem	DLK500 FRYKA	FRYKA Kältetechnik
Magnetrührer	Nuova II	Bioblock Scientific
Magnetständer	DynaMag <sup>™</sup> -2 Magnet	Thermo Fisher
Mikroskop	Axiovert 40C	Carl Zeiss
Netzgerät	PowerPac 1000	BioRad
pH-Meter	Digital-pH-Meter 646	Knick
Photometer	NanoPhotometer P-Class P300-30	Implen
Pipettierhilfen	Mehrkanalpipetten	Brand, Eppendorf
	Mikropipetten	Gilson
	Multipipette® pro	Eppendorf
	Pipetboy	Integra Biosciences
Pipettierroboter	JANUS	Perkin Elmer
Schüttelinkubator		INFORS AT
Schüttler	VIBRAX-VXR	IKA
Schwenktisch	Polymax 1040	Heidolph
Sequenzierer	ABI PRISM 310 Genetic Analyser	Applied Biosystems
	ABI PRISM 310 XL Genetic Analyser	
	ABI PRISM 3730 Genetic Analyser	
	GS Junior	Roche
	HiSeq 2000	Illumina
	HiSeq 4000	Illumina
	MiSeq	Illumina
	SOLiD 5500xl Genetic Analyzer	Applied Biosystems
Sterilbank	LaminAir® HB2448	Heraeus
Thermocycler	Biometra TRIO	Analytik Jena
	iCycler®	BioRad
	Mastercycler® ep Gradient S	Eppendorf
Thermomixer	TS1 ThermoShaker	Biometra
Vakuumkonzentrator	RVC-2-18	Christ
Vakuumsystem zur DNA-Präparation	QIAvac 24 Plus	Qiagen
Vortexgerät	Vortex-Genie® 2	Bender Hobein
Wärmehaube	Certomat®H	B. Braun Biotech
Wasserbad	Тур В	Lauda
Zellzähler	NucleoCounter® NC-250 <sup>TM</sup>	ChemoMetec A/S
Zentrifugen	Megafuge® 1.0	Hereaus
	Mikro 220 R	Hettich
	MiniSpin® plus	Eppendorf
	Multifuge 1L-R	Hereaus
	Rotilabo <sup>®</sup> -Mini-Zentrifuge	Carl Roth

Rotina 420

Hettich

# 2.14Softwareprogramme und Datenbanken

Programm	Quelle
Adobe Photoshop CS4	Adobe Systems
Alamut	Interactive Biosoftware
BoxPlotR	http://shiny.chemgrid.org/boxplotr/
Coffalyser.Net	MRC-Holland
Clustal Omega	http://www.ebi.ac.uk/Tools/msa/clustalo/
CorelDRAW Graphics Suite X6	Corel
Elektronenmikroskopie-Datenbank	https://www.ebi.ac.uk/pdbe/emdb/
EndNote X7	Thomson Reuters
Ensembl Genome Browser	http://www.ensembl.org/index.html
ESEFinder	rulai.cshl.edu/
ExAC-Browser	http://exac.broadinstitute.org/
ExPASy	http://web.expasy.org/translate/
Fanconi Anemia Mutation Database	http://www.rockefeller.edu/fanconi/
GeneSplicer	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml
GS Run Browser	Roche
Human Splicing Finder	http://www.umd.be/HSF3/
HOPE	http://www.cmbi.ru.nl/hope/
In-Fusion® Online Tools	http://www.clontech.com/US/Products/Cloning_and_ Competent_Cells/Cloning_Resources/Online_In-Fusion_Tools
MaxENTScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
Microsoft Office 2010	Microsoft
MiSeq Reporter	Illumina
MPLUS AV	Phoenix Flow Systems
MutationTaster	http://www.mutationtaster.org/
NCBI	http://www.ncbi.nlm.nih.gov/
NextGENe	Softgenetics
NextGENe Viewer	Softgenetics
NNSPLICE	http://www.fruitfly.org/seq_tools/splice.html
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/
Primer-BLAST	http://www.ncbi.nlm.nih.gov/tools/primer-blast/
Rescue ESE	http://genes.mit.edu/burgelab/rescue-ese/
Sequencher 4.7 Demo	Genes Codes Corporation
SIFT	http://sift.jcvi.org/
UCSC Genome Browser	http://genome.ucsc.edu/

# 3. Methoden

# 3.1 Allgemeine Methoden der Zellkultur

# 3.1.1 Kultivierung von adhärenten und Suspensionszellen

Alle Zellen wurden in Begasungsbrutschränken bei 37 °C, 5 % CO<sub>2</sub> und einer relativen Luftfeuchtigkeit von 96–100 % kultiviert. Zugesetztes FCS wurde vor der Verwendung hitzeinaktiviert (56°C, 20 min). Arbeiten mit Zellkulturen fanden unter sterilen Bedingungen an Sterilwerkbänken statt.

Primäre Fibroblasten wurden in Amniopan, T-Antigen-transformierte Fibroblasten in MEM mit 10 % FCS und HAP1-Zellen in IMDM mit 10 % FCS und 1 % PenStrep kultiviert. Ein Mediumwechsel fand zweimal pro Woche statt. Sobald die Zellen eine Konfluenz von etwa 75–90 % erreicht hatten, wurden sie zur Weiterkultivierung gesplittet. Dazu wurde das Medium abgenommen, die Zellen mit 5 ml 1x PBS gewaschen und anschließend 5 ml Trypsinlösung zugegeben. Bis zum Ablösen der Zellen wurden diese im Brutschrank inkubiert und die Zellsuspension in Falcons überführt. Nach Sedimentation der Zellen wurde der Überstand verworfen, das Zellpellet in frischem Medium aufgenommen und auf neue Zellkulturflaschen mit den entsprechenden Medien aufgeteilt.

Zur Kultivierung von EBV-transformierten (Epstein-Barr-Virus) lymphoblastoiden Zelllinien (LCL, *lymphoblastoid cell line*) wurde RPMI-Medium mit 15 % FCS verwendet. Die Zellen wurden ein- bis zweimal pro Woche, je nach Zellwachstum, geteilt.

# 3.1.2 Kryokonservierung und Auftauen von Zellen

Suspensionszellen oder durch Trypsinisierung abgelöste adhärente Zellen wurden zunächst durch Zentrifugation für 8–10 min bei 209 x g sedimentiert und der Überstand verworfen. Das Zellpellet wurde in 1,8 ml Einfriermedium (4 °C) resuspendiert und in ein steriles Kryoröhrchen überführt. Dieses wurde über Nacht in einem mit Isopropanol gefüllten Einfrierbehältnis bei –80 °C gelagert, um eine schrittweise Reduktion der Temperatur zu erreichen. Am nächsten Tag konnten die Kryoröhrchen für eine dauerhafte Lagerung in Tanks mit Flüssigstickstoff überführt werden.

Zur Weiterkultivierung kryokonservierter Zellen wurden die Aliquots in einem Wasserbad bei 37 °C zügig erwärmt und in vorgewärmtes Zellkulturmedium überführt. Am folgenden Tag wurde das Medium gewechselt, um das im Einfriermedium enthaltene und bei höheren Temperaturen zelltoxische DMSO zu entfernen.

# 3.1.3 Bestimmung der Zellzahl

Mit Hilfe einer Fuchs-Rosenthal-Zählkammer wurde die Anzahl der Zellen/ml einer Zellsuspension bestimmt. Dazu wurden drei Großquadrate der Zählkammer ausgezählt und der Mittelwert der Zellzahl mit dem Faktor 5000 multipliziert.

# 3.2 Gentransfer

# 3.2.1 Klonierung

#### a. Restriktionsverdau

Zum Einbringen der XPF-Wildtypsequenz (wt) sowie der Mutation c.1765C>T (mut) in einen lentiviralen Vektor wurden verschiedene Klonierungsstrategien angewendet. Um die Mutation durch Mutagenese (siehe 3.6.5c) in die codierende Sequenz von XPF einzufügen, wurde die XPF-cDNA aus dem Vektor pWPXL:XPFwt916-HA herausgeschnitten und in den deutlich kleineren Vektor pBlueScript II SK (+) kloniert. Dazu wurden beide Vektoren für 1 h bei 37 °C mit Restriktionsenzymen verdaut (siehe Ansatz), gefolgt von einer Hitzeinaktivierung für 20 min bei 80 °C und einer Pause bei 4 °C.

1 μl (10 U)
1 μl (10 U)
1 µg
5 µl
50 µl

Zur Überprüfung der Produktgrößen wurden diese auf ein Agarosegel aufgetragen.

# b. Ligation

Um die XPF-cDNA mit dem verdauten pBlueScript-Zielvektor zu ligieren, wurden die Restriktionsprodukte zunächst auf ein TAE-Agarosegel aufgetragen, die Banden von Interesse ausgeschnitten und aufgereinigt (siehe 3.6.2b). Für die Ligation wurden in einem Standardansatz (20 µl) zu 2 µl 10x T4 DNA-Ligase-Puffer 50 ng Vektor, das Insert in 3-fach molarem Überschuss sowie 1 µl T4 DNA-Ligase zugegeben. Der Ansatz wurde durch Pipettieren gemischt und für 10 min bei Raumtemperatur (RT) inkubiert und bei 4 °C bis zur Transformation (siehe 3.2.2) aufbewahrt.

### c. InFusion-Klonierung

Um DNA-Sequenzen über Homologe Rekombination in einen linearisierten Vektor einzubringen, wurde das In-Fusion HD Cloning Kit genutzt. Als Voraussetzung müssen beide DNA-Fragmente linearisiert sein und jeweils 15 bp Homologie an den Enden aufweisen. Zunächst wurde hierfür der Vektor durch einen Restriktionsverdau mit zwei verschiedenen Restriktionsenzymen linearisiert, um eine Religation zu vermeiden. Mit Hilfe des InFusion Primer Design Tools wurden Primer erstellt, die jeweils am 5'-Ende einen Überhang von 15 bp besaßen, die komplementär zu den Enden des linearisierten Vektors waren. Über eine PCR-Reaktion wurden diese Sequenzen an das Insert angehängt und anschließend mit der eigentlichen InFusion-Reaktion nach den Angaben im Protokoll verfahren.

# 3.2.2 Transformation

Zur Vervielfältigung wurden Plasmide nach der Klonierung in chemokompetente Bakterien eingebracht. Für Standardklonierungen wurden One Shot TOP10 Chemically Competent E. coli sowie XL10-Gold Ultracompetent Cells verwendet. Bei In-Fusion-Klonierungen erfolgte die Transformation in Stellar Competent Cells, wohingegen lentivirale Plasmide nach der Ligation des Inserts mit dem Zielvektor in MAX Efficiency DH5a Competent Cells eingebracht wurden. Dabei wurde jeweils nach den Angaben der Hersteller verfahren. Für alle Transformationen galt, dass die kompetenten Zellen auf Eis aufgetaut und mit der entsprechenden Menge Plasmid-DNA vorsichtig gemischt wurden. Der jeweilige Ansatz wurde zunächst für 30 min auf Eis inkubiert, ehe ein Hitzeschock für 30–45 sec bei 42 °C erfolgte. Nach zweiminütiger Inkubation auf Eis wurde LB-Medium zugegeben und der Ansatz 1 h bei 37 °C inkubiert, bis er auf Antibiotikahaltigen LB-Platten ausplattiert wurde. Die Platten wurden bei 37 °C über Nacht inkubiert, so dass am nächsten Tag Klone gepickt werden konnten.

### 3.2.3 Transduktion

Bei einigen Patientenlinien fand bereits im Vorfeld ein Ausschluss bzw. eine Komplementationsgruppen-Bestimmung durch retrovirale Transduktion im Labor von Herrn Prof. Dr. Helmut Hanenberg (Medizinische Fakultät, Heinrich-Heine-Universität, Düsseldorf) nach Standardprotokollen statt [Hanenberg et al., 1997; Hanenberg et al., 2002].

Weiterhin erfolgte eine Verpackung lentiviraler Plasmide sowie die lentivirale Transduktion von transformierten Fibroblasten mit Hilfe des *Lenti-X Expression Systems (EF1alpha Version)* nach den Angaben im Protokoll. Zur Selektion wurden die Zellen in Puromycin-haltigem (1 µg/ml) MEM + 10 % FBS kultiviert.

# 3.3 Immunfluoreszenz

Mit Hilfe der indirekten Immunfluoreszenz wurden Antigene auf Einzelzellebene detektiert. Bei dieser Methode bindet zunächst ein spezifischer Primärantikörper an das Antigen und wird durch einen Fluoreszenz-markierten Sekundärantikörper sichtbar gemacht.

Zum Nachweis verschiedener DNA-Reparaturfoci wurden jeweils 200 000–300 000 Zellen in *Slide flasks* eingesät. Zur Induktion des DNA-Schadens wurden die Zellen mit 40 ng/µl MMC für ca. 16 h im Brutschrank bei 37 °C inkubiert. Der Überstand wurde verworfen, die Zellen mit PBS gewaschen und anschließend für 15 min bei RT mit 4% Paraformaldehyd fixiert. Von nun an erfolgten alle Inkubationsschritte auf einem Schwenktisch. Nach der Fixierung wurde die Zellmembran mit eiskaltem 100% Methanol für 30 min auf Eis permeabilisiert. Unspezifische Bindungsstellen wurden durch eine Inkubation

mit 20 % iger FCS/PBS-Lösung blockiert. Danach folgte die Inkubation mit dem Primärantikörper für 1 h bei RT. Nach dreimaligem Waschen mit PBS für 15 min bei RT wurde der Fluoreszenz-markierte Sekundärantikörper zugegeben und für 30 min bei RT im Dunkeln inkubiert. Nach drei Waschschritten mit PBS für 20 min bei RT im Dunkeln konnte der Glasobjektträger von der Zellkulturkammer entfernt werden. Es wurden einige Tropfen des Eindeckelmediums *ProLong Gold Antifade Mountant with DAPI* auf den Objektträger aufgebracht und ein Deckglas aufgelegt. Das Medium verhindert sowohl das Ausbleichen des Fluoreszenz-Signals und färbt zum anderen die Zellkerne durch das enthaltene DAPI an. Die Präparate wurden bis zur Auswertung an einem inversen Fluoreszenz-Mikroskop im Dunkeln bei RT gelagert.

## 3.4 Zellzyklus-Analyse am Durchflusszytometer

Zur Untersuchung der Zellzyklusverteilung wurden LCLs für 48 h mit 15 ng/ml MMC und Fibroblasten mit 10 ng/ml MMC inkubiert. Die Zellzyklusanalyse fand durch durchflusszytometrische Fluoreszenzmessungen am BD LSR II statt. Zur Bestimmung des DNA-Gehaltes wurde die stöchiometrische Bindung von DAPI an die DNA genutzt. Dazu wurde das Zellpellet in Färbepuffer resuspendiert, 2 µg/ml DAPI zugegeben und 30 min bei 4 °C im Dunkeln inkubiert.

Um den Anteil an vitalen und apoptotischen Zellen sowie die Zellzyklusverteilung einer Probe zu bestimmen, wurde zunächst 8 µg/ml Hoechst 33342 zugegeben. Dieser Farbstoff diffundiert durch die Plasmamembran lebender und fixierter Zellen und bindet an die kleine Furche der DNA. Durch die stärkere Chromatin-Kondensation zeigen apoptotische Zellen ein stärkeres Signal. Die Proben wurden für 30 min bei 37 °C im Dunkeln inkubiert und durch Zentrifugation erneut sedimentiert. Das Pellet wurde 5 min auf Eis gekühlt, in 1 µg/ml Propidiumiodid-Lösung resuspendiert und für 1 min inkubiert. Dieser Farbstoff kann die perforierte Zellmembran von toten, nicht aber von vitalen Zellen durchdringen. Die Proben wurden am Durchflusszytometer gemessen und die Daten mit Hilfe der Software MPLUS AV ausgewertet.

# 3.5 Zellviabilitäts-Test

# 3.5.1 MMC-Sensitivität

Um die Sensitivität von Zellen gegenüber MMC zu ermitteln, wurden diese mit unterschiedlichen MMC-Konzentrationen über einen bestimmten Zeitraum kultiviert und anschließend einer Lebend-Tot-Messung unterzogen. Die Daten wurden ebenfalls mit Hilfe der Software MPLUS AV ausgewertet.

Für die Messung von LCLs wurden 250 000 Zellen in T25-Zellkulturflaschen in RPMI + FCS eingesät und Konzentrationen von 0–300 ng/µl MMC zugegeben. Nach einer Inkubation von acht Tagen wurden die Zellen am BD LSR II gemessen. Die Lebend-Tot-Messung fand, wie unter 3.4 beschrieben, durch die Färbung der Zellen mit Hoechst 33342 und Propidiumiodid statt.

Bei Fibroblasten-Kulturen wurden je 10 000–20 000 Zellen/Well einer 6-Well-Platte in MEM + FBS eingesät. Am nächsten Tag wurde 0–100 ng/µl MMC zu den Zellen pipettiert und die Platten für acht Tage inkubiert. Die Zellen wurden abgelöst, sämtlicher Überstand in Falcons gesammelt und für 8 min bei 209 x g abzentrifugiert. Das Zellpellet wurde in 200 µl MEM + FBS vorsichtig resuspendiert. 19 Volumenanteile Zellsuspension wurden mit einem Volumenanteil Solution 6 gemischt, auf ein NC-Slide aufgetragen und am NucleoCounter mit dem *Vitality Assay*-Programm gemessen. Das in der Färbelösung enthaltene Propidiumiodid diente zum Anfärben toter Zellen, wohingegen der Farbstoff VitaBright-48<sup>TM</sup>(VB-48<sup>TM</sup>) zur Bestimmung vitaler Zellen in Abhängigkeit ihres Gehalts an freien Thiolen diente. Eine Abnahme des zellulären Thiolgehalts weist auf einen Zusammenbruch des mitochondrialen Transmembranpotentials hin und ist neben der Chromatin-Kondensation sowie der DNA-Fragmentierung ein frühes Anzeichen für Apoptose. Die Daten wurden direkt am Gerät ausgewertet.

#### 3.5.2 UV-Sensitivität

Um die UV-induzierte Inhibition des Wachstums von primären und transformierten Fibroblasten zu untersuchen, wurden Aliquots von 5 000 Zellen/Well in 6-Well-Platten ausgesät. Nach zwei Tagen wurden die Zellen mit PBS gewaschen und mit 0-12,5 J/m<sup>2</sup> UV-Licht bestrahlt. Nach fünf Tagen wurden die Zellen für 3 h in [Methyl-3H]-Thymidin-haltigem Medium inkubiert, mit PBS gewaschen und für 15 min in [Methyl-3H]-Thymidin-freiem Medium kultiviert, bevor die Überlebensrate bestimmt wurde. Der Versuch wurde im Labor von Herrn Dr. Nicolaas G. J. Jaspers (Department of Genetics, Erasmus University Medical Center, Rotterdam) durchgeführt und ausgewertet, wie auch der UDS (unscheduled DNA synthesis)- und RRS (recovery of RNA sysnthesis)-Test. Dazu wurden Wildtyp-Fibroblasten mit Polystyren-Beads für drei Tage inkubiert. Anschließend wurden sie mit Patientenfibroblasten gemischt und auf Deckgläsern ausgesät. Nach zwei Tagen wurden die Zellen mit PBS gewaschen, mit 16 J/m<sup>2</sup> bestrahlt und für 3 h in 0,1 µM 5-Ethynyl-2'-desoxyuridin (EdU)-haltigem Medium inkubiert. Die Zellen wurden erneut mit PBS gewaschen, in EdU-freiem Medium für 15 min inkubiert, wieder gewaschen und mit 3,7 % Formaldehyd mit 0,5 % Triton für 15 min fixiert. Zur Fluoreszenzfärbung wurden die Zellen 30 min in einem Puffer mit 10 mM CuSO4 und Alexa Fluor 594 Azid (Click-iT; Invitrogen) inkubiert und anschließend mit Vectashield Antifade Mounting Medium with DAPI (Vector Laboratories) eingedeckt. Für RRS-Analysen wurden die Zellen analog zum UDS-Assay behandelt. Allerdings wurden die Zellen 16 h nach der Bestrahlung für 2 h in 0,1 µM 5-Ethynyl-uridin (EU)-haltigem Medium inkubiert.

# 3.6 Allgemeine molekulargenetische Methoden

## 3.6.1 Isolation von Nukleinsäuren

### a. Isolation genomischer DNA

Zur Isolation genomischer DNA (gDNA) aus Vollblut, LCLs und Fibroblasten, wurde entweder die modifizierte Aussalzmethode nach Miller [Miller et al., 1988] angewendet oder erfolgte mit Hilfe des GeneJET Genomic DNA Purification Kits.

Bei der Aussalzmethode wurde Blut mit dem dreifachen Volumen an Lysispuffer (4 °C) gemischt und bis zur vollständigen Lyse für etwa 15 min auf Eis inkubiert. Nach Zentrifugation für 10 min bei 700 x g, zweimaligem Waschen des Zellpellets durch Vortexen mit 5–10 ml Lysispuffer und erneuter Zentrifugation, erfolgte die Lyse der kernhaltigen Lymphozyten. Dazu wurde das Zellpellet in 5 ml SE-Puffer resuspendiert und 130 µl Proteinase K (10 mg/ml) sowie 250 µl 20 % SDS zugegeben. Nach einer Inkubation für 3 h bei 55 °C oder über Nacht bei 37 °C wurden 2,5 ml SE-Puffer sowie 2 ml 6 M NaCl zugegeben. Der Ansatz wurde gevortext, bis sich eine milchige Trübung einstellte. Durch Zentrifugation für 15 min bei 2700 x g wurden die übrigen Zellbestandteile sedimentiert. Der Überstand wurde in ein neues Falcon überführt und die hochmolekulare DNA durch Zugabe des doppelten Volumens an 100 % Ethanol gefällt. Mit Hilfe einer gebogenen Pasteurpipette wurde der DNA-Faden kurz in 70 % Ethanol getaucht, bei RT getrocknet und in 200–600 µl ddH<sub>2</sub>O über Nacht auf dem Drehrad gelöst. LCLs und Fibroblasten wurden analog behandelt, hier fielen allerdings die Schritte der Lyse und des Waschens mit Lysispuffer weg. Stattdessen wurden die Zellen bei 150 x g für 10 min sedimentiert und das Pellet mit 5 ml PBS gewaschen. Der Überstand wurde verworfen, che die Resuspension mit SE-Puffer erfolgte.

Das GeneJET Genomic DNA Purification Kit wurde nach den Angaben des Herstellers verwendet. Allerdings wurde bei der Verarbeitung von Vollblut ein RNase A-Verdau nach der Inkubation mit Proteinase K solution eingefügt. Dazu wurde dem Ansatz 20 µl RNase A solution zugegeben, durch Vortexen gemischt und für 10 min bei RT inkubiert. Zur Elution der DNA wurden 50 µl ddH<sub>2</sub>O zugegeben.

Zur Isolation von gDNA aus LCLs und Fibroblasten wurde zusätzlich das *Quick-gDNA MiniPrep* nach den Angaben des Herstellers genutzt, allerdings erfolgte die Elution der DNA mit 20-50 µl ddH<sub>2</sub>O.

# b. Isolation von RNA

Um die Gesamt-RNA aus Zellen zu isolieren, wurde ein Drittel einer gut wachsenden LCL- oder primären Fibroblastenkultur abgenommen, durch Zentrifugation für 10 min bei 209 x g pelletiert und mit PBS gewaschen. Nach erneuter Zentrifugation und Verwerfen des Überstandes wurde weiter nach den Angaben des Herstellers mit dem *High Pure RNA Isolation Kit* gearbeitet. Die isolierte RNA wurde in 50 µl RNase-freiem ddH<sub>2</sub>O eluiert. Um Allele zu stabilisieren, wurden Zellen für 4,5 h mit 250 µg/ml Cycloheximid (CHX) inkubiert, bevor diese zur RNA-Isolation verwendet wurden. CHX inhibiert die Translation durch Bindung an die 60S-Untereinheit von Ribosomen und kann so u. a. zur kontrollierten Blockierung der Proteinneusynthese in der Zellkultur sowie zur Detektion kurzlebiger Proteine genutzt werden.

#### c. Isolation von Plasmid-DNA aus Bakterien

Die Extraktion von Plasmiden aus verschiedenen Bakterienkulturen basierte auf dem Prinzip der alkalischen Lyse der Zellen und Bindung der freigesetzten DNA an eine Säule aus adsorbtiven Material. Durch aufeinanderfolgende Waschschritte und Zentrifugation bzw. über Vakuum wurde die gebundene Plasmid-DNA gewaschen und schließlich in ddH<sub>2</sub>O eluiert. Die Isolation erfolgte jeweils nach den Angaben des Protokolls des Herstellers. Für analytische Zwecke wurde das *Zyppy Plasmid Miniprep Kit* verwendet, für größere Ausbeuten, beispielsweise nach der Retransformation eines Plasmides, das *Plasmid Plus Maxi Kit*.

#### 3.6.2 Aufreinigung von DNA

# a. Ethanol-Fällung

Eine klassische Methode zur Aufreinigung von DNA mit gleichzeitiger Konzentrierung stellt die Ethanol-Fällung dar. Dazu wurde ein Volumenanteil DNA-Lösung mit 2,5 Volumenanteilen 100 % Ethanol und 0,1 Volumenanteilen 3 M Natriumacetat gemischt und für 15 min bei 24 500 x g zentrifugiert. Der Überstand wurde verworfen, das Pellet mit 5 Volumen 70 % Ethanol unter denselben Bedingungen gewaschen. Der Überstand wurde vorsichtig abgenommen und verworfen. Das Pellet wurde bei RT getrocknet, bis das zurückgebliebene Ethanol verdampft war und anschließend in einem entsprechenden Volumen an Lösungsmittel, wie ddH<sub>2</sub>O oder Formamid, aufgenommen.

## b. Aufreinigung durch Bindung an eine Silica-Membran

Silica-basierte Membranen können vielfältig zur Aufreinigung von DNA verwendet werden und finden in vielen Kitsystemen Anwendung. Sie binden DNA reversibel bei hohen Salz- oder Alkohol-Konzentrationen und geben sie bei niedrigen Konzentrationen wieder ab. Die verschiedenen Kitsysteme wurden, wenn nicht anders angegeben, entsprechend der Angaben des Herstellers angewendet.

Mit Hilfe des DNA Clean and Concentrator Kit konnte DNA gleichzeitig aufgereinigt und konzentriert werden. Nach gelelektrophoretischer Auftrennung von DNA-Fragmenten, beispielsweise PCR-Produkten, wurden sie für bestimmte Anwendungen aus dem Gel ausgeschnitten. Zur Aufreinigung der DNA aus den Gelstücken wurden das GeneJET Gel Extraction Kit sowie das NucleoSpin Gel and PCR Clean-up Kit verwendet. Zudem erfolgte die Aufreinigung fragmentierter DNA nach der Nebulisierung (siehe 3.7.2) mit dem MinElute PCR Purification Kit, das ebenfalls Silica-Säulen nutzt. Später wurde bei der Probenvorbereitung für NGS das QLAquick PCR Purification Kit zur Aufreinigung der amplifizierten DNA nach der emPCR (Emulsions-PCR) eingesetzt.

#### c. Enzymatische Aufreinigung

Zur Vorbereitung von PCR-Produkten für die Sanger-Sequenzierung wurden diese enzymatisch aufgereinigt. Durch die im *ExoSAP-IT* (Affymetrix) enthaltene Exonuklease I (Exo) werden einzelsträngige Produkte, wie Primer, abgebaut. Das Enzym *Shrimp Alkaline Phosphatase* (SAP) hydrolisiert nicht eingebaute dNTPs (Desoxyribonukleosidtriphosphat). Es wurde folgender Ansatz gewählt:

PCR-Produkt	1-3 µl
ExoSAP-IT	0,5 µl
ad ddH <sub>2</sub> O	7 µl

Die Reaktion wurde für 15 min bei 37 °C im Cycler inkubiert, gefolgt von einer Hitzeinaktivierung für weitere 15 min bei 80 °C. Zur Aufbewahrung wurden die Proben bei 4 °C gelagert.

### d. Aufreinigung durch paramagnetische Beads

Neben der Ethanol-Fällung wurden Sequenzier-Reaktionen auch mit Hilfe von Magnetbeads von störenden Substanzen wie Primern, Enzymen oder Salzen befreit. Dabei bindet DNA reversibel an paramagnetische Beads und Verunreinigungen können mit Ethanol weggewaschen werden. Durch Verwendung eines Pipettier-Roboters konnte dieser Vorgang automatisiert ablaufen. Dazu wurde das *Agencourt CleanSEQ System* benutzt. Weiterhin wurden paramagnetische *Agencourt AMPure XP* Beads zur Entfernung kleinerer DNA-Fragmente nach der Nebulisierung verwendet (siehe 3.7.2). Streptavidin-gekoppelte Dynabeads M-270 wurden nach der emPCR genutzt, um die angereicherte Library zu selektieren und aufzureinigen (siehe 3.7.3).

# 3.6.3 Quantifizierung von DNA

Zur Quantifizierung von DNA wurden sowohl photometrische als auch fluorometrische Verfahren angewendet. Um gleichzeitig die Konzentration und Reinheit von DNA-Proben zu bestimmen, wurden diese am Nanophotometer gemessen. Dieses detektiert das Absorptionsmaximum einer Probe bei verschiedenen Wellenlängen. Aus dem Verhältnis der Extinktionen kann man so neben der Konzentration auch auf den Reinheitsgrad einer Probe schließen, da beispielsweise das Absorptionsmaximum von Nukleinsäuren bei 260 nm, das von Proteinen bei 280 nm liegt. Verunreinigungen können jedoch die Konzentrationswerte verfälschen.

Eine höhere Messgenauigkeit kann man mit einer fluorometrischen Messung erzielen, da hierzu Moleküle wie doppelsträngige DNA (dsDNA) indirekt über die Verwendung eines Fluorophors nachgewiesen werden. Dabei ist zunächst die Erstellung einer Standardkurve nötig, wobei bekannte Nukleinsäure-Konzentrationen, sogenannte Standards, mit dem jeweiligen Fluoreszenz-Farbstoff versetzt werden und die relative Fluoreszenz (RFU) gemessen wird. Zur Anwendung kamen das *Qubit dsDNA BR Assay Kit*, der RL *Standard* des *GS FLX Titanium Rapid Library Preparation Kits* sowie das *QuantiFluor dsDNA System*, jeweils nach den Angaben im Protokoll.

#### 3.6.4 cDNA-Synthese

Zur Synthese komplementärer DNA (cDNA) aus isolierter Gesamt-RNA (siehe 3.6.1b) wurde das Enzym *SuperScript II Reverse Transkriptase* genutzt. Die verwendeten Oligo(dT)<sub>12-18</sub>-Primer binden an das Poly(A)-Segment der mRNA, so dass diese selektiv aus dem RNA-Pool synthetisiert und amplifiziert werden kann. Dafür wurde folgender Ansatz pipettiert:

Oligo(dT) <sub>12-18</sub> -Primer	2 µl
Gesamt-RNA	20 µl
ddH <sub>2</sub> O	24 µl

Der Ansatz wurde für 10 min bei 70 °C im Cycler inkubiert. Anschließend wurde auf Eis folgendes dazu pipettiert:

5x First Strand Buffer	8 µl
DTT	4 μl
dNTPs	2 µl

Nach einer Inkubation des Ansatzes für 2 min bei 42 °C wurden direkt im Cycler 1,3 µl des Enzyms zugegeben. Der Ansatz wurde für weitere 50 min bei 42 °C inkubiert, ehe ein Inaktivierungsschritt bei 70 °C für 18 min folgte.

### 3.6.5 PCR-Amplifikation

Eine Polymerase-Kettenreaktion (*polymerase chain reaction*; PCR) ermöglicht eine selektive und exponentielle Amplifikation einer bestimmten DNA-Sequenz mit Hilfe von thermostabilen Polymerasen [Mullis und Faloona, 1987; Saiki et al., 1988]. Die Reaktion findet in der Regel in drei Schritten unterschiedlicher Temperatur statt.. Diese untergliedern sich in eine Denaturierung, um die DNA-Doppelhelix zu trennen, einen *Annealing*-Schritt, bei dem sich die Primer an komplementäre Sequenzen der DNA-Einzelstränge anlagern, sowie der Synthese durch die Polymerase. Diese verlängert die Primer durch den Einbau von dNTPs komplementär zum DNA-Strang. Eine exponentielle Vervielfältigung wird durch eine zyklische Wiederholung dieser Schritte erreicht. Vielfach müssen die Bedingungen der Reaktion angepasst werden. Dabei spielt die Schmelztemperatur der Primer oder auch die Länge des DNA-Fragments eine wichtige Rolle. Optimierungen können durch die Anzahl der Zyklen, der verwendeten Polymerase oder durch bestimmte Zusätze vorgenommen werden.

#### a. Standard-PCR

Zur Amplifikation wurden verschiedene Polymerasen mit den zugehörigen Puffern nach den Angaben des Herstellers verwendet. Bei Standard-PCRs kamen dabei die *Taq DNA Polymerase* sowie die *BIO-X-ACT Short DNA Polymerase* zum Einsatz. Weiterhin wurde die *Q5 High-Fidelity DNA Polymerase* und, vor allem für längere und GC-reiche Sequenzen, die *Phusion High-Fidelity DNA Polymerase* verwendet.

#### b. Amplifikation gesamt-genomischer DNA

Zur Amplifikation von gDNA-Proben, die von Patienten stammten, von welchen kein weiteres Zellmaterial mehr vorlag, wurde das *illustra GenomiPhi V2 DNA Amplification Kit* verwendet. Dieses beinhaltet die Phi29-Polymerase, welche eine 3'-5'-Exonuklease-Aktivität und damit eine hohe Prozessivität besitzt. Im Gegensatz zu anderen Polymerasen bricht sie die Amplifikation nicht ab, wenn sie auf ein bereits repliziertes Fragment stößt, sondern verdrängt dieses und setzt die Replikation fort. Dadurch entsteht am Ende der PCR-Reaktion ein Gemisch aus verschieden langen Fragmenten. Diese besitzen eine durchschnittliche Größe von 10 Kilobasen (kb) und decken das Genom zu 99,82 % ab [Paez et al., 2004]. Im Reaktionsansatz sind Hexamer-Primer enthalten, die sich unspezifisch an die DNA anlagern können. Die Reaktion lief wie folgt ab:

Original-DNA (50 ng/µl)	1 µl
Sample buffer	9 µl

Die Probe wurde für 3 min bei 95 °C denaturiert und dann auf 4 °C abgekühlt. Anschließend wurde auf Eis folgendes dazu pipettiert:

Reaction buffer	9 µl
Enzyme Mix	1 µl

Der Ansatz wurde für 90 min bei 30 °C inkubiert und die Exonuklease-Aktivität der Polymerase durch eine Inkubation bei 65 °C für 10 min inaktiviert.

#### c. Mutagenese und DpnI-Verdau

Zum Einbringen einer spezifischen Mutation in eine DNA-Sequenz, beispielsweise der XPF-Mutation c.1765C>T im pBlueScript-Vektor (siehe 3.2.1), wurde die Mutagenese mit Hilfe einer Two-Stage PCR nach Wang und Malcom [Wang und Malcolm, 2002] durchgeführt. Zunächst wurden zwei Extensions-Reaktionen in separaten Tubes angesetzt (Phase 1). Dabei enthielt ein Tube nur den Forward-, das andere nur den Reverse-Primer.

10x Pfu Polymerase Reaction Buffer	5 µl
100 ng Template	2 µl
10 pmol Primer	1 µl
dNTP-Mix (0,2 mM/dNTP)	1 µl
ddH <sub>2</sub> O	ad 49 µl

Nach einer Denaturierung des Ansatzes bei 95 °C für 3 min wurden je 2,5 U *PfuUltra DNA Polymerase* zugegeben. Beide Ansätze wurden im Cycler bei folgendem Programm inkubiert:

95 °C	30 sec	Г
55 °C	1 min	- 4 Zyklen
68 °C	2x Plasmidlänge in kb	
4 °C	Pause	

In Phase 2 wurden beide Reaktionsansätze gemischt und im Cycler mit obigem Programm inkubiert, allerdings nun für 16 Zyklen. Anschließend wurde die Reaktion bis zum DpnI-Verdau bei 4 °C aufbewahrt. Das Restriktionsenzym erkennt eine spezifische Sequenz und schneidet, wenn diese methyliert vorliegt, wie es für das Ausgangsplasmid der Fall ist. So bleibt am Ende der Reaktion nur das Mutationstragende Plasmid übrig. Dazu wurden 10 U DpnI zum Reaktionsansatz gegeben und für 1 h bei 37 °C inkubiert. Anschließend konnte mit der Transformation (siehe 3.2.2) fortgefahren werden.

# d. MLPA

Durch eine MLPA ist es möglich, Deletionen oder auch Duplikationen eines Gens oder einzelner Exons zu detektieren [Schouten et al., 2002]. Sie basiert auf der Hybridisierung von gDNA mit Sequenz-spezifischen Sonden, die jeweils aus zwei Oligonukleotiden bestehen und an benachbarte Sequenzen binden. Durch eine Ligation werden die Sonden zunächst verknüpft, bevor eine Amplifikation mittels Sondenspezifischer Primer stattfindet. Durch kapillarelektrophoretische Auftrennung werden die MLPA-Fragmente analysiert und die Daten mit Hilfe einer Software ausgewertet. Bei einer homozygoten Deletion kann die Ligation für kein Allel erfolgen. Daher kann kein Fragment amplifiziert und detektiert werden. Ist hingegen nur ein Allel von einer Deletion betroffen, kann nur das nicht betroffene amplifiziert werden. Bei der Auswertung mit einer entsprechenden Software erscheint dieses Fragment als reduzierte Peakfläche, im Vergleich zu einer gesunden Kontrolle.

Zur Untersuchung von Deletionen des FANCA-Gens wurde das SALSA MLPA Kit P031/P032. Um Deletionen in BRCA1 und BRCA2 auszuschließen, wurde ebenfalls eine MLPA durchgeführt, unter Verwendung des SALSA MLPA P002 probemix und des SALSA MLPA P045 probemix. Zur Auswertung wurde das Programm Coffalyser.Net genutzt.

#### 3.6.6 Größenauftrennung von DNA-Fragmenten

#### a. Gelelektrophoretische Auftrennung

Zur Größenauftrennung von DNA-Fragmenten wurde eine Gelelektrophorese durchgeführt. Die Methode beruht auf der elektrophoretischen Mobilität von Fragmenten, welche größenabhängig unter Einfluss eines elektrischen Feldes durch ein Gel wandern, das in einer Pufferlösung liegt.

Abhängig von der Fragmentgröße und der späteren Anwendung wurden 1-2 % ige TBE- oder TAE-Gele hergestellt. Dabei diente Ethidiumbromid oder *HDGreen* zum Anfärben der Nukleinsäuren. Diese Farbstoffe wurden in das noch flüssige Gel gegeben. Die Auftrennung der Proben erfolgte bei einer Spannung von 150 V. Bei jedem Lauf wurde auch ein Größenstandard zur Abschätzung der Fragmentgröße aufgetragen.

#### b. Kapillarelektrophorese

Eine weitere analytische Trennmethode von DNA-Fragmenten stellt die Kapillarelektrophorese dar. Die Trennung findet hierbei in einem dünnen Kapillarrohr in einer Elektrolytlösung statt. Zur Qualitätsprüfung der DNA-Fragmente wurden die Proben zum einen nach der Library-Vorbereitung auf *High Sensitivity DNA Chips*, zum anderen jeweils nach der LM-PCR auf *DNA 7500 Chips* am Bioanalyzer 2100 gemessen. Eine Fragment-Analyse nach einer MLPA-Reaktion wurde auf einem ABI3100 durchgeführt. Ebenso erfolgte eine kapillarelektrophoretische Auftrennung bei der Sanger-Sequenzierung.

# 3.6.7 Methoden zur Sequenzierung

### a. Sanger-Sequenzierung

Die Sequenzierung nach Sanger [Sanger et al., 1977] beruht auf der Didesoxymethode, die auch als Kettenabbruch-Methode bezeichnet wird. Für die Sequenzierung von aufgereinigten DNA-Fragmenten oder Plasmiden (siehe 3.6.2) wurde das *BigDye Terminator v1.1 Cycle Sequencing Kit* verwendet. Im Gegensatz zur PCR wird pro Ansatz nur ein Primer eingesetzt, so dass es zur kontinuierlichen Synthese eines Einzelstranges kommt. Im verwendeten Puffer sind neben dNTPs auch Fluoreszenz-markierte ddNTPs (Didesoxynukleotid-Triphosphate) enthalten. Sobald ein solches Nukleotid während der Strangsynthese eingebaut wird, kommt es zum Abbruch der Kettenverlängerung. Der Grund liegt in der fehlenden 3<sup>c</sup>-OH-Gruppe, die zur Verknüpfung mit der Phosphatgruppe des nächsten Nukleotids nötig wäre. Ein typischer Sequenzieransatz nach einer ExoSAP-Aufreinigung (siehe 3.6.2c) setzte sich wie folgt zusammen:

ExoSAP-Verdau	7 µl
Sequenzierpuffer	2 µl
BigDye	1 µl
Primer	0,5 μl

Die Aufreinigung erfolgte wie unter 3.6.2a oder d beschrieben. Die Proben wurden an einem Kapillarsequenzierer analysiert, der die Proben zum einen elektrophoretisch auftrennt und zum anderen durch Laser die Fluorochrome der ddNTPs anregt und detektiert. Die Auswertung der generierten Chromatogramm-Daten erfolgte mit Hilfe der Sequencher 4.7 Demo-Software.

# b. NGS

Durch die Entwicklung von verschiedenen Verfahren der Hochdurchsatz-Sequenzierung können parallel in einem einzigen Lauf Millionen von DNA-Fragmenten analysiert werden. Zur Sequenzierung mittels NGS wurden für diese Arbeit verschiedene Systeme und Ansätze verwendet.

Sequenzierungen mit selbstdesignten Genpanels (siehe 2.4) wurden am GS Junior vorgenommen. Bei diesem System wird darauf abgezielt, dass je ein DNA-Molekül der vorbereiteten DNA-Bibliothek an ein sogenanntes *Capture Bead* bindet. Dazu wird eine definierte Menge an DNA mit den *Capture Beads* und allen notwendigen PCR-Reagenzien in einem Emulsionsöl vermischt. In dieser Wasser-Öl-Emulsion entstehen kleine Mikroreaktoren, in welchen die gebundenen DNA-Moleküle klonal amplifiziert werden (emPCR). Im Anschluss wird die Emulsion aufgebrochen, die *Capture Beads* angereichert und mit Sequenzierprimern hybridisiert. Diese werden zusammen mit *Packing Beads* auf eine Pikotiterplatte aufzentrifugiert. Die Platte besitzt mehrere Millionen kleiner Vertiefungen, in welchen jeweils nur ein *Capture Bead* Platz findet. Dazu kommen Pyrophosphatase- sowie Enzym-Beads. Im Anschluss kann die eigentliche Sequenzierung im Gerät stattfinden, die auf dem Prinzip der Pyrosequenzierung beruht und zu den *Sequenzierby-Synthesis*-Methoden zählt. Sequenziell werden Fluoreszenz-markierte dNTPs zugegeben. Durch den Einbau eines komplementären Nukleotids durch die Polymerase wird Pyrophosphat freigesetzt. Dieses wird durch das

Enzym ATP-Sulfurylase in Adenosintriphosphat (ATP) umgesetzt, welches wiederum als Substrat für die Luziferase-vermittelte Umsetzung von Luziferin zu Oxyluziferin dient. Dabei wird Licht freigesetzt, das durch eine CCD (*charge-coupled device*)-Kamera detektiert wird. Dieses Lichtsignal ist proportional zur Anzahl der eingebauten Nukleotide. Bei Homopolymeren wird es allerdings ab einer bestimmten Länge schwierig, die Lichtintensität mit der genauen Anzahl an angebauten Nukleotiden zu korrelieren. Dies stellt eine Schwäche des Systems dar. Dem gegenüber stehen große Leseweiten von 450–550 bp.

Anreicherungen mit Hilfe des *TruSight Cancer*-Panels wurden am MiSeq sequenziert. Dies wurde nicht selbst durchgeführt. Exom-Anreicherungen wie auch -Sequenzierungen wurden von kommerziellen Dienstleistern ausgeführt (siehe 2.4). Dabei kamen sowohl die Sequenziersysteme des SOLiD 5500xl Genetic Analyzer als auch des HiSeq 2000 sowie 4000 zum Einsatz.

Die Systeme von Illumina beruhen ebenfalls auf der *Sequencing-by-Synthesis*-Methode. Die Amplifikation der Library sowie die Sequenzierung finden auf beschichteten Glasobjektträgern, sogenannten *Flow Cells*, statt. Über spezifische Adaptoren werden die Moleküle der Library kovalent auf der Glasoberfläche gebunden. Über Brücken-Amplifikation werden zunächst Cluster, Kopien desselben DNA-Fragmentes, erzeugt. Für die Sequenzierung werden modifizierte ddNTPs genutzt, die mit Fluorophoren gekoppelt sind. Pro Sequenzierzyklus kann nur ein Nukleotid eingebaut werden, da die Fluorophore als 3<sup>c</sup>-Terminatoren wirken. Nach der Signaldetektion wird das Fluorophor abgespalten, so dass die Base für den nächsten Synthesezyklus zugänglich wird. Mit dem MiSeq lassen sich so maximal 15 Gb, mit dem HiSeq 2000 1 500 Gb an Sequenzieren, die eine maximale Länge von 2 x 300 bp bzw. 2 x 150 bp aufweisen.

Die klonale Amplifikation des SOLiD-Systems ähnelt der emPCR. Ein maßgeblicher Unterschied besteht allerdings in der Art der Sequenzierung. Diese findet, wie bei den Illumina-Systemen, auf einem Glasobjektträger statt, beruht jedoch auf dem Prinzip der Ligation. Zunächst bindet ein Sequenzierprimer, bevor eine Mischung aus 3<sup>c</sup>-Fluoreszenz-markierten 8-mer-Oligonukleotiden zugegeben wird. Wenn diese an die DNA binden, werden sie an den Primer ligiert und das Fluoreszenz-Signal wird von einem Laser detektiert. Die ersten beiden Basen des Oligonukleotids definieren dabei die Farbe. Es wird also zunächst ein Farbcode generiert, der erst später in eine Nukleotidsequenz übersetzt wird. Das Oligonukleotid wird dann zwischen der fünften und sechsten Base gespalten, so dass das Fluorophor entfernt wird und in einer nächsten Runde eine weitere 8-mer-Sonde binden und daran ligiert werden kann. Nach einigen Zyklen werden der Universalprimer und die daran ligierten Sonden durch Aufschmelzen der Bindungen von der Matrize getrennt. Es folgen Zyklen mit weiteren Universalprimern, die jeweils um eine Base verschoben sind (n-1, n-2, n-3, usw.). So wird jede Base zweimal sequenziert und mit zwei unabhängigen Fluoreszenz-Signalen codiert. Es werden große Datenmengen generiert, allerdings ist die Leseweite mit maximal 85 bp vergleichsweise kurz.

# 3.7 Spezielle Methoden: Anreicherung mit FA-Genpanels und NGS

# 3.7.1 Sondendesign zur Erstellung eines FA-Genpanels

Zur Anreicherung von FA- und ausgewählten Kandidaten-Genen wurde eine *in-solution* Anreicherung durch das *SeqCap* EZ-System von Roche NimbleGen gewählt. Entsprechend der Angaben im Protokoll zur Sondenerstellung ("*NimbleGen Sequence Capture Custom Designs Guide to Submitting Your Target Sequence; NimbleGen*"; Roche NimbleGen) wurde die *Gene Sorter-* und *Table Browser*-Funktion des UCSC-Browsers genutzt, um eine bed-Datei mit den Koordinaten der Zielregion zu erstellen. Diese umfasste zunächst die codierenden Exons der Zielgene, plus je 60 bp *up-* und 10 bp *downstream* davon, welche durch überlappende Sonden abgedeckt wurden (*SeqCap* EZ *Choice Library*). Später wurde die Sondendichte in schlecht abgedeckten Regionen erhöht sowie bekannte intronische Mutationen in das Design aufgenommen (*SeqCap* EZ *Developer Library*). Diese beinhalteten eine beschriebene pathogene, intronische Insertion eines Alu-Elementes in Intron 4 von *FANCD2* (c.274-57\_-56insinvAluYb8nt36\_319+dup c.274-69\_-57). Diese wurde durch Sanger-Sequenzierung bestimmt und zusammen mit der entsprechenden Wildtypsequenz zum Sondendesign übermittelt. Nach der Erstellung von potentiellen Sonden wurden diese, entsprechend der Angaben im Protokoll ("*Guide for Reviewing and Approving Custom Designs - NimbleGen; Sequence Capture Array and SeqCap EZ Choice Library*"; Roche NimbleGen) überprüft, bevor sie zur Herstellung freigegeben wurden.

Die Anzahl an ausgewählten Kandidatengenen variierte dabei je nach Kenntnisstand, wie auch die Anzahl an bekannten FA-Genen. Eine Übersicht über die verschiedenen Designs ist in Tabelle 3 dargestellt.

**Tabelle 3:** Übersicht verschiedener Kriterien, die zur Erstellung der Sondenlibraries berücksichtigt wurden. Eine bestimmte Mutation (Insertions-Mutation eines Alu-Elementes) in *FANCD2* wurde speziell angereichert oder auch bekannte pathogene Mutationen weit im Intron in die Anreicherung eingeschlossen. Bei allen Designs wurden die codierenden Exons ausgewählter Gene sowie je 60 bp 5' und 10 bp 3' davon angereichert. "\*": Anzahl der zum Zeitpunkt der Designerstellung bekannten FA-Gene; "x": im Design enthalten; "-": nicht im Design berücksichtigt.

Sonden-Library	OID38614	OID41218	OID42419	OID43014
Referenzgenom (UCSC-Version)	GRCh37 (hg19)	GRCh37 (hg19)	GRCh38 (hg38)	GRCh37 (hg19)
FA-Gene*	15	16	17	19
Kandidatengene	40	39	43	41
FANCD2 Alu-Mutation	-	Х	х	х
Intronische Mutationen	-	Х	Х	Х

#### 3.7.2 Probenvorbereitung zur Anreicherung

## a. Erstellung einer Rapid Library

Die Fragmentierung und Vorbereitung der Patienten-DNA zur Sondenhybridisierung orientierte sich am Rapid Library Preparation Method Manual von Roche und ist in Abbildung 9 zur Übersicht schematisch dargestellt. Je zwei DNA-Proben wurden parallel angereichert. Die eingesetzte DNA-Menge variierte zwischen 500 ng und 3 µg. Bei der Herstellung einer Probenlibrary wurde mittels Nebulisierung doppelsträngige DNA durch hohen Druck des eingeleiteten inerten Stickstoff-Gases (30 psi/2,1 bar) zufällig in Stücke von 50–900 bp fragmentiert.



Abbildung 9: Übersicht über die Herstellung einer Rapid Library. Die genomische DNA wird zunächst durch Nebulisierung in zufällig große Stücke fragmentiert (oben rechts). Nach der Erzeugung von *blunt ends* werden die 5'-Enden der DNA-Fragmente phosphoryliert (P), wohingegen den 3'-Enden ein A-Überhang angefügt wird (unten links). Anschließend können die MID-Adapter (rot und grün) ligiert werden (unten rechts). Diese enthalten, neben einer spezifischen MID-Sequenz, die Schlüsselsequenz für Rapid Libraries (GACT) sowie die Adaptersequenzen A und B, welche die Amplifikation und Sequenzierung der Fragmente ermöglichen. Durch einen Fluoreszenzfarbstoff (FAM) ist zudem eine Quantifizierung der Library möglich.

Nach einer Aufreinigung (siehe 3.6.2b) werden zunächst die Enden der DNA-Fragmente für die Ligation mit den MID (*Multiplex Identifier*)-Adaptern zum Multiplexing vorbereitet. Eine T4 DNA-Polymerase erzeugt zuerst *blunt ends.* Im Anschluss werden die 5'-Enden der DNA-Fragmente unter Verwendung einer Polynukleotid-Kinase phosphoryliert und den 3'-Enden durch die Taq-Polymerase ein Adenosin-Überhang angefügt, woran die Y-förmigen doppelsträngigen MID-Adapter während des folgenden Schrittes ligiert werden können. Diese Adaptersequenzen bestehen zum einem aus einer spezifischen 11 bp langen MID-Sequenz, welche eine Unterscheidung der gepoolten Proben nach der Sequenzierung erlaubt. Weiterhin ist eine Schlüsselsequenz (GACT für Rapid Libraries) enthalten. Diese unterscheidet die Library von den bei der Sequenzierung mitgeführten internen Kontrollsequenzen, welche durch die Kontroll-Beads eingeführt werden (CATG für Kontroll-Typ I, ATGC für Kontroll-Typ II). Schließlich sind die 454-Adapter A und B enthalten, deren Sequenzen sich nur geringfügig unterscheiden und komplementär zu den Primern der LM-PCR (*Ligation-mediated* PCR) und Sequenzierung sind. Somit können beide Stränge eines Fragmentes sequenziert werden. Zusätzlich befindet sich am 5'-Ende der Adapter-A-Sequenz ein Fluoreszenzfarbstoff, der eine Quantifizierung der Library nach deren Aufreinigung ermöglicht.

Durch eine Aufreinigung der Proben mit SPRI (*Solid Phase Reversible Immobilisation beads*)-Beads (siehe 3.6.2d) werden im Anschluss kleine Fragmente entfernt, um eine optimale Größenverteilung der Fragmente von etwa

600–900 bp zu erreichen. Die Elution der DNA-Fragmente von den dazu verwendeten SPRI-Beads erfolgte mit ddH<sub>2</sub>O statt, wie im Protokoll angegeben, mit TE-Puffer, da das darin enthaltenen EDTA mit Mg<sup>2+</sup> des PCR-Puffers Komplexe bilden und somit die Amplifikation während der LM-PCR beeinträchtigen würde.

Zur Qualitätsbestimmung der fragmentierten DNA-Libraries wurden die Proben entweder auf *High Sensitivity DNA Chips* (Agilent) am Bioanalyzer gemessen oder auf einem 1,2 %igem TBE-Gel aufgetrennt. Letzteres geschah analog zu den Angaben im Protokoll, das die Benutzung des *FlashGel Systems* (Lonza) beschreibt. Es wurden je 5 µl Library und zur Größenabschätzung ebenfalls 5 µl 100 bp- sowie 1 kb-Ladder aufgetragen. Die Auftrennung erfolgte bei 150 V für etwa 25-30 min.

Die Quantifizierung der Library erfolgte am QuantiFluor-ST. Eine weitere Abweichung vom Protokoll bestand darin, dass die Libraries nach der Quantifizierung am QuantiFluor-ST-Fluorometer mit ddH<sub>2</sub>O auf eine RFU von etwa 150 eingestellt wurden, was etwa einer Konzentration von 1\*10<sup>9</sup> Molekülen/µl entspricht, statt durch Verdünnungen mit TE-Puffer Arbeitslösungen mit 1\*10<sup>7</sup> Molekülen/µl herzustellen.

# b. LM-PCR und Sonden-Hybridisierung

Zunächst wurde nach den Angaben im NimbleGen SeqCap EZ Library LR User's Guide (Version 2.0) verfahren. Aufgrund unzureichender Amplifikation wurde, nach Rücksprache mit Roche, ein abgeändertes Protokoll erarbeitet. Dieses orientierte sich im Wesentlichen am Double Capture-Protokoll (NimbleGen SeqCap EZ Rapid Library Small Target Capture LR), welches als alternatives Protokoll für schwierige oder kleinere Targets erschienen ist. Ein Vergleich der wesentlichen Schritte in beiden Protokollen ist in Abbildung 10 zur Übersicht dargestellt. Weiterhin werden die Abänderungen im Folgenden beschrieben.

Zunächst wurden je 50 µl der unter 3.7.2a vorbereiteten Proben für eine LM-PCR zur Amplifikation der Proben-Library eingesetzt. Die verwendeten Primer binden dabei spezifisch an die zuvor an die Fragmente ligierten Adapter. Nach Zugabe von je 50 µl LM-PCR-Mastermix wurde der Ansatz durch Pipettieren gemischt und im Cycler inkubiert. Die Konzentration der Primer wurde von 40 µM auf 10 µM reduziert. Der LM-PCR-Ansatz setzte sich pro Reaktion folgendermaßen zusammen:

FastStart High Fidelity Buffer w/18 mM MgCl <sub>2</sub>	10 µl
DMSO	2 µl
PCR Grade Nucleotide Mix	2 µl
10 μM 454 Rapid-A Oligo	10 µl
10 μM 454 Rapid-B Oligo	10 µl
PCR grade water	15 µl
FastStart High Fidelity Enzyme Blend	1 µl

Das PCR-Programm für die sogenannte Pre-Capture LM-PCR-Reaktion war wie folgt:



Die darauf folgenden Schritte waren bis zur Vorbereitung der Proben für die Hybridisierung mit den Angaben im *User's Guide* identisch. Vor dem Hybridisieren wurde allerdings die Konzentration der amplifizierten Proben-Library zusätzlich am Qubit gemessen und mit den dabei bestimmten Werten weitergearbeitet. Je 500 ng Proben-Library wurden mit 2,25 µl statt 4,5 µl der jeweiligen *SeqCap EZ*-Library zur Hybridisierung verwendet Das fehlende Volumen wurde mit ddH<sub>2</sub>O aufgefüllt und die Proben über Nacht bei 47 °C im Cycler inkubiert. Am nächsten Tag wurde die PCR-Reaktion mit Hilfe von Streptavidin-Dynabeads aufgereinigt und die Proben einer weiteren LM-PCR unterzogen. Der Ansatz sowie



Abbildung 10: Vergleich der Abläufe des Standard-Protokolls mit dem *Double Capture*-Protokoll.

das PCR-Programm waren mit der vorherigen Reaktion identisch, allerdings wurde die Zyklenzahl auf fünf verringert. Danach wiederholte sich die Hybridisierung sowie die Aufreinigung der Probe ein weiteres Mal, bis sich, analog zum *User's Guide*, eine weitere LM-PCR-Reaktion mit 15 Zyklen anschloss. Die quantitative PCR entfiel.

## 3.7.3 Klonale Amplifikation durch emPCR

Die emPCR dient der klonalen Amplifikation der DNA-Library als Vorbereitung zur Sequenzierung. Dazu gibt man die einzelsträngige DNA-Library zu einem Überschuss an *Capture Beads*. Zusammen mit den PCR-Reagenzien wird diese Mischung in ein Emulsionsöl gegeben. Durch Emulgieren entstehen so etwa 10<sup>6</sup> Mikroreaktoren pro ml, in welchen die PCR ablaufen kann. Die Bindung des DNA-Moleküls an ein *Capture Bead* wird dabei über Oligonukleotide vermittelt, welche die komplementäre Sequenz zur Adapter-B-Sequenz aufweisen und mit der Oberfläche des *Capture Beads* verbunden sind. Nach der Amplifikation werden durch Aufbrechen der Emulsion und Denaturierung die DNA-gekoppelten Beads der Aufreinigung und Anreicherung zugänglich gemacht (*Bead recovery*). Dabei nutzt man Biotin-gekoppelte Primer, welche an die freien Enden der Adapter-A-Sequenzen binden können. Durch Zugabe von Streptavidin-gekoppelten Magnetbeads können so über einen Magneten die DNA-Beads isoliert werden, welche eine ausreichende

Anzahl an DNA-Molekülen tragen, was rund 10% der eingesetzten *Capture Beads* entspricht. Eine Übersicht der einzelnen Schritte ist in Abbildung 11 dargestellt.



Abbildung 11: Übersicht der emPCR zur klonalen Amplifikation der DNA-Library. *Capture Beads* (braun) werden mit der vorbereiteten, denaturierten Probenlibrary sowie den PCR-Reagenzien und dem Emulsionsöl gemischt. Im Wasser-Öl-Gemisch bilden sich dabei kleine Mikroreaktoren aus. Über die Adapter-B-Sequenz kann das DNA-Molekül komplementär an die Linker (hellgrün) der *Capture Beads* binden und wird durch die Polymerase (gelb) vervielfältigt. Nach dem Aufbrechen der Emulsion werden die DNA-*Capture Beads* aufgereinigt und angereichert. Nach einem Denaturierungsschritt können Biotin-gekoppelte Enrichment-Primer (orange) an die Adapter-A-Sequenzen binden, an welche wiederum Streptavidin-gekoppelte magnetische *Enrichment Beads* (blau) binden.

Die Vorbereitung und Durchführung der emPCR richtete sich nach den Angaben im *emPCR Amplification Method Manual (Lib-L).* Zunächst wurde die Konzentration der angereicherten Library mit dem *QuantiFluor dsDNA System* fluorometrisch bestimmt. Mit Hilfe folgender Formel konnte schließlich die Anzahl der Moleküle/µl in der angereicherten Library berechnet werden:

$$Moleküle/\mu l = \frac{Konzentration \left[\frac{ng}{\mu l}\right] \times 6,022 \times 10^{23} \left[\frac{Moleküle}{mol}\right]}{656,6 \times 10^9 \left[\frac{g}{mol}\right] \times Amplikonlänge [bp]}$$

Die Formel berücksichtigt dabei neben der gemessenen Konzentration auch die Avogadro-Konstante, welche die Anzahl der Moleküle/mol angibt und mit 6,022 x 10<sup>23</sup> angegeben ist. Weiterhin fließt im Nenner das durchschnittliche Molekulargewicht von Nukleotidpaaren sowie die Amplikonlänge ein. Ausgehend von der berechneten Anzahl an Molekülen/µl wurden Verdünnungsreihen hergestellt, um Ausgangslösungen mit einer Konzentration von 1\*10<sup>6</sup> bzw. 1\*10<sup>5</sup> Molekülen/µl für die emPCR zu erhalten.

Um anschließend das Volumen zu bestimmen, welches von der angereicherten Library für die emPCR einzusetzen ist, wurde folgende Formel verwendet:

$$\mu l \text{ DNA Library} = \frac{\text{gewünschte Anzahl an Molekülen pro Bead} \times (10 \times 10^6 \text{ Beads})}{\text{Konzentration der Library}} \frac{\text{Moleküle}}{\mu l}$$

Die Zentrifugationen zum Waschen der Beads direkt nach der emPCR-Reaktion wurden, anders als angegeben, bei 906 x g für jeweils 6 min durchgeführt.

# 3.7.4 Sequenzierung am GS Junior

Die Vorbereitungen zur Sequenzierung fanden größtenteils gemäß der Beschreibung im *Sequencing Method Manual* statt. Abweichungen gab es in der Menge an zugegebenem Puffer BB2 zu den gewaschenen *Packing Beads.* Statt 200 µl wurden 220 µl zugegeben. Ebenso erhöhten sich die Volumina bei Abschnitt 3.2.3.4 des Manuals. Der Ansatz änderte sich wie folgt:

Polymerase	44 µl
Polymerase Cofactor	22 µl
BB2	72 µl

Die Menge an *Packing Beads*, die zu oben dargestellten Ansatz gegeben wurde, erhöhte sich von 175 auf 185 µl. Zur Sequenzierung wurden folgende Parameter ausgewählt:

Run Group	Applications II
Number of cycles	200 cycles, approximately 500 bases
Run processing type	Full processing for Rapid, Paired End or cDNA Rapid libraries

#### 3.7.5 Auswertung von NGS-Daten

Eine erste Auswertung der GS Junior-Sequenzdaten erfolgte direkt am Sequenzierer mit dem integrierten GS Run Browser. Im Falle des MiSeq stand dafür die MiSeq-Reporter-Software zur Verfügung. Mit Hilfe beider Programme konnten so erste Rückschlüsse sowohl auf die Qualität des Laufes als auch der generierten Daten gezogen werden. Im Fall des GS Run Browsers lassen verschiedene Filter auf Probleme während der Anreicherung oder des Sequenzierlaufes selbst schließen. Dies wird beispielsweise durch die mitgeführten Kontrollen möglich. So sind in den Kontrollen des Typs II (Schlüsselsequenz ATGC) DNA-Sequenzen enthalten, die so designt sind, dass bei jedem Nukleotid-Fluss mindestens ein Signal erzeugt wird. Besondere Bedeutung kommt der Filter-Registerkarte zu, die eine Statistik über Reads angibt, welche aufgrund mangelnder Qualität ausgesondert wurden.

Die Daten aller NGS-Sequenzierungen wurden anschließend mit der Analyse-Software NextGENe von Softgenetics ausgewertet. Dazu mussten die Rohdaten zunächst in ein FASTA-Standardformat konvertiert werden. Im Falle von Multiplex-Proben konnte diesem Schritt eine Zuordnung der Reads anhand ihres MID-Barcodes voran geschaltet werden. Die Software schloss Reads mit niedriger Qualität von weiteren Analysen aus. Nach der Konvertierung wurden die Daten gegen das entsprechende Referenzgenom alignt und konnten schließlich analysiert werden. Durch die Auswahl verschiedener Filter konnten z. B. stille Mutationen aus der

Analyse ausgeschlossen werden oder, bei bekannter Verwandtschaft der Eltern eines Patienten, zunächst ausschließlich homozygote Mutationen betrachtet werden. Durch die Auswahl einer ROI, die beispielsweise den Koordinaten der Zielregionen des SeqCap EZ-Panels entsprach, konnten spezifisch nur diese Regionen angezeigt und ausgewertet werden. Falsch gemappte Reads wurden so nicht weiter in Betracht gezogen. Um eine Filterung der Varianten vorzunehmen, wurde die Frequenz zur Nachweisbarkeit von Varianten auf 15 % gesetzt. Diese wurde nur dann weiter erniedrigt, wenn keine potentiell pathogene Mutation bestimmt werden konnte. Im Fall der GS Junior-Sequenzierungen wurde eine minimale Abdeckung von 1x, bei allen anderen Sequenzierungen von 10x gewählt. Synonyme Mutationen wurden nur dann näher analysiert, wenn sie an essenziellen Spleißstellen lagen. Trunkierende oder Frameshift-Mutationen wurden zunächst als pathogen eingestuft und ebenso wie Missense-Mutationen oder potentielle Spleißmutationen einer in silico-Analyse unterzogen. Dazu wurde in erster Linie das Softwareprogramm Alamut genutzt, welches auf verschiedene Datenbanken, Mutations- (Align-GVGD, MutationTaster, SIFT) sowie Spleißvorhersageprogramme (u. a. MaxEntScan, NNSPLICE, ESEfinder) zurückgreift. Weiterhin werden Daten zur Allelfrequenz angegeben und es bestehen Links zu weiteren Programmen wie PolyPhen-2- oder dem ExAc-Browser. Varianten, die nach dieser Analyse als pathogen eingestuft wurden, wurden anschließend per Sanger-Sequenzierung auf gDNA-Ebene validiert. Ferner wurden potenzielle Spleißmutationen auf cDNA-Ebene auf einen möglichen Effekt hinsichtlich des aberranten Spleißverhaltens untersucht, sofern entsprechendes Patientenmaterial vorhanden war.

# 3.8 Proteinanalytische Methoden

#### 3.8.1 Proteinextraktion

Um die Monoubiquitinierung des FANCD2-Proteins zu induzieren, wurden Zellen für 16 h mit 40 ng/µl MMC inkubiert. Die Zellen wurden gegebenenfalls abgelöst und bei 209 x g für 8–10 min sedimentiert, das Pellet mit kaltem PBS gewaschen und bei 4 °C erneut zentrifugiert. Der Überstand wurde verworfen und das Gewicht des Pellets bestimmt. Pro 125 mg Proteinpellet wurden 50 µl *IP Lysis/Wash Buffer* sowie *Protease Inhibitor Cocktail* im Verhältnis 1:100 dazu gegeben und das Pellet resuspendiert. Die Suspension wurde 5 min auf Eis inkubiert und währenddessen gelegentlich gemischt. Anschließend wurde die Probe bei 16 000 x g für 10 min bei 4 °C zentrifugiert und der Überstand in ein neues, vorgekühltes Eppi überführt.

Zur Extraktion von cytoplasmatischen Proteinen wurde das NE-PER Nuclear and Cytoplasmic Extraction Kit verwendet. Nachdem der cytoplasmatische Extrakt isoliert und in ein neues Eppi überführt worden war, wurde mit dem Zellpellet die Extraktion der Zellkern- und Chromatin-spezifischen Proteine fortgeführt. Dazu wurde das Subcellular Protein Fractionation Kit nach den Angaben des Herstellers angewendet. Der dabei isolierte Membranextrakt wurde für weitere Analysen nicht benötigt und daher verworfen.

## 3.8.2 Quantifizierung von Proteinextrakten

Die Konzentrationsbestimmung von Proteinen fand analog zur DNA-Quantifizierung (siehe 3.6.3) statt. Zur fluorometrischen Messung wurde hier das *Qubit Protein Assay Kit* verwendet.

# 3.8.3 SDS-Gelelektrophorese

Die denaturierende gelelektrophoretische Auftrennung dient der Analyse von Proteinen. Zur Probenvorbereitung wurden die Extrakte (30-50 µg) zunächst mit 4x NuPAGE LDS Sample Buffer sowie 10x NuPAGE Sample Reducing Agent gemischt und für 10 min bei 70 °C denaturiert. Nach einer Abkühlung wurden die Proben auf kommerzielle Minigele aufgetragen. Die Auftrennung erfolgte in einer, mit den entsprechenden Puffern gefüllten, Elektrophorese-Kammer. Dabei war die Kammer an ein Kühlsystem angeschlossen. Die Elektrophorese fand bei 150 V für einige Stunden oder über Nacht bei 70-80 V statt, bis die gewünschte Auftrennung erreicht war. Zur Abschätzung der Größen wurde immer ein vorgefärbter Proteinstandard auf das Gel mit aufgetragen.

### 3.8.4 Proteintransfer

Der Transfer des aufgetrennten Proteingemisches auf eine PVDF-Membran erfolgte per Elektrotransfer entweder über das *Tank-Blot-* oder das *Dry-Blot-*System.

Beim *Tank-Blot*-System wurde zunächst die Membran kurz in 100 % Methanol aktiviert und in Transferpuffer überführt. Ebenso wurden das Gel sowie alle verwendeten Schwämme und Filterpapiere in Transferpuffer äquilibriert bzw. befeuchtet. Der Transfer mit angeschlossener Kühlung erfolgte entweder bei 75 V für mindestens 2 h oder bei 30 V über Nacht. Beim *Dry-Blot* wurde das iBlot2 Geltransfer-System verwendet, wobei die enthaltene PVDF-Membran bereits aktiviert und gebrauchsfertig ist.

### 3.8.5 Immunblot-Analyse

Nach dem Transfer müssen zunächst die freien Bindungsstellen auf der Membran blockiert werden, um die spezifische Bindung eines Antikörpers an das entsprechende Antigen zu ermöglichen. Dazu wurde die Membran für 1 h bei RT in 5% Milch/PBS-T inkubiert. Danach folgte die Inkubation mit dem Primärantikörper für mindestens 1 h bei RT oder bei 4 °C über Nacht. Nach dreimaligem Waschen mit PBS-T wurde ein HRP-gekoppelter Sekundärantikörper für 1 h bei RT auf die Membran gegeben. Anschließend wurde die Membran vier Mal für mindestens 15 min in PBS-T gewaschen. Zur Detektion der gebundenen Antikörper mussten die beiden Lösungen des *Immobilon Western Chemiluminescent HRP Substrats*, eine Peroxidase und ein Luminol, zu gleichen Volumina gemischt und auf die Membran aufgetragen werden. Die Membran wurde durch das Auflegen einer Folie gegen Austrocknung geschützt. Die Detektion der Chemilumineszenz erfolgte an einem Imaging-System über eine integrierte CCD-Kamera.

# 3.8.6 Entfernen von Antikörperbindungen an Membranen

Um gebundene Antikörper nach der Immundetektion von der Membran zu entfernen, damit diese für weitere Analysen mit anderen Antikörpern verwendet werden konnte, wurde sie für 15 min bei 37 °C mit Restore Immunblot Stripping Buffer inkubiert. Nach einem Waschschritt mit PBS-T wurde die Membran wieder in 5 % Milch/PBS-T für 30 min bis 1 h bei RT oder über Nacht bei 4 °C inkubiert. Erneute Immundetektionen konnten folgen.

# 4. Ergebnisse

Insgesamt wurden 55 verschiedene FA-Patienten mit vier unterschiedlichen Versionen der *SeqCap EZ*-Library analysiert. Mit Hilfe des *TruSight Cancer*-Panels wurden 19 und mittels WES sechs Patienten auf pathogene Mutationen in FA-Genen untersucht. Zunächst soll die Etablierung des selbst erstellten FA-Genpanels und die dabei erzielten Ergebnisse erläutert werden, ehe Alternativen in Form des kommerziellen Krebspanels sowie die Resultate der Gesamtexom-Sequenzierung gezeigt werden. Weiterhin schließt sich eine Charakterisierung bestimmter Patienten bzw. Komplementationsgruppen an.

# 4.1 Etablierung eines FA-Genpanels

# 4.1.1 Abdeckung der ROI durch Sonden

Zur Anreicherung wurden insgesamt vier verschiedene SeqCap EZ-Panels erstellt, welche unterschiedliche Zielregionen umfassten. Auf einige Charakteristika dieser Sondenlibraries soll im Folgenden eingegangen werden.

Bei der Erstellung des ersten Designvorschlages wurden zunächst stringente Bedingungen zur Auswahl der Sonden gewählt, so dass eine Sonde nur exakt einmal im Genom an der entsprechenden Zielsequenz binden durfte. Die so generierten Sonden wurden im Anschluss *in silico* auf die Abdeckung der Zielregionen hin überprüft. Für die *SeqCap EZ Choice Library* ergaben sich dabei 37 Sequenzabschnitte, welche weder direkt noch durch flankierende Sonden abdeckbar waren und später vermutlich nicht hätten angereichert werden können. Dies entsprach knapp 5% der gesamten ROI (Tabelle 4). Darunter befanden sich auch ein Kandidatengen (*FOXO3*), welches unter diesen Bedingungen komplett bei der Anreicherung gefehlt hätte, wie auch acht Exons des FA-Gens *FANCD2*.

**Tabelle 4:** Vergleich der möglichen Abdeckung bei unterschiedlichen Vorgaben zur Erstellung der SeqCap EZ Choice Library (OID38614). Unter stringenten Bedingungen waren deutlich mehr Basen der Zielregionen nicht abgedeckt als wenn die Sonden an bis zu maximal fünf Regionen im Genom binden durften (maximum matches 5; mm5). "Offset" gibt die Verlängerung der Sonden auf beiden Seiten an, um eine zusätzliche Abdeckung der Targets durch die Sequenzierung zu simulieren. Kleinere Zielregionen wurden auf eine minimale Länge von 100 bp verlängert, um eine effektive Auswahl an Sonden zu ermöglichen. Überlappende Regionen wurden zusammengefasst. Daraus ergibt sich die Anzahl an Zielregionen.

Parameter	OID38614 (stringent)		OID38614 (mm5)	
Offset [Basen]	0	100	0	100
Zielregionen	858	858	858	858
Target-Basen	224 408	224 408	224 608	224 608
Abgedeckte Target-Basen	213 495	218 483	223 088	224 512
Abgedeckte Target-Basen [%]	95,14	97,36	99,32	99,96
Nicht abgedeckte Target-Basen	10913	5 925	1 520	96
Nicht abgedeckte Target-Basen [%]	4,86	2,64	0,68	0,04

Eine Nachbesserung dieses ersten Designvorschlages erlaubte schließlich, dass eine Sonde an bis zu fünf Stellen im Genom binden konnte, sofern für eine Region unter stringenteren Bedingungen keine Sonden erstellt werden konnten (*maximum matches 5*; mm5; Tabelle 4). Dies bedeutet, dass pro Sonde bis zu fünf Insertionen, Deletionen oder Fehlpaarungen erlaubt waren. Dadurch blieben lediglich noch vier Regionen bestehen, welche nicht direkt durch Sonden abgedeckt werden konnten. Bei drei dieser Regionen war die Möglichkeit gegeben, diese durch Sonden in unmittelbarer Nähe indirekt abzudecken. Für die vierte Region, ein alternatives Exon des *BRCA1*-Gens (NM\_007300), wurde der Target-Abschnitt verlängert und speziell Sonden für die flankierenden Bereiche der eigentlichen Zielsequenz generiert, um eine indirekte Abdeckung zu ermöglichen. Abbildung 12 zeigt die Abdeckung von *BRCA1* mit Sonden in der Ansicht des UCSC-Browsers sowie das beschriebene alternative Exon. Weiterhin wird anhand der Abbildung deutlich, dass im Bereich des beschriebenen alternativen Exons von *BRCA1* ein Pseudogen (*PGOHUM00000250838*) gelistet ist.



**Abbildung 12:** Darstellung der Sondenverteilung der SeqCap EZ Choice Library (OID38614) im mm5-Design in der Ansicht des UCSC-Browsers. Sonden sind als schwarze Balken dargestellt. Darunter sind die Exons (blaue Balken) der verschiedenen Isoformen von BRCA1 abgebildet. Das Referenztranskript (NM\_007294) ist blau hinterlegt (links). Pfeilspitzen geben die Leserichtung des Genes an. Das nicht direkt durch Sonden abdeckbare alternative Exon sowie die erstellten flankierenden Sonden sind rot hinterlegt. Das in dieser Region gelistete Pseuodgen PGOHUM0000250838 ist als roter Balken unter den Transkripten dargestellt. Darunter befindet sich der GC-Gehalt der Region, dargestellt in Prozent, sowie die Position von Repeat-Elementen.

Für alle weiteren Designs wurde ebenfalls die nicht-stringente Variante gewählt. Bei den letzten beiden Designs waren dabei 20 (mm20) statt 5 (mm5) maximale Fehlpaarungen seitens NimbleGen vorgegeben. Eine Übersicht über die entsprechende Abdeckung dieser Libraries ist in Tabelle 5 zusammengefasst. Daran ist abzulesen, dass die Anzahl der Zielbasen von Design zu Design jeweils leicht anstieg, da unterschiedliche Zielregionen gewählt worden waren. Allen vier Panels war ein Genset aus 47 Genen gemeinsam (siehe Anhang). Hinzu kamen unterschiedliche Kandidaten- bzw. neue FA-Gene. Die angereicherten Gene waren jeweils zwischen OID38614 und OID41218 sowie zwischen OID42419 und OID43014 identisch.
**Tabelle 5:** Vergleich der möglichen Abdeckung bei unterschiedlichen Vorgaben zur Erstellung der einzelnen SeqCap EZ Developer Libraries. Bei der Auswahl der Sonden waren bis zu fünf (mm5) bzw. 20 (mm20) Fehlpaarungen erlaubt Unter "Offset" ist die Anzahl an Basen angegeben, um welche die Sonden auf beiden Seiten verlängert wurden, um die zusätzliche Abdeckung der Targets durch die Sequenzierung zu simulieren. Kleinere Zielregionen wurden auf eine minimale Länge von 100 bp verlängert, um eine effektive Auswahl an Sonden zu ermöglichen. Überlappende Regionen wurden zusammengefasst. Daraus ergibt sich die Anzahl an Zielregionen.

Parameter	OID41218 (mm5)		OID42419 (mm20)		OID43014 (mm20)	
Offset in Basen	0	100	0	100	0	100
Zielregionen	872	872	903	903	900	900
Target-Basen	227 257	227 257	227 645	227 645	228 688	228 688
Abgedeckte Target-Basen	225 178	226 932	223 553	227 186	224 936	228 403
Abgedeckte Target-Basen [%]	99,09	99,86	98,20	99,80	98,40	99,90
Nicht abgedeckte Target-Basen	2 079	325	4 092	459	3 752	285
Nicht abgedeckte Target-Basen [%]	0,91	0,10	1,80	0,20	1,60	0,10

Die Developer-Libraries unterschieden sich von der Choice-Library durch die Aufnahme verschiedener intronischer Mutationen, welche als pathogen beschrieben wurden. Dazu wurde die FA-Mutationsdatenbank (Rockefeller-Universität) auf solche Varianten hin näher analysiert. Aktuell (Stand Januar 2017) sind in darin über 900 individuelle Mutationen gelistet, wovon rund ein Viertel große Insertionen oder Deletionen ausmachen. Substitutionen oder kleinere Indel-Mutationen im Intron haben mit 16 % den geringsten Anteil, der Rest entfällt auf Mutationen in codierenden Exons. Zur genaueren Analyse der in der FA-Mutationsdatenbank aufgeführten intronischen Mutationen wurden diese auf den jeweiligen Abstand zum Exon untersucht. Abbildung 13 stellt die Verteilung dieser Abstände grafisch dar.



Abbildung 13: Häufigkeitsverteilung intronischer Mutationen in FA-Genen. Die Varianten wurden nach ihrem Abstand zum Exon in Klassen aufgeteilt und entsprechend der Häufigkeit aufgetragen.

Anhand des Histogramms ist zu erkennen, dass der Hauptteil der Mutationen die kanonischen Spleißdonorund -akzeptorstellen, je zwei Basenpaare 3' bzw. 5' des Exons, betrifft. Die Einstellungen für das Sondendesign waren jeweils so gewählt, dass zusätzlich zu den codierenden Exons je 60 bp *upstream* und 10 bp *downstream* enthalten waren. Damit waren per se bereits 85 % aller intronischen Mutationen enthalten. Dieser Wert bezieht sich auf den aktuellen Kenntnisstand und kann als Größenordnung für alle Panels angesehen werden. Bei der am weitesten im Intron liegenden Mutation, die zur Anreicherung aufgenommen wurde, handelt es sich um die Basensubstitution c.375-2003C>G in *FANCL*, welche aberrantes Spleißen zur Folge hat [Chandrasekharappa et al., 2013]. Für nicht alle der in der FA-Mutationsdatenbank aufgeführten intronischen Mutationen war allerdings ein pathogener Effekt nachgewiesen worden.

# 4.1.2 Anreicherung und Sequenzierung von Patienten bekannter Komplementationsgruppe

Zu Beginn wurden FA-Patienten mit bereits bekannten Mutationen unter Verwendung einer selbst erstellten *SeqCap EZ Choice Library* angereichert und am GS Junior sequenziert. Dies geschah, um die Methode zu testen und soweit zu etablieren, dass eine Mutationsanalyse mit Proben von Patienten unbekannter Komplementationsgruppe möglich war. Des Weiteren sollte überprüft werden, inwieweit die ROI tatsächlich abgedeckt wird und ob bereits bekannte Mutationen in den Daten bestätigt werden können. Dazu wurden vier FA-Patienten unterschiedlicher Komplementationsgruppen untersucht (Tabelle 6).

Tabelle 6: FA-Patienten mit bekannten Mutationen zur Etablierung der FA-Panelanreicherung. Jeweils zwei FA-Patienten mit zuvor bestätigten biallelischen Mutationen wurden zur Anreicherung ausgewählt und am GS Junior sequenziert. Die jeweilige Komplementationsgruppe sowie die compound-heterozygoten Mutationen sind, zusammen mit dem entsprechenden Effekt auf Proteinebene, aufgelistet.

ID	Komplementations-	Mutation A	Allel 1	Mutation Allel 2		
ID	gruppe	cDNA	Protein	cDNA	Protein	
FA-01	FA-D2	c.2204G>A	p.R735Q	c.3888+2T>G	p.I1260_K1283del	
FA-02	FA-Q	c.689T>C	p.L735P	c.2371_2398dup	p.1800Tfs*24	
FA-03	FA-P	c.1093del	p.Q365Sfs*32	c.1163+3dup	p.R317_F387del	
FA-04	FA-L	c.217-20T>G	p.R73_L91del	c.935G>A	p.C312Y	

Es waren sowohl ein *Upstream*- (FA-L) als auch zwei *Downstream*-Patienten, welche den Komplementationsgruppen FA-P und FA-Q angehören, enthalten. Ferner wurde für die Analyse ein Patient der Komplementationsgruppe FA-D2 ausgewählt, von welcher eine Beeinträchtigung der Mutationsanalyse durch das Vorhandensein von Pseudogenen bekannt ist. Zudem wurde bei der Auswahl der Patienten auf eine Diversität der Mutationen geachtet. So waren neben Basensubstitutionen auch Duplikationen, eine Deletion sowie ein Basenaustausch im Intron (c.217-20T>G in *FANCL*) enthalten, von welcher ein Effekt auf das aberrante Spleißen bekannt ist.

#### a. Erstellen einer Probenlibrary

Zur selektiven Anreicherung der Probenlibrary musste zunächst die Menge der einzusetzenden DNA optimiert werden. Diese war im Protokoll mit 500 ng angegeben. Bei der fluorometrischen Messung am Bioanalyzer konnte bei dieser Ausgangsmenge nur eine sehr niedrige Signalintensität von < 10 FU (*fluorescence unit*) gemessen werden (Abbildung 14A), was deutlich von den Angaben des Protokolls abwich. Dieses Ergebnis ließ auf eine geringe Menge an Probenlibrary schließen. Dennoch konnte man anhand der Messung erkennen, dass die Fragmentierung und Größenselektion erfolgreich und die Größenverteilung der Fragmente gleichmäßig war. So sollte die durchschnittliche Fragmentgröße laut Protokoll in einer Größenordnung von 600–900 bp liegen, mit einer Größenverteilung von 500-1250 bp. Zudem sollten nur weniger als 10 % der Fragmente eine Größe von weniger als 350 bp aufweisen. Diese Vorgaben ließen sich bereits größtenteils erfüllen.



**Abbildung 14:** Bioanalyzer-Ergebnis der Fragmentierung von Probe FA-01 während der Optimierung des Protokolls. **A:** Es wurden zunächst 500 ng DNA zur Nebulisierung und Erstellung einer Probenlibrary eingesetzt. Eine gleichmäßige Größenverteilung von etwa 500–1500 bp ist ersichtlich, das Maximum der Fluoreszenzeinheiten (FU) liegt aber deutlich unter 10. **B:** Beim Einsatz von 3 µg DNA zur Nebulisierung lag das Maximum hingegen bei über 160 FU. Die beiden flankierenden Peaks stellen jeweils interne Marker des Bioanalyzer-Chips dar.

Eine drastische Erhöhung der Ausgangsmenge auf 3 µg ergab eine ausreichende Konzentration an fragmentierter DNA, wieder bestimmt anhand der Fluoreszenzintensität der Bioanalyzer-Messung (Abbildung 14B). Die Größenverteilung war dabei mit der vorherigen Messung vergleichbar. Bei beiden in Abbildung 14 gezeigten Elektropherogrammen liegt das Maximum bei etwa 900 bp, die Fragmentgrößen umfassen dabei einen Bereich von rund 500–2000 bp. Ein vergleichbares Ergebnis zeigte sich in der Regel auch bei den folgenden Anreicherungen, sofern während der Libraryerstellung keine Probleme auftraten, und war somit reproduzierbar. Eine Reduktion der Ausgangsmenge auf 2 µg erbrachte eine ähnlich starke Fluoreszenzintensität, so dass mit dieser Ausgangsmenge fortan gearbeitet wurde.

## b. Amplifikation der Probenlibrary durch LM-PCR

Während der Anreicherung fanden, je nach verwendetem Protokoll, zwei bis drei klonale Amplifikationsschritte der verschiedenen DNA-Proben mittels LM-PCR statt. Eine Überprüfung der LM- PCR-Produkte mithilfe von Bioanalyzer-Messungen ergab zu Beginn der Etablierung eine unzureichende Amplifikation der Proben. Wie in Abbildung 15A ersichtlich, war nur eine minimale Signalintensität messbar. Dagegen zeigte sich ein zusätzlicher Peak in unmittelbarer Nähe des internen Standards, was auf eine Primer-Dimerbildung schließen ließ.



**Abbildung 15:** Unzureichende Amplifikation der gDNA-Fragmente durch die LM-PCR. **A:** Mit dem im Protokoll angegebenen LM-PCR-Ansatz wurde nach der ersten Amplifikationsreaktion am Bioanalyzer eine Primer-Dimerbildung (Pfeil) deutlich. Ein Peak zwischen 500 und 1 000 bp zeigt eine schwache Amplifikation der Probe an. Die flankierenden Peaks bei 50 und 10 380 bp sind interne Größenstandards. **B:** Bei einer Agarose-Gelelektrophorese nach der LM-PCR waren deutlich nicht-inkorporierte Primer (roter Rahmen) zu erkennen. Aufgetragen wurden ein 100 bp-Größenstandard, die LM-PCR-Produkte der Proben FA-01 und FA-02 sowie ein LM-PCR-Kontrollansatz ohne DNA (H<sub>2</sub>O). Die Größen des Markers sind in Basenpaaren (bp) angegeben. **C:** Nach Verringerung der eingesetzten Primerkonzentration ging die Dimerbildung (Pfeil) zurück und die Amplifikation wurde deutlich gesteigert.

Zunächst wurde am PCR-Ansatz keine Änderung vorgenommen, sondern verschiedene Parameter der Amplifikationsreaktion sowie die Aufreinigung nach jeder PCR-Reaktion auf ihre Effektivität hin überprüft. Dabei konnten die Enzymaktivität der verwendeten Polymerase wie auch die Elution der Proben von den verwendeten Silica-Säulen als mögliche Fehlerquelle ausgeschlossen werden. Eine Gelelektrophorese der LM-PCR-Reaktion machte hingegen sichtbar, dass eine erhebliche Menge an nicht eingebauten Primern nach der Amplifikation im Reaktionsgemisch vorhanden war (Abbildung 15B). Aus diesem Grund wurde die Primerkonzentration der 454 Rapid-A- und -B-Primer von 40 µM auf 10 µM verringert, was im Anschluss auch zu einer effizienteren Amplifikation führte (Abbildung 15C). Folglich wurde für jede LM-PCR-Reaktion die Konzentration der Primer entsprechend verringert. Anhand der Bioanalyzer-Messung wie auch durch die Gelelektrophorese in Abbildung 15 wird deutlich, dass sich die Spanne der Größenverteilung verkleinert hatte. So reichte sie ab diesem Schritt nur noch von etwa 500–1000 bp, statt wie zuvor von 500–2000 bp.

# c. Optimierung der emPCR und Sequenzierung

## Anreicherung der Capture Beads

Zunächst musste die einzusetzende Menge an DNA-Molekülen pro *Capture Bead* für die emPCR bestimmt werden. Dieses Verhältnis ist essenziell für eine monoklonale Amplifikation im Wasser-Öl-Gemisch und folglich auch für die Generierung einer adäquaten Menge von Reads am GS Junior, welche die nötigen Qualitätsanforderungen erfüllen.

Für die erste emPCR des Probenpools aus FA-01 und FA-02 wurden zwei Moleküle der angereicherten Rapid Library-DNA pro Capture Bead eingesetzt, da dieses Verhältnis laut Protokoll optimaler Weise in einer Anreicherung von etwa 5-20 % während der emPCR resultiert, was als Ziel einer erfolgreichen emPCR anzusehen ist. Unter Verwendung des Bead Counters zeigte sich jedoch, dass das gewählte Verhältnis von DNA-Molekülen pro Capture Bead mit 2:1 zu hoch war. Die Zahl der angereicherten und zurückgewonnenen DNA-Capture Beads überstieg deutlich die Marke von 2\*106 Beads und damit die Obergrenze der Anreicherung von 20 %. Damit galt die Anreicherung als fehlgeschlagen. Mit der Sequenzierung wurde trotzdem fortgefahren, um Erfahrung mit den Abläufen und den Sequenzierergebnissen zu sammeln. Als Konsequenz aus dem ersten Versuch wurde die eingesetzte Molekülmenge für die zweite Anreicherung um 30 % reduziert, was 1,4 DNA-Molekülen pro Capture Bead entsprach. Dennoch wurde bei der Rückgewinnung der angereicherten DNA-Capture Beads die obere Marke von 2\*106 Beads am Bead Counter deutlich überschritten. Daher wurde das Verhältnis während der Anreicherung von Patienten unbekannter Komplementationsgruppe weiter angepasst. Eine Reduktion auf 0,6 DNA-Moleküle pro Capture Bead ergab schließlich ein Ergebnis im erwünschten Bereich von 0,5-2\*106 angereicherter Beads. Dieses Verhältnis wurde für die meisten Anreicherungen gewählt. Dennoch wurden weitere Ratios getestet und so ein Minimum von 0,09 DNA-Molekülen pro Capture Bead bestimmt, um eine Anreicherung von mindestens 5 % zu erreichen. Bei diesem Verhältnis lag die Zahl an zurückgewonnenen DNA-tragenden Capture Beads knapp oberhalb der 0,5\*106-Marke.

# Sequenzierung am GS Junior

Zur Sequenzierung wurden jeweils rund 0,5\*10<sup>6</sup> zurückgewonnene DNA-*Capture Beads* eingesetzt. Bei der ersten Sequenzierung (FA-01 und FA-02), bei welcher die Anreicherung als nicht erfolgreich anzusehen war, konnten rund 50000 Reads hoher Qualität generiert werden (Tabelle 7). Dies entspricht nur etwa der Hälfte der Anzahl an Reads, die in den Protokollen für *Rapid Libraries* angegeben sind. Zur Übersicht sind einige Rohdaten der ersten beiden Sequenzierungen am GS Junior in Tabelle 7 wiedergegeben.

Anhand der Tabelle lässt sich erkennen, dass sich die Anzahl an Reads mit hoher Qualität (*passed filter wells*) beim zweiten Sequenzierlauf am GS Junior um etwa 20 % auf knapp 60 000 steigern ließ. Insgesamt betrachtet war die Effizienz der Sequenzierung aber nicht erhöht. Dies wird ersichtlich, wenn man Zahl der Wells mit gültiger Library-Schlüsselsequenz (*key pass wells*) betrachtet und mit der Anzahl an Reads vergleicht, welche die

Qualitätsprüfung bestanden hatten. Während im ersten Sequenzierlauf noch ein Viertel aller Reads für die weitere Auswertung verfügbar waren, betrug das Verhältnis beim zweiten Lauf nur noch rund 21 %. Damit wurde die Notwendigkeit nach einer weiteren Optimierung der emPCR zusätzlich unterstrichen.

Tabelle 7: Ergebnisse der ersten Sequenzierungen (Projekt 1 und 2) des GS Run Browsers. Jedes Well mit eindeutigem Lichtsignal wurde detektiert (*raw wells*) und auf eine gültige 4-Basenpaar-Anfangssequenz (GACT für die Library) überprüft (*key pass wells*). Die Anzahl an hochqualitativen Reads (*passed filter wells*) sowie die Anzahl der darin enthaltenen Basen (*total bases*) sind angegeben. Herausgefiltert wurden alle Reads mit zu vielen negativen (*failed dot*) oder positiven (*failed mixed*) Nukleotid-Inkorporationen. Die Anzahl an Reads, die eine Mindestlänge aufgrund von 3'-Resektion zur Qualitätsoptimierung (*failed short quality*) oder durch die Entfernung der angehängten Primersequenz (*failed short primer*) unterschritten hatten, sind ebenfalls angegeben.

Projekt	FA- Patienten	Raw wells	Key pass wells	Passed filter wells	Total bases	Failed dot	Failed mixed	Failed short quality	Failed short primer
1	01/02	213 189	192 085	49 903	20 585 257	9 515	36 212	92 801	74
2	03/04	293 088	283 971	59 629	21 278 367	20 687	41 975	158 953	6

Trägt man den Prozentsatz der verworfenen (*failed*) Reads und derer, welche die Qualitätsprüfung bestanden hatten (*passed filter wells*), in einem gestapelten Säulendiagramm grafisch auf, wird die Verringerung der Effizienz im zweiten Lauf weiter deutlich (Abbildung 16). So ist in Abbildung 16A klar die Abnahme an qualitativ hochwertigen Reads erkennbar. Betrachtet man die verschiedenen Kategorien des GS Run Browsers, anhand derer Reads herausgefiltert wurden, ist nur eine Verbesserung bei der Anzahl der *failed mixed*-Reads zu verzeichnen. Entfiel bei der ersten Anreicherung noch fast jeder fünfte Read auf diese Gruppe, waren es im zweiten Lauf nur knapp 15 %. Dagegen war die Zahl der Reads in der *failed dot*-Gruppe leicht angestiegen, bei denen die Software zu wenige Nukleotid-Einbauten pro Nukleotidfluss detektiert hatte. Den größten Anteil an verworfenen Reads hatte allerdings die Gruppe der *short quality*-Reads. Darin enthalten sind Reads, die zur Verbesserung ihrer Qualität soweit getrimmt wurden, dass sie eine Mindestlänge unterschritten und deshalb aussortiert wurden.

Vergleicht man die Histogramme von Abbildung 16 (B+C), welche die Verteilung der Readlängen beider Sequenzierungen darstellen, so fällt beim zweiten Lauf eine erhöhte Zahl an kurzen Reads von < 150 bp sowie ein zusätzlicher Peak bei etwa 60 bp auf. Dies spiegelte sich bei der Bestimmung der durchschnittlichen Readlänge wider, welche mit 357 bp deutlich kleiner war als beim ersten Lauf, bei dem sie mit 413 bp bestimmt wurde.



Abbildung 16: Vergleich verschiedener Qualitätsmerkmale der ersten beiden Sequenzierungen am GS Junior. A: Nur ein geringer Anteil der Reads mit gültiger Schlüsselsequenz hatte alle Qualitätsfilter bestanden (*passed filter wells*; blau). Den geringsten Anteil hatten Reads mit zu vielen negativen (rot; *failed doi*), gefolgt von Reads mit zu vielen positiven Nukleotid-Einbauten (*failed mixed*; grün). Die meisten verworfenen Reads hatten aufgrund von 3'-Resektionen zur Qualitätsoptimierung eine Mindestlänge unterschritten (*short quality*; orange). Die Anzahl der *Short primer*-Reads (Mindestlänge nach Entfernung der Primersequenz unterschritten) war zu gering, um hier sichtbar dargestellt zu werden. **B** + **C**: Größenverteilung der Reads der ersten beiden Sequenzierläufe mit angereicherter DNA von FA-01/-02 (**B**) sowie von FA-03/-04 (**C**). Dargestellt ist die Häufigkeit der Readlängen ab einer Größe von 40 bp. Durch gestrichelte Linien sind die durchschnittliche Länge (rot), der Median (blau) sowie der Modalwert (grün) eingetragen.

#### Auswertung der NGS-Daten mit Hilfe der NextGENe-Software

Im Anschluss an die Sequenzierung wurden die auswertbaren Reads zur Mutationsanalyse mit Hilfe der NextGENe-Software analysiert (Tabelle 8). Bei beiden Projekten war der Anteil der beiden angereicherten Patienten an der Gesamtzahl der Reads vergleichbar. Eine Steigerung der auswertbaren Rohdaten machte sich bei FA-03 und FA-04 in einer erhöhten Abdeckung bemerkbar. Da alle vier Patienten keinen bekannten konsanguinen Hintergrund hatten, wurden unbekannte heterozygote Varianten weiter nach der Art der Veränderung unterschieden. Deutlich fällt hier die hohe Rate von 21–46 % an Indel-Mutationen unter allen detektierten unbekannten Varianten auf (Tabelle 8). Bei näherer Betrachtung konnte ein erheblicher Anteil dieser Mutationen innerhalb oder angrenzend an Homopolymer-Abschnitten nachgewiesen werden, weshalb viele dieser Varianten als Sequenzierfehler vernachlässigbar sind.

**Tabelle 8:** Auswertung der ersten beiden Sequenzierläufe am GS Junior mit Hilfe der NextGENe Software v2.2. Die detektierten Varianten wurden durch die Software mit der dbSNP-Datenbank abgeglichen, um SNPs von unbekannten Varianten (UVs) zu unterscheiden, welche als potentielle Mutationen anzusehen sind. Alle Angaben zur Abdeckung sowie den detektierten Varianten beziehen sich auf die ROI. Als essenzielle Spleißstellen wurden die kanonischen Spleißdonor- und -akzeptorstellen zwei Basenpaare 3<sup>c</sup> bzw. 5<sup>c</sup> eines Exons untersucht.

Auswertung	FA-01	FA-02	FA-03	FA-04
Gesamtzahl konvertierter Reads	20 1 37	25 869	26 446	31 245
Alignte Reads	18 823	24 154	23 155	27 804
Durchschnittliche Abdeckung (x)	6,2	7,1	9,5	12,3
Gesamtzahl Varianten	85	117	300	154
Bekannte SNPs (dbSNP 134)	59	87	151	88
UVs in Exons	27	29	73	65
UVs an essenziellen Spleißstellen	1	1	2	2
Homozygote UVs	2	1	5	2
Heterozygote UVs	27	31	175	66
Synonyme UVs	5	4	25	3
Missense UVs	3	5	30	12
Nonsense UVs	1	1	24	2
Indel-UVs	20	24	139	54

Trotz der geringen durchschnittlichen Abdeckung von maximal 12,3x (Tabelle 8) konnten alle bereits bekannten krankheitsverursachenden Mutationen der vier Patienten in den Sequenzierungsdaten nachgewiesen werden. Alle dieser pathogenen Varianten wurden von der Software als Mutationen erkannt und im von der Software erstellten Mutationsbericht gelistet. Abbildung 17 zeigt die Duplikation in Exon 11 des *ERCC4*-Gens bei Patient FA-02.



Abbildung 17: Nachweis der 28 bp-Duplikation c.2371\_2398dup im *ERCC4/FANCQ*-Gen bei FA-02. A: Darstellung durch das Softwareprogramm NextGENe. Die Duplikation ist blau markiert. B: Im Elektropherogramm der Sanger-Sequenzierung ist die heterozygote Duplikation (unterstrichen), durch die eine Überlagerung der Sequenzen entsteht, zu erkennen.

# 4.2 Analyse von Patienten unbekannter Komplementationsgruppe mit Hilfe von SeqCap EZ-Anreicherungspanels

# 4.2.1 Anreicherung mit verschiedenen SeqCap EZ-Libraries und Sequenzierung am GS Junior

Nach der Etablierung wurden 27 weitere Anreicherungen und Sequenzierungen durchgeführt und damit 51 verschiedene Patienten analysiert. Diese waren zuvor noch durch kein anderes Panel angereichert und per NGS untersucht worden. Einige wichtige Rohdaten der Sequenzierungen sind in Tabelle 9 zusammengefasst

Tabelle 9: Übersicht der GS Run Browser-Daten der unklassifizierten FA-Patienten. Jedes Well mit eindeutigem Lichtsignal wurde detektiert (*raw wells*) und auf eine gültige 4-Basenpaar-Anfangssequenz (GACT für Library) überprüft (*key pass wells*). Die Anzahl an hochqualitativen Reads (*passed filter wells*) sowie die Anzahl der darin enthaltenen Basen (*total bases*) sind angegeben. Herausgefiltert wurden alle Reads mit zu vielen negativen (*failed dot*) oder positiven (*failed mixed*) Nukleotid-Inkorporationen. Die Anzahl an Reads, die eine Mindestlänge aufgrund von 3<sup>c</sup>-Resektion zur Qualitätsoptimierung (*failed short quality*) oder durch die Entfernung der angehängten Primersequenz (*failed short primer*) unterschritten, sind ebenfalls dargestellt. "n.a.": keine Angabe. "\*": der Patient wurde wiederholt angereichert und sequenziert

SeqCap EZ- Panel	Projekt	FA- Patienten	Raw wells	Key pass wells	Passed filter wells	Total bases	Failed dot	Failed mixed	Failed short quality	Failed short primer
	3	05/06	303 473	292 403	182 063	75 816 345	7 853	6737	93 658	2
	4	05*/07	247 230	239 372	133 565	57 825 542	6 954	8 930	88 417	0
	5	07*/08	272 890	262 518	158 867	69 973 671	4 984	20 355	76 439	4
	6	09/10	257 739	246 854	126 608	52 364 334	3 656	36 005	79 585	8
OID	7	11/12	299 959	293 580	203 907	93 476 855	4 513	7 417	76 195	3
01D 38614	8	13/14	284 002	273 181	167 383	73 359 532	5 452	6 988	90 354	2
50014	9	15/16	269 108	260 867	149 104	67 053 108	4 473	8 638	97 232	1
	10	16*/17	275 637	263 496	164 071	71 197 238	4 193	13 711	80 135	2
	11	18/19	261 004	251 523	131 675	54 965 195	2 579	15 177	100 580	2
	12	20/21	212 840	204 363	138 834	62 284 160	3 357	5 919	54 841	0
	13	22/23	268 866	259 442	147 072	63 886 095	5 427	18 825	85 085	6
	14	24/25	248 877	238 823	151 031	n.a.	3 725	7 039	75 067	0
	15	26/27	317 556	309 541	131 611	55 885 170	13 827	19 329	142 652	6
	16	28/29	312 889	302 429	161 702	68 866 312	10752	13 863	114 010	5
OID	17	30/31	248 562	236 926	156 089	68 244 973	4 989	7 467	65 585	5
41218	18	32/33	282 181	263 937	166 401	70680965	20 431	6 295	68 223	4
11210	19	34/35	227 906	218 871	137 746	53 919 220	3 294	6 221	70 430	5
	20	36/37	207 066	196 715	154 360	63 134 826	5 596	4 556	30 336	3
	21	38/39	216 529	207 496	156 773	67 859 244	6 1 5 2	4 329	39 508	0
	22	40/41	305 252	291 112	169 902	70 282 442	27 792	5 912	84 339	1
	23	42/43	273 414	259 743	143 037	27 290 836	9 538	12 780	89 841	7
OID	24	44/45	252 705	243 098	199 610	88 127 773	7 1 3 7	3 098	31 790	6
42419	25	46/47	275 748	264 920	199 546	90 226 369	5 860	4 1 2 3	53 239	4
12117	26	48/49	314 299	304 620	197 520	86 314 364	4 2 47	5 556	95 680	10
	27	50/51	285 790	265 091	191 491	81 111 847	21 109	4 062	44 891	9
OID	28	52/53	288 815	279 807	183 302	81 781 622	5 483	5 968	83 260	5
43041	29	54/55	333 660	324 123	198 958	86 112 157	11 264	8 719	103 003	14

Nicht in Tabelle 9 eingetragen sind einige Merkmale zu den Readlängen, welche ebenfalls durch den GS Run Browser angezeigt werden. So umfasste die größte detektierte Readlänge 1196 bp, wohingegen die durchschnittliche Readlänge 420 bp betrug.

Vergleicht man zunächst den Anteil der Wells mit gültiger Schlüsselsequenz (key pass wells) an allen Wells mit Lichtsignal (raw wells) vor (Tabelle 7) und nach der Etablierung (Tabelle 9) miteinander, lässt sich kein Unterschied zwischen den verschiedenen Projekten erkennen. Demnach hatten zwischen 90 % und 98 % der raw wells eine gültige Erkennungssequenz, so dass diese als key pass wells weiter hinsichtlich bestimmter Qualitätsmerkmale gefiltert werden konnten. Im Schnitt bestand etwa 60% der Wells mit gültiger Schlüsselsequenz alle Filterkriterien (passed filter wells), was eine deutliche Steigerung gegenüber der ersten beiden Sequenzierungen darstellt, bei welchen nur jedes vierte (Projekt 1) bzw. fünfte Well (Projekt 2) zur weiteren Analyse verwendet werden konnte (vgl. Tabelle 7). Dies wird besonders deutlich, wenn man diese Werte grafisch aufträgt (Abbildung 18). Anhand von Abbildung 18A ist zu erkennen, dass die Anzahl der raw und key pass wells aller am GS Junior durchgeführten Sequenzierungen zwar streuen, aber alle innerhalb der sogenannten Whisker-Grenzen des Boxplots liegen, welche hier den 1,5-fachen Interquartilsabstand anzeigen und anhand derer sich Ausreißer bzw. Extremwerte erkennen lassen. Dieses Bild ändert sich erst nach der Filterung der Daten, welche ebenfalls in Abbildung 18A unter passed filter wells dargestellt ist. Hier ist die deutliche Steigerung der Effizienz nach der Etablierung zu erkennen. So liegen die Ergebnisse der ersten beiden Sequenzierungen weit außerhalb des Interquartilsabstandes (Abbildung 18A, blau gefüllte Kreise), womit sie als Ausreißer gelten.



Abbildung 18: Verteilung der Reads vor und nach dem Filtern sowie Darstellung der sequenzierten Basen nach Anreicherung und Sequenzierung mit verschiedenen Libraries. Die Grenzen der Box markieren jeweils die 25. bzw. 75. Perzentile, die Linie im Inneren den Median. Ein schwarzes Kreuz markiert zudem den Durchschnitt. Die äußeren Begrenzungen (Whisker) entsprechen dem 1,5-fachen Interquartilsabstand (nach Tukey). Die Ergebnisse der einzelnen Projekte (A: n = 29; B: n = 28) sind farbig eingezeichnet. Blau: OID38614 (ausgefüllt: Projekt 1 und 2). Grün: OID41218. Orange: OID42419. Rot: OID43014. A: Dargestellt ist die Anzahl der Wells, die mit einem eindeutigen Lichtsignal detektiert (*raw wells*) und auf eine gültige 4-Basenpaar-Anfangssequenz (GACT für Library) geprüft wurden (*key pass wells*). Nach verschiedenen Qualitätsfiltern blieben die Reads übrig, welche zur Mutationsanalyse genutzt werden können (*passed filter wells*). B: Verteilung aller sequenzierten Basen nach dem Filtern und Trimmen der Reads.

Ferner wird aus Abbildung 18B ersichtlich, dass nach der Etablierung durchschnittlich rund 69\*10<sup>6</sup> Basen pro Lauf zur Mutationsanalyse vorhanden waren, während vor der Etablierung im Durchschnitt nur rund 21\*10<sup>6</sup> Basen auswertbar waren (Abbildung 18B). Erneut sind hier nur die ersten beiden Projekte als Ausreißer zu erkennen (Abbildung 18B, blau gefüllte Kreise). Dagegen finden sich die Ergebnisse der mit OID42419 (orange Kreise) und OID43014 (rote Kreise) angereicherten Proben jeweils oberhalb des 75. Perzentils der Boxplots (Abbildung 18A und B). Eine Ausnahme bildete dabei nur Projekt 23 (OID42419). Bei der Anzahl der Wells, die alle Qualitätskriterien erfüllten (*passed filter wells*), lag der Wert im Vergleich mit den anderen Ergebnissen nahe des unteren Quartils der Box (Abbildung 18A) und stellte sogar das Minimum der sequenzierten Basen (Abbildung 18B) dar. Wie bereits im vorherigen Abschnitt erwähnt, war bei diesem Projekt die Umgebungstemperatur sehr hoch und verschiedene Parameter am GS Junior außerhalb des Optimums.

Analog zu Abbildung 16 wurden die Filter genauer analysiert, die alle Reads mit gültiger Schlüsselsequenz zur Qualitätsoptimierung durchlaufen hatten und in einer Abbildung dargestellt (Abbildung 19).



Abbildung 19: Vergleich verschiedener Qualitätsmerkmale aller Reads mit gültiger Library-Schlüsselsequenz bei FA-Patienten mit unbekannten Mutationen (n=51). Blau: Hochqualitative Reads zur weiteren Auswertung (*passed filter wells*); Rot: Reads mit zu vielen negativen Nukleotid-Einbauten (*failed dot*); Grün: Reads mit zu vielen positiven Nukleotid-Einbauten (*failed dot*); Grün: Reads mit zu vielen positiven Nukleotid-Einbauten (*failed dot*); Grün: Reads mit zu vielen positiven Nukleotid-Einbauten (*failed mixed*); Orange: Anzahl an Reads, welche aufgrund von 3'-Resektion zur Qualitätsoptimierung eine Mindestlänge unterschritten hatten (*short quality*). Die Anzahl der *Short primer*-Reads (Mindestlänge nach Entfernung der Primersequenz unterschritten) war zu gering, um hier sichtbar dargestellt zu werden.

Den geringsten Anteil an Reads, welche die Filterkriterien bestanden hatten, wies Projekt 15 (FA-26/FA-27) mit nur 43 % auf, das beste Ergebnis wurde bei Projekt 24 (FA-44/FA-45) mit 83 % erzielt. Die wenigsten

Reads entfielen jeweils auf den "Short Primer"-Filter (Tabelle 9). Weiterhin wurden auch nur sehr wenige Reads durch die "Dots"- und "Mixed"-Filter verworfen. Der Anteil betrug dabei durchschnittlich nur 3 % bzw. 4 %. Im Allgemeinen hatten sich beide Werte nach der Etablierung verbessert, was insbesondere für den "Mixed"-Filter-Anteil gilt. Nach wie vor entfielen die meisten herausgefilterten Reads auf die 3'-Resektion zur Qualitätsoptimierung, welche den anderen besprochenen Filtern nachgeschaltet ist und die Signalintensitäten der Reads beurteilt. Zwar lagen alle Werte noch unterhalb der Ergebnisse von Projekt 1 und 2, dennoch war im Schnitt jeder dritte Read durch das Trimmen verworfen worden.

#### 4.2.2 Alignment mit Hilfe der NextGENe-Software

## a. Abdeckung der gesamten ROI nach der Sequenzierung am GS Junior

Die am GS Junior generierten Rohdaten wurden zum Alignment in die NextGENe-Software geladen und mit deren Hilfe ausgewertet. Durch die Barcode-Sortierung und Datenkonvertierung wurden so zunächst weitere Reads herausgefiltert, welche eine Prüfung der vorgegebenen Qualitätsparameter nicht bestanden hatten. Die übrigen Reads konnten für das Alignment und somit auch zur Mutationsanalyse verwendet werden. In Abbildung 20 ist die Anzahl der konvertierten Reads nach der Etablierung für jeden Patienten dargestellt.



Patient 1 Patient 2

**Abbildung 20:** Anzahl an Reads nach Barcode-Sortierung und Datenkonvertierung (n = 51) durch die NextGENe-Software. Der erstgenannte Patient ist jeweils als blaue, der zweitgenannte Patient als rote Säule dargestellt. Eine schwarze gestrichelte Linie markiert den Mittelwert, die graue Fläche die Standardabweichung.

Anhand der Grafik ist abzulesen, dass im Durchschnitt knapp 80 000 Reads pro Patient zur Mutationsanalyse herangezogen werden konnten. Durch die eingezeichnete Standardabweichung wird deutlich, dass bei fünf Patienten unterdurchschnittlich wenige Reads zur weiteren Analyse zur Verfügung standen. Dazu gehören FA-05 und FA-24, bei welchen bereits nach der ersten LM-PCR-Reaktion kaum noch Produkt bei der Qualitäts- und Quantitätsprüfung nachzuweisen waren. Nach der Barcode-Sortierung entfielen nur 1 395 bzw. 3 322 Reads auf die jeweiligen Patienten. Bei FA-21 wiedersprachen sich die Ergebnisse der Bioanalyzer- und NanoDrop-Messung. Während am Bioanalyzer kein Signal nachweisbar war, wurde durch den NanoDrop eine Konzentration angegeben, die mit dem Ergebnis von Patient FA-20 vergleichbar war, welcher parallel zu FA-21 in Projekt 12 angereichert wurde. Dennoch entfielen von insgesamt 138 834 sequenzierten Reads nur knapp 10 % auf FA-21, wovon schließlich 13 556 Reads erfolgreich ins FASTA-Format zum Alignment konvertiert werden konnten. Im Gegensatz dazu waren bei FA-07 und FA-48 keine Auffälligkeiten während der Anreicherung aufgetreten. Dennoch lag das Ergebnis auch hier weit unterhalb des Durchschnittes. Bei FA-05 sowie FA-07 wurde aufgrund der nicht zufriedenstellenden Abdeckung die Anreicherung und Sequenzierung wiederholt, FA-24 wurde hingegen mit anderen Anreicherungsmethoden sowie einer FANCD2-Immunblotanalyse weiter untersucht. Bei FA-21 sowie FA-48 war die Mutationsanalyse trotzdem erfolgreich (Tabelle 12), weshalb auf eine Wiederholung verzichtet wurde.

Nachdem die konvertierten Reads durch die Software gegen das Referenzgenom gemappt wurden, konnte die gesamte ROI auf ihre Abdeckung überprüft werden. Die Ergebnisse sind in Abbildung 21 zusammengefasst.



**Abbildung 21:** Abdeckung der ROI nach Datenkonvertierung und Alignment der GS Junior-Sequenzierungsdaten von FA-05–FA-55 durch die NextGENe-Software. Die ROI wurde nach dem Mappen der Reads auf ihre durchschnittliche Abdeckung hin überprüft. Die Grenzen der Box markieren jeweils die 25. bzw. 75. Perzentile, die Linie im Inneren den Median. Ein schwarzes Kreuz markiert zudem den Durchschnitt. Die Whisker (horizontale Linien) entsprechen dem 1,5-fachen Interquartilsabstand (nach Tukey). Ausreißer sind als schwarze Kreise außerhalb der Whisker dargestellt. (n = 54).

Anhand des Boxplots in Abbildung 21 ist zu erkennen, dass bis zu einer durchschnittlichen Abdeckung von 10x die Hälfte der Werte sehr nahe am Median lagen, da hier die Box sehr klein ist. Auch die Streuung ist gering, so dass die Whisker sehr klein sind. Mit zunehmender Streuung der Daten werden die Boxen dagegen breiter, was besonders deutlich bei einer Abdeckung von 40x wird. Danach werden die Boxen wieder schmäler. Bei einer Analyse der Ausreißer bei einer einfachen Abdeckung zeigte sich, dass es sich dabei um FA-05, FA-07, FA-21, FA-22, FA-24, FA-42 sowie FA-48 handelt, also größtenteils um Patienten mit auffallend niedriger Anzahl an konvertierten Reads (vgl. Abbildung 20). Betrachtet man die Mittelwerte in Abbildung 21 so ist zu erkennen, dass diese den Verlauf einer Kurve beschreiben, welche mit zunehmender Abdeckung abfällt. Dabei besteht der stärkste Abfall zwischen einer Abdeckung von 20x zu 40x. Der Mittelwert einer 40-fachen Abdeckung der ROI liegt hier bereits bei nur noch 46 %, wohingegen im Schnitt noch 72 % eine Abdeckung von 20x aufwiesen. Will man weniger die Extremwerte der jeweiligen Abdeckung berücksichtigen und nimmt dazu die Mediane als Grundlage einer Kurve heran, so ist hier der Abfall in der Abdeckung der ROI im Vergleich von 40x auf 60x am größten. So lag der Median einer 40-fachen Abdeckung noch bei 51 %, wohingegen er bei 60x auf 21 % abgefallen war. Die höchste durchschnittliche Abdeckung konnte mit 109x bei FA-49 bestimmt werden. Hier wiesen sogar noch knapp 13 % der ROI eine Abdeckung von 200x auf.

#### b. Abdeckung einzelner Zielregionen

Die theoretische Abdeckung von Zielregionen der nicht-stringenten Designoption war durch Roche NimbleGen jeweils mit annähernd 100 % angegeben und nach der Berechnung der Vorhersage wären nur wenige hundert Basen selbst durch die Sequenzierung nicht abdeckbar (Tabelle 5). Zur Überprüfung wurden die einzelnen Targets nach der Sequenzierung auf ihre tatsächliche Abdeckung untersucht. Dies ermöglichte eine Nachbesserung des Designs und eine gezielte Bewertung der nicht abgedeckten Regionen. Sobald mindestens eine Base einer Zielregion nicht durch minimal einen Read abgedeckt war, wurde sie als "Lücke" bewertet. Wenn hingegen ein gesamter Targetabschnitt durch keinen einzigen Read abgedeckt war, wurde dieser als "Totalausfall" aufgelistet. In Abbildung 22 ist der Anteil der jeweils lückenhaften bzw. fehlenden Zielregionen für die einzelnen Libraries grafisch dargestellt.

Anhand der Grafik wird deutlich, dass die Anzahl der Totalausfälle relativ konstant war und der Großteil der Werte unterhalb des Mittelwertes von 1,8 % lag, während dieser Wert für lückenhafte Zielregionen mit 3,5 % bestimmt wurde. Ausreißer sind anhand der Grafik klar erkennbar, beispielsweise FA-01. Dieser fällt insbesondere im Vergleich mit Patient FA-02 auf, welcher parallel dazu angereichert und sequenziert wurde. Obwohl die durchschnittliche Abdeckung während der Etablierungsphase sehr niedrig war (Tabelle 8), konnte dennoch ein Großteil der Zielregion abgedeckt werden. Ferner lässt sich eine unzureichende Abdeckung bei Patient FA-07 nach der ersten Sequenzierung erkennen, weshalb der Patient auch wiederholt angereichert und sequenziert wurde (FA-07\*). Daneben wurden die Patienten FA-05 und FA-16 je zwei Mal angereichert und sequenziert, weil das Ergebnis nach der ersten LM-PCR und damit vor der Sondenhybridisierung nicht zufriedenstellend war. Bei FA-05 spiegelte sich die unzureichende erste Anreicherung deutlich in den lückenhaften Sequenzierdaten wider, so dass dieser erste Versuch komplett von weiteren Analysen ausgeschlossen wurde. Weiterhin war die Anzahl an Lücken bei den Patienten FA-08, FA-07\*, FA-09 sowie FA-10 auffallend hoch. Hierbei ist anzumerken, dass zwischen der Rückgewinnung der angereicherten DNA-*Capture Beads* nach der emPCR und dem Zeitpunkt der Sequenzierung fünf (FA-08/FA-07\*) bzw. zehn (FA-09/FA-10) Tage lagen. Bei allen anderen Anreicherungen wurde hingegen direkt am Tag der DNA-*Capture Bead*-Rückgewinnung am GS Junior sequenziert.



Abbildung 22: Anteil der Targets an der jeweiligen ROI, welche nur lückenhaft (blau) oder überhaupt nicht (rot) durch Reads nach der Sequenzierung aller am GS Junior sequenzierten Patienten abgedeckt waren. "\*" markiert Patienten, welche wiederholt angereichert und sequenziert wurden. Die erste Sequenzierung von FA-05 war fehlgeschlagen, weshalb die Ergebnisse hier nicht einfließen. Vertikale gestrichelte Linien markieren die Grenzen der verschiedenen Libraries. Horizontale gestrichelte Linien geben die Mittelwerte des Anteils an Lücken (blau) und komplett fehlender Targets (rot) an. Die jeweilige Standardabweichung der Mittelwerte ist als blaue (Lücken) bzw. rote Fläche (Totalausfall) eingezeichnet.

Die Anzahl an Lücken sowie vollständig fehlenden Zielregionen waren auch bei den verschiedenen *Developer*-Libraries relativ konstant. Eine klare Ausnahme stellt FA-24 dar, bei welchem etwa ein Drittel aller Targets unzureichend abgedeckt war und fast 14 % in den Sequenzierdaten fehlten. Allerdings war hier bereits während der Anreicherung ein Problem aufgetreten. So war bei der Messung der Probenlibrary am Bioanalyzer nach der ersten LM-PCR-Reaktion kein Fluoreszenzsignal nachweisbar, obwohl am NanoDrop eine Konzentration von 13,5 [ng/µl] bestimmt wurde und somit für eine Anreicherung ausreichte.

Betrachtet man die Ergebnisse der dritten Sondenlibrary (OID42419; FA-42–FA-51), so fallen deutliche Schwankungen bei der Anzahl der Lücken auf. Bei der Sequenzierung von FA-42 und FA-43 ist anzumerken, dass die Raumtemperatur bei der Sequenzierung deutlich erhöht war. Dies hatte zur Folge, dass bestimmte Werte von Sensoren des GS Juniors, wie die CCD- oder PTP-Heiztemperatur, im obersten Maximum bzw. außerhalb davon lagen und der Lauf nicht optimal durchgeführt werden konnte, was sich anhand verschiedener Fehlermeldungen am Sequenzierer zeigte. Einen hohen Anteil an Lücken und Totalausfällen (n

= 65 und n = 51) wies zudem FA-48 auf. Hier waren ähnliche Probleme wie bei FA-24 aufgetreten, welche sich ebenfalls in der Abdeckung der Zielregionen widerspiegelte. Die Probengröße der letzten verwendeten Library (OID43014) war mit vier Patienten sehr klein. Ausreißer waren hier nicht aufgetreten.

Neben dem Anteil an Regionen mit unvollständiger oder fehlender Abdeckung wurden die NGS-Daten auf immer wiederkehrende nicht-abgedeckte Regionen überprüft. Die Übersicht über Zielregionen mit unzureichender Abdeckung, welche für Abbildung 22 erstellt wurde, diente dabei als Ausgangspunkt. Alle Regionen, die bei mehr als der Hälfte der auswertbaren Patienten nicht oder nicht vollständig abgedeckt waren, wurden zusammengetragen. Diese Targets wurden im nächsten Panel, sofern noch enthalten, mit einer höheren Konzentration an Sonden versehen, um die Abdeckung dieser Regionen zu verbessern. In Tabelle 10 sind diese Targets für jedes Panel aufgelistet. Man erkennt, dass sich die Anzahl auf je 25 bis 29 Targets beschränkt.

**Tabelle 10:** Übersicht über die Regionen, welche jeweils bei mehr als der Hälfte der per FA-Genpanel angereicherten und am GS Junior sequenzierten Patienten durch Lücken unterbrochen waren oder komplett fehlten. Für jedes dieser Targets sind das Chromosom (Chr) auf dem das Gen liegt, das jeweils betroffene codierende Exon (*coding sequence*; CDS) sowie der GC-Gehalt der Zielregion gelistet. Die verschiedenen Libraries sind mit der Anzahl der jeweils angereicherten Patienten angegeben. "x": die Region war bei > 50% der analysierten Patienten nicht oder unzureichend abgedeckt. "o": das Target war bei > 50% der analysierten Patienten angegeben war im entsprechenden Panel nicht enthalten.

Chr	Gen	CDS	GC [%]	OID38614 (n = 25)	OID41218 (n = 18)	OID42419 (n = 10)	OID43014 $(n = 4)$
1	FAAP20	1	84,1	0	0	х	х
1	APITD1	1	65,6	x	х	х	Х
1	SFPQ	1	74,2	х	х	х	x
2	REV1	3	32,1	x	х	х	0
2	RIF1	28	24,6	О	х	-	-
3	ATRIP	1	73,9	х	х	х	х
3	ATR	27	36,2	x	х	х	x
3	ATR	41	33,0	х	х	О	О
3	HES1	4	70,1	х	х	х	х
4	CTBP1	1	76,6	х	х	х	x
4	HELQ	10	33,8	О	х	О	Х
4	HELQ	11	30,1	x	х	х	Х
5	RAD50	2	33,1	-	-	х	О
5	RAD50	17	29,1	-	-	О	Х
5	RAD50	22	49,7	-	-	х	х
6	FANCE	1	77,4	х	х	х	x
6	MMS22L	17	28,5	х	х	х	х
6	FOXO3	1	72,3	х	х	-	-
6	REV3L	1	66,2	-	-	x	х
6	REV3L	4	32,9	-	-	х	О
6	REV3L	18	28,0	-	-	х	х
7	MLL3	1	67,2	Х	Х	-	-
8	TONSL	1	75,0	х	х	х	х

## Ergebnisse

8	TONSL	2	76,0	х	x	х	х
8	TONSL	3	74,3	х	х	0	Х
8	TONSL	26	73,2	х	0	х	0
11	MUS81	1	73,8	х	х	0	0
14	FANCM	7	32,0	х	х	X	х
14	FANCM	17	33,7	0	0	х	0
14	RAD51B	4	32,2	х	0	х	0
14	XRCC3	3	70,4	0	х	О	0
16	PALB2	4	39,2	х	х	х	0
16	SLX1B	2	69,1	-	-	X	х
16	SLX1B	3	74,1	-	-	х	х
16	FANCA	1	77,3	х	х	х	х
17	BRIP1	16	34,2	х	х	х	х
17	FAAP100	1	76,3	х	х	х	х
17	FAAP100	2	74,0	х	х	х	х
17	STRA13	1	72,0	Х	X	0	0

Bereits bei der Auswertung der ersten Library war ein hoher Anteil an codierenden ersten Exons aufgefallen, von denen ein hoher GC-Gehalt und damit einhergehende Schwierigkeiten auch von der Routine-Analyse mittels Sanger-Sequenzierung bekannt sind. Daher wurde für alle Targets, welche in Tabelle 10 gelistet sind, der prozentuale Anteil an Guanin und Cytosin bestimmt, um mögliche Rückschlüsse auf die fehlende Abdeckung zu ziehen. Der GC-Gehalt der in der Tabelle gelisteten ersten Exons betrug zwischen 66 % und 84 % und lag somit deutlich über dem Durchschnitt von 41 %. Insgesamt hatten über die Hälfte der problematischen Regionen einen hohen GC-Gehalt. Bei den restlichen Regionen lag dieser bei 25-50 %.

Nachdem diese Targets identifiziert waren, wurde für jede neu designte Library die Konzentration an Sonden in dieser Region um das Sechs- bis Zehnfache erhöht. Beim Vergleich des Ergebnisses zwischen der Choiceund der ersten Developer-Library fällt dabei zunächst keine Verbesserung auf. Stattdessen waren nun drei zusätzliche Regionen bei über der Hälfte aller analysierten Patienten nicht abgedeckt, die diesen Schwellenwert zuvor nicht überschritten hatten. Dabei handelte es sich um Exon 28 im *RIF1*-Gen, Exon 10 in *HELQ* sowie Exon 3 in *XRCC3*. Nur Exon 26 in *TONSL* war besser abgedeckt und nicht mehr gelistet. *FANCE* war insgesamt gesehen ebenfalls besser abgedeckt. So fehlte Exon 1 bei diesem Gen unter Verwendung der Choice-Library bei 24 von 25 Patienten komplett und war nun immerhin bei 17 von 18 Patienten vorhanden, aber mit Lücken unterbrochen. Beim Erstellen des Designs der zweiten Developer-Library (OID42419) wurde die Anzahl aller Replikate um den Faktor 10 erhöht. Dennoch zeigten alle Targets, die nun mit mehr Sonden abgedeckt waren, keine bessere Abdeckung. Dazu zählen *REV1* Exon 7, *ATR* Exon 27, *HELQ* Exon 11, *MMS22L* Exon 17, *TONSL* Exons 1-3, *PALB2* Exon 4 sowie *BRIP1* Exon 16.

# 4.2.3 Zuordnung der Patienten zu Komplementationsgruppen

# a. MLPA- und Immunblot-Analyse zur Eingrenzung der Komplementationsgruppe

Sofern geeignetes bzw. genügend Patientenmaterial vorhanden war, wurde zusätzlich zur Anreicherung eine *FANCA*-MLPA-Analyse durchgeführt sowie ein FANCD2-Immunblot angefertigt (Tabelle 11). Diese Versuche halfen dabei, die Komplementationsgruppe weiter einzuschränken oder die Ergebnisse der Sequenzierung zu bestätigen. Tabelle 11 zeigt, dass bei insgesamt 43 Patienten das *FANCA*- Gen auf große Deletionen und Duplikationen überprüft wurde. Bei zwei dieser Patienten (Patient FA-06 und FA-19) konnte jeweils eine heterozygote Deletion in *FANCA* festgestellt werden, welche ein bzw. vier Exons umfasste.

**Tabelle 11:** Ergebnis der FANCA-MLPA- sowie FANCD2-Immunblot-Analysen. "ohne Befund": keine Auffälligkeiten. " $\Delta$ ":Deletion "-": nicht durchgeführt. "het": heterozygot.

Patienten- ID	FANCA-MLPA	FANCD2-Immunblot	Patienten- ID	FANCA- MLPA	FANCD2- Immunblot
FA-05	ohne Befund	downstream	FA-31	ohne Befund	-
FA-06	$\Delta$ Exon 40 (het)	-	FA-32	-	-
FA-07	ohne Befund	downstream	FA-33	ohne Befund	upstream
FA-08	-	<i>domnstream</i> , Geschwisterkind FANCD2-defizient	FA-34	ohne Befund	upstream
FA-09	ohne Befund	FANCD2-defizient	FA-35	ohne Befund	upstream
FA-10	ohne Befund	upstream	FA-36	-	-
FA-11	-	-	FA-37	ohne Befund	-
FA-12	ohne Befund	upstream	FA-38	ohne Befund	-
FA-13	ohne Befund	upstream	FA-39	ohne Befund	upstream
FA-14	ohne Befund	upstream	FA-40	ohne Befund	-
FA-15	ohne Befund	-	FA-41	ohne Befund	upstream
FA-16	ohne Befund	downstream	FA-42	ohne Befund	upstream
FA-17	-	upstream	FA-43	ohne Befund	upstream
FA-18	ohne Befund	upstream	FA-44	ohne Befund	-
FA-19	$\Delta$ Exons 11-14 (het)	upstream	FA-45	ohne Befund	upstream
FA-20	ohne Befund	upstream	FA-46	-	downstream
FA-21	ohne Befund	upstream	FA-47	ohne Befund	-
FA-22	ohne Befund	upstream	FA-48	ohne Befund	-
FA-23	ohne Befund	upstream	FA-49	ohne Befund	downstream
FA-24	ohne Befund	FANCD2-defizient	FA-50	ohne Befund	-
FA-25	ohne Befund	-	FA-51	ohne Befund	-
FA-26	ohne Befund	-	FA-52	-	-
FA-27	-	FANCD2-defizient	FA-53	ohne Befund	-
FA-28	ohne Befund	-	FA-54	ohne Befund	-
FA-29	ohne Befund	downstream	FA-55	ohne Befund	-
FA-30	ohne Befund	upstream			

Eine weitere Klassifizierung wurde durch FANCD2-Immunblots erreicht. Insgesamt acht FA-Gene bilden zusammen eine funktionelle, FANC-D2/I-spezifische E3-Ligase. Ist diese defekt oder liegen Mutationen in FANCI vor, wird FANCD2 nicht mehr monoubiquitiniert, d.h. im Immunblot ist in diesem Fall die modifizierte Variante nicht mehr nachweisbar und der Patient wird einer Upstream-Komplementationsgruppe zugeordnet. Bei Mutationen in FANCD2 ist hingegen kein bzw. nur sehr schwach Restprotein nachweisbar, wohingegen bei Mutationen in einem Gen nach der FANCD2-Monoubiquitinierung (downstream) beide Isoformen detektiert werden können. Bei 30 Patienten konnte eine lymphoblastoide oder transformierte Fibroblasten-Zelllinie für diese funktionelle Untersuchung herangezogen werden. Zwei Drittel dieser Patienten ließen sich demnach einer Upstream-Gruppe zuordnen, wohingegen sieben Patienten FANCD2-Doppelbanden aufwiesen (Tabelle 11). Bei drei Patienten fehlten beide Banden, was als klarer Hinweis auf biallelische Mutationen in FANCD2 galt. Dabei handelte es sich um FA-09, FA-24 sowie FA-27. Bei einem weiteren Patienten, FA-08, existierten bereits mehrere FANCD2-Immunblots eines betroffenen Geschwisterkindes (FA-08.2), welches aufgrund der FANCD2-Defizienz der Komplementationsgruppe FA-D2 zugeordnet worden war. Ein Beispiel für solch einen FANCD2-Immunblot ist in Abbildung 23 dargestellt. Darauf ist zu erkennen, dass bei dem Geschwisterkind FA-08.2, im Vergleich zu einem bereits bekannten FA-D2-Patienten, Restprotein nachweisbar war. Im Gegensatz dazu waren bei der Analyse einer transformierten Fibroblastenkultur von FA-08 beide FANCD2-Banden deutlich zu erkennen (Abbildung 23). Zusätzlich waren noch weitere Banden einer geringeren Größe erkennbar. Um Mutationen in FANCD2 nachweisen oder ausschließen zu können, wurden die cDNAs beider Geschwister überprüft. Allerdings konnte dabei keine potentiell pathogene Mutation detektiert werden.



Abbildung 23: FANCD2-Immunblot der Geschwister FA-08 (links) und FA-08.2 (rechts) sowie von FA-12. Bei einer Kontrolle ist deutlich nicht-ubiquitiniertes (FANCD2-S) und ubiquitiniertes (FANCD2-L) FANCD2-Protein zu erkennen, welches bei einem bekannten FA-D2-Patienten (FA-D2) nicht nachweisbar war. RAD50 diente als Ladungskontrolle.

#### b. Erfolgreiche Komplementationsgruppen-Zuordnung

37 der 51 am GS Junior sequenzierten Patienten konnten eindeutig einer Komplementationsgruppe zugeordnet sowie die jeweils zugrunde liegenden Mutationen identifiziert werden. Bis auf FA-36, bei dem das komplette Material zur Anreicherung verwendet worden war, konnten bei allen klassifizierten Patienten die Mutationen per Sanger-Sequenzierung validiert werden. Allerdings konnte durch Sequenzierung der elterlichen DNAs von FA-36 zumindest die Segregation der Mutation c.1761-2A>C in *FANCG* bestätigt werden. Die parentale Segregation der Mutationen konnte bei weiteren 25 Fällen verifiziert werden ( $\triangleq$  67,6%), wohingegen bei drei Patienten nur ein Elternteil und bei den übrigen zehn Patienten kein Material der Eltern zur Verfügung stand. In Tabelle 12 sind die Ergebnisse der Mutationsanalyse aller zugeordneten Patienten aufgelistet. Darin ist neben dem Effekt auf Proteinebene auch vermerkt, ob die jeweilige Mutation bereits einen Eintrag in der FA-Datenbank hat und somit schon bei einem anderen FA-Patienten beschrieben wurde.

Pathogene Mutationen konnten in insgesamt zehn verschiedenen Komplementationsgruppen bestimmt werden. Ein Großteil entfiel dabei auf die Untergruppen FA-A (19 Patienten), FA-C (fünf Patienten) sowie FA-G (vier Patienten). Des Weiteren konnte bei drei Patienten die Komplementationsgruppe FA-D2 (FA-09, FA-24, FA-27) auf molekularer Ebene bestätigt werden. Dabei waren die Mutationen von FA-27 bereits im Vorfeld bekannt und der Patient war lediglich zur Bestätigung der Anreicherung der Alu-Mutation in *FANCD2* in die Analyse eingeschlossen worden. Zwei neue FA-I-Patienten (FA-33, FA-43) konnten ebenfalls identifiziert werden sowie je ein Patient der seltenen Komplementationsgruppen FA-E, -F, -L, -N und -T

Aus Tabelle 12 ist abzulesen, dass insgesamt 54 verschiedene Varianten detektiert werden konnten, wovon 19 in homozygotem Zustand vorlagen. Den geringsten Anteil hatten dabei Missense-Mutationen (n = 10). Des Weiteren konnten zwölf Nonsense- sowie 14 Indel-Mutationen bestätigt werden. Dabei umfasste die Deletion c.4268\_4368+37del, die heterozygot bei FA-13 sowie FA-45 nachgewiesen wurde, beinahe das komplette Exon 43 in *FANCA*. In der MLPA-Analyse war diese Deletion bei beiden Patienten zuvor nicht nachgewiesen worden. Ebenso konnte diese Deletion, wie auch die Deletion der Exons 11-14 in *FANCA* bei FA-19, nicht mithilfe des CNV-Tools der Auswertesoftware detektiert werden, welche zum Nachweis von Kopienzahl (*copy number variation*; CNV)-Varianten in den NGS-Daten genutzt werden kann. Den größten Anteil (n = 18) an allen nachgewiesenen Mutationen hatten Spleißmutationen. Bei drei bereits bekannten und z. T. auch in der Literatur beschriebenen Varianten konnte ein neuer Effekt auf Proteinebene gezeigt werden. Dazu zählt die 1-bp-Deletion c.3639del in *FANCA*, welche von Balta et al. [Balta et al., 2000] bei einem türkischen Patienten als Frameshift-Mutation mit dem Effekt p.E1214Rfs\*33 beschrieben wurde. Dieselbe Mutation wurde bei Patientin FA-13 detektiert, welche ebenfalls türkischer Abstammung ist. Eine Analyse der cDNA der Patientin ergab jedoch, dass es zum Skipping von Exon 37 kommt (Abbildung 24A), was eine Verschiebung des Leserasters sowie ein vorzeitiges Stoppcodon (p.F1210Yfs\*10) zur Folge hat.

Eine nähere Untersuchung der mutierten Sequenz mit Hilfe der Mutationsanalyse-Software Alamut, welche u. a. auf das Programm ESEFinder zugreift, ergab, dass es durch die Deletion zu einer Veränderung verschiedener exonischer *Splicing Enhancer*-Motive (ESE) kommt (Abbildung 24B).

Tabelle 12: Ergebnis der Mutationsanalyse unter Verwendung der Auswertesoftware von NextGENe. ID: Patienten-ID; FA-Gruppe: Komplementationsgruppen-Zuordnung; Ref: Referenz.
LOVD-ID: Die jeweilige ID der Mutation aus der Fanconi-Anämie-Mutationsdatenbank (http://www.rockefeller.edu/fanconi/) ist gelistet. Die in der Datenbank vorangestellten "0" jeder
ID wurden hier zur Übersichtlichkeit weggelassen. "-": keine Angabe. "?": unklar.

ID	FA-		Mutation A	llel 1		Mutation Allel 2				
ID	Gruppe	cDNA	Protein	LOVD-ID	Ref.	cDNA	Protein	LOVD-ID	Ref.	
FA-06	FA-A	c.856C>T	p.Q286*	FANCA_64	[Wijker et al., 1999]	c.4010+1_4010+18d el	p.D1312Vfs*26	FANCA_294	[Wijker et al., 1999]	
FA-09	FA-D2	c.982C>T	p.R328*	-	Neu	c.3707G>A	p.[H1229Wfs*4;H122 9Efs*7]	FANCD2_2	[Kalb et al., 2007]	
FA-12	FA-C	c.1585A>C	p.T529P	-	Neu	c.1585A>C	p.T529P		Neu	
FA-13	FA-A	c.3639del	p.F1210Yfs*10	FANCA_259	[Balta et al., 2000]	c.4268_4368+37del	p.A1423Efs*15	FANCA_543	[Moghrabi et al., 2009]	
FA-14	FA-A	c.893+1G>A	p.?	-	Neu	c.3391A>G	p.T1131A	FANCA_241	[Levran et al., 2005a]	
FA-15	FA-A	c.3348+1G>A	p.R1117Ifs*23	-	Neu	c.3348+1G>A	p.R1117Ifs*23	-	Neu	
FA-17	FA-A	c.1567-1G>T	p.V523Kfs*79	-	Neu	c.4009A>T	p.D1312Vfs*26	-	Neu	
FA-18	FA-A	c.2851C>T	p.R951W	FANCA_205	[Levran et al., 2005a]	c.2222+1G>T	p.R741Sfs*2	-	Neu	
FA-20	FA-C	c.45G>A	p.W15*	-	Neu	c.67del	p.D23Ifs*23	FANCC_6	[Strathdee et al., 1992a]	
FA-21	FA-A	c.1015G>A	р.А339Т	-	Neu	c.1015G>A	р.А339Т	-	Neu	
FA-22	FA-A	c.2316+1G>A	p.Q742*	-	Neu	c.2316+1G>A	p.Q742*	-	Neu	
FA-23	FA-C	c.165+1G>T	р.?	FANCC_24	-	c.165+1G>T	p.?	FANCC_24	-	
FA-24	FA-D2	c.696-121C>G	p.S232insQNNF *	FANCD2_2 7	[Kalb et al., 2007]	c.696-121C>G	p.S232insQNNF*	FANCD2_2 7	[Kalb et al., 2007]	
FA-25	FA-G	c.787C>T	p.Q263*	-	Neu	c.787C>T	p.Q263*	-	Neu	
FA-26	FA-A	c.2528A>C	p.Y843S	-	Neu	c.3391A>G	p.T1131A	FANCA_241	[Levran et al., 2005a]	
FA-27	FA-D2	c.3453_3456del	p.N1151Kfs*46	FANCD2_2 2	[Kalb et al., 2007]	c.274-5756insinv AluYb8 nt36_ 319+dup c.274-69 57	p.192Yfs*7	FANCD2_8	[Kalb et al., 2007]	
FA-28	FA-A	c.2638C>T	p.R880*	FANCA_499	-	c.2638C>T	p.R880*	FANCA_499	-	
FA-30	FA-F	c.193C>T	p.Q65*	FANCF_6	-	c.604del	p.L202*	FANCF_11	-	
FA-31	FA-C	c.1642C>T	p.R548*	FANCC_5	[Murer- Orlando et al., 1993]	c.1642C>T	p.R548*	FANCC_5	[Murer- Orlando et al., 1993]	

FA-32	FA-N	c.945_954del	p.P316Lfs*3	-	Neu	c.945_954del	p.P316Lfs*3	-	Neu
FA-33	FA-I	c.2957_2969del	p.V986Afs*39	-	Neu	c.3041G>A	p.C1014Y	-	Neu
FA-34	FA-G	c.313G>T	p.E105*	FANCG_15	[de Winter et al., 1998]	c.1182_1192delinsC	p.E395Wfs*5	FANCG_46	[Auerbach et al., 2003]
FA-35	FA-A	c.1361_1374deli nsGAG	p.A454Gfs*20	-	Neu	c.1361_1374delinsG AG	p.A454Gfs*20	-	Neu
FA-36	FA-G	c.1761-2A>C	p.?	-	Neu	c.1761-2A>C	p.?	-	Neu
FA-37	FA-L	c.1A>G	p.M1?	-	Neu	c.1096_1099dup	p.T367Nfs*13	FANCL_3	[Ali et al., 2009]
FA-38	FA-A	c.2796G>A	p.W932*	-	Neu	c.2796G>A	p.W932*	-	Neu
FA-39	FA-E	c.350_351del	p.V117Afs*11	-	Neu	c.491T>C	p.L164P	-	Neu
FA-40	FA-A	c.4261-2A>C	p.L1421Sfs*7	-	Neu	c.4261-2A>C	p.L1421Sfs*7	-	Neu
FA-41	FA-C	c.382del	p.D128Ifs*16	-	Neu	c.521+1G>A	p.M153Nfs*23	FANCC_28	-
FA-42	FA-A	c.2557C>T	p.R853*	FANCA_399	-	c.2557C>T	p.R853*	FANCA_399	-
FA-43	FA-I	c.2636+4A>G	p.D820_R879del	-	Neu	c.3041G>A	p.C1014Y	-	Neu
FA-45	FA-A	c.3919_3923dup	p.L1308Ffs*3	-	Neu	c.4268_4368+37del	p.A1423Efs*15	FANCA_543	[Moghrabi et al., 2009]
FA-46	FA-A	c.1304G>A	p.R435H	FANCA_599	[Moghrabi et al., 2009]	c.1304G>A	p.R435H	FANCA_599	[Moghrabi et al., 2009]
FA-48	FA-A	c.2606A>C	p.Q869P	FANCA_189	[Levran et al., 2005a]	c.2779-1G>T	p.?	-	Neu
FA-49	FA-T	c 65+600_468+38 4del	p.?	UBE2T_4	[Rickman et al., 2015]	c.384+4A>G	p.[G96_I128del; p.S129Vfs*23]	-	Neu
FA-50	FA-A	c.1267C>T	p.Q423*	FANCA_355	[Callen et al., 2005]	c.1267C>T	p.Q423*	FANCA_355	[Callen et al., 2005]
FA-51	FA-A	c.3520_3522del	p.W1174del	FANCA_253	[Levran et al., 1997]	c.3520_3522del	p.W1174del	FANCA_253	[Levran et al., 1997]
FA-54	FA-G	c.1649del	p.T550Ifs*9	FANCG_11	[Demuth et al., 2000]	c.1649del	p.T550Ifs*9	FANCG_11	[Demuth et al., 2000]

An diese Motive können verschiedene Serin-Arginin-reiche (SR) Proteine binden, welche essenzielle Spleißfaktoren darstellen. Durch die Deletion des Thymins an Position c.3639 fällt jeweils ein Sequenzmotiv für die Spleißfaktoren SF2/ASF, SC35, SRp40 sowie SRp55 weg, was eine mögliche Ursache des Exon-Skippings darstellt.



Abbildung 24: Alternatives Spleißen bei FA-13 aufgrund einer 1-bp-Deletion in *FANCA*. A: c.3639del führt auf cDNA-Ebene zum Skipping von Exon 37, nachgewiesen anhand von cDNA einer lymphoblastoiden Linie (LCL) von FA-13 im Vergleich zu einer Kontrolle. B: Übergang von IVS 36 zu Exon 37 (grau). Die kryptische Spleißakzeptorstelle (AG) in Intron 36 ist fett gedruckt. Die deletierte Base (rot) ist Teil von Erkennungssequenzen verschiedener Spleißfaktoren. Die Serin-Arginin-reichen (SR)-Proteine sind mit dem im ESEFinder angegebenen Score für die Wildtypsequenz (oberhalb des Exons) sowie für c.3639del (unterhalb des Exons) entsprechend zu ihrer Erkennungssequenz angegeben.

Im Gegensatz zur vorherigen Mutation war für c.3707G>A in *FANCD2* bereits eine Auswirkung auf das aberrante Spleißen nachgewiesen worden [Kalb et al., 2007]. Die Sequenzierung der cDNA von Patientin FA-09 ergab neben der beschriebenen Deletion von 44 bp (c.3684\_3727del) noch ein weiteres Spleißprodukt, das durch das Skipping der ersten 31 bp von Exon 37 zustande kommt (Abbildung 25A und B).



Abbildung 25: Nachweis von aberrantem Spleißen bei bereits beschriebenen Mutationen. A: Schematische Übersicht verschiedener aberranter Spleißprodukte, welche durch c.3707G>A (rot) in *FANCD2* hervorgerufen werden. Schwarze gestrichelte Linien oberhalb der Exons symbolisieren das Spleißen im Wildtyp (WT). Unterhalb ist das Spleißmuster bei Vorliegen der Mutation als gestrichelte und/oder gepunktete Linie dargestellt. Veränderungen durch die Mutation sind in rot hervorgehoben. B: Durch Sanger-Sequenzierung konnten die verschiedenen Spleißprodukte, wie in (A) dargestellt, verifiziert werden. C: Der Effekt von c.521+1G>A in *FANCC* konnte bei FA-41 als Skipping von Exon 6 bestimmt werden.

Eine weitere Mutation, c.521+1G>A in *FANCC*, fand sich bereits zwei Mal in der FA-Rockefeller-Datenbank (FANCC\_000028, Tabelle 14), allerdings war noch kein Effekt der Mutation bestimmt worden. Da es sich hierbei um die kryptische Spleißdonorstelle in Intron sechs handelt, war eine Auswirkung auf das normale Spleißmuster zu erwarten. Diese Annahme konnte durch die Sequenzierung der cDNA von FA-41 bestätigt werden, indem das Skipping von Exon 6 nachgewiesen werden konnte (Abbildung 25C).

## c. Charakterisierung neuer Mutationen in verschiedenen FA-Genen

Neben einer Vielzahl bereits bekannter Mutationen fanden sich 27 neue Mutationen im untersuchten Patientenkollektiv. Diese wurden durch *in silico*-Analyse auf ihre Pathogenität überprüft. Im Fall einer potentiellen Spleißmutation wurde diese, sofern geeignetes Patientenmaterial zur RNA-Isolation vorhanden war, auf cDNA-Ebene weiter charakterisiert.

## Missense-Mutationen

Zu einem Aminosäureaustausch aufgrund einer Basensubstitution kam es auf 15 Allelen bei 13 Patienten, wovon über die Hälfte der Mutationen (n = 6/10) bislang bei keinem FA-Patienten beschrieben worden war. Diese sind zur Übersicht in Tabelle 13 aufgeführt.

ID	Gen	Exon	cDNA	Protein	MutationTaster (p-value)	SIFT (score)	PolyPhen-2 (score)
FA-21	FANCA	12	c.1015G>A	р.А339Т	Disease causing (1)	Deleterious (0)	Probably damaging (0,996)
FA-26	FANCA	27	c.2528A>C	p.Y843S	Disease causing (0,99)	Deleterious (0,01)	Possibly damaging (0,753)
FA-12	FANCC	15	c.1585A>C	p.T529P	Disease causing (0,691)	Tolerated (0,07)	Probably damaging (0,940)
FA-39	FANCE	2	c.491T>C	p.L164P	Disease causing (1)	Deleterious (0)	Probably damaging (0,953)
FA-33, FA-43	FANCI	28	c.3041G>A	p.C1014Y	Disease causing (1)	Deleterious (0)	Probably damaging (0,969)
FA-37	FANCL	1	c.1A>G	p.M1?	Disease causing (1)	Damaging (0)	Benign (0)

Tabelle 13: Neue Missense-Mutationen. Die Klassifizierung der Varianten durch drei verschiedene Mutations-Vorhersageprogramme ist angeführt. Der vom jeweiligen Programm bestimmte Score ist in Klammern angegeben.

Vier dieser Varianten betreffen hoch konservierte Aminosäuren in den Upstream-Genen FANCA, FANCE und FANCI. So zeigte sich bei Patient FA-21 ein homozygoter Austausch eines Alanins nach Threonin (p.A339T) in Exon 12 von FANCA, welches bis zum Ortholog des Zebrafisches (*Danio rerio*) konserviert ist (Abbildung 26A), wohingegen das veränderte Tyrosin an Position 843 (p.Y843S) in Exon 27 (FA-26) nur bis zur Ebene des Kugelfisches (*Tetraodon nigorividis*) vorkommt (Abbildung 26B).

	Exon 12							Exon 27											
FANCA	A 336	S	D	A	¥ 340	Q	М	Q	R	835	F	C	Т	Å	A 840		S	Y	
<b>V</b> Orthologues (Source: Ensembl)	000																		
Human	A	S	D	A	v	Q	м	Q	R	K	F	C	T	A	A	1	s	Y	L
Chimp	A	s	D	A	v	Q	м	Q	R	К	F	C	т	A	A	1	5	Y	L
Northern white-cheeked gibbon					r –														Ē
Olive baboon	A	S	D	A	v	Q	M	Q	R	K	F	C	Т	A	A	1	5	Y	
Rat	V	5	D	A	1	Q	м	Q	K	К	F	C	т	A	A	1	s	Y	
Mouse	v	s	D	A	1	Q	м	Q	ĸ	N	F	C	т	A	A	٧	5	Y	L
Dog	V	S	D	A	v	R	L	Q	R	K	F	C	т	A	A	1	S	Y	
Platypus	A	5	D	A	1	R	М	Q	R	R	F	C	Т	A	A	٧	5	Y	L
Chicken	¥	S	D	A	1	н	K	Q	R	R	F	C	Т	A	A	v	s	Y	L
Frog	v	S	D	A	1	Q	C	Q	A	K	F	C	т	A	A	1	S	Y	
Tetraodon	v	S	D	A	v	A	L	Q	K	R	F	C	L	L	A	v	C	Y	
Zebrafish	v	S	D	A	1	A	v	0	s	C	L	C		v	C	v	C	V	Ē

Abbildung 26: Hoch konservierte Aminosäuren in Exon 12 und 27 von FANCA, welche bei zwei FA-A-Patienten durch eine Missense-Mutation verändert sind. Dargestellt ist die Ansicht des Softwareprogramms Alamut, welches u. a. auf die Datenbank Ensembl zugreift. Je höher die Konservierung, desto dunkler sind die Aminosäuren hinterlegt. Bei FA-21 kommt es in Exon 12 durch die homozygote Mutation c.1015G>A zum Austausch eines Alanins (roter Rahmen) gegen ein Threonin. Bei FA-26 führt dagegen die heterozygote Substitution c.2528A>C zum Austausch eines konservierten Tyrosins in Exon 27 (p.Y843S).

Im N-terminalen Bereich von FANCE konnte auf dem maternalen Allel einer indischen Patientin (FA-39) der Austausch einer hochkonservierten Aminosäure nachgewiesen werden. So wird durch die Substitution c.491C>T (Abbildung 27A) in Exon 2 ein Leucin durch ein Prolin ersetzt.



Abbildung 27: Missense-Mutationen in weiteren FA-Genen. A: Die Substitution c.491T>C (Stern) in *FANCE* konnte bei FA-39 sowie deren Mutter heterozygot nachgewiesen werden. B: Das Cystein an Position 1014 in FANCI (roter Rahmen) ist hoch konserviert und befindet sich in einer  $\alpha$ -helikalen Solenoid-Domäne (grün). Alle konservierten Aminosäuren sind gelb hinterlegt. Verändert nach [Joo et al., 2011]. C: NextGENe-Darstellung der homozygoten Missense-Mutation c.1585A>C in *FANCC* bei FA-12. Angegeben ist die revers-komplementäre Sequenz. D: Bestätigung von c.1A>G (Stern) in *FANCL* bei FA-37. Die Mutter war hier Überträgerin der Mutation. Das Startcodon ist grau hinterlegt. "Ex": Exon.

Des Weiteren sind beide neu identifizierten FA-I-Patientinnen (FA-33 und FA-43) heterozygot für die Missense-Mutation c.3041G>A (rs140404896), welche zum Austausch eines konservierten Cysteins durch ein Prolin in der Solenoid-3-Domäne führt (p.C1014Y) (Abbildung 27B). Im Browser des *Exome Aggregation* 

*Consortium* (ExAC) findet sich diese Variante, als einzige der neu beschriebenen Missense-Mutationen, mit einer Allelfrequenz von 7,42\*10-5 gelistet.

Nur eine moderate Konservierung weist hingegen die Aminosäure Threonin in Exon 15 des *FANCC*-Gens an Position 529 auf. Der zugrunde liegende Basenaustausch c.1585A>C (rs587778326) wurde mit einer Abdeckung von 67x homozygot in FA-12 detektiert und ist in Abbildung 27C dargestellt. Der Patient war zuvor mittels FANCD2-Immunblot-Analyse als *upstream* eingestuft worden (Abbildung 23). Die parentale Segregation der Mutation konnte hier bestätigt werden. Dies galt ebenso für Patient FA-37, für den die Mutation c.1A>G in *FANCL* dem maternalen Allel zugeordnet wurde (Abbildung 27D). Eine verlässliche Vorhersage durch SIFT und PolyPhen-2 war hier nicht gegeben, da beide Programme den Verlust des Startsequenz zur Translation verloren geht, kann diese Mutation als krankheitsverursachend eingestuft werden.

# Nonsense-Mutationen

Substitutionen, die ein Basentriplett zu einem Stopp-Codon ändern, werden von MutationTaster automatisch als krankheitsverursachend eingestuft. Bei den anderen beiden verwendeten Mutationsvorhersage-Programmen, SIFT und PolyPhen-2, ist die Vorhersage eines solchen Austausches nicht möglich.

Die meisten Nonsense-Mutationen (n = 5/12) konnten in *FANCA* nachgewiesen werden. Darunter befand sich eine unbeschriebene homozygote Substitution in Exon 29 (c.2796G>A), welche bei FA-38, einer Patientin syrischer Abstammung, identifiziert werden konnte (Tabelle 12). Abbildung 28A zeigt diese Mutation, welche zum Austausch eines Tryptophans gegen ein Stoppcodon an Aminosäureposition 932 führt. Die Segregation der Mutation innerhalb der Familie konnte ebenfalls bestätigt werden.

Da bei den Eltern von FA-25 asiatischer Abstammung ein konsanguiner Hintergrund vorlag, wurde auch hier nach homozygoten Varianten gefiltert. Im Vorfeld war die Diagnose FA aufgrund von klinischen Merkmalen gestellt und durch die Untersuchung des G2-Phase-Arrests per Durchflusszytometrie gesichert worden. Als besondere Merkmale wurde ein duplizierter Daumen sowie die Graufärbung der Haare im Kleinkindesalter beschrieben, welche später aber nachdunkelten. In *FANCG* fand sich die homozygote Mutation c.787C>T (rs149721361) (Abbildung 28B), die bereits in Exon 7 zu einem vorzeitigen Stopp-Codon (p.Q263\*) führt und bei beiden Eltern heterozygot vorlag. Diese Variante hatte als einzige der neu entdeckten Nonsense-Mutationen eine dbSNP-Nummer sowie einen Eintrag im ExAC-Browser. Dort wird die Variante einmal heterozygot bei einer Person aus Südasien aufgeführt und mit einer Allelfrequenz von 8,238\*10<sup>-6</sup> angegeben. Weiterhin konnte bei FA-20 aus Deutschland in *FANCC* eine Basensubstitution ermittelt werden, welche auf Proteinebene zu einem vorzeitigen Abbruch der Translation führt. In diesem Fall (c.45G>A) kommt es bereits schon in Exon 2 zu einem Stoppcodon (p.W15\*). Die Mutation war in den NGS-Daten 56x abgedeckt. Nur wenige Basen weiter *domnstream* konnte zudem die zweite Mutation, c.67del, auf denselben Reads detektiert werden (Abbildung 28C). Anhand der NGS-Daten wird deutlich, dass die Mutationen auf unterschiedlichen Allelen liegen, da auf einem Read jeweils nur eine Mutation vorkommt. Eine parentale Zuordnung war aus Ermangelung elterlicher DNA allerdings nicht möglich. Anders verhielt es sich bei einer weiteren deutschen Patientin, FA-09. Durch einen FANCD2-Immunblot war diese vor der Anreicherung schon durch eine Defizienz an FANCD2-Protein aufgefallen (Abbildung 28D). Neben einer paternalen Missense-Mutation (c.3707G>A) in Exon 37 von *FANCD2* konnte die compound-heterozygote Nonsense-Mutation c.982C>T (p.R328\*) in Exon 12 eindeutig dem maternalen Allel zugeordnet werden, wohingegen das paternale Allel die Wildtypsequenz an dieser Stelle aufwies (Abbildung 28E).



**Abbildung 28:** Neu identifizierte Nonsense-Mutationen in verschiedenen FA-Genen. **A:** Ansicht der homozygoten Nonsense-Mutation c.2796G>A (blau) in Exon 29 von *FANCA* bei FA-38 in der NextGENe-Software (oben). Dargestellt ist die komplementäre Sequenz. Sanger-Sequenzierung von DNA aus Blut bestätigte die Mutation (Stern) (unten). **B:** Segregation der Substitution c.787C>T (Stern) in *FANCG* bei FA-25, welche ebenfalls zu einem vorzeitigen Stopp führt. **C:** Bei FA-20 konnte heterozygot eine Substitution (c.45G>A; blau) sowie eine 1 bp-Deletion (c.67del; rosa) in Exon 2 von *FANCC* detektiert werden. Abgebildet ist die Reverssequenz in der Darstellung der NextGENe-Software. **D:** Patientin FA-09 (roter Rahmen) war per FANCD2-Immunblot der Komplementationsgruppe FA-D2 zugeordnet worden. **E:** Eine heterozygote Nonsense-Mutation (c.982C>T) (Stern) in *FANCD2* konnte in gDNA aus primären Fibroblasten (Fib) bei FA-09 detektiert und die familiäre Segregation per Sanger-Sequenzierung bestätigt werden. "Ex": Exon.

# Indel-Mutationen

Die Mehrzahl aller identifizierten Indel-Mutationen (n = 12/14) hatte auf cDNA-Ebene eine Verschiebung des Leserasters zur Ursache. Eine Ausnahme bildet dabei die große Deletion in *UBE2T* von FA-49, welche die Exons zwei bis sechs umfasst und auch das in Exon zwei befindliche Startcodon betrifft (Tabelle 12). Zudem wurde eine *in-frame* Deletion der Aminosäure Tryptophan an Position 1174 in Exon 36 des *FANCA*-Gens bei FA-51 homozygot nachgewiesen. Mit der Vorhersage bezüglich der Pathogenität von Indel-Mutationen verhält es sich wie bei den Nonsense-Mutationen.

Sechs Indel-Mutationen waren vorher noch bei keinem FA-Patienten beschrieben worden und werden im Folgenden kurz erläutert. In *FANCA* konnte zunächst eine homozygote Indel-Mutation in Exon 15 (c.1361\_1374delinsGAG) bei FA-35 bestimmt werden. Die Deletion von neun Nukleotiden mit gleichzeitiger Insertion der Sequenz GAG war von der NGS-Mutationsanalysesoftware zwar erkannt worden (Abbildung 29A). Im Mutationsbericht der Software war sie allerdings als c.1361\_1370del aufgeführt, da die Insertion von drei Basen nicht mit berücksichtigt wurde. Durch Sanger-Sequenzierung ließ sich die Mutation richtig definieren. Neben der bei bereits zwei FA-L-Patienten beschriebenen Duplikation c.1096\_1099dup (FANCL\_000003), die auch bei FA-37 heterozygot nachgewiesen wurde (Tabelle 12), war c.3919\_3923dup in *FANCA* bei FA-45 die einzige pathogene Duplikation, welche bei den untersuchten Patienten gefunden wurde (Abbildung 29B).

Daneben konnten vier kleinere Deletionen in der untersuchten FA-Patientengruppe ermittelt werden. Dazu zählt c.382del in *FANCC* (p.D128Ifs\*16), welche heterozygot bei einer aus Afrika stammenden Patientin (FA-41) detektiert wurde. Die Bestätigung der Mutation durch Sanger-Sequenzierung ist in Abbildung 29C dargestellt. Bei FA-39 konnte in Exon 2 von *FANCE* neben der bereits beschriebenen Missense-Mutation c.491T>C die Deletion c.350\_351del bestätigt werden. Die Deletion betrifft eine kurze, 2-bp-Repeatsequenz. Diese Mutation fand sich allerdings nicht in den NGS-Daten wieder (Abbildung 29D), sondern wurde erst durch Sanger-Sequenzierung der Patienten-cDNA nachgewiesen (Abbildung 29E). Wie sich herausstellte, handelte es sich bei der zur Anreicherung verwendeten DNA nicht um die der Indexpatientin, sondern um ein nicht-betroffenes Geschwisterkind, das lediglich heterozygoter Träger der Missense-Mutation ist. Aus einer lymphoblastoiden Linie der Patientin wurde gDNA isoliert und diese Mutation verifiziert (Abbildung 29E). Weiterhin konnte die Deletion dem paternalen Allel zugeordnet werden.

Auch bei FA-33 finnischer Abstammung konnte die zweite compound-heterozygote Mutation in Form einer Deletion identifiziert werden. Hier war bereits die Missense-Mutation c.3041G>A (p.C1014Y) in *FANCI* dem paternalen Allel zugeordnet worden, wohingegen die Deletion c.2957\_2969del (rs762390984) in Exon 27 von der Mutter der Patientin übertragen wurde. In den Daten des *1000 Genomes Projects* findet sich diese Sequenzveränderung nicht, im ExAC-Browser ist sie hingegen 24 Mal, jeweils heterozygot, gelistet. Am häufigsten (n = 19) wurde diese Variante in Probanden finnischer Abstammung (n = 6614) nachgewiesen, was einer Allelfrequenz von 2,873\*10-3 entspricht. Betrachtet man die Mutation der finnischen Personen im IGV-

alignt wurde. Ein Ausschnitt dieser Ansicht ist in Abbildung 29E dargestellt. A Position 16:89,851,355 16:89,851,360 16:89,851,375 16:89,851,365 16:89,851,370 15< 460 455 s G s R Translation G G G G G Reference Consensus G G G G č č Pile-Up

Viewer, so ist zu erkennen, dass die Deletion im Viewer nicht entsprechend der tatsächlichen Bruchpunkte



Abbildung 29: Neu identifizierte Indel-Mutationen in verschiedenen FA-Genen. A: Die Indel-Mutation c.1361\_1374delinsGAG (blau) in Exon 15 von *FANCA* wurde bei FA-35 homozygot nachgewiesen. Abgebildet ist die Mutation in der Darstellung der NextGENe-Software, welche die revers-komplementäre Sequenz angibt. B: Eine Duplikation von 5 bp (unterstrichen) konnte in Exon 39 von *FANCA* bei FA-45 nachgewiesen werden. C: Bei FA-41 fand sich dagegen eine 1 bp-Deletion in *FANCC* (c.382del; unterstrichen). D: In den NGS-Daten konnte c.350\_351del in *FANCE* (roter Rahmen) nicht identifiziert werden, da es sich um das nicht betroffene Geschwisterkind handelte. Durch die Analyse der cDNA und gDNA einer lymphoblastoiden Zelllinie (LCL) der Index-Patientin (FA-39) konnte die compound-heterozygote Mutation identifiziert werden. E: IGV-Ansicht zu c.2957\_2969del (roter Rahmen) in *FANCI* bei einem finnischen Probanden, der im ExAC-Browser mit dieser Mutation heterozygot gelistet ist. Grau: Abdeckung. Rot: Forward-Reads. Blau: Revers-Reads. Schwarze Linie: Deletion. Orange: Austausch C>G. F: Ansicht der Anreicherungsdaten zu FA-32 im IGV-Viewer. Dargestellt ist die Deletion c.945\_954del (roter Rahmen) in *PALB2*. Farbgebung: siehe E. "Ex": Exon.

Auch die letzte neue Indel-Mutation, c.945\_954del (rs781282667) in *PALB2*, fand sich heterozygot im ExAC-Browser wieder und war dort mit einer Allelfrequenz von 8,25\*10-6 angegeben. Bei den Eltern von FA-32 war eine Konsanguinität bekannt und erwartungsgemäß lag diese Mutation homozygot vor (Abbildung 29F). Des Weiteren war im Vorfeld aufgrund des schweren Phänotyps (u.a. Wilms-Tumor im Kleinkindesalter) der Verdacht auf Mutationen in *PALB2* gestellt worden, was anhand der NGS-Datenanalyse bestätigt werden konnte.

# Spleißmutationen

Am häufigsten (n = 18) konnten Mutationen nachgewiesen werden, die sich auf den Spleißvorgang auswirken und so zu einer veränderten mRNA-Synthese führen. Darunter befanden sich elf Substitutionen, die noch nicht in der internationalen FA-Mutationsdatenbank gelistet waren und in Tabelle 14 zusammengefasst sind. In zwei Fällen war kein Patientenmaterial zur RNA-Isolation vorhanden, so dass ein Effekt auf das Spleißen nicht nachgewiesen werden konnte. Dazu gehört die Veränderung der kanonischen Spleißakzeptorsequenz in Intron 28 von *FANCA*, verursacht durch c.2779-1G>T (FA-48). Zudem konnte der Effekt von c.1761-2A>C in *FANCG* (rs765150956) nicht bestimmt werden (FA-36). Diese Variante fand sich im ExAC-Browser bei einer Person heterozygot aufgeführt, die, wie FA-36, südasiatischer Abstammung war.

ID	Gen	Exon/ Intron	cDNA	Protein	RNA	Bemerkung		
FA-14	FANCA	IVS10	c.893+1G>A	p.?	r.[888_893delins893 +87_893+215;?]	Insertion Exon 10a & Exonisierung IVS 10a		
FA-17	FANCA	IVS16	c.1567-1G>T	p.V523Kfs*79	r.1567_1576del	Skipping von 10 bp von Exon 17		
FA-18	FANCA	IVS24	c.2222+1G>T	p.R741Sfs*2	r.2222_2223ins2222 +1_2222+304	Exonisierung IVS 24		
FA-22	FANCA	IVS25	c.2316+1G>A	p.?	[r.2222_2223ins2222 +1_2222+304;r.231 6_2317ins2316+1_2 316+141]	Exonisierung IVS 24 & IVS 25		
FA-48	FANCA	IVS28	c.2779-1G>T	p.?	r.?	-		
FA-15	FANCA	IVS33	c.3348+1G>A	p.R1117Ifs*23	r.3348_3349ins3348 +1_3348+205	Exonisierung von 205 bp von IVS 33		
FA-17	FANCA	Exon 40	c.4009A>T	p.D1312Vfs*26	r.3935_4010del	Exon 40 Skipping		
FA-40	FANCA	IVS42	c.4261-2A>C	p.L1421_D1427 delinsSPSPVV*	r.4260_4261ins4261- 66_4261-1	Exonisierung von 66 bp von IVS 42		
FA-36	FANCG	IVS13	c.1761-2A>C	p.?	r.?	-		
FA-43	FANCI	IVS24	c.2636+4A>G	p.D820_R879del	r.2457_2636del	Exon 24 Skipping		
FA-49	FANCT	IVS5	c.384+4A>G	p.G96_I128del	[r.286_384del;r.384_ 385ins384+1_384+1 42]	Exon 5 Skipping & Exonisierung von IVS 5		

Tabelle 14: Übersicht über neu identifizierte Spleißmutationen in FA-Genen.

Wie alle der in Tabelle 14 aufgelisteten neuen Spleißmutationen wurden die beiden nicht näher untersuchten Varianten von MutationTaster als krankheitsverursachend eingestuft (p-value  $\geq 0,995$ ). Verschiedene

Spleißvorhersageprogramme hatten jeweils den Wegfall der kanonischen Spleißakzeptorstelle angezeigt (Abbildung 30A+B). Wie in Abbildung 30A dargestellt, wurde für c.2779-1G>T in *FANCA* von drei Programmen, die mit der Alamut-Software verlinkt sind, die Aktivierung einer kryptischen Akzeptorstelle in Exon 29 angegeben.



**Abbildung 30:** Spleißvorhersage durch die Software Alamut. **A:** Für c.2779-1G>T in *FANCA* gaben alle fünf Vorhersageprogramme den Verlust des kryptischen Spleißakzeptors an. Durch drei Programme wurde die Entstehung einer alternativen Akzeptorstelle vorhergesagt und bewertet. **B:** Bei c.1761-2A>C in *FANCG* wurde ebenfalls von allen Programmen der Verlust der Spleißakzeptorstelle angegeben.

Für die übrigen neun Varianten konnte eine Auswirkung auf das Spleißen durch die Analyse der cDNA bestätigt werden. Dazu zählt die Substitution c.893+1G>A (IVS 10) in *FANCA*, welche heterozygot bei FA-14 türkischer Abstammung sowie einem ebenfalls betroffenen Geschwisterkind nachgewiesen werden konnte. Für die kanonische Spleißdonorstelle von Intron 10 ist bereits der Austausch G>T (FANCA\_000069) bei italienischen Patienten in der Literatur beschrieben [Savino et al., 1997; Savino et al., 2003]. Darin wird neben der Insertion des alternativ gespleißten Exons 10a (*Accession No.* AJ131188), welches eine Größe von 129 bp besitzt, die Exonisierung von 86 bp von Intron 10a (Sequenz zwischen Exon 10 und 10a) beschrieben. Das Vorhandensein von Exon 10a wurde auch bei Kontrollen gezeigt, was durch eigene cDNA-Analysen bestätigt werden konnte. Analog zur bei Savino et al. 2003 [Savino et al., 2003] angegebenen PCR konnten bei FA-14 ebenfalls drei PCR-Produkte bestimmt werden. Diese wurden jeweils aus dem Agarosegel ausgeschnitten, aufgereinigt und sequenziert. Neben der Wildtypsequenz, welche dem kleinsten PCR-Produkt entsprach, ließ sich die Exonisierung von Intron 10a bestätigen (Produkt mittlerer Größe) (Abbildung 31A). Allerdings kam es schon sechs Basen weiter *upstream* als zuvor beschrieben zur Überlagerung der Sequenzen. Dies lässt sich auf die Verwendung eines kryptischen Spleißdonors in Exon 10 zurückführen (Abbildung 31A+C), welcher von verschiedenen Vorhersageprogrammen (MaxEntScan, GeneSplicer, Human Splicing Finder) als

kryptische Spleißstelle angegeben war. Eine Erhöhung des Scores für diese Stelle durch die Mutation c.893+1G>A hatte jedoch keines der Programme angegeben. Die komplette Sequenz des größten PCR-Produktes war nicht auswertbar, der Übergang von Exon 11 zu Exon 10a konnte mit dem Revers-Primer jedoch nachgewiesen werden (Abbildung 31B). Zusätzlich wurde die Exonisierung von Intron 10a bestimmt.



Abbildung 31: Aberrantes Spleißen durch c.893+1G>A in *FANCA*, das bei FA-14 nachgewiesen werden konnte. A: Bei der Isolation und Sequenzierung des PCR-Produktes mittlerer Länge (M) konnte neben der Wildtypsequenz die Insertion von Exon 10a nachgewiesen werden. Dabei kommt es zum Skipping der letzten sechs Basen von Exon 10 (unterstrichen). B: Das längste Produkt ("L") zeigte neben der Insertion von Exon 10a die Exonisierung von Intron 10a. C: Eine Übersicht des beschriebenen aberranten Spleißens ist schematisch zusammengefasst. In der unteren Hälfte der Darstellung ist das alternative Spleißen durch die Wildtypsequenz abgebildet, welches auch die Insertion von Exon 10a beinhaltet (graue gestrichelte Linie). Die Substitution von c.893+1G>A (Pfeil) führt zur Aktivierung einer kryptischen Spleißstelle in Exon 10 (rot) und hat die Insertion von Exon 10a (rote gestrichelte Linie) oder zusätzlich die Exonisierung von Intron 10a (gepunktete rote Linie) zur Folge. Verändert nach [Savino et al., 2003].

Wie aus Tabelle 14 ersichtlich und in Abbildung 32 dargestellt, konnten drei weitere neue Substitutionen an kanonischen Spleißdonorstellen in *FANCA* identifiziert werden, welche jeweils die Exonisierung intronischer Sequenzen zur Folge haben. Die Diagnose FA war bei FA-18 anhand von peripheren Blutlymphozyten gestellt worden, bei der es keine Hinweise für eine Mosaikbildung gab. Zur Anreicherung war deshalb gDNA aus einer Blutprobe verwendet worden, die wenige Monate nach der Diagnosestellung zur Subtypisierung eingesandt wurde. Erst nachdem in den NGS-Daten nur eine als pathogen eingestufte heterozygote Mutation in *FANCA* gelistet und diese Missense-Mutation (c.2851C>T) per Sanger-Sequenzierung bestätigt worden war, wurde DNA aus primären Fibroblasten auf weitere Varianten überprüft. Anhand dieser Probe konnte c.2222+1G>T (rs775388912) als zweite compound-heterozygote Mutation festgestellt (Abbildung 32A) und der Nachweis für ein bestehendes Mosaik im hämatopoetischen System erbracht werden. Beide Mutationen konnten formal jeweils einem paternalen Allel zugeordnet werden. In den Spleißvorhersageprogrammen war durch diese Mutation, neben dem Verlust der kanonischen Spleißdonorstelle, auch der Ausfall einer SF2/ASF-Spleißenhancerstelle angegeben. Auf cDNA-Ebene konnte schließlich die Exonisierung der Intron 24-Sequenz nachgewiesen werden (Abbildung 32B). Die Exonisierung eines ganzen Introns wird ebenfalls

durch c.2316+1G>A ausgelöst, welche homozygot bei FA-22 in den NGS-Daten gelistet war. Hier konnte auf cDNA-Ebene die homozygote Exonisierung von Intron 25 (Abbildung 32C) nachgewiesen werden. Zusätzlich wurde die heterozygote Exonisierung von Intron 24 sowie einer Teilsequenz von Intron 24 ab Position c.2222+102 festgestellt (Abbildung 32D). Bei der NGS-Analyse von FA-15 konnte in Intron 33 mit c.3348-1G>A eine weitere Spleißdonor-Veränderung in *FANCA* belegt werden. Diese Substitution hat die Aktivierung einer kryptischen Spleißdonorstelle in Intron 33 zur Folge (Abbildung 32E), was zur Exonisierung von 205 bp führt (r.3348\_3349ins3348+1\_3348+205).



**Abbildung 32:** Spleißdonormutationen in *FANCA*. **A:** Bei der Sequenzierung von DNA aus Blut und Fibroblasten (Fib) von FA-18 konnte die Reversion der Substitution c.2222+1G>T (Stern) nachgewiesen werden. **B:** Auf cDNA-Ebene wurde als Folge dieser Spleißmutation die Exonisierung von Intron 24 bestimmt. **C:** Die Substitution c.2316+1G>A (Stern) in Intron 25 von *FANCA* konnte bei FA-22 verifiziert werden. Durch cDNA-Sequenzierung wurde die Exonisierung der IVS 25-Sequenz nachgewiesen. **D:** Daneben konnte die Exonisierung von IVS 24 belegt werden. Dabei fiel ein weiteres Spleißprodukt auf, das sich als r.2222\_2223ins2222+102\_2213-1 herausstellte und als heterozygote Überlagerung erkennbar ist. **E:** Bei FA-15 führt die homozygote Mutation c.3348+1G>A (Stern) in *FANCA* zur Exonisierung der ersten 205 bp von IVS 24. "Ex": Exon. "IVS": Intron. "LCL": Lymphoblastoide Zelllinie.

Ferner konnten in *FANCA* zwei Mutationen bestimmt werden, welche die kanonische Spleißakzeptorsequenz verändern. Dazu zählt c.1567-1G>T in Intron 16, die bei FA-17 in den NGS-Daten bestimmt wurde. Spleißvorhersageprogramme hatten die potentielle Nutzung einer kryptischen Spleißakzeptorstelle innerhalb von Exon 17 angezeigt (Abbildung 33A).



**Abbildung 33:** Spleißakzeptormutationen in *FANCA*. **A:** Von verschiedenen Vorhersageprogrammen wurde als Folge von c.1567-1G>T in Intron 16 die Aktivierung eines kryptischen Spleißakzeptors in Exon 17 angegeben. Dargestellt ist die Ansicht in Alamut. Das Skipping von 10 bp (rot hinterlegt) konnte auf cDNA-Ebene bestätigt werden. **B:** Die Mutation c.1567-1G>T (Stern) war nicht in einer lymphoblastoiden Linie (LCL), aber in DNA aus primären Fibroblasten (Fib) nachweisbar. **C:** Dagegen wurde die Mutation c.4009A>T in Exon 40 bei FA-17 in beiden Zelltypen bestätigt. **D:** Schematische Übersicht des aberranten Spleißens, ausgelöst durch c.4009A>T. Wildtypisches Spleißen ist in der unteren Hälfte abgebildet. Durch die Substitution in Exon 40 (Pfeil) kommt es sowohl zum Skipping von Exon 40 (c.3035\_4010del) (rote gestrichelte Linie) als auch zur Aktivierung einer kryptischen Spleißakzeptorstelle in Exon 41 (rot), welche das Skipping von Exon 40 sowie der ersten 23 bp von Exon 41 zur Folge hat (rote gepunktete Linie). **E:** Bestätigung des unter (**D**) beschriebenen aberranten Spleißens bei FA-17 aufgrund von c.4009A>T. **F:** Bei FA-40 konnte homozygot die Exonisierung von 66 bp von IVS 42 beobachtet werden, ausgelöst durch c.4261-2A>C (Stern). "Ex": Exon. "IVS": Intron.

Die Sequenzierung von cDNA aus primären Fibroblasten von FA-17 bestätigte das Skipping der ersten 10 bp von Exon 17. Dies war allerdings erst nach der Stabilisierung der Allele durch Zugabe von CHX möglich, da sonst die Sequenz aufgrund zu schwacher Signalintensitäten nicht auswertbar war. In einer lymphoblastoiden Linie zeigte sich die Mutation revertiert (Abbildung 33B). Ob diese Reversion auch im Blut vorlag bzw. sich dort durchgesetzt hatte, konnte nicht ermittelt werden, da keine weiteren Blutproben von FA-17 zur Verfügung standen. In den NGS-Daten des Patienten war zudem die als Missense-Mutation bewertete Variante c.4009A>T in Exon 40 gelistet. Diese war sowohl in der lymphoblastoiden Linie als auch in primären Fibroblasten verifizierbar (Abbildung 33C). Da die Mutation die 5'-Spleißkonsensus-Sequenz betrifft, wurde auch der entsprechende Bereich der cDNA auf Veränderungen überprüft. Tatsächlich konnte das Skipping von Exon 40 detektiert werden (Abbildung 33D+E), womit die Mutation als Spleißmutation einzuordnen ist. Daneben zeigte sich ein weiteres Spleißprodukt, das sich als c.3035\_4033del bestimmen ließ und durch die Verwendung eines kryptischen Spleißaktzeptors in Exon 41 zustande kommt (Abbildung 33D+E). Bei der zweiten Spleißakzeptormutation handelte es sich um c.4261-2A>C, welche bei FA-40 homozygot im Mutationsbericht der NGS-Auswertesoftware NextGENe aufgeführt war. Durch Sanger-Sequenzierung konnte die familiäre Segregation verifiziert und durch cDNA-Sequenzierung die Exonisierung von 66 bp aus IVS 42 als Resultat der veränderten Spleißakzeptorsequenz aufgezeigt werden (Abbildung 33F). Anhand von Tabelle 14 wird deutlich, dass zwei weitere heterozygote Mutationen je vier Basenpaare 3' vom jeweiligen Exon in detektiert wurden, für die ein Effekt auf das aberrante Spleißen nachgewiesen werden konnte. Dabei handelt es sich um c.2636+4A>G in FANCI sowie um c.384+4A>G in UBE2T/FANCT. Auf diese Spleißmutationen soll in den folgenden Abschnitten gesondert eingegangen werden.

#### Alternatives und aberrantes Spleißen bei FANCI und Charakterisierung neuer FANCI-Mutationen

Bei FA-43, einer Patientin mit klassischen phänotypischen Merkmalen wie Kleinwuchs und Daumenfehlbildungen, aber nur leichten hämatologischen Auffälligkeiten, war zunächst die paternal vererbte Missense-Mutation c.3041G>A in *FANCI* aufgefallen, welche auch schon für FA-33 bestimmt wurde (Tabelle 13). In unmittelbarer Nähe der Substitution wurde in den NGS-Daten die heterozygote Deletion c.3058+1del angezeigt, welche durch Sanger-Sequenzierung nicht bestätigt wurde. Dagegen konnte c.2636+4A>G sowohl bei FA-43 als auch bei deren Mutter verifiziert werden. Das Programm RESCUE-ESE zeigte als Resultat den Verlust einer Spleißenhancer-Hexamersequenz an. Durch cDNA-Sequenzierung konnte eine Veränderung des kryptischen Spleißens nachgewiesen werden, welches nicht nur bei der Patientin, sondern auch bei verschiedenen Kontrollen bestimmt werden konnte (Abbildung 34A). Unter den Kontrollen befanden sich auch verschiedene FA-I-Patienten, die im Vorfeld durch Immunblot-Analysen oder klassische Sanger-Sequenzierung der Komplementationsgruppe zugeordnet worden waren. Die Mutationen dieser Patienten sind in Tabelle 15 zusammengetragen. Zudem wurden zwei bereits veröffentlichte FA-I-Patienten, EUFA961 [Dorsman et al., 2007] und F010191 [Sims et al., 2007], als Kontrollen verwendet. Auch hierzu finden sich die Mutationen in Tabelle 15 wieder.

Patienten-		Mutation Allel 1		Mutation Allel 2					
ID	cDNA	Protein	Bemerkung	cDNA	Protein	Bemerkung			
FA-33	c.2957_2969de 1	p.V986Afs*39	-	c.3041A>G	p.C1014Y	-			
FA-43	c.2636+4A>G	p.D820_R879del	Exon 24 Skipping	c.3041A>G	p.C1014Y	-			
FAI-01	c.881A>T	p.L253*	Exon 10 Skipping	c.1992+1G>A	p.L631_L664del	Exon 20 Skipping			
FAI-02.1	c.3853C>G	p.R1285G	-	c.3853C>G	p.R1285G	-			
FAI-02.2	c.3853C>G	p.R1285G	-	c.3853C>G	p.R1285G	-			
FAI-03	c.3907G>T	p.E1303*		c.3949C>T	p.Q1317*	-			
FAI-04	c.3293_3295de 1	p.E1098del	-	c.3058+4A>G	p.F1003Rfs*9	Exon 28 Skipping			
EUFA961	c.2572C>T	p.H858_R879del	<i>in-frame</i> Deletion	c.3437_3455de 1	p.E1117fs	Exon 32 Skipping			
F010191	c.1840C>T	p.R614*	-	c.3895C>T	p.R1299*	-			

Tabelle 15: FANCI-Mutationen aller neu zugeordneten Patienten dieser Komplementationsgruppe. FA-33 und FA-43 wurden durch NGS-Analysen identifiziert, die anderen Patienten durch klassische Mutationsanalyse-Methoden.

Bislang findet sich aus Tabelle 15 nur c.3058+4A>G sowie der Austausch eines Arginins gegen Glutamin an Position 1285 in der internationalen FA-Mutationsdatenbank wieder. Ein Effekt der potentiellen Spleißmutation c.3058+4A>G ist darin nicht angegeben. Durch Sequenzierung der cDNA von FAI-04 konnte das Skipping von Exon 28 als Effekt der Substitution in Intron 28 belegt werden. Dem in der FA-Mutationsdatenbank gelisteten Aminosäureaustausch p.R1285G lag die Substitution c.3854G>A zugrunde, während bei den Geschwistern FAI-02.1 und FAI-02.2 homozygot der Austausch c.3853C>G nachgewiesen werden konnte, welcher zur selben Aminosäuresubstitution führt (Tabelle 15).

Anhand des in Abbildung 34A dargestellten Agarosegels sowie der Elektropherogramme in Abbildung 34B wird ersichtlich, dass bei FA-43 quantitativ gesehen das Skipping von Exon 24 deutlich stärker war als bei den Kontrollen. Die Pathogenität der biallelischen *FANCI*-Mutationen bei FA-43 konnte durch einen FANCI-Immunblot bestätigt werden, da hierdurch die fehlende Ubiquitinierung des Proteins gezeigt werden konnte. Gleichzeitig war neben der FANCI-S-Bande eine zweite Bande verminderter Größe zu erkennen, wie es auch bei FAI-01 der Fall war (Abbildung 34C). Eine Ubiquitinierung des FANCI-Proteins war bei allen untersuchten FA-I-Patienten ebenfalls nicht nachweisbar. Zudem war bei FAI-03 die FANCI-S-Bande leicht in ihrer Größe vermindert. Allen FA-I-Patienten war gemeinsam, dass keine biallelischen Nullmutationen vorlagen und überall Restprotein vorhanden war (Abbildung 34C+D).


Abbildung 34: Charakterisierung von *FANCI*-Mutationen. A: Die Gelelektrophorese einer cDNA-PCR mit Primern in Exon 23 und 25 (Produktgröße 373 bp) macht deutlich das Skipping von Exon 24 bei FA-43 sowie schwach bei allen weiteren Proben erkennbar (roter Rahmen), wodurch ein Produkt von 193 bp entsteht. Bei EUFA961 liegt heterozygot eine Deletion von 66 bp in Exon 24 vor, nachweisbar durch eine weitere Bande bei 307 bp. Die Größen des 100 bp-Standards (Ladder) sind links neben dem Gelbild angegeben. B: Nachweis des Exon-24-Skippings bei FA-01 sowie FAI-43 und einer Kontrolle. Bei FA-43 dominiert die Sequenz des Skippings. C: Ein FANCI-Immunblot zeigt bei allen FA-I-Patienten sowie einem FA-C-Patienten die fehlende Monoubiquitinierung (FANCI-L). Bei FA-43 (roter Rahmen) ist ein verkürztes Protein zu erkennen. Bei allen FA-I-Patienten ist Restprotein vorhanden. Vinculin diente als Ladungskontrolle. D: Schematische Übersicht der FANCI-Proteinstruktur und zu den neu identifizierten Mutationen der sieben Patienten aus sechs Familien. Funktionelle Domänen sind in der Proteinstruktur farbig abgesetzt und die entsprechenden Aminosäuren (*amino acid*; aa) sind dazu angegeben. Unterhalb des Schemas sind die unbeschriebenen Mutationen mit dem jeweiligen Effekt auf die Proteinlänge dargestellt. Größere Deletionen sind als rotes Rechteck an der betroffenen Stelle eingezeichnet. Mutationen, welche nur eine Aminosäure betreffen, sind mit einem roten Pfeil markiert. "Ub": Ubiquitinierung. "SIM": *SUMO-like domain-interacting motif.* "NLS": Kernlokalisierungs-Signal (*nuclear localization signal*). "ARM": Armadillo. "EDGE": Aminosäureabfolge (Glutaminsäure-Asparaginsäure-Glycin-Glutaminsäure).

#### Identifizierung und Charakterisierung eines FA-T-Patienten

Patient FA-49 wurde als erstes Kind deutscher Eltern mit einem Anhängsel an einem Daumen, Café-au-lait-Flecken sowie einem Ventrikelseptumdefekt geboren. In der zweiten Lebensdekade wurde er erstmals aufgrund einer Thrombozytopenie und Makrozytose sowie einer Leukopenie auffällig, wobei keine weiteren phänotypischen Auffälligkeiten festgestellt werden konnten. Beide Eltern sind gesund, die Großmutter sowie eine Großtante mütterlicherseits waren an einem Mamma- bzw. Uteruskarzinom erkrankt wohingegen beim Großvater väterlicherseits ein Pankreaskarzinom diagnostiziert worden war.

Eine Untersuchung peripherer Lymphozyten auf einen erhöhten spontanen ( $\Sigma G2/GF = 0,430$ ) sowie MMCinduzierten G2-Phase-Arrest ( $\Sigma G2/GF = 0,581$ ) bestätigte die Verdachtsdiagnose FA. In den NGS-Daten fand sich homozygot die potentielle Spleißmutation c.384+4A>G im UBE2T/FANCT-Gen (Abbildung 35A), welche per Sanger-Sequenzierung bestätigt werden konnte. Eine Überprüfung der parentalen Segregation zeigte, dass nur der Vater, nicht aber die Mutter, heterozygoter Überträger der Mutation war (Abbildung 35B).



Abbildung 35: Spleißmutation in *UBE2T* bei FA-49 und ihr Effekt auf cDNA- und Proteinebene. A: In den NGS-Daten wurde die potentielle Spleißmutation c.384+4A>G (blau) in Intron 5 bei FA-49 homozygot angezeigt. Dargestellt ist die Ansicht im NextGENe-Viewer, welcher die revers-komplementäre Sequenz angibt. B: Bestätigung der Mutation (Stern) per Sanger-Sequenzierung. Der Vater war heterozygoter Überträger, die Mutter zeigte die Wildtypsequenz an dieser Stelle. C: Die gelelektrophoretische Auftrennung einer cDNA-PCR mit Primern in Exon 3 und 7 zeigte drei deutliche Banden beim Patienten (776 bp, 633 bp und 534 bp) (roter Rahmen). Bei der Kontrolle war schwach eine zweite Bande bei 776 bp zu erkennen. Die Größen des 100 bp-Standards (Ladder) sind links angegeben. D: Die Sequenzierung des PCR-Produkts bestätigte alternatives Spleißen bei FA-49. "Ex": Exon. "IVS": Intron.

Durch Untersuchung der cDNA konnte die Mutation als Spleißmutation bestätigt werden. Bei der elektrophoretischen Auftrennung des PCR-Produktes war bei FA-49 jeweils eine Bande zu erkennen, die größer (776 bp) bzw. kleiner (534 bp) als die zu erwartende Bande der Wildtypsequenz (633 bp) war (Abbildung 35C). Bei der Kontrolle konnte man ebenso schwach eine Bande bei 776 bp erkennen, allerdings

nur, wenn eine große Menge des PCR-Produktes aufgetragen wurde. Bei der Sequenzierung der Kontrolle konnte dennoch nur die Wildtypsequenz nachgewiesen werden (Abbildung 35D). Dagegen konnte beim Patienten sowohl die Exonisierung von Intron 5 ( $\triangleq$  776 bp) als auch das Skipping von Exon 5 ( $\triangleq$  534 bp) neben Vorhandensein der Wildtypsequenz ( $\triangleq$  633 bp) verifiziert werden (Abbildung 35D).

Als mögliche Ursache für das vermeintlich homozygote, möglicherweise aber hemizygote Vorliegen der Spleißmutation wurde eine große Deletion in *UBE2T* angenommen. Für einen der zwei bisher beschriebenen Patienten war eine große Deletion, entstanden durch die Rekombination zweier Alu-Elemente in *UBE2T*, bereits beschrieben [Rickman et al., 2015; Virts et al., 2015]. Durch die Amplifikation der cDNA mit Primern in der 5'- und 3'-UTR konnte beim Vater von FA-49 sowie einer Kontrolle nur eine Bande der Größe von knapp 800 bp detektiert werden. Beim Patienten wie auch bei dessen Mutter war zusätzlich eine zweite Bande bei etwa 250 bp vorhanden. Durch Gelisolation und Sequenzierung dieser kleineren Bande konnte schließlich die Deletion c.-64\_468del bestätigt werden (Abbildung 36A).



Abbildung 36: Deletion in *UBE2T* bei FA-49. A: Auf cDNA-Ebene konnte bei FA-49 nach der Gelisolation verschiedener PCR-Produkte die Deletion c.-64\_468del ermittelt werden. Die Mutter erwies sich als Überträgerin. B: Gelelektrophorese einer gDNA-PCR zum Nachweis der Rekombination zweier Alu-Elemente in IVS 1 und IVS 6, welche die Deletion zur Folge haben. Links sind die Größen des 100 bp-Standards (Ladder) angegeben. Das Prinzip der PCR ist in (C) abgebildet (verändert nach [Rickman et al., 2015]).

Codierende Exons sind dunkelgrau, nicht-codierende Exons sowie die UTRs hellgrau dargestellt. Primer sind schematisch als blaue Pfeile abgebildet. Der Übergang der Introns 1 und 6 zum dazwischen liegenden Alu-Element konnte per Sequenzierung bestätigt werden ( $\mathbf{D}+\mathbf{E}$ ). 732 ist heterozygoter Träger dieser Deletion und diente als Positivkontrolle.

Mit Hilfe einer spezifischen PCR, welche nur die Amplifikation eines durch Deletion verkürzten Templates ermöglicht, konnte auf gDNA-Ebene sowohl bei FA-49, bei dessen Mutter, als auch bei einer Positivkontrolle ein Produkt bestätigt werden (Abbildung 36B+C). Bei dieser Kontrolle handelt es sich um eine revertierte lymphoblastoide Zelllinie des bei Rickman et al. 2015 [Rickman et al., 2015] und Virts et al. 2015 [Virts et al., 2015] beschriebenen FA-T-Patienten. In dieser Zelllinie wie auch im Blut des Patienten war nicht die Deletion, sondern die zweite compound-heterozygote Mutation revertiert, bei der es sich um eine Duplikation handelt (c.110-280\_468+264dup). Diese revertierte LCL wird im Folgenden als 732 bezeichnet.

Per Sanger-Sequenzierung konnte die Rekombination zweier Alu-Elemente in Intron 1 und 6 bestätigt werden, wodurch es zur Deletion der dazwischen liegenden Sequenz kommt (Abbildung 36D+E). Allerdings ist auf Abbildung 36B zu erkennen, dass auch beim Vater eine schwache Bande bei der Gelelektrophorese erkennbar war. Kontaminationen von daneben liegenden Gelspuren sowie des PCR-Ansatzes konnten ausgeschlossen werden. Das schwache PCR-Produkt ließ sich jedoch nicht sequenzieren. Eine Untersuchung des UBE2T-Proteins durch eine Immunblot-Analyse ergab, dass bei FA-49 noch schwach Restprotein auf Höhe des wildtypischen Proteins nachweisbar war (Abbildung 37A). Zudem wurde ein FANCD2-Blot zur Bestätigung der *Upstream*-Komplementationsgruppe durchgeführt. Allerdings konnten hier beide Banden detektiert werden (Abbildung 37B). Eine Positivkontrolle in Form von Protein eines anderen UBE2T-Patienten stand dabei nicht zur Verfügung, es konnte nur die revertierte Linie von Patient 732 verwendet werden.





Abbildung 37: Immunblot-Analysen mit FA-49. A: Ein Antikörper gegen UBE2T wies bei FA-49 Restprotein nach. Zudem sind unspezifische Banden zu erkennen B: Beim Nachweis der FANCD2-Monoubiquitinierung konnten bei einer revertierten UBE2T-Zelllinie (732) wie auch bei FA-49 und dessen Vater die nicht-ubiquitinierte Isoform (FANCD2-S; 155 kDa) sowie die monoubiquitinierte Form (FANCD2-L; 162 kDa) detektiert werden. Vinculin diente jeweils als Ladungskontrolle.

#### d. Nicht zugeordnete Patienten trotz Mutationsanalyse

Von insgesamt 51 unklassifizierten Patienten, die mithilfe des FA-Genpanels angereicherten und am GS Junior sequenziert wurden, war bei zwei Patienten (FA-05, FA-11) der Verdacht auf FA bereits auf funktioneller Ebene ausgeschlossen worden. Hier diente die Mutationsanalyse dem Ausschluss von pathogenen Varianten in bekannten FA- oder bestimmten Kandidatengenen. Dagegen konnten 11 weitere Patienten nicht oder nicht eindeutig einer FA-Komplementationsgruppe zugeordnet werden, obwohl durchflusszytometrische Untersuchungen die Diagnose FA bestätigt hatten. Die Ergebnisse zu diesen Patienten sollen im Folgenden erläutert werden.

Bei sechs Patienten ohne Klassifizierung konnte keine potentiell pathogene Mutation bestimmt werden. Dabei handelte es sich um FA-08, FA-29, FA-47, FA-52, FA-53 sowie FA-55. Drei dieser Patienten (FA-29, FA-47, FA-55) ergab die Untersuchung peripherer Blutlymphozyten am Durchflusszytometer ein grenzwertiges bzw. negatives Ergebnis (Abbildung 38), wohingegen primäre Fibroblastenkulturen eine Hypersensitivität gegenüber MMC aufwiesen (Tabelle 16).



Abbildung 38: Zellzyklusanalyse des G2-Phasen-Anstieges PHA-stimulierter peripherer Blutlymphozyten potentieller Mosaikpatienten nach 72-stündiger Inkubation ohne (links) bzw. mit 10 ng/ml MMC (rechts) sowie BrdU. Zur Zellfärbung wurde EtBr sowie HOECHST 33258 verwendet. Aufgetragen ist das Verhältnis der Summe der G2-Phasen zur Wachstumsfraktion gegen den Anteil nicht-proliferierender Zellen. Blau: FA-29; Grün: FA-47; Rot: FA-55; Graue Kreise: Kontrollen. Graue Rauten: FA-Patienten

Deshalb wurde hier jeweils eine Reversion im Blut vermutet und DNA aus primären Fibroblasten zur Mutationsanalyse herangezogen. Dennoch konnten keine potentiell krankheitsverursachenden Varianten festgestellt werden. Für FA-29 wurde später ein WES durchgeführt, dessen Ergebnisse gesondert in Abschnitt 4.4.4a beschrieben werden.

Fibroblasten ±MMC	G1-Phase	S-Phase	G2-Phase
Normale Kontrolle –MMC	86,8	16,7	12,4
Normale Kontrolle +MMC	79,5	6,3	14,2
FA-Kontrolle –MMC	78,7	3,8	17,5
FA-Kontrolle +MMC	52,3	4,8	42,9
FA-29 – MMC	80,2	10,2	9,6
FA-29 +MMC	60,2	6,7	33,1
FA-47 –MMC	61,2	8,1	30,7
FA-47 +MMC	50,4	6,3	43,3
FA-55 –MMC	65,3	11,7	23,8
FA-55 +MMC	46,7	5,1	48,2

**Tabelle 16:** Zellzyklusverteilung potentieller Mosaikpatienten. Je eine Fibroblastenkultur blieb unbehandelt oder wurde für 48 h mit 12 ng/ml MMC inkubiert. Zur Messung am Durchflusszytometer wurden die Zellen mit DAPI gefärbt. Zum Vergleich ist das Ergebnis einer normalen Kontrolle sowie eines bekannten FA-Patienten angegeben.

Nur jeweils eine monoallelische, potentiell pathogene Mutation konnte bei der Auswertung der NGS-Daten von FA-07 und FA-16 detektiert und per Sanger-Sequenzierung bestätigt werden. Unterschiedliche Mutationsvorhersage-Programme stuften diese beiden Varianten übereinstimmend als pathogen ein (Tabelle 17). Dabei handelt es sich um die heterozygote Substitution c.1972C>T in *FANCM*, die bei FA-07 bestätigt wurde und auf Proteinebene ein vorzeitiges Stoppcodon zur Folge hat (p.R658\*). Zum anderen konnte bei FA-16 die monoallelische Missense-Mutation c.1765C>T (p.R589W) in *FANCQ* nachgewiesen werden. Beide Patienten wurden später noch mit dem *TruSight Caner*-Panel angereichert oder durch WES analysiert.

Bei FA-10, FA-19 sowie FA-44 konnten dagegen lediglich verschiedene Varianten unbekannter Signifikanz (VUS) ermittelt werden, welche in Tabelle 17 aufgelistet sind. Die Einstufung der Veränderungen durch verschiedene Mutationsvorhersage-Programme war nicht eindeutig sondern z. T. sogar widersprüchlich. Beispielsweise wurde die Missense-Mutation c.280A>T in *FANCB*, welche bei FA-10 hemizygot vorlag (Abbildung 39A), sowohl als Polymorphismus als auch als schädlich eingestuft (Tabelle 17). Die formale Bestätigung der Segregation zeigte, dass die Mutter nicht Trägerin der Mutation ist (Abbildung 39B) und die Variante folglich *de novo* entstanden sein muss. Da DNA aus Blutlymphozyten zur Anreicherung verwendet worden war, wurde im Anschluss auch DNA aus primären Fibroblasten auf die Variante hin untersucht um auszuschließen, dass es sich um einen Effekt der entsprechenden Linie handelt. Dabei konnte die Mutation auch in diesem Gewebetyp bestätigt werden (Abbildung 39B).

ID	Gen	cDNA	Protein	dbSNP	Ab- deckung	Mutation Taster (p-value)	SIFT (score)	PolyPhen- 2 (score)
FA-07	FANCM	c.1972C>T	p.R658*	rs368728 266	13x	Disease causing (1)	-	-
FA-10	FANCB	c.280A>T	p.194F	-	7x	Polymor- phism (1)	Dele- terious (0,01)	Possibly damaging (0,847)
FA-16	FANCQ	c.1765C>T	p.R589W	rs14710 5770	60x	Disease causing (1)	Dele- terious (0)	Probably damaging (1,0)
FA-19	FANCA	c.894- ?_1359+?del	p.F299Pfs* 72	-	-	-	-	-
FA-44	FANCA	c.3782T>C	p.F12618	-	27x	Polymor- phism (1)	Dele- terious (0,01)	Possibly damaging (0,773)
FA-44	FANCL	c.238C>G	p.L80V	rs56351 3081	33x	Disease causing (1)	Dele- terious (0)	Probably damaging (0 <b>,</b> 999)

**Tabelle 17:** Mono- und biallelische Varianten in FA-Patienten, die nicht eindeutig einer Komplementationsgruppe zugeordnet werden konnten. Die Abdeckung der Variante in der NextGENe-Software ist angegeben, ebenso die Klassifizierung durch drei verschiedene Vorhersageprogramme. Der jeweils vom Vorhersageprogramm bestimmte Score ist in Klammern eingetragen. "-": keine Angabe.

Dagegen war bei FA-19 bereits durch eine MLPA-Analyse die heterozygote Deletion der Exons 11-14 in *FANCA* nachgewiesen worden (Tabelle 11), woraufhin dieses Gen auf weitere Varianten untersucht wurde. Allerdings fand sich bei der Durchsicht der GS Junior-Daten keine pathogene Veränderung, sodass die cDNA analysiert wurde. Dabei zeigte sich neben der bereits detektierten Deletion das Skipping von Exon 30 (r.2853\_2981del) (Abbildung 39C), welches auf Proteinebene zu einer *in-frame* Deletion von 43 Aminosäuren führt. Jedoch konnte dieses aberrante Spleißverhalten auch bei Kontrollen beobachtet werden (Abbildung 39C), weshalb eine krankheitsverursachende Wirkung als unwahrscheinlich eingestuft wurde. Bei einem FANCA-Immunblot konnte dagegen kein FANCA-Protein bei FA-19 nachgewiesen werden (Abbildung 39D), weshalb der Patient vermutlich dennoch der Komplementationsgruppe FA-A zuzuordnen ist.

Bei FA-44 konnten hingegen gleich mehrere Varianten in den NGS-Daten bestimmt werden, deren Vorhersage hinsichtlich der Pathogenität teilweise unterschiedlich ausfiel (Tabelle 17). Dabei handelt es sich zum einen um die homozygote Missense-Mutation c.3782C>T in Exon 38 von *EANCA*, welche zum Austausch von Phenylalanin zu Serin an Position 1261 führt. Wie in Abbildung 39E ersichtlich, konnte die Variante per Sanger-Sequenzierung bestätigt werden. Sie findet sich bislang nicht in der FA-Mutationsdatenbank der Rockefeller-Universität aufgeführt. Zum anderen fand sich die heterozygote Substitution c.238C>G in *FANCL* (Abbildung 39F). Diese wurde von allen Vorhersageprogrammen als pathogen klassifiziert, wie in Tabelle 17 dargestellt. Allerdings konnte keine zweite Mutation in *FANCL* in den NGS-Daten gefunden werden. Elterliche DNA zur Bestätigung der Segregation beider Mutationen lag nicht vor. Aus diesen Gründen galt dieser Fall als nicht eindeutig geklärt.



Abbildung 39: Varianten unbekannter Signifikanz, die bei FA-10, FA-19 sowie FA-44 bestimmt werden konnten. A: Ansicht der Mutation c.280A>T in *FANCB* im NextGENe-Analyseprogramm (revers-komplementäre Sequenz) nach Anreicherung und Sequenzierung von gDNA aus peripheren Blutlymphozyten von FA-10. B: Die Mutation (Stern) konnte auch in gDNA aus primären Fibroblasten des Patienten, nicht aber bei dessen Mutter nachgewiesen werden. C: Die Sequenzierung der *FANCA* cDNA-Sequenz wies bei FA-19 das Skipping von Exon 30 nach, welches auch bei Kontrollen detektiert werden konnte. D: Durch einen FANCA-Immunblot konnte die Defizienz des Proteins bei verschiedenen Patienten detektiert werden, darunter auch bei FA-19 (roter Rahmen). E: Bei FA-44 fand sich die homozygote Substitution (c.3782T>C, Stern) in  $\blacklozenge$  bei der Analyse von DNA aus primären Fibroblasten ("Fib"). F: Zudem konnte die heterozygote Missense-Mutation c.238C>G (Stern) in *FANCL* verifiziert werden. "Ex": Exon.

#### 4.3 Mutationsanalyse von FA-Patienten mit Hilfe eines kommerziellen Krebspanels

Mit Hilfe des *TruSight Cancer*-Panels wurden 19 verschiedene Patienten angereichert. Darunter befanden sich fünf Patienten, welche zuvor schon mit Hilfe einer *SeqCap EZ*-Library angereichert und am GS Junior sequenziert worden waren. Drei dieser Patienten dienten dabei als Positivkontrollen (FA-02, FA-13 und FA-34). Hier waren die Mutationen bereits bekannt und das *TruSight*-Panel sollte damit auf die Einsatzmöglichkeit in der FA-Diagnostik getestet werden. Dagegen war bei FA-07 und FA-24 die Komplementationsgruppen-Zuordnung trotz FA-Panelanalyse nicht erfolgt. Alle Sequenzierungen fanden am MiSeq statt und auf den entsprechenden Datensatz soll zunächst im Folgenden eingegangen werden.

#### 4.3.1 Alignment mit Hilfe der NextGENe-Software

#### a. Abdeckung der gesamten ROI nach der Sequenzierung am MiSeq

Ähnlich zum GS Run Browser von Roche kann mit Hilfe der MiSeq-Reporter-Software von Illumina eine Rohdatenanalyse erfolgen. Im Bericht dieser Software wurde der Prozentsatz an Basen mit einer Sequenzgenauigkeit von 99,9 % (Q30-Wert) aller mittels MiSeq sequenzierten Patienten mit Werten zwischen 92,8 und 96,8 % angegeben. Weiterhin wurde die mediane Readlänge mit 120 bis 150 bp bestimmt. Beide genannten Werte lagen für FA-07 und FA-24 nicht vor, da der vollständige Report der Software nicht verfügbar war.

Die Anzahl der am MiSeq generierten Reads sowie die Verteilung der Reads nach Konvertierung und Alignment durch die Auswertesoftware NextGENe ist für alle 19 FA-Patienten in Abbildung 40A dargestellt.



Abbildung 40: Übersicht zu ausgewählten Rohdaten der 19 Patienten, die mit dem *TruSight Cancer*-Panel angereichert und am MiSeq sequenziert wurden. A: Anzahl der Reads nach der Sequenzierung am MiSeq sowie der Datenkonvertierung und Alignment durch die NextGENe-Software. Die Grenzen der Box markieren jeweils die 25. und 75. Perzentile, die Linie im Inneren den Median. Ein schwarzes Kreuz markiert den Durchschnitt. Die Whisker (gestrichelte vertikale Linien) entsprechen dem 1,5-fachen Interquartilsabstand (nach Tukey). Die einzelnen Patienten sind als schwarze Kreise eingetragen. **B:** Aufgetragen ist die durchschnittliche Abdeckung der ROI des *TruSight Cancer*-Panels. Erklärung der Symbole: siehe **A**.

Nur wenige Reads wurden aufgrund der Filtereinstellungen, die in der Auswertesoftware bei der Datenkonvertierung und dem Alignment gesetzt wurden, herausgefiltert. Im Durchschnitt konnten von 3,17\*10<sup>6</sup> Reads, welche am MiSeq sequenziert wurden, 3,06\*10<sup>6</sup> Reads alignt werden. Die durchschnittliche Abdeckung der Zielregionen in den NGS-Daten der einzelnen Patienten ist in Abbildung 40B grafisch aufgetragen. Diese reichte von 380x bis 891x, wobei alle Werte innerhalb des 1,5-fachen Interquartilsabstand lagen.

Betrachtet man verschiedene Grenzwerte zur durchschnittlichen Abdeckung (Abbildung 41), so erkennt man, dass bis zu 200x die Werte aller Anreicherungen sehr nahe beieinander lagen. Im Mittel lag bei 96,3 % der Zielregion eine 100-fache Abdeckung vor und selbst bei 200x waren noch 90 % abgedeckt. In der Darstellung der Daten als Boxplot sind dennoch einzelne Ausreißer zu erkennen (Abbildung 41). Dabei handelt es sich jeweils um FA-07 sowie bei 100x und 200x zusätzlich um FA-24. Dennoch lagen die Ergebnisse sehr nahe beieinander, der Abstand zum unteren Whisker beträgt nur maximal drei Prozent. Bei einer höheren Abdeckung war die Streuung hingegen deutlich größer und die Boxen deshalb auch breiter. Auch in Abbildung 41 würden sich, ähnlich wie bei Abbildung 20, die Mittelwerte zu einer abfallenden Kurve verbinden lassen. Die deutlichste Abnahme hinsichtlich der durchschnittlichen Abdeckung wäre demnach von



200x zu 400x auszumachen. Bei einer 1000-fachen Abdeckung waren im Durchschnitt noch immer 13,3 % der ROI enthalten.

Abbildung 41: Abdeckung der ROI nach Datenkonvertierung und Alignment der 19 am MiSeq sequenzierten FA-Patienten. Die ROI wurde nach dem Mappen der Reads auf ihre Abdeckung überprüft. Die Grenzen der Box markieren jeweils die 25. bzw. 75. Perzentile, die Linie im Inneren den Median. Ein schwarzes Kreuz markiert zudem den Durchschnitt. Die Whisker (gestrichelte vertikale Linien) entsprechen dem 1,5-fachen Interquartilsabstand (nach Tukey). Ausreißer sind als schwarze Kreise außerhalb der Whisker dargestellt.

#### b. Abdeckung einzelner Zielregionen

Analog zur Auswertung der Sequenzierungsdaten des GS Juniors wurde die Abdeckung der ROI der MiSeq-Daten auf Lücken und Ausfälle einzelner Targets überprüft. Insgesamt umfasste die Zielregion des Krebspanels 1 737 Regionen. Eine Lücke in einem Read wurde auch hier als Unterbrechung von mindestens einer Base gewertet. In Abbildung 42 ist der jeweilige Prozentsatz der lückenhaften oder vollständig fehlenden Zielregionen dargestellt. Im Durchschnitt waren nur 0,58 % der Targets durch eine Lücke unterbrochen. Die Anzahl an Zielregionen, welche durch keine Base abgedeckt war, lag mit knapp 0,2 % noch niedriger. Alle Werte lagen konstant bei unter 1 % und es gab zwischen den verschiedenen Projekten keine Ausfälle zu verzeichnen.



Abbildung 42: Anteil der Targets an der ROI des *TruSight Cancer*-Panels, welche nur lückenhaft (blau) oder überhaupt nicht (rot) durch Reads nach der Sequenzierung abgedeckt waren. "\*" markiert Patienten, welche zuvor bereits durch ein *SeqCap EZ*-Panel angereichert und am GS Junior sequenziert worden waren. Horizontale gestrichelte Linien geben die Mittelwerte des Anteils an Lücken (blau) und komplett fehlenden Targets (rot) an. Die jeweilige Standardabweichung der Mittelwerte ist als blaue (Lücken) bzw. rote (Totalausfall) Fläche eingezeichnet.

Nur wenige Targets zeigten wiederkehrend eine unzureichende Abdeckung. Regionen, welche bei  $\geq 50$  % der Patienten Lücken aufwiesen oder komplett fehlten, sind in Tabelle 18 zusammengefasst. Dabei fällt auf, dass kein einziger Totalausfall einer kompletten Region zu verzeichnen war. Größtenteils besaßen die unvollständig abgedeckten Bereiche einen hohen GC-Gehalt. Lediglich ein FA-Gen, *FANCE*, war in diesen Regionen enthalten.

**Tabelle 18:** Übersicht über Regionen, welche bei mehr als der Hälfte der analysierten Patienten (n = 19) durch Lücken unterbrochen waren (Lücke) oder komplett fehlten (x). Für jede dieser unzureichend abgedeckten Zielregionen ist die chromosomale Position (Chr = Chromosom), das Gen sowie das jeweils betroffene Exon (CDS = *coding sequence*) angegeben. Der GC-Gehalt der Region sowie die Häufigkeit einer lückenhaften Abdeckung bzw. des kompletten Fehlens der Zielregion wurden bestimmt.

Chr	<b>Chr Position Start</b>	Chr Position Ende	Gen	CDS	GC [%]	<b>Σ(</b> Lücke)	<b>Σ(</b> x)
2	48 010 373	48 010 633	MSH6	1	71,6	17	0
2	96 930 876	96 931 120	TMEM127	1	72,7	14	0
4	41 747 824	41 748 340	PHOX2B	3	72,3	19	0
5	176 562 105	176 563 032	NSD1	1	42,7	10	0
6	35 420 323	3 5420 571	FANCE	1	79,1	17	0
11	2 905 900	2 906 720	CDKN1C	1	76,9	19	0
11	32 456 246	32 456 892	WT1	1	72,8	16	0
19	33 792 244	33 793 321	CEBPA	1	74,8	19	0
21	36 164 432	36 164 908	RUNX1	8	71,5	19	0

#### 4.3.2 Bestätigung bereits zuvor bekannter Mutationen

Wie anfangs erwähnt waren fünf Patienten bereits mit einem SeqCap-Panel angereichert und am GS Junior sequenziert worden, wobei bei drei Patienten damit auch die jeweils compound-heterozygot vorliegenden

Mutationen bestimmt werden konnten. Dazu gehören FA-02 (FA-Q), FA-13 (FA-A) sowie FA-34 (FA-G) (Tabelle 6 und Tabelle 12). In Tabelle 19 sind die Ergebnisse zur Abdeckung und Allelfrequenz der Mutationen dieser drei Patienten zusammengefasst. Wie es sich bereits in den bisher beschriebenen Daten widerspiegelte, war die Abdeckung am MiSeq generell und auch für die betrachteten Mutationen um ein Vielfaches höher als am GS Junior.

Tabelle 19: Übersicht zur Abdeckung und Allelfrequenz der pathogenen Mutationen von drei Patienten, die auf verschiedene Weise angereichert und sequenziert wurden. Bei Indel-Mutationen wurde jeweils die minimale und maximale Abdeckung bzw. Allelfrequenz angegeben.

ID Gen			SeqCap E2	Z/GS Junior	TruSight Cancer/MiSeq		
		cDNA	Abdeckung	Allelfrequenz	Abdeckung	Allelfrequenz	
<b>EA 02</b>	E ANCO	c.689T>C	15	40	353	56	
<b>FA-02</b> <i>FAN</i>	FANCQ	c.2371_2399dup	19	58	424	11	
FA 13	EANCA	c.3639del	24	44	208	42	
<b>FA-13</b> <i>F</i> /-	TANCA	c.4268_4368+37del	41-47	43-49	38-174	17-79	
EA 34	EANCC	c.313G>T	26	23	444	35	
гл-34	TZINCG	c.1182_1192delinsC	22-23	44-46	170-177	57-59	

Für FA-13 und FA-34 konnten die bekannten Mutationen durch die Anreicherung mit dem Krebspanel und MiSeq-Sequenzierung bestätigt werden. Allerdings wurde auch hier die heterozygote Deletion c.4268\_4368+37del nicht durch das CNV-Tool der NextGENe-Software detektiert (vgl. 4.2.3b). Die heterozygote Duplikation c.2371\_2399dup in *FANCQ* bei FA-02 wurde hingegen in den NGS-Daten der *TruSight Cancer*-Anreicherung als Sequenzierfehler markiert und nicht im Report aufgeführt, obwohl dieselben Einstellungen zur Mutationsanalyse wie bei der Auswertung der GS Junior-Daten gewählt worden waren. Zur Erkennung einer Variante waren hier jeweils als Schwellenwert 15% eingestellt worden. Erst durch ein Herabsetzen des Grenzwertes wurde die Duplikation erkannt, da die Allelfrequenz der Mutation bei einer Abdeckung von 424x nur knapp 11% betrug.

# 4.3.3 Nachweis von Mutationen, die mittels SeqCap EZ-Panelsequenzierung nicht detektiert wurden

Bei der zum Zeitpunkt der Analyse 38 Jahre alten Patientin FA-07 war die Diagnose FA anhand einer Blutprobe durch eine durchflusszytometrische Untersuchung des spontanen und induzierten G2-Phase-Arrestes bestätigt worden. Neben klassischen Merkmalen, wie ein niedriges Geburtsgewicht (ca. 2 600 g), Pigmentierungsstörungen der Haut und Panzytopenie, lagen auch ein Tonsillenkarzinom sowie Lungenkrebs vor. Letzterer war zuvor durch Radio- sowie Chemotherapie, bei welcher Cisplatin zum Einsatz kam, behandelt worden. In den Sequenzdaten des GS Juniors war bereits die heterozygote Nonsense-Mutation c.1972C>T (p.R658\*) in *FANCM* bestimmt worden (Tabelle 17), die auch in den NGS-Daten des Krebspanels detektiert wurde. Diese Variante findet sich in der LOVD-Datenbank einmal bei einem Patienten mit Polycythaemia vera gelistet, der zunächst eine Anämie und später eine AML entwickelte. Durch eine somatisch erworbene uniparentale Disomie (UPD) des langen Arms von Chromosom 14 war dieser Patient homozygot für die Nonsense-Mutation in *FANCM* geworden [Harutyunyan et al., 2011]. Zudem ist die Variante acht Mal im ExAC-Browser, jeweils heterozygot, gelistet. Für nicht-finnische Europäer ergab sich dadurch eine Allelfrequenz von 1,05\*10<sup>4</sup>. Neben dieser Mutation konnte in den MiSeq-Daten eine weitere heterozygote Mutation in *FANCM* nachgewiesen werden. Dabei handelt es sich um c.1241dup, welche zum Frameshift und einem vorzeitigen Stoppcodon führt (p.N414Kfs\*12). Für diese Duplikation konnte die Mutter als Überträgerin bestätigt werden. DNA des Vaters lag zur Segregationsanalyse nicht vor. Ein FANCD2-Immunblot ordnete die Patientin einer *Downstream*-Komplementationsgruppe zu, so dass die im Panel enthaltenen FA-*Downstream*-Gene noch genauer analysiert wurden. Dabei fiel auf, dass sowohl in *BRCA1* als auch in *BRCA2* eine Reihe von aufeinander folgenden SNPs homozygot vorlag (Abbildung 43). Eine MLPA-Analyse beider Gene zeigte keine Veränderungen. Bevor noch weitere *Downstream*-Gene einer näheren Überprüfung unterzogen wurden, wurden die Daten des Krebspanels mit den WES-Daten verglichen. Die Ergebnisse werden in 4.4.3 beschrieben.





Abbildung 43: Übersicht homozygoter SNPs in *BRCA1* und *BRCA2* bei FA-07. Nach der Anreicherung mit dem Krebspanel und Sequenzierung am MiSeq fielen in *BRCA1* und *BRCA2* jeweils eine Reihe homozygoter SNPs bei FA-07 auf (rote Pfeile), welche von heterozygoten SNPs flankiert wurden (schwarze Pfeile). Das jeweilige Referenztranskript ist mit angegeben. Alle Exons sind nummeriert. Nicht codierende Exons sowie die UTR sind dunkel gefüllt.

Auch FA-24, ein Patient pakistanischer Abstammung, war zuvor mit Hilfe des *SeqCap EZ*-Panels angereichert und am GS Junior sequenziert worden. Allerdings war die Anreicherung fehlgeschlagen, so dass die NGS-Daten für eine Mutationsanalyse nicht nutzbar waren. Zudem lag die Information vor, dass dieser Patient extern bereits auf alle bekannten FA-Gene gescreent worden war und dabei keine Mutation nachgewiesen werden konnte. Daher wurde parallel zur Anreicherung mit dem *TruSight Cancer*-Panel ein WES durchgeführt sowie ein FANCD2-Immunblot mit einem Proteinlysat der Patientenzelllinie angefertigt. Dieser zeigte das Fehlen beider Banden an, so dass die Daten des Krebspanels auf Mutationen in *FANCD2* überprüft wurden. Tatsächlich fand sich an Position c.696-121 die homozygote Substitution C>G, welche in der Literatur bereits bei einem indischen Patienten beschrieben worden war [Kalb et al., 2007]. Durch die Mutation kommt es zur Aktivierung einer kryptischen Spleißdonor- sowie -akzeptorstelle, welche auf cDNA-Ebene die Exonisierung von 34 bp (c.695\_696ins696-126\_696-159) zur Folge hat, was auch bei FA-24 bestätigt werden konnte.

#### 4.3.4 MLPA- und Immunblot-Analyse zur Eingrenzung der Komplementationsgruppe

Die übrigen 14 Patienten waren zuvor noch keiner Mutationsanalyse unterzogen worden. Insofern entsprechendes Patientenmaterial vorhanden war, wurden eine durchflusszytometrische Untersuchung zur Bestimmung der G2-Phase-Blockierung, eine *FANCA*-MLPA-Analyse sowie ein FANCD2-Immunblot durchgeführt. Die Ergebnisse sind zur Übersicht in

Tabelle 20 eingetragen. Bei zwei Patienten (FA-67, FA-69) wurden vorab keine durchflusszytometrischen Analysen durchgeführt, die Diagnose FA basierte hier auf klinischen Merkmalen. Eine zugesandte B-Lymphozyten-Zelllinie von FA-69 erwies sich hingegen bei dieser Untersuchung als nicht MMC-sensitiv. Außer gDNA aus Blut stand von FA-69 kein weiteres Material für die Durchflusszytometrie oder Mutationsanalyse zur Verfügung.

**Tabelle 20:** Ergebnisse zur Durchflusszytometrie, *FANCA*-MLPA- sowie FANCD2-Immunblot-Analyse (FANCD2-IB). Unter "Flow" ist das Zellmaterial eingetragen, welches einen spontanen sowie MMC-induzierten G2-Phase-Arrest des Zellzyklus zeigte. "\*" zeigt dabei ein Mosaik im hämatopoetischen System an. "\*\*" weist darauf hin, dass die Untersuchung von Fibroblasten ein negatives Ergebnis ergeben hatte. "ohne Befund": keine Auffälligkeiten. "." nicht durchgeführt. LCL: Lymphoblastoide Zelllinie.

ID	Flow	FANCA- MLPA	FANCD2- IB	ID	Flow	FANCA- MLPA	FANCD2- IB
FA-56	Fibroblasten*	ohne Befund	-	FA-63	Blut**	ohne Befund	downstream
FA-57	Blut	ohne Befund	-	FA-64	Amnionzellen	ohne Befund	-
FA-58	Blut	ohne Befund	-	FA-65	Fibroblasten	ohne Befund	-
FA-59	Blut	ohne Befund	-	FA-66	Fibroblasten	-	-
FA-60	Fibroblasten*	-	-	FA-67	-	-	-
FA-61	Blut*	ohne Befund	-	FA-68	Blut	ohne Befund	upstream
FA-62	Blut	ohne Befund	-	FA-69	LCL negativ	ohne Befund	-

Die Untersuchungen von FA-56, FA-60 sowie FA-61 am Durchflusszytometer ergaben hingegen, dass vermutlich ein somatisches Mosaik im hämatopoetischen System vorliegt. Die Zellzyklusanalyse PHAstimulierter Blutlymphozyten mit Hilfe der BrdU/Hoechst-Methode ergab bei der Auswertung des Verhältnisses der Summe der G2-Phasen zur Wachstumsfraktion (*growth fraction*; GF) Werte von 0,172 (FA-56), 0,118 (FA-60) sowie 0,109 (FA-61) und lag damit jeweils im Bereich normaler Kontrollen (Abbildung 44). Eine bei FA-61 extern durchgeführte Chromosomenbruchanalyse hatte allerdings zuvor eine erhöhte Bruchrate ergeben und die Anreicherung erfolgte, bevor eine Blutprobe auf einen G2-Phase-Anstieg untersucht werden konnte.



Abbildung 44: Zellzyklusanalyse zur Bestimmung des G2-Phasen-Arrestes PHA-stimulierter peripherer Blutlymphozyten. Aufgetragen ist das Verhältnis der Summe der G2-Phasen zur Wachstumsfraktion gegen den Anteil nicht-proliferierender Zellen nach 72-stündiger Inkubation ohne (links) bzw. mit 10 ng/ml MMC (rechts), bestimmt mit der BrdU-HOECHST-Methode. Grün: FA-56; Blau: FA-60; Rot: FA-61; Schwarz: FA-63; Graue Kreise: Kontrollen. Graue Rauten: FA-Patienten.

Die Analyse kultivierter Fibroblasten von FA-56 ergab einen, im Vergleich zu Kontrollen, unverhältnismäßig hohen Anstieg der G2-Phase als Reaktion auf MMC von 15,5 % (spontan) auf 52,6 % (mit MMC) (Tabelle 21). Ein vergleichbares Ergebnis wurde ebenso für FA-60 bestimmt (Tabelle 21). Im Gegensatz dazu zeigte die Untersuchung peripherer Blutlymphozyten von FA-63 einen hohen spontanen G2-Phase-Arrest ( $\Sigma$ G2/GF: 0,409) (Abbildung 44), welcher bei der Analyse von Fibroblastenkulturen nicht bestätigt werden konnte (Tabelle 21). Außer für FA-66 und FA-67 lag bei allen untersuchten Patienten das Ergebnis einer *FANCA*-MLPA-Analyse vor, die jeweils keine Auffälligkeiten zeigte. Dagegen war nur bei zwei Patienten ein FANCD2-Immunblot angefertigt worden. Demzufolge war FA-63 *downstream* einzuordnen, wohingegen bei FA-68 nur die nicht-ubiquitinierte Bande von FANCD2 nachgewiesen werden konnte.

Fibroblasten ± MMC	G1-Phase [%]	S-Phase [%]	G2/M-Phase [%]
Normale Kontrolle –MMC	86,6	6,7	6,5
Normale Kontrolle +MMC	79,5	6,3	14,2
FA-Patient –MMC	78,7	3,8	17,5
FA-Patient +MMC	52,3	4,8	42,9
FA-56 –MMC	66,6	17,9	15,5
FA-56 +MMC	37,3	10,1	52,6
FA-60 –MMC	60,2	21,9	17,9
FA-60 +MMC	43	13,5	43,5
FA-63 –MMC	67,1	8,4	24,5
FA-63 +MMC	64,8	13,5	21,7

**Tabelle 21:** Ergebnis der Zellzyklusverteilung bei Fibroblastenkulturen von FA-56, FA-60 sowie FA-63. Je eine Kultur blieb unbehandelt oder wurde für 48 h mit 12 ng/ml MMC inkubiert. Zellen wurden vor der Messung am Durchflusszytometer mit DAPI gefärbt. Zum Vergleich ist das Ergebnis einer normalen Kontrolle sowie eines bestätigten FA-Patienten mit eingetragen.

### 4.3.5 Erfolgreiche Komplementationsgruppen-Zuordnung

Mit Hilfe des Krebspanels konnten neun FA-Patienten einer Komplementationsgruppe zugeordnet werden. Die Ergebnisse der Mutationsanalyse sind in Tabelle 22 zusammengefasst. Die Segregation der Mutationen konnte nur bei FA-58 und FA-68 bestätigt werden. Für alle anderen Patienten stand kein Material der Eltern zur Verfügung. Wie man der Tabelle entnehmen kann, konnten fünf neue Mutationen bestimmt werden, welche bislang nicht in der internationalen FA-Mutationsdatenbank eingetragen sind. Der Effekt von c.521+1G>A in *FANCC* (FA-66) war bereits bei FA-41 auf cDNA-Ebene als Skipping von Exon 6 nachgewiesen worden (Abbildung 25).

Tabelle 22: Ergebnis der Mutationsanalyse unter Verwendung der Auswertesoftware von NextGENe. ID: Patienten-ID; FA-Gruppe: Komplementationsgruppen-Zuordnung; "Ref.": Referenz. "LOVD-ID": Die jeweilige ID der Mutation aus der Fanconi-Anämie-Mutationsdatenbank (http://www.rockefeller.edu/fanconi/). Die in der Datenbank vorangestellten "0" jeder ID wurden hier zur besseren Übersicht weggelassen. "-": keine Angabe.

ID	FA-		Mutation A	Allel 1		Mutation Allel 2			
ID	Gruppe	DNA	Protein	LOVD-ID	Ref.	DNA	Protein	LOVD-ID	Ref.
FA-57	FA-A	c.121C>T	p.Q41*	-	-	c.121C>T	p.Q41*	-	-
FA-58	FA-A	c.2535_2536del	p.C846Qfs*20	FANCA_180	[Ameziane et al., 2008]	c.2535_2536del	p.C846Qfs*20	FANCA_180	[Ameziane et al., 2008]
FA-59	FA-D2	c.3996G>T	p.Q1332H	-	-	c.3996G>T	p.Q1332H	-	-
FA-62	FA-B	c.592C>T	p.Q198*	-	-	-	-	-	-
FA-65	FA-I	c.3006+3A>G	p.R964_Q1002d el	FANCI_17	[Dorsman et al., 2007]	c.3006+3A>G	p.R964_Q1002del	FANCI_17	[Dorsman et al., 2007]
FA-66	FA-C	c.521+1G>A	p.M153Nfs*23	FANCC_28	-	c.521+1G>A	p.M153Nfs*23	FANCC_28	-
FA-67	FA-F	c.484_485del	p.L162Dfs*103	FANCF_4	[de Winter et al., 2000a]	c.484_485del	p.L162Dfs*103	FANCF_4	[de Winter et al., 2000a]
FA-68	FA-A	c.862G>T	p.E288*	FANCA_65	[Morgan et al., 1999]	c.3783delC	p.F1262Sfs*4	-	-
FA-69	FA-P	c.1744C>T	p.R582*	-	-	c.2137C>T	p.R713*	-	-

Unter den neuen Mutationen befanden sich zwei Varianten in *FANCA*. So war bei FA-57 mit VACTERL-H-Assoziation die homozygote Nonsense-Mutation c.121C>T (p.Q41\*) in den NGS-Daten gelistet, welche bereits in Exon 2 ein vorzeitiges Stoppcodon zur Folge hat (Abbildung 45A). Im ExAC-Browser war diese Variante mit einer Allelfrequenz von 8,243\*10<sup>-6</sup> angegeben. Daneben fand sich in Exon 38 die compoundheterozygote 1-bp-Deletion c.3783del bei FA-68. Die deletierte Base ist Teil eines Triplettrepeats und führt zur Verschiebung des Leserahmens, wodurch ebenfalls ein vorzeitiges Stoppcodon entsteht (p.F1262Sfs\*4). Die Mutation konnte dem Vater zugeordnet werden (Abbildung 45B), wobei die Mutter sowie nicht betroffene Geschwister die Wildtypsequenz zeigten.



Abbildung 45: Unbeschriebene Mutationen in FA-Genen, die mit Hilfe des *TruSight Cancer*-Panels nachgewiesen werden konnten. A: Im Blut von FA-57 wurde die neue Nonsense-Mutation c.121C>T in Exon 2 von *FANCA* homozygot detektiert (blau/Stern). Oben ist die revers-komplementäre Sequenz in der Ansicht der NextGENe-Software dargestellt, darunter das Elektropherogramm der Sanger-Sequenzierung (reverse Sequenz). B: Die Segregation von c.3783del (unterstrichen) in *FANCA* konnte bei FA-68 bestimmt werden. C: Die Analyse von DNA aus Blut von FA-62 erbrachte den Nachweis der hemizygoten Nonsense-Mutation c.592C>T (Stern) (p.Q198\*) in *FANCB*. D: Bei FA-69 war die heterozygote Substitution c.1744C>T in *SLX4* in gDNA aus Blut nachweisbar (Stern), nicht aber in einer lymphoblastoiden Zelllinie (LCL). E: Im Blut von FA-69 wurde die neue Nonsense-Mutation c.2137C>T in Exon 10 von *SLX4* heterozygot bestimmt(blau/Stern). Oben ist die revers-komplementäre Sequenz in der Ansicht der NextGENe-Software dargestellt, darunter das Elektropherogramm der Sanger-Sequenzierung, "Ex": Exon.

Weiterhin konnten insgesamt drei neue Nonsense-Mutationen in den seltenen Komplementationsgruppen FA-B und FA-P detektiert werden. Dabei war bei FA-62, einem männlichen Patienten, der hemizygote Austausch c.592C>T in *FANCB* nachweisbar, welcher in den NGS-Daten 788x abgedeckt war und durch Sanger-Sequenzierung bestätigt werden konnte (Abbildung 45C). Diese Substitution führt an Position 198 in Exon 3 zu einem vorzeitigen Stoppcodon. Dagegen waren bei FA-69 zwei heterozygote Nonsense-Mutationen in Exon 8 (c.1744C>T) und Exon 10 (c.2137C>T) (rs760126773) in *FANCP* im Mutationsbericht der Analysesoftware aufgeführt. Beide Substitutionen konnten durch Sanger-Sequenzierung von DNA aus Blut bestätigt werden (Abbildung 45D+E). Die Mutation in Exon 8 war in einer lymphoblastoiden Zelllinie revertiert (Abbildung 45D), welche keine MMC-Sensitivität mehr aufwies.

Eine weitere neue Mutation war zudem im Mutations-Report von FA-59 in *FANCD2* gelistet. Dabei handelt es sich um die homozygote Missense-Mutation c.3996G>T, die zum Austausch eines hochkonservierten Glutamins gegen ein Histidin im C-terminalen Bereich von FANCD2 führt (p.Q1332H), welcher die kürzlich beschriebenen Tower-Domäne im ID-Komplex bildet (Abbildung 46A+B). Somit konnte mit Hilfe von gezielter Anreicherung ein dritter FA-D2-Patient identifiziert werden. Da es sich bei FANCD2 um ein Schlüsselprotein des FA/BRCA-Signalweges handelt, wurden weitere FA-D2-Patienten, welche während der Routinediagnostik oder durch andere Forschungsprojekte identifiziert wurden, näher charakterisiert. Dies soll separat in Abschnitt 4.5 behandelt werden.



Abbildung 46: Eine bisher unbeschriebene Missense-Mutation c.3996G>T in der Tower-Domäne von FANCD2. A: 3D-Rekonstruktion der Kryo-Elektronenmikroskopie-Aufnahme des FANCD2-FANCI-Komplexes (EMDB-8141, nach [Liang et al., 2016]). Die gabelförmige Tower-Domäne ist hervorgehoben. Hierin liegt die bei FA-59 homozygot nachgewiesene Missense-Mutation p.Q1332H. B: Schematische Übersicht der Proteindomänen von FANCD2. Über der Struktur sind wichtige regulatorische Aminosäurereste angegeben, welche durch Phosphorylierung (P) oder Ubiquitinierung (U) modifiziert werden. Die Bezeichnung sowie Position (*amino acids; aa*) der Domänen ist unter der Struktur vermerkt. Die Mutation p.Q1332H ist mit einem Stern markiert. NLS<sup>2</sup>: *Nuclear localization signal.* "CUE": *coupling of ubiquitin to ER degradation.* "PIP": *PCNA-interacting peptide.* "EDGE": Aminosäuren-Abfolge.

## 4.3.6 Nicht zugeordnete Patienten trotz Mutationsanalyse

Über ein Drittel der Patienten (n = 5/14) konnte trotz NGS-Analyse keiner Komplementationsgruppe zugeordnet werden. Dazu zählen die drei mutmaßlichen Mosaikpatienten FA-56, FA-60 sowie FA-61. Bei FA-60 fiel auf, dass alle SNPs in *EANCA* homozygot vorlagen. Allerdings hatte die MLPA-Analyse keine Deletion angezeigt. Dagegen wurden bei FA-61 zwei compound-heterozygote Varianten in *PALB2* ermittelt. Die maternal vererbte Mutation c.2674G>A wurde von allen drei Mutationsvorhersage-Programmen als schädlich eingestuft, während die paternale compound-heterozygote Variante unterschiedlich bewertet wurde (Tabelle 23). Dagegen konnten bei FA-56, FA-63 sowie FA-64 keine pathogenen Mutationen in FA-Genen nachgewiesen werden.

ExAC Mutation SIFT PolyPhen-2 ID cDNA dbSNP Gen Protein (Fre-Taster (score) (score) (p-value) quenz) Probably rs45476 Disease causing Deleterious 6,589\*10-5 FA-61 PALB2 c.2674G>A p.E892K damaging 495 (0,996)(0,02)(0,996)Probably rs45624 Disease causing Tolerated 5,972\*10-3 FA-61 PALB2 c.2794G>A p.V932M damaging 036 (0, 23)(0,717)(0,993)Probably rs18003 Disease causing Deleterious FA-63 FANCC c.584A>T p.D195V 3,083\*10-3 damaging 65 (0,03)(1)(0,971)

**Tabelle 23:** Mono- und biallelische Varianten in FA-Patienten, die nicht eindeutig einer Komplementationsgruppe zugeordnet werden konnten. Die im ExAC-Browser angegebene Allelfrequenz der Variante wie auch die Klassifizierung durch drei verschiedene Vorhersageprogramme sind aufgeführt. Der jeweils vom Vorhersageprogramm bestimmte Score ist in Klammern eingetragen.

# 4.4 Mutationsanalyse mit Hilfe von WES

Sechs Patienten, welche bereits mit mindestens einem Genanreicherungspanel analysiert worden waren, wurden zusätzlich einer Exomanalyse unterzogen. Dabei handelte es sich vornehmlich um Patienten, die einen auffälligen bzw. ungewöhnlichen Phänotyp zeigten. Bei zwei Patienten, FA-16 sowie FA-24, konnte durch das WES eine Komplementationsgruppen-Zuordnung erfolgen. Hier lagen biallelische Mutationen in bereits bekannten FA-Genen vor. Bei FA-07 und FA-10 konnten dagegen keine weiteren potentiell pathogenen Mutationen nachgewiesen, sondern nur bereits bekannte Mutationen in FA-Genen bestätigt werden. Dagegen wurden bei FA-29 sowie FA-63 pathogene Mutationen in Nicht-FA-Genen bestämmt. Bei beiden Patienten schlossen sich weitere Untersuchungen zur Charakterisierung der Mutationen an. Die Ergebnisse zu den verschiedenen Projekten werden in den folgenden Abschnitten näher beschrieben.

Insgesamt wurden somit 69 verschiedene Patienten per NGS analysiert, wovon 64 zuvor keiner Komplementationsgruppe zugeordnet worden waren. Während bei vier Patienten (FA-05, -11, -29 und -63) die Verdachtsdiagnose FA nicht bestätigt werden konnte, blieben 13 Patienten nach allen durchgeführten Analysen weiterhin ohne Komplementationsgruppen-Zuordnung. Diese sind zur Übersicht in Tabelle 24 zusammengefasst. Hier sind noch weitere Untersuchungen, wie beispielsweise WES, zur Klassifizierung nötig.

Tabelle 24: Übersicht der FA-Patienten, die trotz NGS-Analyse nicht klassifiziert werden konnten. Das per Durchflusszytometrie
untersuchte Material mit dem jeweiligen Ergebnis, das Resultat einer FANCA-MLPA sowie die Klassifizierung anhand eines
FANCD2-Immunblots sind eingetragen. Die Methode zur Anreicherung des Patientenmaterials ist weiterhin vermerkt. "ohne
Befund": keine Auffälligkeiten. "A": Deletion "-": nicht durchgeführt. "x": durchgeführt.

<b>D</b>				NG5			
ID	Durchflusszytometrie	rchflusszytometrie MLPA Immur		SeqCap EZ	TruSight Cancer	WES	
FA-07	Blut + Fibroblasten positiv	ohne Befund	downstream	х	х	Х	
FA-08	Fibroblasten positiv	-	downstream	Х	-	-	
FA-10	Blut + Fibroblasten positiv	ohne Befund	upstream	Х	-	х	
FA-19	Blut positiv	$\Delta$ Exon 11-14					
FA-44	Blut (vor Transplantation) + Fibroblasten positiv	ohne Befund	-	x	-	-	
FA-47	Blut negativ, Fibroblasten positiv	ohne Befund	-	х	-	-	
FA-52	Blut positiv	-	-	х	-	-	
FA-53	Fibroblasten positiv	ohne Befund	-	х	-	-	
FA-55	Blut negativ, Fibroblasten positiv	ohne Befund	-	х	-	-	
FA-56	Blut negativ, Fibroblasten positiv	ohne Befund	-	-	Х	-	
FA-60	Blut negativ, Fibroblasten positiv	ohne Befund	-	-	Х	-	
FA-61	Blut negativ, keine Fibroblasten	ohne Befund	-	-	Х	-	
FA-64	Fibroblasten positiv	ohne Befund	-	-	Х	-	

#### 4.4.1 Alignment mit Hilfe der NextGENe-Software

Die Anzahl der Reads, die beim WES generiert wurden und nach Datenkonvertierung und Alignment durch die Auswertesoftware NextGENe für die Mutationsanalyse zur Verfügung standen, ist in Abbildung 47A für alle sechs Patienten dargestellt. Eine Übersicht über die durchschnittliche Abdeckung der einzelnen Patienten findet sich dagegen in Abbildung 47B wiedergegeben.

In beiden Abbildungen fällt auf, dass FA-24 bei allen betrachteten Parametern das schlechteste Ergebnis zeigte. Dieser war als einziger per *NimbleGen SeqCap EZ Human Exome Library-Kit* angereichert und an einem SOLiD 5500xl Genetic Analyzer sequenziert worden. Der Patient mit dem besten Ergebnis, FA-63, zeigte im Vergleich 46 % mehr sequenzierte Reads und eine durchschnittliche Abdeckung von 133x, im Gegensatz von nur 15x bei FA-24. Dementsprechend zeigte die Analyse aller bekannten FA-Gene bei FA-24 eine erhebliche Anzahl an Lücken. Dieser Anteil konnte mit 12 % bestimmt werden, wohingegen bei FA-63, als einziges von allen WES-Projekten, alle codierenden Exons der FA-Gene vollständig abgedeckt waren. Als wiederkehrende Bereiche mit mangelnder Abdeckung in diesen Genen konnten fünf Exons in *FANCD2* (Exon 12, 13, 16, 21, 24) sowie Exon 1 von *EANCE* identifiziert werden, da sie bei mindestens drei der per WES analysierten Patienten nicht oder lückenhaft abgedeckt waren.



Abbildung 47: Übersicht ausgewählter Rohdaten der sechs Patienten, die mit unterschiedlichen Exom-Panels angereichert wurden. A: Anzahl der Reads nach der Sequenzierung am jeweiligen Sequenzierer sowie der Datenkonvertierung und dem Alignment durch die NextGENe-Software. Die Grenzen der Box markieren jeweils die 25. und 75. Perzentile, die Linie im Inneren den Median. Ein schwarzes Kreuz markiert den Durchschnitt. Die Whisker (gestrichelte vertikale Linien) entsprechen dem 1,5-fachen Interquartilsabstand (nach Tukey). Die einzelnen Patienten sind als schwarze Kreise eingetragen, die quantitativen Extreme sind farbig hervorgehoben (rot: FA-24; blau: FA-63). B: Durchschnittliche Abdeckung der ROI der Exomsequenzierungen. Erklärung der Symbole: siehe A.

Im Mittel wurden bei den WES-Projekten 6,23\*10<sup>7</sup> Reads durch die Sequenzierung erzeugt, wovon 6,07\*10<sup>7</sup> konvertiert und 5,76\*10<sup>7</sup> gegen das Referenzgenom alignt werden konnten. Dadurch ergab sich eine durchschnittliche Abdeckung von 74x für die Sequenzierung des gesamten Exoms.

Betrachtet man verschiedene Grenzwerte zur Abdeckung der gesamten ROI (Abbildung 48), so erkennt man auch hier die deutlich reduzierte Abdeckung des Exoms bei FA-24. Nur 72 % der ROI waren 1x und weniger als die Hälfte 10x abgedeckt. Dagegen fiel erneut FA-63 durch eine sehr gute Abdeckung auf. Hier waren noch knapp 20 % der ROI 200x abgedeckt. Generell ist bereits ab einer geringen Abdeckung die Streuung zwischen den einzelnen Projekten groß, was anhand der Breite der Boxen veranschaulicht wird. Unter Berücksichtigung der Mediane war der deutlichste Abfall bei einer Abdeckung von 20x auf 40x zu verzeichnen. Außer bei FA-63 waren kaum Bereiche mit einer Abdeckung von 200x versehen.



Abbildung 48: Abdeckung der ROI (bed-File der entsprechenden Exom-Anreicherung) nach Datenkonvertierung und Alignment der sechs per WES analysierten FA-Patienten. Die ROI wurde nach dem Mappen der Reads auf ihre Abdeckung überprüft. Die Grenzen der Box markieren jeweils die 25. bzw. 75. Perzentile, die Linie im Inneren den Median. Ein schwarzes Kreuz markiert zudem den Durchschnitt. Die Whisker (gestrichelte vertikale Linien) entsprechen dem 1,5-fachen Interquartilsabstand (nach Tukey). Ausreißer sind als Kreise (rot: FA-24; blau: FA-63) außerhalb der Whisker dargestellt.

# 4.4.2 Erfolgreiche Komplementationsgruppen-Zuordnung durch WES: Identifizierung eines dritten FA-Q-Patienten

Anhand von klinischen Merkmalen war die Diagnose FA bei FA-16 aus Großbritannien bereits im Kindesalter gestellt worden. Dazu zählten bilaterale Radialstrahlanomalien mit fehlendem rechten und hypoplastischem linken Daumen sowie einer Thrombozytopenie. Daneben zeigte die Patientin noch eine Vielzahl an weiteren phänotypischen Auffälligkeiten wie Kleinwuchs, Mikrozephalie, Café-au-lait-Flecken, Nierenanomalien und Schwerhörigkeit. Im Alter von 44 Jahren wurde ein milder Typ-1-Diabetes diagnostiziert. Zudem berichtete die Patientin über eine Überempfindlichkeit gegenüber Sonnenlicht, welche bereits im Kleinkindalter aufgefallen war. Durch Chromosomenbruchanalysen im Alter von 21 und 33 Jahren konnte die Verdachtsdiagnose FA bestätigt werden. Der klinische Hintergrund der Familie war unauffällig und bei der Patientin waren auch im Alter von 46 Jahren keine Tumoren diagnostiziert worden. Ebenso lag das Ergebnis einer Knochenmarksuntersuchung vor, welche eine milde Hypoplasie, aber keine klonalen Chromosomenaberrationen ergab. Ein Blutbild bestätigte die Thrombozytopenie und eine Makrozytose.

#### a. Komplementationsgruppen-Zuordnung

Aufgrund der stabilen Blutwerte, auch noch im Alter von 46 Jahren, wurde zunächst das Vorliegen eines somatischen Mosaikes im blutbildenden System untersucht. Durchflusszytometrische Untersuchungen zur Bestimmung des G2-Phase-Arrestes von primären Fibroblastenzellen ergaben, dass sich bereits ohne die Zugabe einer DNA-quervernetzenden Substanz 46,6 % der Zellen in der G2-Phase befanden. Nach Zugabe von MMC stieg der Wert der arretierten Zellen auf 67,2 % an. Bei der Analyse von peripheren Blutlymphozyten konnte ebenfalls ein erhöhter G2-Phase-Arrest nachgewiesen werden. Hier lag das Verhältnis von  $\Sigma$ G2/GF bei 0,515 (0 µg/ml MMC) bzw. 0,585 (10 µg/ml MMC) und damit im Bereich anderer FA-Patienten ( $\Sigma$ G2/GF >0,4). Wie bereits in Tabelle 11 vermerkt, zeigte eine *FANCA*-MLPA-Analyse keine Auffälligkeiten, wohingegen ein FANCD2-Immunblot die Patientin einer *Downstream*-Komplementationsgruppe zuordnete (Abbildung 49A). Anhand der RAD51-Foci-Bildung nach Schadensinduktion konnten zu diesem Zeitpunkt die Untergruppen FA-D1, -N sowie -O ausgeschlossen werden, da diese unauffällig war. Das Ergebnis zu dieser Untersuchung ist in Abbildung 49B dargestellt.

Durch eine Exomsequenzierung von genomischer DNA aus primären Fibroblasten konnten zwei pathogene Mutationen in *FANCQ* nachgewiesen und durch Sanger-Sequenzierung bestätigt werden. Die maternale Mutation c.793-2A>G betrifft die kanonische Spleißakzeptorstelle von Intron 4 (Abbildung 49C). cDNA-Analysen bestätigten aberrantes Spleißen. Durch Ausschneiden der Banden nach der Größenauftrennung in einem Agarosegel und Aufreinigung derselben konnte das Hauptspleißprodukt sequenziert werden (Abbildung 49D). Dabei wurde das Skipping von Exon 5 nachgewiesen, welches einen Frameshift zur Folge hat und zur vorzeitigen Termination der Translation (p.T265Vfs\*13) führt. Diese Mutation war aufgrund mangelnder Abdeckung (4x) nicht bei der Anreicherung mit dem *SeqCap EZ*-Panel nachgewiesen worden. In den Exomdaten betrug die Abdeckung der Substitution dagegen 27x.

Die Missense-Mutation c.1765C>T (rs147105770) (Abbildung 49E), welche in den WES-Daten als zweite Mutation in FANCO gelistet war und von verschiedenen Vorhersageprogrammen als pathogen eingestuft wurde, konnte bereits durch bei der Sequenzierung am GS Junior bestimmt werden (4.2.3d). Diese paternal vererbte Mutation in Exon 8 führt zum Austausch eines hoch konservierten Arginins gegen Tryptophan (p.R589W). Im ExAC-Browser finden sich acht jeweils heterozygote Einträge zu dieser Mutation, die dort mit einer Allelfrequenz von 6,603\*10-5 angegeben ist. Ein pathogener Effekt dieser Mutation ist zudem schon von anderen Erkrankungen bekannt. So wurde sie bereits bei drei Patienten der Xeroderma pigmentosum-Untergruppe F (XP-F), XP24BR, XP32BR und AS871 [Gregg et al., 2011; Fassihi et al., 2016] sowie dem Patienten XPCS1CD mit kombiniertem XP/CS-Phänotyp und Merkmalen von FA [Kashiyama et al., 2013] identifiziert. Als funktionelle Bestätigung der Pathogenität der Mutationen wurden MMC-Überlebenskurven der Patientenzellen erstellt. Durch Komplementation mit Wildtyp-FANCO konnte die Komplementationsgruppe für FA-16 bestätigt werden, da die MMC-Sensitivität der Zellen aufgehoben werden konnte (Abbildung 49F).



**Abbildung 49:** Zuordnung von FA-16 zur Komplementationsgruppe FA-Q. **A:** Ein FANCD2-Immunblot konnte bei FA-16 (roter Rahmen) beide Banden nachweisen. FANCJ und Vinculin dienten als Ladungskontrollen. **B:** Die Ergebnisse der Auszählung von RAD51-Foci nach Schadensinduktion mit 40 ng/µl MMC waren zwischen FA-16 und einer Kontrollzelllinie (Kon) in Fibroblasten (Fib) wie auch Zellen lymphoblastoider Linien (LCL) vergleichbar. **C:** Die Substitution c.793-2A>G (Stern) konnte dem maternalen Allel zugeordnet werden. **D:** Auf cDNA-Ebene führt diese Mutation zu aberrantem Spleißen. Als Hauptspleißprodukt konnte das Skipping von Exon 5 nachgewiesen werden. **E:** Die zweite compound-heterozygote Mutation konnte als c.1765C>T (Stern) in Exon 8 (Ex8) identifiziert werden und wurde vom Vater vererbt. **F:** MMC-Überlebenskurven zeigten, dass die Sensitivität gegenüber der Substanz durch Komplementation mit der Wildtypsequenz (*FANCQ*<sup>WT</sup>), nicht aber durch die Missense-Mutation c.1765C>T (*FANCQ*<sup>MUT</sup>) aufgehoben werden kann. Eine FA-B-Linie diente als Kontrolle, ebenso die Untersuchung mit dem Leervektor (mock). LD50-Werte sind in der entsprechenden Farbe als gestrichelte Linien eingezeichnet. Fehlerbalken kennzeichnen die Standardabweichung aus drei Experimenten.

#### b. Funktionelle Charakterisierung der identifizierten FANCQ-Mutationen

Auf Proteinebene war durch das Skipping von Exon 5 ein verkürztes Protein von 31,4 kDa vorhergesagt. Dieses war durch eine Immunblot-Analyse nicht nachweisbar. Stattdessen konnte Restprotein bei etwa 100 kDa nachgewiesen werden, was der normalen Proteinlänge entspricht (104 kDa), wie in Abbildung 50 ersichtlich. Bei 1333, welcher ebenfalls Träger einer Missense-Mutation (p.L280P) sowie einer trunkierenden Mutation (p.I800Tfs\*24) in *FANCQ* ist, konnte dagegen neben Restprotein bei 100 kDa auch ein verkürztes Proteinprodukt detektiert werden (Abbildung 50A). Bei 1333 handelt es sich um einen von bislang zwei beschriebenen FA-Q-Patienten [Bogliolo et al., 2013], welcher auch zur Etablierung des Anreicherungspanels als Patient FA-02 analysiert wurde. Durch eine Proteinfraktionierung konnte Bei FA-16 weiterhin das Vorhandensein des Proteins im Kern sowie am Chromatin bestätigt werden. Dieses Ergebnis ist in Abbildung 50B dargestellt.

Da es sich bei ERCC4/FANCQ um einen Bestandteil des NER-Signalweges handelt und zudem in Fibroblasten von 1333 eine milde zelluläre UV-Sensitivität beschrieben worden war, wurden für FA-16 ebenfalls Überlebensversuche nach UVC-Bestrahlung durchgeführt. Der LD50-Wert der primären Fibroblasten von FA-16 (2,8 J/m<sup>2</sup>) lag dabei zwischen dem von 1333-Fibroblasten (1,6 J/m<sup>2</sup>) sowie Kontrollzellen (5,8 J/m<sup>2</sup>) und war vergleichbar zu einem XP-F-Patienten mit bekanntem milden XP-Phänotyp und moderater UV-Sensitivität (XP23OS) (Abbildung 50C) [Arase et al., 1979]. Die höchste Sensitivität gegenüber UVC-Bestrahlung zeigte bei diesem Versuch der einzige bisher bekannte XFE-Patient (XP51RO) [Niedernhofer et al., 2006]. Durch weitere Untersuchungen konnte ferner die Beeinträchtigung des GG-NER sowie des TC-NER-Signalweges (vgl. Abschnitt 1.1.2) bestätigt werden. Die Ergebnisse dazu finden sich in Abbildung 50D-F dargestellt. Zur Untersuchung der GG-NER wurde ein UDS-Assay durchgeführt, welcher durch die Verwendung von 5-Ethinyl-2'-desoxyuridin (EdU) den Einbau von Nukleosiden nach Bestrahlung von Zellen mit UVC-Licht nachweisbar macht. Im Vergleich mit einer Kontrolle lag die bestimmte relative Intensität bei der Untersuchung primärer Hautfibroblasten von 1333 bei 13,26 % (SEM: ± 10,89 %) sowie 16,10 % (SEM: ±7,28 %) bei FA-16 (Abbildung 50D+E). Ein vergleichbares Ergebnis konnte beim RRS-Test ermittelt werden, bei welchem EU (5'-Ethinyluridin) zum Nachweis der Transkriptions-gekoppelten NER eingesetzt wurde. Hier betrug die Intensität 17,21 % (SEM: ± 5,5 %) bei 1333 und 16,75 % (SEM: ±6,96 %) bei FA-16, wobei die Kontrolle auf 100 % gesetzt wurde (Abbildung 50D+F).

Ein Vergleich der MMC-Sensitivität von FA-16 mit beiden bei Bogliolo et al. 2013 [Bogliolo et al., 2013] beschriebenen FA-Q-Patienten, FA104 sowie 1333, zeigte zudem, dass der LD50-Wert von FA-16 intermediär war. Die mittlere letale Dosis lag bei FA104 bei 21 nM, für FA-16 konnte diese mit 37 nM MMC bestimmt werden. Dagegen lag der LD50-Wert von 1333 bei 74 nM, wohingegen er bei einer Kontrolle erst bei über 200 nM MMC ermittelt werden konnte. Die MMC-Überlebenskurven dieser Zelllinien sind in Abbildung 50G dargestellt.



**Abbildung 50:** Charakterisierung der *EANCQ*-Mutationen von FA-16. **A:** Ein FANCQ-Immunblot konnte Restprotein bei FA-16 (roter Rahmen) nachweisen. Bei 1333 ist zusätzlich ein verkürztes Produkt <100 kDa zu erkennen. Tubulin diente als Ladungskontrolle. **B:** In allen drei Fraktionen (Cyt = Cytoplasma; Nuc = Nucleus; Chr = Chromatin) konnte bei FA-16 FANCQ-Protein nachgewiesen werden (roter Rahmen). Tubulin, YY1 sowie Histon H3 dienten jeweils als Ladungskontrolle. **C:** Immortalisierte (imm) und primäre (prim) Fibroblasten von 1333 (blau) und FA-16 (rot) zeigten eine erhöhte Sensitivität gegen UV-Licht, ähnlich einer XP-F-Zelllinie (XP23OS; grün). Am sensitivsten waren Zellen des XFE-Patienten XP51RO (orange). Gestrichelte Linien markieren die LD50-Werte. Dargestellt sind Mittelwerte  $\pm$  SEMs. **D:** Der Einbau von Ethinyldesoxyuridin (EdU) wurde beim UDS-Assay zum Nachweis der Globalen-Genom-Nukleotidreparatur (oben) genutzt. 5'-Ethinyluridin (EU) diente zur Bestimmung der RRS (unten), um die Transkriptions-gekoppelte NER nachzuweisen. Die Intensitäten der beiden Tests sind für 1333 (blau) und FA-16 (rot) sowie eine Kontrolle (grau) in **E** bzw. **F** (Mittelwerte  $\pm$  SEMs) aufgetragen. Als weitere Kontrollen dienten ein XP-F- (grün) sowie der einzige bekannte XFE-Patient (orange) im UDS-Assay. **G:** Untersuchung der MMC-Sensitivität aller drei FA-Q-Patienten (1333: blau; FA-16: rot; FA104: grün) durch Inkubation von LCLs mit unterschiedlichen Konzentrationen an MMC (Mittelwerte  $\pm$  SEMs). Gestrichelte Linien markieren die jeweiligen LD50-Werte. "SEM": Standardfehler.

#### 4.4.3 Bestätigung bereits bekannter Mutationen bei FA-07, FA-10 sowie FA-24

Der Phänotyp sowie die Ergebnisse zur Anreicherung von FA-07 mit dem Krebspanel wurden unter 4.3.3 bereits näher beschrieben. Dabei wurden zwei compound-heterozygote Varianten in *FANCM* entdeckt. In den WES fanden sich ebenfalls beide Mutationen wieder. Dabei war die 1-bp-Duplikation c.1241dup 74x, die Nonsense-Mutation c.1972C>T 126x abgedeckt und somit deutlich niedriger als in den MiSeq-Daten. Da ein FANCD2-Immunblot beide Banden auf der richtigen Höhe nachgewiesen hatte, wurden zudem alle bis dato bekannten *Downstream*-Gene näher untersucht und mit den Ergebnissen der MiSeq-Sequenzierung verglichen. In den MiSeq-Daten waren Indel-Mutationen in *BRCA2*, *BRIP1*, *PALB2* sowie *ERCC4* enthalten, welche in den WES-Daten nicht im Mutationsbericht gelistet waren. Da diese Varianten meist in unmittelbarer Nähe von Homopolymeren detektiert worden waren, wurden diese als Sequenzierfehler angesehen und nicht weiter verfolgt. Übereinstimmend fanden sich in beiden Datensätzen zwei heterozygote Missense-Mutationen in verschiedenen Genen (Tabelle 25), welche per Sanger-Sequenzierung bei der Patientin bestätigt wurden.

**Tabelle 25:** Monoallelische Varianten in bekannten FA-Genen bei FA-07. Die Abdeckung der Variante in der NextGENe-Software ist angegeben, ebenso die Klassifizierung durch drei verschiedene Mutationsvorhersageprogramme. Der jeweils vom Vorhersageprogramm bestimmte Score ist in Klammern eingetragen.

Gen	cDNA	Protein	dbSNP	Ab- deckung	Mutation Taster (p-value)	SIFT (score)	PolyPhen-2 (score)
RAD51C	c.29T>C	p.M10T	rs45476495	19x	Disease causing (0,999)	Tolerated (0,3)	Probably damaging (0,959)
SLX4	c.3812C>T	p.S1271F	rs38101813	19x	Polymorphism (0,661)	Deleterious (0,03)	Probably damaging (0,96)

Dabei handelte es sich um c.29T>C (p.M10T) in RAD51C sowie um die Variante c.3812C>T (rs38101813) in Exon 12 von *SLX4*. Beide Missense-Mutationen wurden von den verwendeten Mutationsvorhersage-Programmen unterschiedlich eingestuft (Tabelle 25). Eine zweite Mutation war in keinem der beiden Gene enthalten und es konnte auch kein Kandidatengen bestätigt werden.

Anders als FA-07 zeigte FA-10 einen klassischen FA-Phänotyp. Zum Zeitpunkt der WES war der Patient sieben Jahre alt und wies neben postnatalem Minderwuchs, Mikrophthalmie und Mikrozephalie weitere angeborene Fehlbildungen wie beidseitige Daumenhypoplasie, Hüftdysplasie, Radialstrahl- sowie Nierenanomalien, Pigmentstörungen der Haut und eine Thrombopenie auf. Außer der bereits bestätigten *de novo*-Substitution c.280A>T in *EANCB* konnte in den bekannten *Upstream*-FA-Genen keine pathogene Mutation nachgewiesen werden.

Die WES-Ergebnisse von FA-24, welcher am SOLiD 5500xl sequenziert wurde, waren nicht zufriedenstellend (vgl. Abbildung 47 sowie Abbildung 48). Parallel zum WES wurde der Patient auch per FANCD2-Immunblot und Krebspanel-Sequenzierung untersucht. Nachdem durch die Immunblot-Analyse eine Defizienz des Patienten für FANCD2 festgestellt wurde und durch die *TruSight Cancer*-Panelanalyse die Mutation c.696-

121C>G in *FANCD2* bereits bestätigt worden war, wurden die Exomdaten auf diese Mutation hin überprüft. Dabei konnte sie mit einer Abdeckung von 10x ebenfalls homozygot nachgewiesen werden.

#### 4.4.4 Nachweis pathogener Mutationen in anderen Genen der DNA-Reparatur

#### a. NTHL1-Mutationen bei einer Familie mit multiplen Tumorerkrankungen

Von FA-29 türkischer Abstammung war bekannt, dass der Patient sowie dessen jüngerer Bruder bereits in jungen Jahren an Krebs erkrankt waren. So war bei FA-29 im Alter von 29 Jahren ein Mundbodenkarzinom festgestellt worden, welches operativ entfernt und die Region zusätzlich bestrahlt wurde. Ein Jahr später wurde erstmals ein Myelodysplastisches Syndrom (MDS) diagnostiziert, welches mit einer Neutro- sowie Thrombopenie einherging. Des Weiteren entwickelte sich eine Anämie (Hb = 7,5 g/dl) und der Patient verstarb schließlich mit 34,5 Jahren. Bei seinem Bruder (FA-29.2) war mit 23 Jahren ein Plattenepithelkarzinom an der Zungenspitze diagnostiziert worden. Auf eine präoperative Radio- und Chemotherapie, bei welcher auch Cisplatin zum Einsatz kam, hatte FA-29.2 mit einer aplastischen Myelosuppression reagiert und war schließlich an Multiorganversagen verstorben. Eine weitere Familienanamnese ergab später, dass deren Mutter mit 50 Jahren an einem Endometriumkarzinom erkrankt war. Bei einem weiteren Geschwister (FA-29.3) war zudem im Alter von 24 Jahren ein Hirntumor diagnostiziert worden. Der Vater ist bis zum jetzigen Zeitpunkt gesund und auch bei den Großeltern war keine Krebserkrankung beschrieben worden. In Abbildung 51 ist der Stammbaum der Familie dargestellt, der nach der Familienanamnese erstellt werden konnte.



Abbildung 51: Stammbaum des Patienten FA-29. Angegeben ist das aktuelle Alter bzw. das Alter zum Todeszeitpunkt. Phänotypische Auffälligkeiten sind, soweit bekannt, vermerkt. Ein "?" gibt an, dass hierzu keine Information vorlag.

Wie bereits in Abschnitt 4.2.3a beschrieben, war FA-29 als *Downstream*-Patient klassifiziert worden (Tabelle 11). Die durchflusszytometrische Bestimmung des G2-Phasen-Anstieges von peripheren Blutlymphozyten hatte zudem ein Ergebnis im Bereich normaler Kontrollen ergeben ( $\Sigma$ G2/GF spontan = 0,198; induziert = 0,313), wohingegen bei primären Fibroblastenkulturen ein überproportionaler Anstieg nach Behandlung mit 12 ng/ml MMC von 10,2% auf 33,2% nachweisbar war (Abbildung 52). Die Anreicherung mit dem *SeqCap EZ*-Panel hatte allerdings keine pathogene Mutation in bekannten FA-Genen nachweisen können, weshalb DNA aus primären Fibroblasten zum WES verschickt wurde.



**Abbildung 52:** Zellzyklusanalyse primärer Fibroblastenkulturen zum Nachweis eines G2-Phasen-Arrestes. Nach Inkubation mit 12 ng/ml (+) MMC für 48 h konnte bei Kontrollzellen (links) keine Akkumulation in der G2-Phase nachgewiesen werden, wohingegen Zellen eines *FANCA*-Patienten sowie von FA-29 eine Arretierung zeigten. Zur Kontrolle wurden die Zellen ohne MMC (-MMC) am Durchflusszytometer gemessen. Die Färbung der Zellen erfolgte mit DAPI, die Zellzyklusphasen sind oben angegeben.

In den Daten des WES von FA-29 bestätigte sich zunächst der Ausschluss von Mutationen in bekannten FA-Genen. Stattdessen konnte eine homozygote Substitution in Exon 3 von *NTHL1* (NM\_002528) detektiert werden, welche zur Unterbrechung des Leserahmens durch ein Stoppcodon führt (p.W182\*). Die Mutation c.545G>A konnte per Sanger-Sequenzierung sowohl in primären Fibroblasten, als auch in gDNA aus einer lymphoblastoiden Zelllinie des Patienten bestätigt werden (Abbildung 53A), welche nicht sensitiv auf MMC reagierte. Das Gen codiert für das Protein NTHL1, welches eine von bislang elf bekannten Glykosylasen der Basenexzisionsreparatur darstellt. Eine homozygote Nonsense-Mutation in diesem Gen wurde kürzlich bei drei unterschiedlichen Familien mit adenomatöser Polyposis und kolorektalen Karzinomen beschrieben [Weren et al., 2015]. In einer Tumorgewebeprobe von FA-29.2 war die Substitution gleichfalls homozygot. Bei der Überprüfung der Segregation fiel weiterhin auf, dass der Vater heterozygot, die Mutter sowie die Schwester (FA-29.3) ebenfalls homozygot für c.545G>A waren (Abbildung 53B).



Abbildung 53: Homozygote Nonsense-Mutation in der BER-Glykosylase NTHL1. A: c.545G>A (Stern) konnte sowohl in MMCsensitiven primären Fibroblasten (Fib) als auch in einer nicht-sensitiven lymphoblastoiden Zelllinie (LCL) homozygot nachgewiesen werden. B: Die Sanger-Sequenzierung aller Familienmitglieder bestätigte den Vater als heterozygoten Überträger (Stern). Die Mutter sowie beide Geschwister (FA-29.2 und FA-29.3) waren wie FA-29 ebenfalls homozygot (Stern). Vom verstorbenen Bruder FA-29.2 stand nur eine Tumorprobe zur Verfügung. C: Ein NTHL1-Immunblot konnte bei FA-29 und dessen Mutter kein Protein nachweisen (roter Rahmen). Tubulin diente als Ladungskontrolle. Eine HAP1-Zelllinie und eine gesunde Kontrolle waren ebenso aufgetragen wie eine HAP1-Zelllinie mit NTHL1-*Knockout* (HAP1<sup>NTHL1-/-</sup>). D: Der Knockout wurde auch auf gDNA in der HAP1 <sup>NTHL1-/-</sup>Linie bestätigt. Deletierte Basen sind unterstrichen. "Ex": Exon. "IVS": Intron.

Für ein mögliches verkürztes Protein war eine Größe von 20 kDa vorhergesagt worden, wohingegen das Produkt normaler Länge ein Molekulargewicht von 34 kDa aufweist. Durch eine Immunblot-Analyse, die in Abbildung 53C dargestellt ist, war jedoch kein verkürztes Protein nachweisbar. So konnte beim Vater, dem heterozygoten Träger der Nonsense-Mutation sowie einer Kontrolle Protein auf normaler Höhe bestätigt werden. Bei der Mutter wie auch bei FA-29 war hingegen keine Bande nachweisbar. Als Kontrollzelllinie diente dabei eine kommerzielle HAP1-Knockout-Zelllinie für NTHL1 (HAP1NTHL1-/-). Bei dieser war durch das CRISPR-Cas9-System eine Frameshift-Mutation eingeführt worden, welche auf DNA-Ebene als c.457\_461del in Exon 3 bestätigt wurde (Abbildung 53D). Die 5 bp-Deletion hätte nach Vorhersage ein verkürztes Protein (p.R153Afs\*15) von 19,5 kDa zur Folge, was ebenfalls nicht verifiziert werden konnte, obwohl das Epitop des verwendeten Antikörpers die N-terminalen Aminosäuren 1-150 umfasste.

# b. Nachweis compound-heterozygoter *RPA1*-Mutationen bei einem Patienten mit FAähnlichem Phänotyp

Ebenso wie bei FA-29 war bei FA-63 schon in jungen Jahren Krebs diagnostiziert worden. Mit 18 Jahren wurde ein Osteosarkom im linken distalen Femur festgestellt, an dessen Folgen der Patient verstarb. Weiterhin wurden Organfehlbildungen in Form eines angeborenen Herzfehlers sowie eine beidseitige Nierendysplasie

diagnostiziert. Eine milde Leukopenie sowie Neutropenie waren während der Aufnahme des Patienten zur Behandlung des Osteosarkoms ebenso festgestellt worden wie eine mäßiggradige Hypoplasie des Knochenmarkes. Eine Blutprobe erwies sich bei der Bestimmung der G2-Phasen-Arretierung als MMCsensitiv, was in Abbildung 44 dargestellt ist. Dabei war eine hohe spontane Arretierung der Zellen in der G2-Phase festgestellt worden, welche nach Zugabe von MMC nicht deutlich anstieg. Stattdessen konnte ein hoher Zerfall beobachtet werden. Aufgrund dieses Ergebnisses wurde die Subtypisierung eingeleitet und DNA aus Blut mittels des *TruSight Cancer*-Panels angereichert. Wie in Abschnitt 4.3.6 beschrieben, konnte dabei die heterozygote paternale Mutation c.584A>T (FANCC\_000019) in *FANCC* bestätigt werden (

Tabelle 23). Durch einen FANCD2-Immunblot mit Proteinlysat einer lymphoblastoiden Zelllinie wurde der Patient einer *Downstream*-Komplementationsgruppe zugeordnet und DNA aus Blut zum WES gegeben. Währenddessen konnte eine Fibroblastenkultur des Patienten hinsichtlich einer G2-Phasen-Anstieg vor und nach der Behandlung mit 12 ng/ml MMC untersucht werden, wobei hier die Zellen nicht arretierten, was gegen die Verdachtsdiagnose FA sprach.

Die Anreicherung erfolgte mittels *SureSelect Human All Exome v5*-Kit, die Sequenzierung fand am HiSeq 4000 statt. Die Sequenzdaten wurden im Anschluss wie üblich mit der NextGENe-Software ausgewertet. Unter den insgesamt über 166 000 gelisteten Varianten konnte, außer c.584A>T in *FANCC*, keine weitere Mutation in einem bekannten FA-Gen bestimmt werden. Von über 25 700 Varianten innerhalb der codierenden Sequenz hatten Missense-Mutationen mit 45 % den größten Anteil. Unter diesen konnten zwei compoundheterozygote Varianten in *RPA1* ermittelt werden, die in Tabelle 26 mit der Einstufung durch Mutationsvorhersage-Programme eingetragen sind. Das Gen RPA1 codiert für die 70 kDa große Untereinheit des heterotrimeren Replikationsproteins A (RPA), welches während der Replikation und bei DNA-Schäden an einzelsträngige Intermediate bindet, diese stabilisiert und verschiedene Proteine zur weiteren Prozessierung aktiviert bzw. rekrutiert, wie beispielsweise RAD51 und RAD52 [Sleeth et al., 2007].

**Tabelle 26:** *RPA1*-Mutationen, welche durch WES bei FA-63 nachgewiesen werden konnten. Die Abdeckung der Variante in der NextGENe-Software ist angegeben, ebenso die Klassifizierung durch drei verschiedene Vorhersageprogramme. Der jeweils vom Vorhersageprogramm bestimmte Score ist in Klammern eingetragen.

cDNA	Protein	dbSNP	Abdeckung	MutationTaster (p-value)	SIFT (score)	PolyPhen-2 (score)
c.1082G>C	p.W361S	-	170x	Disease causing (1)	Deleterious (0)	Probably damaging (0,996)
c.1165C>T	p.R389W	rs202068855	46x	Disease causing (1)	Deleterious (0,02)	Probably damaging (0,993)

Zunächst wurde die Segregation beider Mutationen durch Sanger-Sequenzierung überprüft. Dabei konnte c.1082G>T (p.W361S) in Exon 11 dem paternalen, c.1165C>T (p.R389W) in Exon 12 dem maternalen Allel zugeordnet werden (Abbildung 54A-C). Zudem konnte ein gesundes Geschwister als heterozygoter Träger der Missense-Mutation in Exon 12 bestätigt werden, wohingegen in Exon 11 dort die Wildtypsequenz vorlag.



Abbildung 54: Compound-heterozygote *RPA1*-Mutationen bei FA-63. A: Durch Sanger-Sequenzierung konnte der Vater als heterozygoter Überträger der Missense-Mutation c.1082G>C (Stern) in Exon 11 des *RPA1*-Gens bestätigt werden. B: Die Mutter erwies sich dagegen als Überträgerin der Substitution c.1165C>T (Stern) in Exon 12. C: Exon- und Proteinstruktur von *RPA1*. In blau sind die Exons der 1 851 bp langen codierenden Sequenz dargestellt. Mit roten Pfeilen sind die bei FA-63 identifizierten Mutationen eingezeichnet. Gestrichelte Linien markieren die Grenzen verschiedener RPA1-Proteindomänen sowie die entsprechenden Bereiche der codierenden Sequenz. In grau sind die vier DNA-Bindedomänen (DBD) von RPA1 dargestellt, welche über flexible Linkerregionen (schwarze Linien) verbunden sind. In DBD-A und -B konnten vier hochkonservierte Aminosäuren (Pfeile) nachgewiesen werden, welche für die Bindung an ssDNA essenziell sind. In rot ist eine bei FA-63 veränderte Aminosäure (aa) hervorgehoben (p.W361S). "Ex": Exon.

Die Substitution c.1082G>C findet sich bislang nicht im ExAC-Browser gelistet, wohingegen c.1165C>T dort mit einer Allelfrequenz von 1,433\*10-3 angegeben ist. Dabei wurde letztgenannte Missense-Mutation bereits drei Mal homozygot bei verschiedenen Individuen finnischer Abstammung nachgewiesen. Die betroffene Aminosäure zeigt eine moderate Konservierung, welche bis zum Zebrafisch reicht, während W361 hochkonserviert ist. Für diese Aminosäureposition wurden von verschiedenen Arbeitsgruppen bereits Untersuchungen zur funktionellen Beeinträchtigung von RPA während der Replikation und Rekombination angefertigt. So zeigten verschiedene Doppelmutanten essenzieller DNA-Bindedomänen (DBD) von RPA1 einen Defekt der DNA-Reparatur [Haring et al., 2008; Hass et al., 2012]. Diese DBD vermitteln die Bindung des Heterotrimers an ssDNA. In RPA1 finden sich vier dieser Domänen, welche als DBD-A, -B, -C sowie -F bezeichnet werden und in Abbildung 54C schematisch dargestellt sind. Die beiden zentralen DNA-Bindedomänen, DBD-A und -B, bilden die DNA-Bindungsstelle des RPA-Komplexes und sind für die hohe Affinität zur Bindung einzelsträngiger DNA verantwortlich. Beide Domänen enthalten eine Reihe polarer und aromatischer Aminosäuren, welche für die Bindung an ssDNA entscheidend sind [Bochkarev et al., 1997]. Darunter finden sich vier hochkonservierte aromatische Aminosäuren. In DBD-A sind dies F238 und F269,

bei DBD-B W361 sowie F386. Eine weitere funktionelle Charakterisierung fand in der vorliegenden Arbeit nicht statt.

## 4.5 Sequenzanalysen bei FANCD2

#### 4.5.1 Pseudogen-Analyse

Von *FANCD2* war im Vorfeld bekannt, dass durch das Vorhandensein zweier Pseudogene die Mutationsanalyse erschwert wird [Kalb et al., 2007; Ameziane et al., 2012; Knies et al., 2012]. Bei Pseudogenen handelt es sich um Kopien eines funktionellen Gens, welche durch Duplikation oder Transposition entstanden sind. Durch Nonsense- oder Indel-Mutationen haben diese Kopien allerdings ihre Funktionalität verloren. Aufgrund der Sequenzähnlichkeit zum funktionellen Gen werden diese Regionen bei der Panelanreicherung ebenfalls angereichert und anschließend gegen die Referenzsequenz gemappt. Dabei kann zunächst nicht unterschieden werden, ob es sich bei einer Variante um eine tatsächliche Mutation oder um eine der Pseudogensequenz entstammenden Veränderung handelt. Ein kompletter Ausschluss der Pseudogene bei der Erstellung von Sonden zur Anreicherung ist allerdings nicht möglich, da sonst auch zahlreiche Exons von *FANCD2* nicht durch Sonden abgedeckt wären und somit in den NGS-Daten fehlen würden.

In der aktuellen Pseudogendatenbank (*Build* 83) wurden insgesamt fünf Pseudogene für *FANCD2* (ENSG00000144554) auf dem Plusstrang von Chromosom 3 gelistet. Dabei handelt es sich um *PGOHUM00000299338* (*FANCD2P1*), *PGOHUM00000299341* (*FANCD2P2*), *PGOHUM00000299342* (*FANCD2P3*), *PGOHUM00000305819* (*FANCD2P4*) sowie *PGOHUM00000305820* (*FANCD2P5*). Eine genauere Analyse der aufgeführten Pseudogene zeigte, dass *FANCD2P4* und *FANCD2P5* gegen Regionen komplexer Alleldiversität mappen und alternativen Loci von Chromosom 3 (CHR\_HSCHR3\_1\_CTG1) zugeordnet wurden. Ein Sequenzvergleich unter den Pseudogenen zeigte weiterhin, dass *FANCD2P5* mit *FANCD2P3* identisch ist und *FANCD2P4* zu 99,82 % mit *FANCD2P2* übereinstimmt. Deshalb wurden *FANCD2P4* sowie *FANCD2P5* von der weiteren Analyse ausgeschlossen und nur mit den übrigen drei Pseudogenen weitergearbeitet. Eine Übersicht dazu findet sich in Tabelle 27 wieder.

**Tabelle 27:** Übersicht zu FANCD2-Pseudogenen. Angegeben ist die chromosomale Position (hg38) ("Chr" = Chromosom), die Größe sowie Übereinstimmung der Sequenzen mit FANCD2. Zudem ist die Pseudogenklasse vermerkt. Die Daten beruhen auf www.pseudogene.org/.

Gen	Chr	Start	Stopp	Strang	Größe [bp]	Sequenzähnlichkeit zu <i>FANCD2</i> [%]	Klasse
FANCD2	3	10 026 429	10 101 930	+	75 501	100	-
FANCD2P1	3	9 992 874	9 994 323	+	1 449	93,48	Unbestimmt
FANCD2P2	3	11 873 631	11 885 678	+	12 047	90,89	Dupliziert
FANCD2P3	3	11 886 770	11 891 579	+	4 809	92,82	Dupliziert

In Hinblick auf die genomische Struktur lassen sich Pseudogene verschiedenen Klassen zuordnen. So weisen prozessierte Pseudogene keine Introns mehr auf, wohingegen durch Duplikation entstandene Pseudogene eine typische Exon-Intron-Struktur zeigen. In diese Gruppe fallen *FANCD2P2* sowie *FANCD2P3*. Daneben gibt es noch Pseudogene, welche sich keiner dieser Klassen eindeutig zuordnen lassen, da die Sequenzen zu stark degradiert sind. Darunter fällt *FANCD2P1*, welches mit 1,5 kb das kleinste der identifizierten Pseudogene ist. Durch eine BLAST-Analyse der Sequenzen konnte *FANCD2P1* den Exons 15 und 16 zugeordnet werden, inklusive der Sequenz von Intron 15. Die Sequenzhomologie von *FANCD2P2* zu *FANCD2* erstreckte sich hingegen von Intron 18 bis Intron 25. Direkt daran schloss die Sequenz von *FANCD2P3* an, welche Exon 26 bis Exon 28, einschließlich der dazwischen liegenden Intronsequenzen, umfasst (Abbildung 55A+B).



Abbildung 55: Chromosomale Position von *FANCD2* und dessen Pseudogenen (hg38). A: Das Ideogramm weist auf die Lokalisation von *FANCD2* auf dem kurzen Arm von Chromosom 3 hin (rote Markierung). B: Der Abschnitt von Chromosom 3 von Position 9 990 000-11 895 000 ist schematisch dargestellt. Darunter befinden sich die Genstrukturen der Pseudogene 1-3 sowie *FANCD2*, entsprechend ihrer Position auf Chromosom 3 angeordnet. Gefülte Boxen symbolisieren Exons (blau: Pseudogen; grau: *FANCD2*), gestrichelte Linien dagegen Introns (blau: Pseudogen; schwarz: *FANCD2*). Pfeile markieren die Leserichtung. Über den Exons der Pseudogene sind die homologen Exons von *FANCD2* angegeben

#### 4.5.2 Bestätigung hypomorpher Mutationen

Zusätzlich zu den bereits beschriebenen FA-D2-Patienten FA-09, FA-24 sowie FA-59 konnten 24 weitere Patienten aus 21 Familien identifiziert werden. Eine Übersicht dieser neuen FA-D2-Patienten und deren Mutationen ist in Tabelle 28 eingetragen sowie in Abbildung 56 dargestellt. Insgesamt fanden sich 24 verschiedene Mutationen, wovon 11 bisher bei keinem FA-D2-Patienten beschrieben wurden. Bei über 40 % aller nachgewiesenen Mutationen (n = 10/24) konnte eine Auswirkungen auf das aberrante Spleißen bestätigt werden, darunter auch bei zwei von drei neuen Mutationen. So führt die Substitution c.3467-2A>G in Intron 34 zum Verlust der kanonischen Spleißakzeptorsequenz, was das Skipping von Exon 35 zur Folge hat (r.3467\_3560del). Diese heterozygote Mutation konnte bei FAD2-02 dem paternalen Allel zugeordnet werden und in gDNA aus Blut sowie Fibroblasten verifiziert werden. In einer lymphoblastoiden Zelllinie war diese Substitution allerdings nur marginal nachweisbar. Der Frameshift resultiert in einer vorzeitigen Termination der Translation (p.A1156Vfs\*11), was sich in einem instabilen Transkript auf cDNA-Ebene widerspiegelte. Im Gegensatz dazu führt die homozygote Substitution c.2975A>G zum Skipping der 117 bp von Exon 30, was zu einer in-frame Deletion von 39 Aminosäuren führt (p.A954\_K992del). Das Spleißvorhersageprogramm Human Splicing Finder hatte den Verlust der wildtypischen Spleißdonorstelle, NNSPLICE die Entstehung einer neuen Spleißakzeptorstelle durch die Mutation vorhergesagt. Die Substitution konnte bei FAD2-04.1 homozygot nachgewiesen werden. Von FAD2-20 stand kein Material zur Verfügung, um den Effekt der Deletion c.1947+2del nachweisen zu können, welche den kanonischen Spleißdonor von Intron 21 betrifft.

Nach Spleißmutationen waren Indel-Mutationen die häufigste Art von Varianten, welche bei den hier untersuchten FA-D2-Patienten bestimmt werden konnte. Von insgesamt sechs Indel-Mutationen waren vier noch nicht beschrieben. Darunter befand sich nur eine Duplikation, welche bei FAD2-15 aus Frankreich heterozygot bestätigt werden konnte. Für c.4336\_4337dup wurde ein Frameshift vorausgesagt, welcher ein verlängertes Protein zur Folge hat (p.S1446Rfs\*28). Die Deletion eines Triplettrepeats in Exon 5 führt dagegen zur i*n-frame* Deletion p.E95del und wurde bei FAD2-05 heterozygot auf gDNA-Ebene bestimmt. Dagegen resultiert die 2-bp-Deletion c.1632\_1633del von FAD2-12 unmittelbar in einem Stoppcodon (p.N545\*). Die Deletion eines Thymins an Position c.2534 in Exon 18, die bei FAD2-24 homozygot bestätigt wurde, hat stattdessen einen Frameshift zum Ergebnis, welcher ebenfalls zu einer vorzeitigen Termination der Translation führt.

ID	Mutation 1			Mutation 2			Hadmark
	DNA	Protein	Ref.	DNA	Protein	Ref.	Herkunnt
FAD2-01	c.1948-16T>G	p.E650*	[Kalb et al., 2007]	c.1948-16T>G	p.E650*	[[Kalb et al., 2007]	vermutlich Türkei
FAD2-02	c.2444G>A	p.R815Q	[Kalb et al., 2007]]	c.3467-2A>G	p.A1156Vfs*11	-	Frankreich/ Schweiz
FAD2-03	c.1948-16T>G	p.E650*	[Kalb et al., 2007]	c.1948-16T>G	p.E650*	[Kalb et al., 2007]	Türkei
FAD2-04.1	c.2975A>G	p.A954_K992del	-	c.2975A>G	p.A954_K992del	-	Türkei?
FAD2-04.2	c.2975A>G	p.A954_K992del	-	c.2975A>G	p.A954_K992del	-	vermutlich Türkei
FAD2-05	c.238_285del	p.E95del	-	c.2443C>T	p.R815W	-	Deutschland
FAD2-07.1	c.2396C>A	p.A799D	-	c.3803G>A	p.W1268*	[Kalb et al., 2007]	Serbien
FAD2-07.2	c.2396C>A	p.A799D	-	c.3803G>A	p.W1268*	[Kalb et al., 2007]	Serbien
FAD2-09	c.1948-16T>G	p.E650*	[Kalb et al., 2007]	c.1948-16T>G	p.E650*	[Kalb et al., 2007]	vermutlich Türkei
FAD2-11	c.1948-6C>A	p.E650*	[Kalb et al., 2007]	c.1948-6C>A	p.E650*	[Kalb et al., 2007]	Deutschland
FAD2-12	c.1632_1633del	p.N545*	-	c.1948-6C>A	p.E650*	[Kalb et al., 2007]	Deutschland
FAD2-13	c.2444G>A	p.R815Q	[Kalb et al., 2007]	c.2715+1G>A	p.E906Lfs*4	[Kalb et al., 2007]	;
FAD2-15	c.782A>T	p.S232Rfs*6	[Kalb et al., 2007]	c.4336_4337dup	p.S1446Rfs*28	-	Frankreich
FAD2-16	c.274-57_274- 56ins298	p.I92Yfs*7*	[[Kalb et al., 2007]	c.2444G>A	p.R815Q	[Kalb et al., 2007]	Portugal
FAD2-17	c.782A>T	p.S232Rfs*6	[Kalb et al., 2007]	c.1370T>C	p.L457P	[Kalb et al., 2007]	Spanien
FAD2-18	c.782A>T	p.S232Rfs*6	[Kalb et al., 2007]	c.784- ?_1134+?dup	p.V262_L378dup	[Kalb et al., 2007]	Italien
FAD2-19	c.2444G>A	p.R815Q	[Kalb et al., 2007]	c.2444G>A	p.R815Q	[Kalb et al., 2007]	Frankreich
FAD2-20	c.2444G>A	p.R815Q	[Kalb et al., 2007]	c.1947+2del	p.?	-	Frankreich
FAD2-21.1	c.1321_1322del	p.V379_K515del	[Kalb et al., 2007]	c.2444G>A	p.R815Q	[Kalb et al., 2007]	Frankreich
FAD2-21.2	c.1321_1322del	p.V379_K515del	[Kalb et al., 2007]]	c.2444G>A	p.R815Q	[Kalb et al., 2007]	Frankreich
FAD2-22	c.2444G>A	p.R815Q	[Kalb et al., 2007]	c.2444G>A	p.R815Q	[Kalb et al., 2007]	Deutschland/ Frankreich
FAD2-23	c.782A>T	p.S232Rfs*6	[Kalb et al., 2007]	c.782A>T	p.S232Rfs*6	[Kalb et al., 2007]	Frankreich
FAD2-24	c.2534del	p.V845Gfs*4	-	c.2534del	p.V845Gfs*4	-	?
FAD2-25	c.3453_3456del	p.N1151Kfs*46	[Kalb et al., 2007]	c.782A>T	p.S232Rfs*6	[Kalb et al., 2007]	Frankreich

Tabelle 28: Neu zugeordnete FA-D2-Patienten. "ID": Patienten-ID. "Ref.": Referenz. "-": keine Angabe. "?": unbestätigt/unbekannt. In den Familien 4, 7 sowie 21 konnten je zwei betroffene Kinder mit den entsprechenden Mutationen bestätigt werden.


Abbildung 56: Schematische Übersicht aller bislang identifizierten *FANCD2*-Mutationen. Das nicht-codierende Exon 1 ist grau, die Transkriptionsstartstelle mit einem Pfeil markiert. Einzelne Exons sind zur Übersichtlichkeit nummeriert. Über der Exonstruktur der codierenden Sequenz sind alle im Verlauf dieser Arbeit identifizierten Mutationen eingetragen. Bereits in der FA-Datenbank vermerkte Varianten sind unterhalb der cDNA angegeben. Intronische Mutationen mit einer Auswirkung auf das aberrante Spleißverhalten sind als ungefüllte, exonische Spleißmutationen als gefüllte Pfeilspitzen dargestellt. Gefüllte Rauten symbolisieren Nonsense-, gefüllte Kreise Missense-Mutationen. Indel-Mutationen <5 bp sind als ungefüllte Quadrate dargestellt, größere Indels dagegen als gefüllte Rechtecke. Sofern eine Mutation homozygot nachgewiesen wurde, ist ein "H" vermerkt. War die Mutation vorher noch nicht beschrieben, trägt sie ein "N".

Unter insgesamt fünf verschiedenen Missense-Mutationen waren drei neue Varianten zu finden. Neben der bei FA-59 beschriebenen Missense-Mutation c.3996G>T (p.Q1332H) in Exon 41 konnten zwei weitere in Exon 26 identifiziert werden. So wurde bei den Geschwistern FAD2-07.1 und FAD2-07.2 die heterozygote Substitution c.2396C>A festgestellt, welche eine hochkonservierte Aminosäure betrifft (p.A799D). Einen Eintrag im ExAC-Browser gibt es zu dieser Variante nicht. Dagegen war der Basenaustausch c.2443C>T (rs763185403), welcher bei FAD2-05 heterozygot bestimmt wurde, mit einer Allelfrequenz von 1,71\*10-<sup>5</sup> hinterlegt. Auf Proteinebene führt diese Missense-Mutation zur Substitution eines Arginins gegen ein Tryptophan (p.R815W). Die bereits veröffentlichte Mutation c.2444G>A (rs766567785) führt ebenfalls zum Austausch des Arginins an Position 815, allerdings gegen ein Glutamin. In der hier untersuchten Patientengruppe konnte diese Mutation sechs Mal heterozygot sowie zwei Mal homozygot bestimmt werden (Abbildung 56, Tabelle 28) und ist insgesamt betrachtet die am häufigsten nachgewiesene Mutation bei FA-D2-Patienten (n = 14). Im ExAC-Browser ist diese Substitution mit einer Allelfrequenz von 1,46\*10-4 angegeben.

Von insgesamt 23 der 27 neuen FA-D2-Patienten lagen zudem Informationen bezüglich des Phänotyps vor. Dabei zeigte jeder Patient mindestens zwei Auffälligkeiten. Am häufigsten traten Kleinwuchs (83 %) sowie Mikrozephalie (78 %) auf. Bei der Untersuchung der Organe wurden bei knapp der Hälfte der Patienten Nierenanomalien (n = 11) und bei gut einem Viertel (n = 6) angeborene Herzfehler festgestellt. Daumenfehlbildungen (11/23), Pigmentierungsstörungen (10/23) sowie Radialstrahlanomalien (9/23) waren ebenfalls häufig. Bei sieben Patienten wurden zusätzlich Ohrfehlbildungen bzw. Schwerhörigkeit und bei jeweils 26 % auch eine Mikrophthalmie oder psychomotorische Entwicklungsstörung diagnostiziert. Ferner waren bei neun Patienten hämatologische Auffälligkeiten angegeben. Normale Blutwerte zeigten sich bei vier Patienten mit Mosaikstatus im hämatopoetischen System. Bei FAD2-02 waren bis zum Alter von 40 Jahren keine Veränderungen des Blutbildes aufgetreten, obwohl sich im Blut keine Reversion einer Mutation nachweisen lies. Dagegen war die Spleißmutation c.3467-2A>G in einer aus Blut etablierten lymphoblastoiden Zelllinie revertiert. Der Patient war erst durch die Diagnose eines Plattenepithelkarzinoms im linken Oberkiefer im Alter von 30 Jahren als möglicher FA-Patient aufgefallen. Außerdem war bei FAD2-20 ein Karzinom in der Mundhöhle im Alter von 24 Jahren festgestellt worden.

# 5. Diskussion

In der vorliegenden Arbeit wurde zunächst ein FA-spezifisches Genpanel etabliert und dieses zur molekularen Diagnosestellung bei FA-Patienten verwendet. Damit konnte ein Großteil der Exons bekannter FA-Gene sowie patientenspezifische Mutationen in intronischen Sequenzabschnitten angereichert werden. Allerdings zeigten sich Schwächen bei der verwendeten Sequenzierungstechnologie in Form von teilweise unzureichender Abdeckung der Zielregionen. Als Alternative zum FA-Panel wurde ein kommerzielles Genpanel zur Mutationsanalyse verwendet, welches eine Vielzahl der bekannten FA-Gene enthält und mit einem anderen Sequenziersystem analysiert wurde. Zusätzlich wurde bei einigen FA-Patienten, deren Mutationen durch die Gen-Anreicherungspanels nicht geklärt werden konnten, zusätzlich ein WES durchgeführt. Ein größerer Datendurchsatz der für das kommerzielle Genpanel und WES verwendeten Sequenziermethoden erlaubte eine bessere Abdeckung der jeweils angereicherten Zielregionen. Das WES zeigte seine Stärke vor allem bei der Identifizierung von Mutationen in Genen außerhalb des FA/BRCA-Signalweges. Durch die Kombination aller Anreicherungsstrategien konnte eine molekulare Diagnose bei einer Vielzahl der untersuchten Patienten gestellt werden.

## 5.1 Etablierung eines Anreicherungspanels zur NGS-Analyse von FA-Patienten

Der erste Teil der Arbeit befasste sich mit der Etablierung eines spezifischen Genpanels, um die Routinediagnostik, bei welcher zur Mutationsanalyse vorwiegend Sanger-Sequenzierungen durchgeführt werden, durch eine Hochdurchsatz-Sequenzierung aller bekannten FA-Gene in einem Ansatz zu unterstützen oder in Teilbereichen zu ersetzen. Durch die Aufnahme von Kandidatengenen in das Panel sollte zudem eine gleichzeitige Untersuchung potentieller neuer FA-Gene ermöglicht werden.

Die Verwendung eines kundenspezifischen Genpanels machte eine Anpassung der Protokolle zur Anreicherung nötig, wie sich bereits bei der Erstellung der Probenlibrary zeigte. Dabei erwies sich die anfangs eingesetzte Menge an DNA als zu gering, da während der Bearbeitung der Proben ein Teil der fragmentierten DNA verloren ging und somit keine ausreichende Konzentration des Templates für die folgende LM-PCR zur Verfügung stand. Es blieb bereits bei der Fragmentierung der DNA durch Nebulisierung einiges Material in dem dazu verwendeten Gefäß zurück. Weiterhin ging bei der Aufreinigung durch die SPRI-Bead-Selektion Material verloren, so dass sich die Menge an Template für die LM-PCR immer weiter verringerte. Die Ausgangsmenge vor der Fragmentierung musste deshalb um das Vierfache erhöht werden, um zufriedenstellende Ergebnisse zu erhalten, auf deren Grundlage weitergearbeitet werden konnte. Allerdings stellte die Ausgangsmenge bei den meisten hier analysierten Patienten keinen limitierenden Faktor dar, da genügend Material zur Verfügung stand. Ein Problem wäre die Ausgangsmenge bei Proben mit begrenzter Zugänglichkeit, wie z. B. bei Tumorproben. In solchen Fällen müsste ein anderes Protokoll zur Vorbereitung der Probenlibrary verwendet werden. Für die Fragmentierung stünden verschiedene Alternativen zur Nebulisierung zur Auswahl, wie etwa die Sonifikation oder die Verwendung bestimmter Enzyme. Insbesondere bei der Fragmentierung durch enzymatischen Verdau, welcher entweder durch Endonukleasen erfolgen oder durch Transposase-Homodimere vermittelt werden kann, wird deutlich weniger DNA benötigt [Goryshin et al., 1998; Syed et al., 2009; Adey et al., 2010]. Die Homodimere werden beispielsweise zur Fragmentierung der DNA für die Anreicherung mit dem TruSight Cancer-Panel verwendet. In einem als Tagmentierung bezeichneten Schritt wird dabei die DNA an zufälligen Stellen fragmentiert und gleichzeitig werden Adapter an die entstehenden Fragmente ligiert. Die erhebliche Zeitersparnis durch diese Methode stellt, neben dem geringen Probeneinsatz, einen weiteren Vorteil dieses Protokolls dar. Allerdings sollte man bedenken, dass eine geringe Ausgangsmenge zur Erstellung einer Probenlibrary auch Nachteile mit sich bringen kann. Um auf eine ausreichende Menge an Probenlibrary zu kommen, muss die fragmentierte Probe stärker amplifiziert werden, was wiederum zur Einschränkung der Komplexität der Library führen kann [Parkinson et al., 2012]. In einer Studie von Knierim et al. 2011 [Knierim et al., 2011] wurden die verschiedenen Möglichkeiten zur Fragmentierung durch Nebulisierung, Sonifikation und enzymatischen Verdau verglichen. Dabei konnten keine größeren Unterschiede hinsichtlich der Effizienz festgestellt werden. Allerdings resultierte der Endonuklease-Verdau in einer höheren Anzahl an künstlich erzeugten Indel-Mutationen. In der vorliegenden Arbeit konnten durch die Nebulisierung reproduzierbare Ergebnisse erzielt werden, was ein entscheidender Grund dafür war, diese Methode beizubehalten.

Neben der eingesetzten Menge an DNA zur Fragmentierung bedurfte auch die zur Amplifikation der Probenlibrary durchgeführte LM-PCR-Reaktion einer Anpassung. Es zeigte sich, dass im Reaktionsansatz ein Überschuss an Primern enthalten war, der nachweislich die Effizienz der PCR beeinträchtigte Durch Verringerung der eingesetzten Primermenge konnte ein Rückgang an Primerdimeren wie auch eine Steigerung der Amplifikation nachgewiesen werden. Eine zu niedrige Templatekonzentration als Grund für die geringe Amplifikation erscheint eher unwahrscheinlich, da schon sehr früh während der Etablierungsphase der Methode die Ausgangsmenge zur Fragmentierung erhöht wurde.

Des Weiteren konnte gezeigt werden, dass ein optimiertes Verhältnis von angereicherten DNA-Molekülen zu *Capture Beads* für die Erzeugung qualitativ hochwertiger Reads von entscheidender Bedeutung ist. Bei einer zu hoch gewählten Ratio ist es wahrscheinlicher, dass mehrere einzelsträngige DNA-Fragmente der Probenlibrary an ein *Capture Bead* binden, die in der darauf folgenden emPCR als Template dienen. Eine monoklonale Amplifikation je eines DNA-Moleküls pro Well wäre somit nicht gegeben. Beim seriellen Fluss der Nukleotide über die Pikotiterplatte würden bei Wells mit mehreren DNA-Molekülen mehrere Nukleotideinbauten verzeichnet und diese Wells von der Software aussortiert, wie es auch bei den ersten eigenen Sequenzierungen beobachtet wurde. Durch eine Anpassung der Ratio konnte dieser Anteil an verworfenen Reads mit gültiger Schlüsselsequenz von anfangs durchschnittlich 23 % (Projekte 1 und 2) auf rund 4 % gesenkt werden. Im selben Filterungsschritt wurden zudem die Wells der Pikotiterplatte erkannt, die mehr als ein *Capture Bead* beinhalteten. Dies trat auf, wenn auf die Platte zu viele angereicherte Beads zur Sequenzierung geladen wurden. Eine Abschätzung der zur Sequenzierung eingesetzten Beadmenge war nur durch den *Bead Counter* möglich, was jedoch schr subjektiv war, da es sich um eine rein optische Abschätzung

der Beadmenge ohne eine Skalierung handelte. Unterschiede in der Anzahl der jeweils eingesetzten angereicherten Beads waren dabei nicht zu vermeiden und trugen, besonders während der ersten Sequenzierungen, zu einem hohen Anteil an verworfenen Reads bei. In Hinblick auf die zur Sequenzierung vorbereitete Library konnte festgestellt werden, dass eine Lagerung von mehreren Tagen, wie sie laut des offiziellen Protokolls möglich ist, mit einer Erhöhung von Lücken in der ROI einherging, weshalb eine Sequenzierung möglichst zeitnah durchgeführt werden sollte. Dies lässt sich damit erklären, dass die DNA nach der emPCR denaturiert wird und im einzelsträngigen Zustand deutlich instabiler ist.

Obwohl während der ersten Sequenzierungen noch nicht alle Parameter optimal gewählt worden waren und die durchschnittliche Abdeckung nur bei knapp 9x lag, konnten dennoch die pathogenen biallelischen Varianten der vier untersuchten Patientenproben (FA-01–FA-04) detektiert werden. Dies spricht für die Spezifität und Sensitivität des Anreicherungspanels. So wurden schließlich weitere 51 DNAs angereichert und sequenziert, von welchen ein Großteil einer FA-Komplementationsgruppe zugeordnet und die kausalen Mutationen identifiziert werden konnten. Auch für das *TruSight Cancer*-Panel erfolgte zunächst eine Validierung durch die Analyse von drei Patienten mit bekannten Mutationen, die alle bestätigt werden konnten. Allerdings mussten die Einstellungen zur Allelfrequenz bei FA-02 erst deutlich erniedrigt werden, ehe die Mutation (c.2371\_2399dup in *FANCQ*) als solche erkannt wurde. Vermutlich war das andere Allel durch bevorzugte Amplifikation in der Probenlibrary überrepräsentiert, so dass die Mutation in einer Region mit über 400-facher Abdeckung nicht bestimmt werden konnte. Daher sollte der Schwellenwert zur Allelfrequenz bei der Datenauswertung bedacht und ein Datenset gegebenenfalls nachanalysiert werden, wenn nur eine oder keine pathogene Mutation mit den gewählten Einstellungen nachgewiesen werden kann.

## 5.2 Verschiedene NGS-Strategien zur Mutationsanalyse von FA-Genen

#### 5.2.1 Anreicherung ausgewählter Zielregionen

Die Einführung von NGS-Methoden hat in nur drei Jahren zur Identifizierung von sechs weiteren Genen beigetragen, welche mit FA in Verbindung gebracht werden [Bogliolo et al., 2013; Hira et al., 2015; Sawyer et al., 2015; Wang et al., 2015; Bluteau et al., 2016; Park et al., 2016]. Eine konventionelle FA-Diagnostik unter Zuhilfenahme der Sanger-Sequenzierung wird bei steigender Anzahl an bekannten FA-Genen immer zeitaufwändiger und teurer. Dies gilt insbesondere, wenn der Patient einer seltenen Komplementationsgruppe angehört. Da es bislang kein kommerziell erhältliches FA-Genpanel gibt, wurde im Rahmen dieser Arbeit ein FA-spezifisches *SeqCap EZ*-Panel erstellt, das in mehreren leicht abgewandelten Versionen zur Sequenzierung der DNA von FA-Patienten verwendet wurde. Mit Hilfe des Panels konnte flexibel auf die Identifizierung neuer FA-Gene reagiert werden. Zudem konnten, anders als beim *TruSight Cancer*-Panel oder beim WES, auch spezifische pathogene Mutationen außerhalb des Bereiches der codierenden Sequenzen angereichert und untersucht werden. Die bekannte Insertion eines Alu-Elementes, welche bislang bei insgesamt drei FA-D2-

Patienten identifiziert werden konnte, war darin ebenso enthalten wie verschiedene intronische Mutationen. So ließ sich ein möglichst umfassendes Spektrum an bekannten FA-Mutationen erfassen. Die spezifische Anreicherung der Alu-Mutation konnte durch die Untersuchung eines bekannten FA-D2-Patienten, der heterozygoter Träger dieser Mutation ist, bestätigt werden. Da rund 100 bp auf beiden Seiten der Sonden durch die Sequenzierung erfasst wurden, waren auch kleinere Introns in den Sequenzdaten enthalten, was auch für die anderen durchgeführten Anreicherung ausgenommen. Die Anreicherung der gesamten Intronsequenz ausgewählter Gene wäre prinzipiell machbar, würde allerdings zu Lasten der allgemeinen Abdeckung gehen. Mit einem GS Junior wäre dies nicht möglich, da dieser zum einen zu wenige Daten generiert und zum anderen seit Ende des vergangenen Jahres die Produktion von Sequenzierreagenzien für das Gerät eingestellt wurde.

Im Gegensatz zur FA gibt es kommerzielle Anreicherungspanels für zahlreiche Krankheiten oder damit verbundenen Signalwege. Da der FA/BRCA-Reparaturweg mit vielen anderen Signalwegen der DNA-Reparatur vernetzt ist bzw. überlappt, wurde ein Panel gesucht, das eine möglichst große Schnittmenge mit den bekannten FA-Genen aufweist. Durch das *TruSight Cancer*-Panel, welches zur molekulargenetischen Diagnostik von Krebserkrankungen wie Brust- und Eierstockkrebs oder auch Darmkrebs verwendet wird, werden insgesamt 94 Gene sowie zusätzliche 284 SNPs angereichert, die mit einer Prädisposition für verschiedene Krebserkrankungen einhergehen. Unter diesen Genen finden sich 17 der 21 bekannten FA-Gene, lediglich *RAD51/FANCR*, *UBE2T/FANCT*, *XRCC2/FANCU* sowie *REV7/FANCV* sind nicht enthalten. Diese Gene codieren für die zuletzt identifizierten FA-Komplementationsgruppen, für welche bislang nur ein bis maximal drei Patienten bekannt sind [Ameziane et al., 2015; Hira et al., 2015; Wang et al., 2015; Bluteau et al., 2016; Park et al., 2016]. Eine Nachanalyse dieser Gene durch Sanger-Sequenzierung ist daher nötig, sofern alle anderen Komplementationsgruppen ausgeschlossen werden können.

Durch das vorgegebene Sondenset wurden auch hier nur die codierenden Exons der verschiedenen Gene angereichert. Patientenspezifische tief intronische Mutationen waren somit von einer Anreicherung ausgeschlossen. Zudem erscheint es unwahrscheinlich, dass die *FANCD2*-Alu-Insertion mit dieser Anreicherungsstrategie abgedeckt worden wäre, da Alu-Elemente aufgrund ihrer hohen Sequenzidentität untereinander wie auch hochrepetitive Sequenzen in der Regel bei der Erstellung von Sonden ausgeschlossen werden. Da durch die Sequenzierung der angereicherten Fragmente auch essenzielle Spleißstellen abgedeckt wurden, konnte dennoch ein Großteil der bekannten Mutationen durch das Krebspanel erfasst werden.

Ähnliches gilt für die verschiedenen Kits zur Exomanreicherung, welche das umfassendste Bild über mögliche pathogene Mutationen in den codierenden Regionen liefern. Mit der Gesamtexomanalyse wurde in dieser Arbeit nicht primär die Anwendbarkeit in der Routinediagnostik untersucht, da Mutationen in den häufigsten FA-Genen bei allen per WES untersuchten Patienten zuvor durch eine Panelsequenzierung ausgeschlossen worden waren. Diese Patienten galten als gute Kandidaten zur Identifizierung von Genen, welche bislang nicht mit FA assoziiert sind, so dass hier auch ein experimenteller Ansatz verfolgt wurde. Bei zwei Patienten konnten mit Hilfe von WES tatsächlich potentiell pathogene Mutationen in RPA1 bzw. NTHL1 identifiziert werden so dass diese, trotz einer auffälligen G2-Phasen-Arretierung in entsprechenden Zelllinien, unabhängig von der Gruppe der FA-Patienten betrachtet wurden. In der Identifizierung von Mutationen in Genen außerhalb des kanonischen FA/BRCA-Signalweges spiegelt sich das Potenzial der WES wider, da rund 95 % aller codierenden Exons abgedeckt werden, welche etwa 85 % aller pathogenen Mutationen enthalten und sich unabhängig von der Fragestellung analysieren lassen [Rabbani et al., 2014]. Gerade bei Krankheiten wie FA, die ein heterogenes klinisches Erscheinungsbild zeigen, stellt dies, in Verbindung mit stetig sinkenden Preisen der verwendeten Reagenzien und vergleichbarem Zeitaufwand, einen großen Vorteil gegenüber der Panelsequenzierung dar. In Hinblick auf den Einsatz zu diagnostischen Zwecken ergeben sich aber auch ethische Probleme in Form von Nebenbefunden bzw. unvorhergesehenen Befunden (incidental findings). Diese können sich unabhängig von der ursprünglichen Fragestellung ergeben und für den Patienten sowie dessen Angehörige weitreichende Folgen haben. Dabei stehen sich das Recht auf Unwissen des Patienten sowie der mögliche Nutzen über das Wissen solcher Varianten gegenüber. Daher ist eine klare Regelung zum Umgang mit diesen Befunden unabdingbar. Eine mögliche Orientierung dazu geben verschiedene veröffentlichte Leitlinien, wie etwa des American College of Medical Genetics and Genomics (ACMG) [Green et al., 2013; Rehm et al., 2013]. Nach einer Überarbeitung finden sich darin aktuell 59 Gene gelistet, für die eine hohe Penetranz bekannt ist und das Wissen über das Vorliegen einer pathogenen Mutation auch einen klinischen Nutzen mit sich bringt [Kalia et al., 2017]. Bekannte krankheitsverursachende Mutationen in diesen Genen oder Varianten, für die eine Pathogenität angenommen wird [Richards et al., 2008], sollen nach den Empfehlungen der ACMG, unabhängig von der initialen Fragestellung, mit in den Befund aufgenommen werden. Um Nebenbefunde zu umgehen, kann die Auswertung von Exom-Daten auch auf bestimmte krankheitsassoziierte Gene beschränkt werden, indem eine vordefinierte Genliste als Region von Interesse vorgegeben wird. Durch die Speicherung der Rohdaten können diese später erneut analysiert und die Untersuchung auf andere Regionen oder das gesamte Exom erweitert werden.

Die Steigerung der Anzahl untersuchter Regionen geht auch mit einer Zunahme an detektierten Varianten einher, welche interpretiert werden müssen. Neben effizienten Filterungsstrategien können diese durch die Trio-Exomanalyse eingeschränkt werden, bei welcher neben dem Patienten auch die Eltern analysiert werden. Insbesondere bei autosomal-rezessiven Erkrankungen lässt sich durch die exomweite Segregationsanalyse die Anzahl an potentiell krankheitsverursachenden Varianten effizient einschränken und auch *de novo*-Varianten lassen sich so schnell bestimmen. In der Literatur findet sich bei einem direkten Vergleich der Aufklärungsquote von Trio- gegen Einzelperson-Exomanalysen eine signifikante Erhöhung von 21 % auf

37 % [Farwell et al., 2015]. Eine weitere Möglichkeit der Exomanalyse, die vorwiegend zu diagnostischen Zwecken eingesetzt werden soll, besteht in der Sequenzierung des sogenannten klinischen Exoms. Mit Hilfe des TruSight One Sequencing Panels werden zur sogenannten klinischen Exomanalyse 4813 Gene angereichert, die mit einem klinischen Phänotyp assoziiert sind. Da nur gut ein Fünftel des gesamten Exoms analysiert wird, kann eine höhere Abdeckung krankheitsrelevanter Gene erzielt als auch eine einfachere Interpretation der Sequenzdaten erreicht werden. Da intronische Varianten selten sind und zudem durch weitere Analysen auf Transkriptebene bestätigt werden müssen, sind Genpanels oder WES der Sequenzierung des gesamten Genoms (WGS) für diagnostische Zwecke weiterhin vorzuziehen. Auch wenn das Exom nur ca. 1 % der genomischen Sequenz ausmacht, so können in diesem Bereich 85 % aller pathogenen Mutationen bestimmt werden [Rabbani et al., 2014]. Auch bei der NGS-Analyse der verschiedenen FA-Patienten konnten die meisten pathogenen Mutationen in den codierenden Exons oder an den Exon-Intron-Grenzen nachgewiesen werden. Zudem wird beim WGS die Interpretation der detektierten Varianten durch die deutliche Steigerung an generierten Daten erschwert. Eine weitere Möglichkeit der Hochdurchsatzsequenzierung besteht in der Transkriptomsequenzierung, die u.a. zur Quantifizierung aller exprimierten Transkripte, zur Untersuchung von Einzelnukleotid-Polymorphismen, Spleißvarianten, posttranskriptionellen Modifikationen oder zur Identifizierung von Isoformen herangezogen werden kann. Letztlich lässt sich allerdings das volle Mutationsspektrum nur durch eine Kombination verschiedener Methoden sicher abdecken.

# 5.2.2 Abdeckung der angereicherten Zielregionen durch unterschiedliche Anreicherungs- und Sequenziermethoden

Im Durchschnitt konnte die Abdeckung der ROI des *SeqCap EZ*-Panels durch die Anpassung der Protokolle sowie Optimierungen im Ablauf von durchschnittlich 12x auf 46x gesteigert werden. Beim WES lag dieser Wert bei 74x und damit noch immer deutlich unter der durchschnittlichen Abdeckung der Krebspanel-Sequenzierung, bei der eine mittlere Abdeckung von 621x erzielt werden konnte. Diese Ergebnisse sagen allerdings wenig über die Anreicherungsmethode selbst aus, sondern resultieren aus den unterschiedlichen Sequenziergeräten, mit welchen die Libraries analysiert wurden. Diese unterscheiden sich deutlich in ihrer Leistungsfähigkeit zur Erzeugung von Sequenzdaten. So beträgt der *Output* an einem GS Junior 30 Mb (Megabasen), wohingegen der Durchsatz eines MiSeq-Sequenzierers von 540 Mb bis 15 Gb reicht, womit maximal 25 Millionen Reads einer Leseweite von 2 x 300 bp produziert werden können. In einem ganz anderen Größenbereich hinsichtlich des Datenoutputs bewegen sich dagegen die Sequenzierer, die zum WES genutzt wurden. Während der Durchsatz des SOLiD 5500xl mit 180 Gb angegeben ist, beträgt dieser beim HiSeq 4000 maximal 1 500 Gb, wodurch bis zu fünf Milliarden Reads (2 x 150 bp) pro Lauf produziert werden können. Entsprechend wurde mit diesen beiden Sequenzierern die niedrigste (15x) sowie die höchste (133x) durchschnittliche Abdeckung bei der Sequenzierung des gesamten Exoms erzielt. Auch wenn sich die beiden in dieser Arbeit verwendeten Panels zur gezielten Anreicherung ausgewählter Gene sowohl in der Anzahl als auch in der Verteilung der Sonden unterscheiden, so sind die Größen der Zielregionen mit rund 230 kb für das *SeqCap EZ*-Panel und 255 kb beim *TruSight Cancer*-Panel gut miteinander vergleichbar. So wäre eine ähnlich gute mittlere Abdeckung wie bei der Krebspanel-Sequenzierung (621x) für Proben vorstellbar, die mit dem FA-spezifischen Genpanel angereichert, aber am MiSeq sequenziert würden.

Indem die Abdeckung jeder einzelnen Zielregion des *SeqCap EZ*-Panels überprüft wurde, konnten Bereiche mit unzureichender Abdeckung identifiziert werden, welche gegebenenfalls per Sanger-Sequenzierung nachanalysiert werden mussten. Daneben konnten die Regionen bestimmt werden, welche wiederholt nicht oder nur unvollständig abgedeckt waren, um beim nächsten Sondendesign eine Nachbesserung vorzunehmen. Es zeigte sich, dass nur ein geringer Prozentsatz wiederkehrend eine unzureichende Abdeckung aufwies und es sich dabei oftmals um sehr GC-reiche Regionen handelte, was ebenso für das *TruSight Cancer*-Panel zutraf. Diese Regionen sind bekannt für eine schwierige Amplifikation und eine niedrige Abdeckung, da sie unter Standard-PCR-Bedingungen meist nicht denaturieren und somit sowohl für Primer als auch die Polymerase nicht zugänglich sind [McDowell et al., 1998; Veal et al., 2012]. Im Fall des *SeqCap EZ*-Panels war durch den Algorithmus des Herstellers die Verteilung der Sonden schon soweit optimiert, dass diese nicht weiter abgeändert werden konnte. Daher konnte nur durch eine Erhöhung der Sondendichte versucht werden, eine Verbesserung der Abdeckung zu erzielen.

Aufgrund der unterschiedlichen Anzahl an Patienten, die mit den verschiedenen Methoden untersucht wurden, ist ein Vergleich zur Abdeckung einzelner auffälliger Zielregionen schwierig. Übereinstimmend fand sich bei allen drei Panelsystemen allerdings Exon 1 von *FANCE* mit einer fehlenden oder lückenhaften Abdeckung bei mindestens der Hälfte der jeweils angereicherten Patienten. Hierzu konnte bei den *SeqCap EZ*-Panels durch die Erhöhung der Sondendichte eine leichte Verbesserung hinsichtlich der Anreicherung verzeichnet werden, indem das Exon zumindest häufiger, wenn auch immer noch lückenhaft, abgedeckt war. Dagegen konnte das erste Exon von *FANCA* mit dem FA-Genpanel nicht angereichert werden, es fehlte in allen Datensätzen. Da Mutationen in *FANCA* die häufigste Ursache der Erkrankung darstellen, sollte dieses Exon bei allen nicht-zugeordneten Patienten nachsequenziert werden. Durch einen höheren Datendurchsatz bei der Sequenzierung an einem anderen Sequenzierer könnte die mangelhafte Abdeckung einzelner Regionen kompensiert und die Sensitivität und Uniformität der NimbleGen-Anreicherung bestätigt werden [Bodi et al., 2013; Samorodnitsky et al., 2015].

Ein weiterer Vorteil einer hohen Abdeckung besteht in der Möglichkeit, Kopienzahl-Analysen durchzuführen. Da FA-A die häufigste Komplementationsgruppe darstellt und es aufgrund von zahlreichen Alu-Repeat-Elementen oftmals zu größeren Deletionen kommt, wird bislang eine FANCA-MLPA-Analyse der weiteren Mutationsanalyse vorangeschaltet. Dies galt auch für die meisten in dieser Arbeit untersuchten Patienten. Weitere FA-Gene werden in der Routinediagnostik nicht per MLPA auf Deletionen oder Duplikationen

untersucht, da diese Veränderungen dort eher selten vorkommen und den Kosten- und Zeitaufwand nicht rechtfertigen würden. Eine Analyse der Kopienzahl von Genen parallel zum NGS war bislang nur durch Array-CGH-Untersuchungen möglich. Tatsächlich wurde die Durchführbarkeit sowie Effizienz der Implementierung von Kopienzahl-Analysen in die Auswertung von NGS-Daten bereits in verschiedenen Studien zu FA gezeigt [Ameziane et al., 2012; Nicchia et al., 2015]. Ein eigener Test des CNV-Tools bei Patienten mit heterozygoten Deletionen von ein bis vier Exons in FANCA (FA-13, FA-19, FA-45), deren DNA mit dem FA-Genpanel analysiert wurde, konnte diese Beobachtung nicht bestätigen. Zwar war der Bereich der Deletion nur mit etwa halb so vielen Reads abgedeckt, allerdings erkannte das CNV-Tool der NextGENe-Auswertesoftware dies nicht als Kopienzahlvariation. Dies könnte auf die generell niedrige Abdeckung des Gens durch die FA-Panel-Anreicherung, auf die Variabilität in der Abdeckung der ROI oder auch auf die gewählten Einstellungen des CNV-Tools zurückzuführen sein. Allerdings konnte die heterozygote Deletion c.4268\_4368+37del in FANCA bei FA-13 mit dem CNV-Tool der Auswertesoftware auch nicht nach der Anreicherung mit dem TruSight Cancer-Panel und der MiSeq-Sequenzierung nachgewiesen werden, obwohl hier die Abdeckung der Region besser als nach der GS-Junior-Sequenzierung war. Dass die Detektion von größeren Deletionen mit Hilfe der verwendeten Auswertesoftware prinzipiell möglich ist, konnte anhand externer Daten von Nicht-FA-Patienten überprüft werden, welche ebenfalls mit dem Krebspanel angereichert und am MiSeq sequenziert wurden.

Das Einsparen der MLPA-Analyse würde eine deutliche Kosten- und Zeitersparnis in der FA-Mutationsanalyse bedeuten. Es zeigte sich allerdings, dass auch hier Deletionen übersehen werden können. So war die *FANCA*-MLPA-Analyse bei FA-13 sowie FA-45, welche beide heterozygote Träger der Deletion c.4268\_4368+37del sind, beide Male ohne Auffälligkeiten. Eine Überprüfung der zur MLPA verwendeten Sonden ergab hierzu, dass die Sonden zum Nachweis von Exon 43 in der 3'UTR liegen. Da dieser Bereich nicht von der Deletion betroffen war, konnte die Deletion, welche beinahe das komplette Exon 43 umfasst, nicht detektiert werden, weshalb hier eine Nachbesserung von Seiten des Herstellers vorgenommen werden sollte.

# 5.3 Komplementationsgruppen-Zuordnung durch die gezielte Anreicherung von FA-Genen

## 5.3.1 Priorisierung, Klassifizierung und Validierung von Varianten

Die gezielte Anreicherung von bekannten FA-Genen durch das *SeqCap EZ*-Panel erwies sich als effektiver Ansatz zur Mutationsanalyse bei FA-Patienten. Von 47 Patienten mit unbestimmten Mutationen konnten mit Hilfe des FA-Panels 37 Patienten (78,7%) einer Komplementationsgruppe zugeordnet und dabei die biallelischen kausalen Mutationen nachgewiesen werden. Mit dem *TruSight Cancer*-Panel wurden insgesamt nur 14 zuvor nicht näher untersuchte FA-Patienten analysiert. Nach dem Ausschluss der Diagnose FA bei FA-63 konnten neun von 13 Patienten (69,2%) zugeordnet werden. Durch WES konnte ein weiterer FA-Patient (FA-16) aufgeklärt werden. Insgesamt konnten mit allen drei Methoden 47 von 60 neuen FA-Patienten (78,3%) klassifiziert und der Nachweis von biallelischen Mutationen erbracht werden. Aufgrund der unterschiedlichen Anzahl an untersuchten Patienten und der verschiedenen Sequenzierungsstrategien ist die Effektivität der einzelnen Anreicherungen nicht direkt miteinander vergleichbar. Im Vergleich zur konventionellen Mutationsanalyse durch Sanger-Sequenzierung war bei den im Rahmen dieser Arbeit durchgeführten Anreicherungsstrategien eine wesentlich kürzere Zeitspanne zur Mutationsanalyse nötig, welche zudem mit einem bedeutend geringerem Arbeitsaufwand verbunden war. Dies machte sich insbesondere bei den Patienten bemerkbar, die nicht der Komplementationsgruppe FA-A angehören, da *FANCA* in der Routinediagnostik als erstes Gen einer Mutationsanalyse unterzogen wird, sofern nicht eindeutige Hinweise auf ein anderes FA-Gen vorliegen.

Obwohl im Fall der Panelanalyse nur eine begrenzte Anzahl von maximal 95 Genen angereichert wurde, war eine Priorisierung der Varianten für eine effiziente Mutationsanalyse hilfreich sowie nötig und eine Voraussetzung für die effektive Analyse der WES-Daten. So wurden im Schnitt in der ROI des Krebspanels etwa 300 Varianten gelistet. Beim FA-Panel war es sogar fast die doppelte Anzahl, was vor allem auf das Vorhandensein von Homopolymer-Problemen der GS Junior-Sequenzierung zurückgeführt wurde. Es wurden häufig Insertionen und Deletionen in Bereichen von Homopolymeren oder Repeatsequenzen beobachtet. Dies stimmt mit zuvor beschriebenen Ergebnissen zur 454 Pyrosequenzierung überein, wobei Homopolymere als Hauptauslöser für Sequenzierfehler, insbesondere für Indel-Mutationen, ausgemacht werden konnten [Gilles et al., 2011; Loman et al., 2012].

Zur Priorisierung werden oftmals Varianten mit einem dbSNP-Eintrag von der weiteren Analyse ausgeschlossen [Sherry et al., 2001; Ng et al., 2010; Chandrasekharappa et al., 2013]. In der vorliegenden Arbeit hatten von den neu zugeordneten Patienten über 40 % der Mutationen einen dbSNP-Eintrag. Sinnvoller erscheint es daher, die Allelfrequenz in die Priorisierung mit einzubeziehen, um seltene Varianten von SNPs zu unterscheiden. Varianten werden als selten eingestuft, wenn der Anteil des selteneren Allels (*minor allele frequency*; MAF) bei unter 1 % liegt. Während man zu diesem Zweck früher meist nur 100 Kontrollpersonen per Sanger-Sequenzierung auf eine Variante hin überprüfte, stehen heute durch verschiedene Datenbanken, welche auf NGS-Daten beruhen, deutlich größere Datensätze über das Vorkommen von Varianten in verschiedenen Populationen zur Verfügung. Ein Beispiel dafür ist der in dieser Arbeit herangezogene ExAC-Browser, in welchem man auf die Exomdaten von über 60 000 nicht verwandter Individuen zurückgreifen kann. Durch die enorme Steigerung der Anzahl verfügbarer Sequenzdaten hat sich auch bei einigen Varianten die Einstufung der Pathogenität verändert. Ein Beispiel findet sich in der Basensubstitution c.3263C>T (p.S1088F) in *EANCA*, die bei mehreren in dieser Arbeit untersuchten FA-Patienten sowohl homo- als auch heterozygot nachgewiesen werden konnte. So wurde sie bei der Erstbeschreibung als Missense-Mutation eingestuft, auch weil sie bei 100 Kontrollen nicht nachgewiesen werden konnte [Wijker et al., 1999]. Dagegen konnte sie bei einer Untersuchung von 88 Proben aus Brustkrebsfamilien ohne *BRCA1*- oder *BRCA2*-Mutationen sowie 300 Kontrollen jeweils mehrfach nachgewiesen werden [Seal et al., 2003] und ist im ExAC-Browser aktuell mit einer Frequenz von 5,1 % angegeben. Demnach kann diese Substitution nicht als seltene Variante, sondern vielmehr als Polymorphismus angesehen werden. Dies macht deutlich, dass ein Eintrag in der FA-Mutationsdatenbank (http://www2.rockefeller.edu/fanconi//) nicht mit einer gesicherten Mutation gleichzusetzen ist und immer weitere Informationen zur Mutation eingeholt werden sollten.

Durch eine Klassifizierung der priorisierten Varianten mittels *in silico*-Vorhersageprogrammen sowie Datenbank- und Literaturrecherchen konnten bei den 47 neuen FA-Patienten insgesamt 66 verschiedene Mutationen bestimmt werden. Bei Mutationen, die keiner funktionellen Untersuchung wie beispielsweise Spleißanalysen unterzogen wurden oder ein Nachweis des entsprechenden Proteins durch eine Immunblot-Analyse erfolgte, bleibt zu bedenken, dass Vorhersageprogramme oder Datenbanken eine Klassifizierung nur stützen, nicht aber die Kausalität einer Mutation beweisen können. Selbst bei trunkierenden Mutationen, die am wahrscheinlichsten pathogen sind, kann beispielsweise ein alternatives Startcodon zur Translation genutzt werden, wodurch der Effekt der Mutation aufgehoben werden kann. Dagegen können auch synonyme Basensubstitutionen aberrantes Spleißen zur Folge haben [Sauna und Kimchi-Sarfaty, 2011; MacArthur et al., 2012].

Zur Bekräftigung der Kausalität konnte bei insgesamt 30 der 47 neu klassifizierten FA-Patienten die Segregation beider Mutationen bestätigt werden, um das Vorhandensein der entsprechenden Mutationen auf unterschiedlichen Allelen nachzuweisen, da hier Material der Eltern zur Verfügung stand. Zudem wurden bei allen Patienten die als pathogen eingestuften Mutationen per Sanger-Sequenzierung validiert. Die Sanger-Sequenzierung gilt als Goldstandard zum Nachweis von Mutationen in Hinblick auf Sensitivität sowie Spezifität und wird daher in vielen Labors zur Validierung von Mutationen, welche per NGS detektiert wurden, verwendet. Allerdings wird zunehmend diskutiert, ob dies noch immer nötig ist, da die Spezifität der NGS-Analysen immer weiter ansteigt und verschiedene Studien zeigen konnten, dass NGS-Daten immer mehr mit der Spezifität von Sanger-Sequenzierungen vergleichbar sind [Sikkema-Raddatz et al., 2013; Strom et al., 2014; Baudhuin et al., 2015; Beck et al., 2016]. Dies gilt allerdings nur für NGS-Daten mit hoher Abdeckung und für Reads mit hohem Qualitätsscore. Setzt man dabei die Filter zu stringent, um falschpositive Varianten herauszufiltern, etwa durch Heraufsetzen der Abdeckung und der Allelfrequenz einer Mutation, besteht auch hier die Gefahr, dass man pathogene Varianten übersieht [Mu et al., 2016]. Für die vorliegenden Daten der eigenen Anreicherungen zeigten sich, v. a. beim SegCap EZ-Panel, auch einige falschpositive Varianten, welche meist in oder angrenzend an Homopolymere oder Pseudogenregionen gelistet waren. Neben diesen schwierig zu sequenzierenden, komplexen DNA-Bereichen kommen solche Sequenzierfehler generell häufig in AT- oder GC-reichen Regionen oder in Abschnitten mit niedriger Abdeckung vor [Beck et al., 2016; Mu et al., 2016]. Eine falsch-negative Mutation konnte dagegen nicht bestimmt, aber auch nicht ausgeschlossen werden. Manche der Sequenzierfehler würden sich durch überlegte bioinformatische Filterungsstrategien lösen lassen. So bleibt die Sanger-Sequenzierung trotzdem ein wichtiger Bestandteil bei NGS-Analysen und v. a. in der Diagnostik, um möglicherweise falsche Ergebnisse auszuschließen, die für den Patienten sowie dessen Familie weitreichende Folgen haben könnten. Als unabhängiger Test ist die Sanger-Sequenzierung nicht nur zur Validierung einer Mutation, sondern auch des Probenmaterials wichtig, um eventuelle Verwechslungen zu prüfen.

#### 5.3.2 Zuordnung der Patienten und Identifizierung neuer Mutationen

Von insgesamt 69 Patienten, die mit Hilfe verschiedener Anreicherungsmethoden durch NGS untersucht wurden, dienten fünf (FA-01, FA-02, FA-03, FA-04, FA-27) zur Bestätigung bekannter Mutationen während der Etablierung der *SeqCap EZ*-Anreicherung. Bei vier weiteren Patienten, FA-05, FA-11, FA-29 sowie FA-63 war hingegen die Diagnose FA durch funktionelle Untersuchungen ausgeschlossen worden. Somit blieben 60 neue FA-Patienten, bei welchen eine Komplementationsgruppen-Bestimmung erfolgen sollte. Davon konnten, wie im obigen Abschnitt bereits erwähnt, bei 47 Patienten biallelische Mutationen in bekannten FA-Genen bestimmt werden und somit eine eindeutige Komplementationsgruppen-Zuordnung erfolgen. Bei den übrigen 13 Patienten besteht weiterhin die Verdachtsdiagnose FA, doch konnten hier nicht oder nicht zweifelsfrei die krankheitsverursachenden Mutationen bestimmt werden. Darunter befindet sich auch FA-19, der Träger einer heterozygoten Deletion in *FANCA* ist und die Defizienz des Proteins beim Patienten durch eine Immunblot-Analyse bestätigt werden konnte. Allerdings konnte hier die zweite compound-heterozygote Mutation nicht detektiert werden.

Der Anteil der einzelnen identifizierten Komplementationsgruppen an der Gesamtzahl der untersuchten FA-Patienten ist in Abbildung 57 zusammengefasst. Daraus wird ersichtlich, dass die in dieser Arbeit klassifizierten Patienten 12 verschiedenen Komplementationsgruppen zugeordnet werden konnten, worin sich die genetische Heterogenität von FA zeigt. Gut zwei Drittel (68,1 %) konnten einer der drei häufigsten Komplementationsgruppen FA-A, -C und -G zugeordnet werden. Mit 46,8 % (n = 22/47 liegt der Anteil an FA-A-Patienten unter den zugordneten Patienten in der vorliegenden Arbeit zwar deutlich unter den in der Literatur angegebenen 60–70 % [Mehta und Tolar, 1993-2017]. Dies erklärt sich aber mit der Voruntersuchung der meisten Patienten per MLPA-Analyse zum Nachweis von Deletionen in *EANCA*, welche einen Großteil der Mutationen in diesem Gen ausmachen. Durch diese Vorauswahl verschob sich das Verhältnis der FA-A-Patienten dieser Studie. Allgemein lässt sich feststellen, dass der Großteil der Patienten (87,2 %) einer *Upstream*-Komplementationsgruppe angehört, was sich mit den Angaben der Rockefeller-FA-Mutationsdatenbank deckt.



**Abbildung 57:** Komplementationsgruppen-Zuordnung der per Genpanel (*SeqCap EZ* und *TruSight Cancer*) oder WES analysierten Patienten mit FA-Diagnose. Der Anteil der jeweiligen Komplementationsgruppe an der Anzahl aller untersuchten neuen FA-Patienten (n = 60) ist in Prozent angegeben, bei 13 Patienten steht eine Klassifizierung mit dem Nachweis biallelischer Mutationen noch aus ("offen").

Obwohl die Grundmechanismen der ICL-Reparatur während der Evolution stark konserviert sind, kann die Reparatur einer DNA-Quervernetzung nur durch ein Zusammenspiel mehrerer Reparaturwege geleistet werden. Das erste Modell der ICL-Reparatur wurde in E. coli als Abfolge von NER und HR beschrieben [Cole, 1973], wohingegen der Vorgang in eukaryotischen Zellen deutlich komplexer ist. Während die meisten niederen Eukaryoten ein vereinfachtes System des FA/BRCA-Signalweges zeigen, findet sich die höchste Stufe der Komplexität bei Vertebraten. Klar ist, dass alle Komponenten des Kernkomplexes, inklusive FAAP20 und FAAP100, einen Beitrag zur Assemblierung, Stabilität und/oder Funktionalität des Komplexes als E3-Ligase leisten [Kim et al., 2012; Huang et al., 2014; Swuec et al., 2017; van Twest et al., 2017], was sich in vivo im Verlust der FANCD2-Monoubiqutinierung bei Mutationen in Genen der einzelnen Upstream-Komplementationsgruppen zeigt. Bei Pulldown-Untersuchungen des rekombinanten Kernkomplexes zeigte sich, dass ausgerechnet der AG20-Subkomplex in vitro nicht zu einer Steigerung der FANCD2-Ubiquitinierung durch den übrigen Kernkomplex beiträgt, obwohl Mutationen in FANCA und FANCG zusammengenommen bei rund drei Viertel der FA-Patienten ursächlich sind [van Twest et al., 2017]. Dagegen finden sich z. B. nur bei 2,1 % der FA-Patienten Mutationen in FANCB und FANCL (Rockefeller-FA-Mutationsdatenbank; Stand Januar 2017), deren Proteinprodukte zusammen mit FAAP100 ein Homodimer bilden und eine zentrale Funktion im Kernkomplex innehaben. Die ungleiche Verteilung der einzelnen Komplementationsgruppen spiegelt somit vermutlich auch die funktionelle Bedeutung des jeweiligen Gens im FA-Signalweg wider. In Hinblick auf die Komplexität des Kernkomplexes ist weiterhin zu berücksichtigen, dass manche FA-Proteine auch Funktionen außerhalb des FA/BRCA-Signalweges innehaben, wie etwa FANCA, das an der Stabilisierung von Replikationsgabeln beteiligt ist [Schlacher et al., 2012].

Die genetische Heterogenität von FA in der eigenen Kohorte wird auch bei der Betrachtung des identifizierten Mutationsspektrums deutlich, da mehr als die Hälfte der Mutationen (n = 34/66) bislang noch bei keinem FA-Patienten beschrieben wurde. Darunter fanden sich je sieben Missense- und Indel-Mutationen, acht Nonsense-Mutationen sowie 12 Mutationen mit einem vorhergesagten Effekt auf das kanonische Spleißverhalten, der für zehn dieser Varianten bestimmt werden konnte. Durch regulatorische Elemente in Intron- und Exonsequenzen werden das sogenannte Spleißosom, ein Komplex aus weit über hundert Proteinen und fünf RNA-Molekülen, sowie RNA-bindende Hilfsproteine an die korrekten Stellen geleitet, um Intronsequenzen aus der prä-mRNA zu entfernen und die Exons zur reifen mRNA zu verbinden. Die Entstehung eines einzigen Transkriptes pro Gen ist dabei selten, stattdessen werden über 90 % der Gene alternativ gespleißt [Pan et al., 2008], was zur Entstehung alternativer Isoformen und damit zur Proteinvielfalt beiträgt. Die Regulation der RNA-Prozessierung ist häufig gewebe- oder entwicklungsspezifisch, wobei eine biologische Bedeutung dieser Transkripte nicht immer bekannt ist. Ein Beispiel für alternatives Spleißen bei FA findet sich im Skipping von Exon 22 in FANCD2, das geringfügig in der cDNA von Kontrollen beobachtet werden kann [Kalb et al., 2007]. Dagegen sind bislang zwei Mutationen in Intron 21 von FANCD2 bekannt, welche die kanonische Spleißakzeptor-Erkennungsstelle stärker beeinträchtigen und Exon-22-Skipping in hohem Maße hervorrufen. Beides, sowohl alternatives als auch aberrantes Spleißen in FANCD2 bei Exon 22 konnte in der vorliegenden Arbeit beobachtet werden. Schätzungsweise 50 % aller Mutationen führen zu einer Störung in der Regulation der RNA-Prozessierung [Lopez-Bigas et al., 2005], was für die untersuchte Patientengruppe durch den hohen Anteil an Spleißmutationen bestätigt wurde. Dadurch zeigt sich, dass die Sequenzierung der cDNA ein entscheidender Schritt ist, um die Pathogenität von Varianten zu bestimmen. Dies ist insbesondere für Mutationen von Bedeutung, welche nicht die kanonischen Spleißdonor- oder Spleißakzeptorstellen betreffen, weit im Intron liegen oder als stille oder Missense-Mutation vorhergesagt werden, aber trotzdem den Spleißvorgang beeinträchtigen.

Unter den bekannten Mutationen (n =32) fanden sich auch einige *Founder*-Mutationen wieder. Dazu zählt beispielsweise die Substitution c.1267C>T in *FANCA*, die bislang nur bei einem deutschen Angehörigen der Sinti nachgewiesen werden konnte [Callen et al., 2005] oder auch c.67del in *FANCC*, welche von einem holländischen Vorfahren abstammt [de Vries et al., 2012]. Da über die geographische Herkunft bzw. den ethnischen Hintergrund der untersuchten Kohorte kaum Informationen vorlagen und keine Haplotyp-Analyse durchgeführt wurde, kann für die entsprechenden Patienten nicht mit absoluter Sicherheit bestätigt werden, dass bei ihnen ein *Founder*-Effekt vorliegt.

## 5.3.3 Einige Patienten blieben trotz gezielter Anreicherung ohne Zuordnung

Wie auch bei anderen Krankheiten, die durch genetische Heterogenität gekennzeichnet sind, blieb die molekulare Analyse mancher FA-Patienten ohne eindeutiges Ergebnis. So betrug die Aufklärungsquote für das *SeqCap EZ*-Panel 78,7 %, während mit Hilfe des Krebspanels rund 70 % der Patienten zugeordnet werden konnten. Eine erfolglose Zuordnung kann mehrere Ursachen haben. Beispielsweise konnten nicht immer alle

Regionen von Interesse erfasst werden, weshalb diese Bereiche durch Sanger-Sequenzierung nachgeprüft werden sollten. Sofern dies nicht für alle Regionen durchgeführt wurde, ist ein vorläufiger Ausschluss des Patienten von der entsprechenden Komplementationsgruppe nicht möglich. Daneben besteht die Möglichkeit, dass das krankheitsverursachende Gen nicht im Panel enthalten war und deshalb nicht angereichert werden konnte. Zum einen kann es sich dabei um eines der neuesten FA-Gene handeln, welches zum Zeitpunkt des Sondendesigns für das SegCap EZ-Panel noch nicht bekannt und auch nicht als Kandidatengen enthalten war. Diese neuen FA-Gene müssten bei den nicht zugeordneten Patienten gegebenenfalls noch nachsequenziert werden. Durch die spezifische Auswahl der Kandidatengene für das SeqCap EZ-Panel sind dies nur maximal zwei Gene. REV7/FANCV müsste bei allen nicht zugeordneten Patienten nachgeprüft werden, wohingegen UBE2T/FANCT nur bei den Patienten überprüft werden müsste, welche mit einer der ersten beiden Sondenlibraries angereichert wurden. Beim TruSight Cancer-Panel sind dies für alle Patienten die neuesten FA-Gene RAD51/FANCR, UBE2T/FANCT, XRCC2/FANCU sowie REV7/FANCV. Zum anderen kann es sein, dass bei den Patienten Mutationen in neuen FA-Genen vorliegen, welche bislang noch nicht identifiziert wurden und auf den Seg Cap EZ-Panels oder dem TruSight Cancer-Panel nicht enthalten waren. Weiterhin ist denkbar, dass krankheitsverursachende Mutationen durch zu stringente Filterungen aussortiert und dadurch von einer weiteren Analyse ausgeschlossen wurden, wie etwa im Bereich von Homopolymeren. Insbesondere bei der GS Junior-Sequenzierung traten in diesen Regionen häufig Sequenzierfehler auf. Dabei ist nicht auszuschließen, dass eine tatsächliche Mutation in diesen Bereichen fälschlicherweise als Sequenzierfehler bewertet oder durch Filterung entfernt wurde. Ferner sind Kopienzahlvarianten nicht auszuschließen wie auch Mutationen in Introns, die einen Effekt auf das aberrante Spleißverhalten besitzen und nicht ins Sondendesign eingeschlossen waren. Spleißmutationen ließen sich zwar durch die Analyse der cDNA bestimmen. Ohne konkreten Anhaltspunkt, wie etwa das Vorliegen einer pathogenen Mutation in einem bestimmten FA-Gen, wäre die Einzelgenanalyse durch Sequenzierung von cDNA aber zu aufwändig. Für eine effiziente Überprüfung der Genexpression aller FA-Gene wäre ein RNA-Seq die effizienteste Methode, um Rückschlüsse auf die Genexpression ziehen zu können, da sich somit das gesamte Transkriptom durch Hochdurchsatzsequenzierung analysieren ließe [Marioni et al., 2008; Morin et al., 2008; Chandrasekharappa et al., 2013].

In Anbetracht der steigenden Anzahl an nicht-synonymen Veränderungen, die mit Hilfe von NGS bestimmt werden, gewinnen Vorhersageprogramme immer mehr an Bedeutung, insbesondere in der klinischen Diagnostik. Während Nonsense- oder Frameshift-Mutationen normalerweise als pathogen eingestuft werden, ist eine Bewertung von Missense-Varianten deutlich schwieriger. Zur Klassifizierung von Varianten wurden daher zahlreiche bioinformatische Programme entwickelt, deren Algorithmen unterschiedliche Informationen zur Vorhersage der Pathogenität nutzen, wie etwa die Proteinstruktur, sofern diese bekannt ist, oder auch die Konservierung von Aminosäuren. Eine Kombination aus verschiedenen Programmen zur Bewertung einer Variante erscheint daher sinnvoll, um die Möglichkeit zu verringern, dass Varianten als falsch positiv oder falsch negativ bewertet werden [Hicks et al., 2011]. Für einige der unklassifizierten Patienten konnten zwar mono- oder auch biallelische Missense-Veränderungen bestimmt werden, doch wurden diese von den drei verwendeten Mutationsvorhersageprogrammen MutationTaster [Schwarz et al., 2014], SIFT [Kumar et al., 2009] sowie PolyPhen-2 [Adzhubei et al., 2010] mit unterschiedlichen Auswirkungen hinsichtlich der Pathogenität bewertet und so als Varianten unbekannter Signifikanz eingestuft. Letztlich kann nur über funktionelle Analysen, wie etwa Komplementationsversuche, die Pathogenität einer Variante überprüft werden, was aber vom Aufwand her nur für ausgewählte Veränderungen in Frage kommt.

In rund 10-30 % aller FA-Patienten kann das Verschwinden einer Mutation im hämatopoetischen System oder in lymphoblastoiden Zelllinien beobachtet werden, was eine Mosaikbildung kennzeichnet und die molekulare Analyse erschwert [Lo Ten Foe et al., 1997; Soulier et al., 2005]. Mosaizismus ist die Koexistenz von Zellen mit unterschiedlicher genetischer Ausstattung innerhalb eines Individuums, hervorgerufen durch postzygotische Mutationen, die an eine Subpopulation adulter Zellen weitergegeben werden. Bei knapp 15 % (n = 9/61) der in dieser Arbeit untersuchten Patienten liegt vermutlich ebenfalls eine Reversion vor, allerdings konnten nur bei vier Patienten die kausalen biallelischen Mutationen bestimmt und so eine Aussage über einen Mosaikstatus getroffen werden. Eine Reversion kann beispielsweise durch eine Rückmutation, eine oder mehrere kompensatorische Mutationen oder Genkonversion entstehen [Lo Ten Foe et al., 1997; Waisfisz et al., 1999; Gross et al., 2002]. Lymphoblastoide Zelllinien können in vitro durch einen erhöhten Selektionsdruck und längerer Kultivierung eine MMC-Resistenz entwickeln. Dagegen kann die Reversion einer Linie auch als Hinweis auf ein partielles Mosaik im hämatopoetischen System des Patienten gedeutet werden. Diese Linien stammen von B-Lymphozyten ab, die in vivo nur eine kurze Lebensdauer aufweisen. Revertierte Zellen, die einen Proliferationsvorteil besitzen, können sich daher in vitro schneller durchsetzen. Für drei Patienten (FA-09, -17, -69) konnte die Reversion einer Mutation nur in EBV-transformierten B-lymphoblastoiden Zelllinien, nicht aber im Blut der Patienten bestimmt werden. Verlaufskontrollen zur Überprüfung des Mosaikstatus im hämatopoetischen System sind hierbei anzuraten. Durch eine Chromosomenbruchanalyse könnte ebenfalls der Mosaikstatus abgeschätzt werden, da sich bei FA-Mosaikpatienten zwei Zellpopulationen unterscheiden lassen, die aus der Mischung von revertierten und FA-typischen Zellen hervorgehen [Oostra et al., 2012]. Da die Ausbildung eines somatischen Mosaikes meist mit einer Verbesserung der Blutwerte einhergeht, ist die Bestätigung des Mosaikstatus auch für die Prognose des Patienten und den weiteren Behandlungsverlauf wichtig. Ein vollständiges Mosaik konnte in dieser Arbeit nur bei einem Patienten (FA-18) auf molekularer Ebene bestätigt werden, während bei weiteren vier Patienten mit Verdacht auf ein Mosaik die Komplementationsgruppen-Zuordnung noch aussteht. Ein Rückschluss auf den exakten Mechanismus der Reversion wurde hier nicht weiter untersucht. Eine kompensatorische Mutation konnte jedoch durch die NGS-Analyse ausgeschlossen werden.

## 5.4 Charakterisierung einzelner Komplementationsgruppen

#### 5.4.1 Erweiterung des Mutationsspektrums in selteneren Komplementationsgruppen

Die Charakterisierung von FA-Mutationen kann sowohl zu einem besseren Verständnis ihrer klinischen und biologischen Bedeutung beitragen als auch Einblicke in die Funktion einzelner FA-Gene in der DNA-Schadensantwort liefern. Ferner können Langzeitstudien über Genotyp-Phänotyp-Korrelationen dazu genutzt werden, den Krankheitsverlauf besser abzuschätzen und die Behandlung zu optimieren. Dies ist insbesondere für Patienten seltener Komplementationsgruppen von großer Bedeutung. Eine direkte Genotyp-Phänotyp-Korrelation für einzelne Komplementationsgruppen ist bislang eher selten möglich. Diese lässt sich hauptsächlich auf einige Downstream-Gruppen anwenden, wie beispielsweise FA-D1 und FA-N. Patienten dieser Komplementationsgruppen zeigen einen sehr schweren FA-Phänotyp mit einer Vielzahl an kongenitalen Fehlbildungen sowie einer hohen Inzidenz an embryonalen Tumoren, die mit einer erhöhten Mortalität im Kindesalter einhergehen [Howlett et al., 2002; Wagner et al., 2004; Alter et al., 2007; Reid et al., 2007; Kopic et al., 2011; Myers et al., 2012]. Anhand der Einträge in der FA-Datenbank der Rockefeller-Universität scheinen für beide Untergruppen fast ausschließlich Nonsense- oder Frameshift-Mutationen ursächlich zu sein. Gleiches gilt für den einzigen FA-N-Patient dieser Studie (FA-32), bei dem mit Hilfe der SeqCap-Panelanreicherung eine homozygote Deletion identifiziert werden konnte, welche zum Frameshift führt (p.P316Lfs\*3). Bei dem Patienten war neben verschiedenen Fehlbildungen ein Wilmstumor diagnostiziert worden und er verstarb bereits im zweiten Lebensjahr. Allerdings ändert sich auch bei dieser Komplementationsgruppe mit der Identifizierung neuer Patienten das Bild über den Schweregrad der Erkrankung in Abhängigkeit von den krankheitsverursachenden Mutationen. So konnte kürzlich erstmals bei einem Geschwisterpaar neben einer Frameshift-Mutation in PALB2 eine hypomorphe Mutation in Form eines in-frame Skippings von Exon 6 nachgewiesen werden, welche zu einer milden Form des FA-N-Phänotyps führte [Byrd et al., 2016].

Ähnliches kann auch für FA-B beobachtet werden. Lange wurde angenommen, dass nur Nullmutationen in *FANCB* zur Entstehung des FA-Phänotyps führen. Dies bestätigte sich auch mit der Identifizierung einer Stoppmutation in FA-62 innerhalb dieser Studie (p.Q198\*). Allerdings sind inzwischen auch Missense-Mutationen bekannt, deren pathogener Effekt entweder durch Komplementationsversuche bestätigt oder aufgrund neuester Erkenntnisse zur BL100-Komplexbildung in einen funktionellen Zusammenhang gebracht werden konnte [Chandrasekharappa et al., 2013; Swuec et al., 2017]. Um eine Pathogenität der bei FA-10 detektierten hemizygoten Missense-Mutation zu überprüfen, müssten demnach ebenfalls funktionelle Analysen, wie etwa *Yeast-Two-Hybrid*-Bindungsstudien mit den Subkomplexpartnern FAAP100 und FANCL [Medhurst et al., 2006; Ling et al., 2007], durchgeführt werden. Von den in der offiziellen FA-Datenbank gelisteten *EANCB*-Mutationen sind etwa ein Fünftel (5/24) *de novo*-Ereignisse. Da die Basensubstitution bei FA-10 sowohl im Blut als auch in den Fibroblasten des Patienten nachgewiesen werden konnte, muss die

Punktmutation in der mütterlichen Keimzelle oder während der frühen Embryogenese entstanden sein. Aktuelle Untersuchungen, welche mit Hilfe von NGS durchgeführt wurden, gehen von einer genomweiten Mutationsrate von etwa 1–1,6\*10-8 pro Generation aus, was rund 50–100 *de novo*-Einzelnukleotidvarianten pro Individuum entspricht [Awadalla et al., 2010; Conrad et al., 2011; Michaelson et al., 2012; Wang und Zhu, 2014]. Das Wissen um die Entstehung einer Mutation bzw. deren Segregation ist insbesondere bei der genetischen Beratung von entscheidender Bedeutung.

Anders als FA-N oder FA-B zählen FA-L und -P zu den seltenen Komplementationsgruppen, von welchen jeweils nur eine einstellige Zahl an FA-Patienten (n = 7; Stand Januar 2017) in der Rockefeller-Mutationsdatenbank geführt werden. Für die Untergruppen FA-L sowie -P gelang durch die gezielte Anreicherung die Zuordnung von jeweils einem neuen Patienten. Aufgrund weniger klinischer Daten können hierzu keine weiteren Aussagen hinsichtlich einer Genotyp-Phänotyp-Korrelation getroffen werden, jedoch wurde das Mutationsspektrum für beide Gruppen erweitert. Beispielsweise war bei FA-37 das kanonische Startcodon von FANCL, der katalytischen Untereinheit des FA-Kernkomplexes, durch eine Substitution verändert. Ein Verlust des initialen Startcodons lässt eine drastische Veränderung der Expression dieses Allels vermuten. Möglicherweise kann auch ein alternatives Startcodon downstream des kanonischen Translationsstartes von der ribosomalen 40S-Untereinheit erkannt und die Translation dort initiiert werden. Sofern dieses im selben Leserahmen wie die kanonische Translationsinitiationsstelle liegt, hätte dies ein Nterminal verkürztes Proteinprodukt zur Folge. Zudem wurde auf dem zweiten Allel die Duplikation c.1096\_1099dup identifiziert, welche mit einer bei Ali et al. veröffentlichten Duplikation identisch ist, die dort fälschlicherweise als c.1095\_1098dup angegeben ist [Ali et al., 2009]. Dabei handelt es sich um eine hypomorphe Mutation im letzten codierenden Exon von FANCL, welche zu einem Frameshift und ein um drei Aminosäuren verlängertes Protein führt (p.T367Nfs\*13). Während der von Ali et al. beschriebene Patient einen milden FA-Phänotyp zeigte, lag für die bei Meetei et al. sowie Chandrasekharappa et al. beschriebenen vier Patienten keine phänotypische Beschreibung vor [Meetei et al., 2003b; Ali et al., 2009; Chandrasekharappa et al., 2013]. Dagegen zeigten zwei bei Vetro et al. 2015 beschriebene Patienten mit biallelischen Funktionsverlust-Mutationen in FANCL einen schweren klinischen FA-Phänotyp und verstarben jeweils kurz nach der Geburt [Vetro et al., 2015]. FA-37 mit biallelischen FANCL-Mutationen scheint einen eher milden Phänotyp zu zeigen, da uns eine Probe zur FA-Diagnostik erst im jugendlichen Alter des Patienten erreichte, weil ein Verdacht auf MDS vorlag. Somit bestätigt sich dem Anschein nach für diese Gruppe, dass eher die Art bzw. Schwere der Mutation den Phänotyp maßgeblich beeinflussen als das betroffene Gen selbst [Neveling et al., 2009].

Diese Beobachtung scheint allerdings auf Patienten mit biallelischen Mutationen in *SLX4* nicht zuzutreffen, da der einzige Patient ohne Restprotein keinen schwereren Phänotyp zeigte als Patienten mit trunkiertem SLX4-Restprotein, bei welchem die Interaktion mit den bekannten Bindungspartnern XPF, MUS81 und/oder SLX1 bestätigt werden konnte [Kim et al., 2011; Stoepker et al., 2011; Schuster et al., 2013]. Mit der Identifizierung von FA-69 sind somit acht FA-P-Patienten aus sechs verschiedenen Familien bekannt. Im Vordergrund steht bei dieser Gruppe der hämatologische Phänotyp mit Knochenmarkversagen. Funktionelle Untersuchungen konnten hier nicht durchgeführt werden, da die lymphoblastoide Zelllinie, anders als die untersuchten Blutlymphozyten, revertiert war. Weil es sich bei beiden Mutationen um Nonsense-Mutationen handelt, ist ein Abbau der mRNA beider Allele durch NMD (*Nonsense-mediated mRNA decay*) wahrscheinlich. Die kritische Region zur Interaktion mit XPF-ERCC1 wäre im trunkierten Restprotein noch enthalten, welche essenziell für die Reparatur von DNA-Quervernetzungen ist [Kim et al., 2013; Hashimoto et al., 2015]. Ob die Struktur-spezifische Endonuklease allerdings noch an den Ort des Schadens durch das verkürzte SLX4-Gerüstprotein koordiniert werden kann, bleibt in diesem Fall unbestimmt.

Für alle zuvor beschriebenen FA-Gene besteht eine eindeutige Assoziation mit der Erkrankung FA, wohingegen FANCM, eines der am stärksten konservierten FA-Gene (Abbildung 7), noch immer kontrovers diskutiert wird [Mehta und Tolar, 1993-2017; Bogliolo und Surralles, 2015]. Nachdem sich die Komplementationsgruppe FA-H als identisch mit FA-A erwies, wurden stringentere Kriterien zur Festlegung neuer Untergruppen vorgeschlagen, wie etwa das Vorliegen von zwei Patientenzelllinien, die sich gegenseitig nicht komplementieren [Joenje et al., 2000]. Da progressives Knochenmarksversagen zu den grundlegenden klinischen Merkmalen der FA zählt und das Manifestationsalter im Median bei etwa acht Jahren liegt, wird den Komplementationsgruppen FA-O, -R und -S ein FA-ähnlicher Phänotyp zugeschrieben, weil dieses Merkmal bei keinem der Patienten zum Zeitpunkt der Beschreibung beobachtet werden konnte [Vaz et al., 2010; Ameziane et al., 2015; Sawyer et al., 2015; Wang et al., 2015]. Zur ersten und bislang einzigen beschriebenen FA-M-Patientin (EUFA867) gibt es hierzu widersprüchliche Angaben. Während in der Erstbeschreibung der Komplementationsgruppe FA-M noch ein FA-typischer Phänotyp mit progressiven Knochenmarksversagen für EUFA867 angegeben wurde [Meetei et al., 2005], beschrieben Singh et al., dass sich bei der Patientin auch mit 26 Jahren noch normale Blutwerte zeigten [Singh et al., 2009]. In dieser zweiten Arbeit konnten, neben den bei Meetei et al. identifizierten biallelischen FANCM-Mutationen, auch compound-heterozygote FANCA-Mutationen nachgewiesen werden [Meetei et al., 2005; Singh et al., 2009]. Die mit FANCA komplementierte Zelllinie von EUFA867 erwies sich allerdings weiterhin als MMC-, sowie Camptothecinund UV-sensitiv. Dagegen konnte die Monoubiquitinierung von FANCD2, wenn auch vermindert, vermittelt werden, was mit früheren Beobachtungen übereinstimmt [Mosedale et al., 2005; Singh et al., 2009]. Durch WES konnten bei FA-07 die per Sequenzierung bestimmten FANCM-Mutationen c.1241dup (p.N414Kfs\*12) und c.1972C>T (p.R589\*) bestätigt und Mutationen in weiteren beschriebenen FA-Genen ausgeschlossen werden. Eine Untersuchung auf Proteinebene zeigte eine verminderte FANCD2-Monoubiquitinierung. Daher scheint bei beiden Patientinnen mit biallelischen FANCM-Mutationen (EUFA687 und FA-07) die Lokalisierung der Kernkomplexproteine an die Schadensstelle vermindert, was sich in der reduzierten FANCD2-Monoubiquitinierung widerspiegelte. Weiterhin konnte auch bei FA-07 ein erhöhter spontaner wie auch induzierter G2-Phase-Arrest ermittelt werden. Eine Segregation konnte bei dieser Patientin nur für die Duplikation nachgewiesen werden, da nur maternale DNA zur Verfügung stand. Somit ist nicht eindeutig geklärt, ob beide Mutationen auf unterschiedlichen Allelen vorliegen.

Da FANCM maßgeblich an der Rekrutierung des Kernkomplexes ans Chromatin, an der Aktivierung des ATR-vermittelten S-Phase-Kontrollpunktes und auch an der Umkehr blockierter Replikationsgabeln beteiligt ist [Collis et al., 2008; Gari et al., 2008; Kim et al., 2008] sollten biallelische Nonsense- oder Funktionsverlustmutationen in *FANCM* einen erheblichen Einfluss auf den Phänotyp zeigen. Dagegen spricht allerdings eine 2014 erschienene Studie über Funktionsverlust-Mutationen in der finnischen Bevölkerung [Lim et al., 2014]. Hier konnten biallelische Mutationen in *FANCM* bei sieben Probanden nachgewiesen werden, die allesamt keinen FA-Phänotyp zeigten. Zudem konnte bei diesen Personen kein erhöhtes Auftreten von Krebserkrankungen oder anderen chronischen Erkrankungen beobachtet werden. Unter diesen Mutationen befand sich auch die mit erhöhtem Brustkrebsrisiko assoziierte Mutation konnte durch Komplementationsversuche in FANCM-defizienten immortalisierten embryonalen Mausfibroblasten bestätigt werden [Peterlongo et al., 2015]. Solange nicht geklärt ist, woher die starke Diskrepanz der Ergebnisse stammt, sollte FANCM nur als FA-assoziiertes Gen angesehen werden.

## 5.4.2 Identifizierung eines vierten UBE2T-defizienten Patienten

Ein häufiger Mechanismus zur Regulation von Proteinen besteht in der reversiblen posttranslationalen Modifikation durch Ubiquitin oder Ubiquitin-ähnliche Proteine, wie SUMO oder NEDD8, die durch eine E1-E2-E3-Enzymkaskade vermittelt wird. Für die Übertragung von Ubiquitin sind bislang zwei Ubiquitinaktivierende E1-Enzyme bekannt [Finley et al., 1984; Handley et al., 1991; Pelzer et al., 2007], wohingegen knapp 40 Ubiquitin-konjugierende E2-Enzyme beschrieben wurden [Stewart et al., 2016]. Dem gegenüber steht eine außergewöhnliche Diversität an über 600 E3-Ubiquitin-Ligasen, welche für die Substrat-Selektivität entscheidend sind [Li et al., 2008].

Im FA/BRCA-Signalweg stellt UBE2T das E2-Enzym dar, welches mit der E3-Ligase FANCL für die Monoubiquitinierung von FANCD2 verantwortlich ist [Machida et al., 2006]. Bislang wurden erst drei Patienten mit compound-heterozygoten Mutationen in *UBE2T* identifiziert, welche jeweils als Funktionsverlust-Varianten eingestuft wurden [Hira et al., 2015; Rickman et al., 2015; Virts et al., 2015]. Mit Hilfe der zuvor durchgeführten Anreicherung und anschließendem NGS wurden in dieser Arbeit bei einem weiteren FA-Patienten (FA-49) Mutationen in *UBE2T* bestimmt. Durch gezielte cDNA-Analysen konnte die von Rickman et al. und Virts et al. beschriebene Deletion c.-64\_468del auf dem maternalen Allel von FA-49 nachgewiesen werden [Rickman et al., 2015; Virts et al., 2015]. Diese wird durch die Rekombination von zwei AluYa5-Elementen in Intron 1 und 6 hervorgerufen und umfasst die Exons 2 bis 7. Da das erste Exon von *UBE2T* nicht codierend ist, ist auch das Startcodon von der Deletion betroffen, weshalb das Allel als Nullallel anzusehen ist [Rickman et al., 2015]. Da beide Alu-Elemente in Intron 1 und 6 in gleicher Orientierung vorliegen und identische Sequenzen besitzen, scheint als Mechanismus, welcher zur Deletion führt, die nichtallelische homologe Rekombination (NAHR) möglich, die auch für eine Vielzahl von FANCA-Deletionen verantwortlich ist [Flynn et al., 2014].

Bei der zweiten Mutation von FA-49 handelt es sich um eine Basensubstitution vier Basen 3' von Exon 5 (c.384+4A>G), die paternal vererbt wurde. cDNA-Analysen bestätigten aberrantes Spleißen als Folge der Mutation. Als Hauptspleißprodukt konnte die Exonisierung von Intron 5 bestimmt werden. Auf gleicher Höhe wie dieses Spleißprodukt konnte bei der Auftrennung des PCR-Produkts bei verschiedenen Kontrollen gleichfalls eine Bande bestimmt werden, wohingegen bei der Sequenzierung nur die Wildtypsequenz bestätigt werden konnte. Alternatives Spleißen tritt bei Kontrollen demnach nur sporadisch auf oder es handelte sich bei der zusätzlichen Bande um ein Artefakt der PCR. Auf Proteinebene wurde für die Exonisierung von Intron 5 ein Frameshift sowie die Insertion von 22 Aminosäuren vorhergesagt (p.S129Vfs\*23). Laut Vorhersage hätte das daraus resultierende Protein eine Größe von 16,9 kDa. Da durch eine Immunblot-Analyse kein Protein auf dieser Höhe machgewiesen wurde, ist das Proteinprodukt entweder instabil oder die mRNA wird durch NMD abgebaut. Durch den posttranskriptionalen Kontrollmechanismus des NMDs können mRNAs mit einem prämaturen Stoppcodon eliminiert und somit der Entstehung potentiell dominantnegativer Proteinprodukte vorgebeugt werden. Durch die Zugabe des NMD-Inhibitors CHX vor der Isolation der RNA könnte überprüft werden, ob die aus der Mutation resultierende mRNA durch NMD abgebaut wird [Carter et al., 1995]. Ferner sollten quantitative mRNA-Analysen zur Bestimmung des Ausmaßes einer möglichen Herunterregulation des Transkriptes herangezogen werden, wie etwa durch Northern Blots, quantitative Real-time PCR oder digitale PCR. Ähnliches gilt für die bei Hira et al. identifizierte Mutation c.179+5G>A in Intron 2, welche in der Originalpublikation fälschlicherweise als c.180+5G>A angegeben wurde [Hira et al., 2015]. Als weiteres aberrantes Spleißprodukt, welches durch die paternale Mutation hervorgerufen wird, konnte bei FA-49 das Skipping von Exon 5 beobachtet werden. Auf Proteinebene führt dies zur in-frame Deletion von 33 Aminosäuren (p.G96\_128del) in der katalytischen Ubiquitin-konjugierenden Domäne (UBC fold), welche insgesamt knapp 150 Aminosäuren umfasst (aa 5-152) und hochkonserviert bei allen Ubiquitin-E2-Enzymen vorliegt. Innerhalb dieser Domäne befinden sich verschiedene essenzielle Aminosäurereste (S5, R6, R9, R60, R99, S101 sowie N103), welche für elektrostatische sowie Wasserstoffbrückenbindungen zwischen UBE2T und FANCL verantwortlich sind. Jedoch scheint nur Arginin an Position 60 essenziell für die Interaktion mit FANCL zu sein, welches von der Deletion nicht betroffen ist [Hodson et al., 2014]. Ob eine Interaktion von UBE2T mit FANCL trotz der inframe Deletion innerhalb der UBC-Domäne möglich ist, könnte durch Bindungsstudien untersucht werden. Da bei der Immunblot-Analyse kein verkürztes Protein im Bereich der vorhergesagten Größe von 19,0 kDa bestimmt werden konnte, muss weiter untersucht werden, ob das trunkierte Protein überhaupt gebildet werden kann und stabil ist.

Da FA-49 durch die maternal vererbte Deletion hemizygot für die Spleißmutation c.384+4A>G ist, bei der Sequenzierung der cDNA aber auch die Wildtypsequenz bestätigt wurde, konnte gezeigt werden, dass die kanonische Spleißdonorstelle in Intron 5 weiterhin durch den Spleißapparat erkannt und genutzt werden kann. Dadurch ist die Möglichkeit gegeben, dass wildtypisches UBE2T-Protein gebildet wird, was auf Proteinebene durch den Nachweis einer schwachen Bande auf gleicher Höhe wie bei Kontrollen bestätigt werden konnte. Zudem konnte in den Patientenzellen von FA-49 die FANCD2-Monoubiquitinierung nachgewiesen werden, die sich auf das funktionale Restprotein normaler Länge zurückführen lässt. Beide FANCD2-Isoformen konnten auch bei der immortalisierten Fibroblasten-Zelllinie AP65P-hTERT des bei Hira et al. 2015 veröffentlichten japanischen Patienten PNG-252 bestätigt werden, wobei der Proteingehalt an UBE2T sogar mit Kontrollen vergleichbar war [Hira et al., 2015]. Da die Missense-Mutation c.4C>G, welche heterozygot bei PNG-252 bestimmt wurde, die Interaktion mit FANCL deutlich schwächt, zeigte sich bei AP65P-hTERT nur eine verminderte Monoubiquitinierung von FANCD2. Ähnliches findet sich auch bei Patienten mit hypomorphen Mutationen in FANCD2 oder FA-Kernkomplex-Proteinen, weshalb die Spleißmutation c.384+4A>G ebenfalls als hypomorph anzusehen ist [Ali et al., 2009; Hartmann et al., 2010]. Neben der Möglichkeit, dass mutiertes oder wildtypisches UBE2T die Monoubiquitinierung von FANCD2 vermittelt, könnten auch andere E2-Enzyme dafür verantwortlich sein, welche die Funktion von UBE2T übernehmen. Es ist bekannt, dass manche E3-Ligasen mit mehreren E2-Enzymen interagieren können [Dodd et al., 2004; Zhang et al., 2005; Christensen et al., 2007; Ambivero et al., 2014]. Für FANCL konnte bereits eine Interaktion mit einem weiteren Ubiquitin-konjugierenden Enzym nachgewiesen werden, welches in vitro auch zur Monoubiquitinierung von FANCD2 genutzt werden kann [Alpi et al., 2008; Zhang et al., 2011]. Dabei handelt es sich um UBE2W, welches allerdings Substrate am N-Terminus ubiquitiniert [Scaglione et al., 2013], was im Fall von FANCD2 zu einer unspezifischen Monoubiquitinierung und damit nicht zur Aktivierung des FA/BRCA-Signalweges führt [Rajendra et al., 2014]. Zudem konnte die Überexpression von UBE2W in der UBE2T-Patientenzelllinie 100166/1 die G2-Phase-Arretierung nicht aufheben [Virts et al., 2015]. Durch die hohe Konservierung der UBC-Domäne sowie einzelner hydrophober Reste an den Interaktionsstellen von FANCL und der UBC-Domäne scheint eine Bindung von weiteren E2-Enzymen an FANCL zur Monoubiquitinierung zumindest möglich [Hodson et al., 2014]. Da UBE2T der bevorzugte E2-Bindungspartner von FANCL ist, sollten Bindungsstudien für die Identifizierung von alternativen E2-Enzymen zur FANCD2-Monoubiquitinierung in einer zellulären Umgebung stattfinden, in der UBE2T fehlt. Dazu würden sich beispielsweise die FANCT-Patientenzelllinien eignen, in denen ein reduziertes UBE2T-Proteinlevel vorliegt. Ferner bleibt für FA-49 zu bestimmen, ob eine Induktion der FANCD2-Focibildung gegeben ist, da diese von der Monoubiquitinierung des Proteins durch den Kernkomplex abhängig ist [Garcia-Higuera et al., 1999]. Eine Quantifizierung der Focibildung wurde nur von Hira et al. vorgenommen und zeigte dabei eine deutliche Reduktion von Foci-positiven Zellen im Vergleich zur komplementierten Patientenzelllinie [Hira et al., 2015]. Allerdings fehlt der Vergleich zu einer Upstream-Patientenlinie sowie einer gesunden Kontrolle, was die Aussagekraft der Analyse einschränkt. Um eine eindeutige Zuordnung des Patienten FA-49 zur Komplementationsgruppe FA-T vorzunehmen, müsste die Zelllinie ebenfalls mit der Wildtypsequenz von UBE2T sowie den verschiedenen Spleißprodukten der Mutation c.384+4A>G komplementiert und die Aufhebung des G2-Phase-Arrestes nach Schadensinduktion untersucht werden.

#### 5.4.3 Fanconi-Anämie bei einer Patientin mit XP-assoziierter Missense-Mutation

Ursprünglich waren *ERCC4/XPF/FANCQ*-Mutationen ausschließlich mit Defekten der NER assoziiert und sowohl für Patienten der Xeroderma pigmentosum-Untergruppe F (XP-F), mit Cockayne-Syndrom als auch mit dem progeroiden XFE-Syndrom beschrieben [Sijbers et al., 1996; Niedernhofer et al., 2006]. 2013 gelang erstmals der Nachweis eines Zusammenhangs von biallelischen *ERCC4/XPF/FANCQ* -Mutationen mit FA, da zwei FA-Q-Patienten sowie ein Patient mit einem kombinierten Phänotyp mit Merkmalen von XP, CS als auch FA beschrieben wurden [Bogliolo et al., 2013; Kashiyama et al., 2013].

Der FA-Q-Patientin, die mit Hilfe von WES innerhalb dieser Arbeit zugeordnet werden konnte, kommt eine besondere Bedeutung zu. Zum einen handelt es sich um eine der ältesten in der Literatur beschrieben FA-Patienten, welche zudem bis zum Beginn der sechsten Lebensdekade weder Knochenmarksversagen noch Tumoren entwickelte. Zum anderen konnte zum ersten Mal eine erhöhte Empfindlichkeit gegenüber Sonnenlicht bei einem FA-Patienten beschrieben werden, welche auf zellulärer Ebene bestätigt wurde. Damit ergibt sich ein einzigartiges neues Merkmal, dass nur bei der Untergruppe FA-Q beobachtet werden kann und zur Klassifizierung neuer FA-Q-Patienten herangezogen werden sollte.

Besonders interessant ist die Tatsache, dass dieselbe Mutation in Kombination mit einer zweiten Mutation ganz unterschiedliche phänotypische Ausprägungen zur Folge hat. So wurde die Missense-Mutation c.1765C>T (p.R589W) in der SF2-Helikase-ähnlichen Domäne sowohl bei der FA-Patientin FA-16, drei XP-F-Patienten (XP24BR, XP32BR, AS871) sowie einem Patienten mit kombinierten XPCS-Phänotyp und Merkmalen von FA (XPCS1CD) bestätigt [Ahmad et al., 2010; Kashiyama et al., 2013; Fassihi et al., 2016]. Durch eine Untersuchung der subzellulären Verteilung von XPF in verschiedenen Patientenzelllinien mit XPF-Missense-Mutationen konnte das Protein vermehrt im Zytoplasma nachgewiesen werden. Dies ist vermutlich auf eine fehlerhafte Faltung des Proteins zurückzuführen, wodurch die Translokation in den Zellkern beeinträchtigt wird [Ahmad et al., 2010]. In dieser Studie fanden sich sowohl XP24BR, XP32BR, AS871 und der einzige bislang bekannte XFE-Patient XP51RO [Niedernhofer et al., 2006], wobei hier die Interaktion von XPF mit seinem heterodimeren Bindungspartner ERCC1 nicht beeinträchtigt war. Dagegen konnte Restprotein bei FA-16 wie auch bei 1333 [Bogliolo et al., 2013] sowohl im Nukleus als auch am Chromatin nachgewiesen werden. Vermutlich kann bei FA-Q-Patienten ein gewisser Anteil von mutiertem Protein korrekt gefaltet werden, wodurch die Translokation in den Zellkern sowie ans Chromatin möglich wird. Denkbar wäre hierzu eine Beteiligung von endogenen Chaperonen, welche dazu beitragen, dass Proteine eine korrekte dreidimensionale Konformation annehmen [Ellis, 1987]. Bei verschiedenen Krankheiten wie etwa Parkinson, Chorea Huntington oder Zystischer Fibrose, die mit einer Fehlfaltung bestimmter Proteine assoziiert sind, wird der Einsatz von molekularen Chaperonen zu Therapiezwecken erforscht [Cummings et al., 2001; Chaudhuri und Paul, 2006; Banning et al., 2016] und wurde bereits zur Aufhebung der fehlerhaften Konformation von XPF bei XP-Patienten diskutiert [Ahmad et al., 2010]. Dabei sollte man allerdings bedenken, dass allen drei FA-Q-Patienten das Vorhandensein einer Missense-Mutation gemeinsam ist, deren Proteinprodukt zwar jeweils in den Zellkern translozieren und durch genügend Restfunktion die Nukleotidexzisionsreparatur, nicht aber die Reparatur von DNA-Quervernetzungen vermitteln kann [Bogliolo et al., 2013]. Bogliolo et al. konnten für beide FA-Q-Patienten zeigen, dass eine Interaktion von ERCC4/XPF/FANCQ trotz der Missense-Mutationen (L230P für FA104; R689S für 1333) sowohl mit ERCC1 als auch mit SLX4/FANCP möglich war [Bogliolo et al., 2013]. Diese Interaktionen können somit nicht maßgeblich für die Ausprägung des FA-Phänotyps bei FA-Q-Patienten verantwortlich sein. Dennoch sollten diese auch für FA-16 noch bestimmt werden, um eine allgemeine Aussage für FA-Q-Patienten in dieser Hinsicht treffen zu können.

Neben einer Missense-Mutation waren beide bei Bogliolo et al. 2013 beschriebenen FA-Q-Patienten Träger von Frameshift-Mutationen, die als Nullallele anzusehen sind [Bogliolo et al., 2013]. Dies gilt auch für die bei FA-16 auf dem maternalen Allel verifizierte Substitution c.793-2A>G, die durch das Skipping von Exon 5 ebenfalls einen Frameshift sowie ein vorzeitiges Stoppcodon hervorruft. Da durch die Immundetektion von XPF kein trunkiertes Protein nachgewiesen werden konnte, wurde die mutierte mRNA entweder durch NMD abgebaut oder das daraus resultierende Protein war instabil. Die Annahme, dass es sich bei dieser Mutation ebenfalls um ein Nullallel handelt, wird zudem durch die Tatsache unterstützt, dass weder ein alternatives Transkript ohne Exon 5 in den Datenbanken gelistet ist, noch ein kryptischer Spleißakzeptor in der Nähe der Mutation identifiziert werden konnte, welcher zur Wiederherstellung des Leserahmens führen würde.

Zusammenfassend lässt sich sagen, dass kein Hotspot für Mutationen in *ERCC4/XPF/FANCQ* identifiziert werden konnte, der einen bestimmten Phänotyp hervorruft (Abbildung 58).



Abbildung 58: Übersicht über ERCC4-Mutationen, die zur Ausprägung verschiedener Phänotypen führen. Die Mutationen sowie der Phänotyp der in dieser Arbeit identifizierten FA-Q-Patientin sind rot markiert. Die Missense-Mutation p.R589W ist zudem grau hinterlegt.

Die unterschiedliche zelluläre Lokalisation von mutiertem XPF-Protein scheint dagegen ein entscheidender Faktor zur Entstehung der unterschiedlichen Krankheitsbilder zu sein, wobei die Mechanismen, die bei FA-Q-Patienten eine Translokation in den Zellkern ermöglichen, weiterhin unbekannt sind.

#### 5.4.4 Mutationen im zentralen Protein des FA-Signalweges: Sequenzanalyse bei FA-D2-Patienten

Bislang sind in der FA-Mutationsdatenbank der Rockefeller-Universität über 1 500 FA-Patienten gelistet (Stand Januar 2017), wovon nur knapp 3 % der Komplementationsgruppe FA-D2 angehören. Die zentrale Rolle der FANCD2-Monoubiquitinerung, welche das Schlüsselereignis im FA-Signalweg darstellt, lässt der Untergruppe FA-D2 eine besondere Bedeutung zukommen. Die letzte umfassendere Analyse von FA-D2-Patienten liegt bereits 10 Jahre zurück und die darin beschriebenen Mutationen machen einen Großteil der in der FA-Mutationsdatenbank gelisteten FANCD2-Varianten aus [Kalb et al., 2007]. Durch NGS sowie klassische Methoden wie Sanger-Sequenzierung oder Immunblot-Analysen konnten hier insgesamt 28 neue FA-D2-Patienen aus 25 unterschiedlichen Familien klassifiziert werden. Die Mutationsanalyse in diesem Gen wird durch das Vorhandensein von Pseudogenen beeinträchtigt, welche eine hohe Sequenzhomologie zum funktionellen Gen aufweisen [Kalb et al., 2007]. Eine genaue Kenntnis über Pseudogensequenzen ist dabei eine Grundvoraussetzung für eine effektive Mutationsanalyse, um beispielsweise durch die Auswahl spezifischer Primer zwischen dem funktionellen Gen und den Pseudogenen zu diskriminieren. Bei der NGS-Analyse stellen Pseudogene ebenfalls eine Herausforderung dar, insbesondere bei der Verwendung von kommerziellen Anreicherungsplattformen. Für das FA-Genpanel wurde versucht, Pseudogenabschnitte von der Anreicherung auszuschließen. Allerdings hätten hierzu die Filter zur Sondenauswahl so stringent gesetzt werden müssen, dass einige Exons des funktionellen Gens überhaupt nicht mehr angereichert worden wären. Ein ähnliches Problem stellte sich beim Alignment der NGS-Daten. Ein konsequenter Ausschluss von Reads mit mehreren Fehlpaarungen geht mit dem Risiko einher, krankheitsverursachende Mutationen im funktionellen Gen ebenfalls zu verwerfen, was auch auf zu stringent gewählte bioinformatische Filterungsstrategien nach dem Alignment zutrifft [Ameziane et al., 2012]. Gleichzeitig sollte bei der Detektion einer potentiell pathogenen Variante in den NGS-Daten die Mutation mit spezifischen gDNA-Primern und/oder auf cDNA-Ebene überprüft werden, um falsch-positive Varianten, die eigentlich einem Pseudogen zugeschrieben werden müssen, zu erkennen [Knies et al., 2012].

Eine genauere Untersuchung von FANCD2-Pseudogenen ergab die Existenz von drei statt zwei voneinander abgegrenzten Regionen, die mit kleinen Unterbrechungen den Exons 15–28 von FANCD2 zugeordnet werden konnten. Dies legt die Vermutung nahe, dass ursprünglich durch Duplikation ein Pseudogen entstanden ist, welches im Laufe der Zeit so viele Mutationen in verschiedenen Bereichen akkumuliert hat, dass diese Regionen in der Pseudogendatenbank als drei voneinander abgegrenzte, eigenständige Pseudogene eingestuft wurden.

Die Diskrepanz in der Anzahl sowie der Lokalisation der FANCD2-Pseudogene lässt sich zum einen durch die Verwendung unterschiedlicher Referenzgenome als auch durch Unterschiede im Alignment zurückführen.

In dieser Arbeit wurde die aktuellste Version des Referenzgenoms (hg38) verwendet, welche auch natürlich vorkommende Varianten im Genom berücksichtigt. Zur Identifizierung der Pseudogene wurde zudem das bioinformatische Analyseverfahren von PseudoPipe herangezogen, mit dem Pseudogene durch ein genomweites Alignment identifiziert sowie klassifiziert werden können [Zhang et al., 2006]. PseudoPipe verwendet dazu BLAST statt BLAT, auf welches in der früheren Studie zurückgegriffen wurde und sich in verschiedenen Aspekten von BLAST unterscheidet. BLAT ist darauf ausgelegt, Sequenzen mit sehr hoher Sequenzidentität zu identifizieren, während BLAST flexibler ist und dadurch eine höhere Sensitivität aufweist [Kent, 2002]. BLAST lässt zudem größere Lücken in Alignments zu, während BLAT homologe Bereiche zwischen zwei Sequenzen zu einem Alignment verbindet.

In der vorliegenden Arbeit wurden die Mutationen der 28 neuen Patienten mit biallelischen Mutationen in FANCD2 charakterisiert. Ähnlich wie bei den anderen Komplementationsgruppen fand sich bei FA-D2 mit 42% ein hoher Anteil an Spleißmutationen sowie ein annähernd gleiches Verhältnis von neuen (n = 11) zu bereits bekannten (n = 13) Mutationen. Die Anzahl an verschiedenen FANCD2-Mutationen konnte dabei auf 51 erweitert werden, welche, zusammen mit den in der FA-Datenbank geführten Patienten, für insgesamt 68 FA-D2-Patienten beschrieben sind. Natürlich ist dies nicht die absolute Zahl an identifizierten FA-D2-Patienten, da nicht alle veröffentlicht oder in der FA-Mutationsdatenbank gelistet sind. Dazu zählt auch FA-01, dessen Mutationen bei Knies et al. 2012 beschrieben wurden [Knies et al., 2012], aber bislang nicht in die Rockefeller-FA-Datenbank aufgenommen wurden. Dennoch bekommt man einen relativ umfassenden Überblick über das Mutationsspektrum charakteristische Merkmale und einige dieser Komplementationsgruppe. Berücksichtigt man alle gelisteten sowie in dieser Arbeit beschriebenen Patienten, hat die Komplementationsgruppe FA-D2 einen Anteil von 4 %, was mit der Schätzung in der Literatur übereinstimmt, welche 3-6% annimmt [Kalb et al., 2007]. Ein geltendes Charakteristikum dieser Komplementationsgruppe, dass keine biallelische Nullmutationen auftreten, konnte ebenso bestätigt werden. Vermutlich sind diese Mutationen embryonal letal, so dass nur hypomorphe Varianten mit dem Überleben vereinbar sind. Diese Annahme wird durch Untersuchungen am Zebrafisch gestützt, die Fancd2 eine wichtige Rolle während der Embryogenese zuschreiben, da ein Knockdown des Genes zur p53-vermittelten Apoptose in verschiedenen Geweben während der Entwicklung führt [Liu et al., 2003]. Diese Beobachtung ist allerdings nicht ohne weiteres auf Säuger übertragbar, da Fancd2-/--Mäuse durchaus lebensfähig sind [Houghtaling et al., 2003]. Je nach genetischem Hintergrund unterschieden sich diese Mäuse allerdings deutlich in ihrem Phänotyp, beispielsweise in Hinblick auf die Geburtenrate oder Entwicklungsverzögerungen. Dies weist auf die Bedeutung von Modifier-Genen hin, welche den Phänotyp maßgeblich beeinflussen und vermutlich auch beim Menschen für die unterschiedlich schweren phänotypischen Ausprägungen trotz identischer Mutationen in FA-Genen verantwortlich sind.

Die Kristallstruktur des murinen FANCI-FANCD2-Heterodimers ergab, dass alle bei Kalb et al. 2007 beschriebenen Missense-Mutationen im Inneren des Komplexes an der Grenzfläche zwischen beiden Proteinen lokalisiert sind. Den betroffenen Aminosäureresten wurde eine stabilisierende Funktion zugeschrieben, so dass der pathogene Effekt dieser Mutationen auf eine Destabilisierung zurückzuführen ist [Joo et al., 2011]. Ähnliches gilt dabei auch für *FANCI*-Missense-Varianten. Zwei der in dieser Arbeit neu identifizierten *FANCD2*-Aminosäuresubstitutionen befinden sich in der HD2-Domäne des Proteins, wie auch die bei FA-01 beschriebene Missense-Mutation. Diese Domäne ist ebenfalls an der Bildung dieser Grenzfläche beteiligt ist. Die dritte neu beschriebene Variante ist hingegen im C-Terminus des Proteins lokalisiert, welcher kürzlich als essenzieller Faktor zur DNA-Bindung des Proteins sowie dessen Monoubiquitinierung identifiziert werden konnte [Liang et al., 2016]. In dieser Studie wurden auch verschiedene Punktmutationen in der Tower-Domäne charakterisiert und es zeigte sich, dass die mutierten Proteine nicht mehr an ICLs rekrutiert werden konnten und die Monoubiquitinierung ausblieb. Beides würde man auch für die neu identifizierte Missense-Mutation p.Q1332H erwarten.

Einen Einblick in die Monoubiquitinierung von FANCD2 durch den Kernkomplex könnte eine nähere Untersuchung der N-terminalen Deletion c.283\_285del von Patient FAD2-05 liefern. Diese ist in einem Bereich lokalisiert, der für die Interaktion mit FANCE essenziell ist (aa 1–291), worüber die Übertragung des Ubiquitinrestes von FANCL auf FANCD2 vermittelt wird [Pace et al., 2002; Gordon et al., 2005; van Twest et al., 2017].

Für 23 der analysierten FA-D2-Patienten lagen Informationen über den Phänotyp vor. Dabei zeigte jeder dieser Patienten mindestens eine Fehlbildung, was auch bereits in der vorherigen Studie zu *FANCD2* gezeigt werden konnte [Kalb et al., 2007]. Damit scheint die Komplementationsgruppe eine Ausnahme von anderen Untergruppen darzustellen, da normalerweise rund ein Drittel aller FA-Patienten keinerlei phänotypische Auffälligkeiten aufweist [Giampietro et al., 1997]. Zwei der FA-D2-Patienten erkrankten an Plattenepithelkarzinomen, was ebenfalls typisch für FA ist. Bei FAD2-02 wurde dies im Alter von 33 Jahren diagnostiziert, was dem Durchschnittsalter bei FA-Patienten entspricht [Kutler et al., 2016], wohingegen bei FAD2-20 Tumoren in der Mundhöhle mit 24 Jahren festgestellt wurden.

Zusammenfassend lässt sich feststellen, dass eine verlässliche Mutationsanalyse in *FANCD2*, wie auch bei anderen FA-Komplementationsgruppen, erst durch eine Kombination von klassischen sowie neuen Techniken erreicht werden konnte. Pseudogensequenzen können insbesondere bei der NGS-Analyse ein Problem darstellen, weshalb eine Überprüfung detektierter Mutationen durch Sanger-Sequenzierung der cDNA erfolgen sollte. Da Spleißmutationen zudem einen hohen Anteil an pathogenen *FANCD2*-Mutationen haben, kann dadurch parallel ein Effekt der Mutation nachgeprüft werden.

### 5.5 Nachweis von Mutationen in weiteren Genen der DNA-Reparatur

Mit Hilfe von WES konnten bei zwei Patienten, für welche initial die Verdachtsdiagnose FA gestellt worden war, Mutationen in DNA-Reparaturgenen identifiziert werden, welche nicht mit FA assoziiert sind. In beiden Fällen steht dabei eine nähere Charakterisierung der Patientenzelllinien bzw. Mutationen allerdings noch aus. Obwohl verschiedene Zelltypen eine Sensitivität gegenüber MMC zeigten, kann hierbei die Verdachtsdiagnose FA nicht aufrechterhalten werden. In diesen Fällen zeigt sich das Potential der WES-Analyse, da potentiell pathogene Mutationen in einem größeren Umfeld (z. B. DNA-Reparatur) als einer einzelnen Krankheits-Entität untersucht werden können.

So wurde bei FA-29, welcher zu Beginn der Analyse als FA-Mosaikpatient eingestuft worden war, eine neue homozygote Nonsense-Mutation (c.545G>A) in der BER-Glykosylase NTHL1 beobachtet, die erstmals 1998 als humanes Homolog der Endonuklease III von E. coli beschrieben wurde [Imai et al., 1998]. Im BER-Signalweg katalysiert NTHL1 vor allem die Exzision von durch Oxidation modifizierten Pyrimidinen, wie etwa Thyminglycol oder Cytosinglycol [Ocampo et al., 2002]. Biallelische Keimbahnmutationen in diesem Gen haben eine Prädisposition für die Entstehung von adenomatöser Polyposis sowie Darmkrebs zur Folge. Bislang sind nur zwei unterschiedliche Mutationen in NTHL1 in diesem Zusammenhang beschrieben worden. Dazu zählt die seltene Nonsense-Mutation c.268C>T (rs150766139) (p.Q90\*), welche mit einer Allelfrequenz von 1,543\*10-3 im ExAC-Browser gelistet ist und in drei nicht miteinander verwandten holländischen Familien bei sieben Patienten homozygot beschrieben wurde [Weren et al., 2015]. Dieselbe Nonsense-Mutation konnte zudem heterozygot bei einer kanadischen Patientin mit deutschen Vorfahren bestimmt werden, bei welcher zusätzlich die heterozygote Basensubstitution c.790+1G>A (rs372946560) am kanonischen Spleißdonor von Intron 4 nachgewiesen werden konnte, welche zu aberrantem Spleißen führt [Rivera et al., 2015]. Die Allelfrequenz dieser Variante ist im ExAC-Browser mit 8,44\*10-6 hinterlegt. Durch WES von gDNA des Patienten FA-29 konnte das Mutationsspektrum in NTHL1 anhand des Nachweises der homozygoten Nonsense-Mutation c.545G>A (p.W182\*) in Exon 3 erweitert werden. Diese seltene Variante findet sich bislang nicht in den ExAC-Browserdaten wieder. Wie für c.268C>T zeigte sich auch hier eine hohe Penetranz der Mutation, da sie in der Familie von FA-29 mit dem Phänotyp kosegregiert. Auf Proteinebene konnte beim Indexpatienten (FA-29) sowie dessen ebenfalls betroffener Mutter als Effekt dieser Nonsense-Mutation eine Defizienz des NTHL1-Proteins festgestellt werden. Ein verkürztes Proteinprodukt der in silico vorhergesagten Größe von 20 kDa konnte bei der Immunblot-Analyse nicht nachgewiesen werden. Vermutlich wird die mRNA durch NMD abgebaut oder das resultierende Proteinprodukt ist instabil. Eine Kombination beider Effekte ist dabei ebenso denkbar.

Allen zuvor identifizierten Patienten mit biallelischen NTHL1-Mutationen war das Vorliegen von kolorektalen adenomatösen Polypen gemeinsam. In den uns vorliegenden Daten der Familie von FA-29 war dies zwar nicht beschrieben, kann aber auch nicht ausgeschlossen werden. FA-29 sowie dessen Bruder (FA-

29.2) waren möglicherweise vor dem Auftreten von adenomatösen Polypen verstorben und es ist nicht bekannt, ob deren Schwester (FA-29.3) sowie die Mutter dahingehend untersucht wurden. Da Polypen eine Vorstufe zur Entstehung von Darmkrebs darstellen können, sollten beide Patienten darauf untersucht werden. Übereinstimmend mit allen bereits bekannten Fällen zeigte sich bei der hier identifizierten Familie ein breites Tumorspektrum, weshalb hier von einem Krebssyndrom gesprochen werden kann [Kuiper und Hoogerbrugge, 2015; Rivera et al., 2015; Weren et al., 2015]. Aktuelle Untersuchungen an Tumorgewebe von FA-29.2 (Roland P. Kuiper, persönliche Kommunikation) scheinen zu bestätigen, dass im somatischen Mutationsspektrum von NTHL1-defizienten Tumorzellen eine hohe Frequenz an C>T-Transitionen beobachtet werden kann, wie sie auch für die bereits veröffentlichten Familien beschrieben wurde [Rivera et al., 2015; Weren et al., 2015]. Dies deutet darauf hin, dass durch Oxidation modifizierte Basen in malignen Zellen nicht effizient repariert werden können und somit vermehrt oxidativem Stress ausgesetzt sind, was schließlich zur Entartung der Zellen führt. Im Gegensatz zu sporadischen Kolonkarzinomzellen konnten diese Veränderungen vornehmlich an nicht-CpG-Stellen nachgewiesen werden (Roland P. Kuiper, persönliche Kommunikation). Diese Mutationssignatur von Tumorzellen kann in Zukunft möglicherweise dazu genutzt werden, um Rückschlüsse auf NTHL1-Mutationen zu ziehen. Für eine solche Mutationssignatur könnten sich auch Therapiemöglichkeiten mit spezifischen Chemotherapeutika ergeben.

Ob die Mutationen in *RPA1*, welche bei FA-63 durch WES identifiziert wurden, tatsächlich für den beobachteten Phänotyp verantwortlich sind, bleibt noch zu klären. Bislang sind Mutationen in *RPA1* noch nicht mit einem bestimmten Krankheitsbild assoziiert, allerdings hat RPA1 aufgrund seiner Beteiligung an essenziellen zellulären Vorgängen eine wichtige Funktion in der Aufrechterhaltung der genomischen Stabilität und der Tumorsuppression inne. Als Untereinheit des heterotrimeren Replikationsproteins A (RPA) ist es beispielsweise an der HR beteiligt. Während der frühen Schadensantwort der HR werden die DNA-Enden am DSB prozessiert. Dadurch entstehen ssDNA-Überhänge, die von RPA gebunden werden [Hustedt und Durocher, 2016]. Im weiteren Verlauf des Reparaturweges vermittelt FANCD1/BRCA2 den Austausch von RPA mit FANCR/RAD51, wodurch ein funktioneller Link zum FA/BRCA-Signalweg besteht.

Beide Mutationen von FA-63 sind in der DBD-B-Domäne des Proteins lokalisiert. Diese bildet mit DBD-A das Kernstück der DNA-Bindungsdomäne des RPA-Heterotrimers, die eine Reihe an polaren Aminosäureresten enthält, welche mit einzelsträngiger DNA in direkten Kontakt treten. Beide Domänen besitzen eine hohe Affinität für einzelsträngige DNA und sind für deren Bindung durch das Heterotrimer ausreichend und essenziell [Walther et al., 1999]. Zusätzlich finden sich darin vier hochkonservierte aromatische Aminosäurereste, F238 und F269 in DBD-A sowie W361 und F386 in DBD-B. Die Mutation einzelner aromatischer Aminosäurereste hat nur einen geringfügigen Einfluss auf die Affinität des Proteinkomplexes, während die Substitution beider konservierter Aromaten in einer der DBDs die Bindung einzelsträngiger DNA beeinträchtigt [Bastin-Shanower und Brill, 2001] und sich deutliche Defekte in Hinblick auf die Funktion des Proteinkomplexes während der DNA-Reparatur zeigen [Haring et al., 2008].

Bei FA-63 konnte die compound-heterozygote Basensubstitution c.1082G>C nachgewiesen werden, welche zum Austausch der aromatischen hydrophoben Aminosäure Tryptophan an Position 361 gegen ein ungeladenes polares Serin führt. Dagegen führt die maternale Substitution c.1165C>T zum Austausch eines Arginins gegen Tryptophan an Position 389 in DBD-B. Eine 3D-Strukturvorhersage dieser Mutation implizierte dabei eine potentielle Destabilisierung dieser zentralen Domäne (R. Kalb, persönliche Kommunikation). Gegen einen pathogenen Effekt dieser Mutation spricht allerdings, dass diese im ExAC-Browser bei drei Individuen finnischer Abstammung homozygot gelistet ist, ohne dass dabei ein pathogener Phänotyp bekannt wäre. Beide Mutationen müssten daher funktionell näher charakterisiert werden, um ihren Einfluss auf die DNA-Bindungsaffinität sowie die Struktur von RPA1 zu bestimmen. Dabei kann man beispielsweise das CRISPR/Cas9-System nutzen, um die Mutationen in humane Zelllinien oder Modellorganismen einzubringen. Interessant wäre auch, ob RPA bei FA-63 noch an der Schadensstelle in DNA-Reparaturfoci lokalisiert ist, was sich anhand von Immunfluoreszenz-Versuchen nachweisen ließe.

Das Vorliegen eines G2/M-Phase-Arrestes in der Patientenzelllinie stimmt dagegen mit Beobachtungen verschiedener RPA1-Doppelmutanten überein, bei welchen jeweils beide aromatische Aminosäurereste einer der DNA-Bindedomänen gegen polare Aminosäuren ausgetauscht worden waren [Haring et al., 2008; Hass et al., 2012]. Durch eine Komplementation der Zellen mit wildtypischer *RPA1*-cDNA könnte untersucht werden, ob die identifizierten Mutationen von FA-63 tatsächlich für die Zellzyklus-Aktivierung verantwortlich sind, sofern ein sicheres Reportersystem geschaffen werden kann.

## Literaturverzeichnis

- Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X, Shendure J. 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol* 11(12):R119.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7(4):248-9.
- Ahmad A, Enzlin JH, Bhagwat NR, Wijgers N, Raams A, Appledoorn E, Theil AF, JH JH, Vermeulen W, NG JJ, Scharer OD, Niedernhofer LJ. 2010. Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. *PLoS Genet* 6(3):e1000871.
- Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP. 2002. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat Genet* 30(2):227-32.
- Ali AM, Kirby M, Jansen M, Lach FP, Schulte J, Singh TR, Batish SD, Auerbach AD, Williams DA, Meetei AR. 2009. Identification and characterization of mutations in FANCL gene: a second case of Fanconi anemia belonging to FA-L complementation group. *Hum Mutat* 30(7):E761-70.
- Ali AM, Pradhan A, Singh TR, Du C, Li J, Wahengbam K, Grassman E, Auerbach AD, Pang Q, Meetei AR. 2012. FAAP20: a novel ubiquitin-binding FA nuclear core-complex protein required for functional integrity of the FA-BRCA DNA repair pathway. *Blood* 119(14):3285-94.
- Alpi A, Langevin F, Mosedale G, Machida YJ, Dutta A, Patel KJ. 2007. UBE2T, the Fanconi anemia core complex, and FANCD2 are recruited independently to chromatin: a basis for the regulation of FANCD2 monoubiquitination. *Mol Cell Biol* 27(24):8421-30.
- Alpi AF, Pace PE, Babu MM, Patel KJ. 2008. Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Mol Cell* 32(6):767-77.
- Alter BP, Rosenberg PS, Brody LC. 2007. Clinical and molecular features associated with biallelic mutations in FANCD1/BRCA2. J Med Genet 44(1):1-9.
- Amann J, Kidd VJ, Lahti JM. 1997. Characterization of putative human homologues of the yeast chromosome transmission fidelity gene, CHL1. J Biol Chem 272(6):3823-32.
- Ambivero CT, Cilenti L, Main S, Zervos AS. 2014. Mulan E3 ubiquitin ligase interacts with multiple E2 conjugating enzymes and participates in mitophagy by recruiting GABARAP. *Cell Signal* 26(12):2921-9.
- Ameziane N, Errami A, Leveille F, Fontaine C, de Vries Y, van Spaendonk RM, de Winter JP, Pals G, Joenje H. 2008. Genetic subtyping of Fanconi anemia by comprehensive mutation screening. *Hum Mutat* 29(1):159-66.
- Ameziane N, Sie D, Dentro S, Ariyurek Y, Kerkhoven L, Joenje H, Dorsman JC, Ylstra B, Gille JJ, Sistermans EA, de Winter JP. 2012. Diagnosis of fanconi anemia: mutation analysis by next-generation sequencing. *Anemia* 2012:132856.
- Ameziane N, May P, Haitjema A, van de Vrugt HJ, van Rossum-Fikkert SE, Ristic D, Williams GJ, Balk J, Rockx D, Li H, Rooimans MA, Oostra AB, Velleuer E, Dietrich R, Bleijerveld OB, Maarten Altelaar AF, Meijers-Heijboer H, Joenje H, Glusman G, Roach J, Hood L, Galas D, Wyman C, Balling R, den Dunnen J, de Winter JP, Kanaar R, Gelinas R, Dorsman JC. 2015. A novel Fanconi anaemia subtype associated with a dominant-negative mutation in RAD51. Nat Commun 6:8829.
- Andreassen PR, D'Andrea AD, Taniguchi T. 2004. ATR couples FANCD2 monoubiquitination to the DNAdamage response. *Genes Dev* 18(16):1958-63.
- Andries K, Verhasselt P, Guillemont J, Gohlmann HW, Neefs JM, Winkler H, Van Gestel J, Timmerman P, Zhu M, Lee E, Williams P, de Chaffoy D, Huitric E, Hoffner S, Cambau E, Truffot-Pernot C, Lounis N, Jarlier V. 2005. A diarylquinoline drug active on the ATP synthase of Mycobacterium tuberculosis. *Science* 307(5707):223-7.
- Apostolou S, Whitmore SA, Crawford J, Lennon G, Sutherland GR, Callen DF, Ianzano L, Savino M, D'Apolito M, Notarangeio A, Memeo E, Piemontese MR, Zelante L, Savoia A, Gibson RA, Tipping AJ, Morgan NV, Hassock S, Jansen S, de Ravel TJ, Van Berkell C, Pronk JC, Easton DF, Mathew CG, Levran O, Verlander PC, Batish SD, Erlich T, Auerbach AD, Cleton-Jansen A-M, Moerland

EW, Cornelisse CJ, Doggett NA, Deaven LL, Moyzis RK. 1996. Positional cloning of the Fanconi anaemia group A gene. *Nat Genet* 14(3):324-328.

- Arase S, Kozuka T, Tanaka K, Ikenaga M, Takebe H. 1979. A sixth complementation group in xeroderma pigmentosum. *Mutat Res* 59(1):143-6.
- Auerbach AD, Greenbaum J, Pujara K, Batish SD, Bitencourt MA, Kokemohr I, Schneider H, Lobitzc S, Pasquini R, Giampietro PF, Hanenberg H, Levran O, International Fanconi Anemia R. 2003. Spectrum of sequence variation in the FANCG gene: an International Fanconi Anemia Registry (IFAR) study. *Hum Mutat* 21(2):158-68.
- Awadalla P, Gauthier J, Myers RA, Casals F, Hamdan FF, Griffing AR, Cote M, Henrion E, Spiegelman D, Tarabeux J, Piton A, Yang Y, Boyko A, Bustamante C, Xiong L, Rapoport JL, Addington AM, DeLisi JL, Krebs MO, Joober R, Millet B, Fombonne E, Mottron L, Zilversmit M, Keebler J, Daoud H, Marineau C, Roy-Gagnon MH, Dube MP, Eyre-Walker A, Drapeau P, Stone EA, Lafreniere RG, Rouleau GA. 2010. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am J Hum Genet* 87(3):316-24.
- Bakker JL, van Mil SE, Crossan G, Sabbaghian N, De Leeneer K, Poppe B, Adank M, Gille H, Verheul H, Meijers-Heijboer H, de Winter JP, Claes K, Tischkowitz M, Waisfisz Q. 2013. Analysis of the novel fanconi anemia gene SLX4/FANCP in familial breast cancer cases. *Hum Mutat* 34(1):70-3.
- Balta G, de Winter JP, Kayserili H, Pronk JC, Joenje H. 2000. Fanconi anemia A due to a novel frameshift mutation in hotspot motifs: lack of FANCA protein. *Hum Mutat* 15(6):578.
- Banning A, Gulec C, Rouvinen J, Gray SJ, Tikkanen R. 2016. Identification of Small Molecule Compounds for Pharmacological Chaperone Therapy of Aspartylglucosaminuria. *Sci Rep* 6:37583.
- Barber LJ, Youds JL, Ward JD, McIlwraith MJ, O'Neil NJ, Petalcorin MI, Martin JS, Collis SJ, Cantor SB, Auclair M, Tissenbaum H, West SC, Rose AM, Boulton SJ. 2008. RTEL1 maintains genomic stability by suppressing homologous recombination. *Cell* 135(2):261-71.
- Bastin-Shanower SA, Brill SJ. 2001. Functional analysis of the four DNA binding domains of replication protein A. The role of RPA2 in ssDNA binding. *J Biol Chem* 276(39):36446-53.
- Baudhuin LM, Lagerstedt SA, Klee EW, Fadra N, Oglesbee D, Ferber MJ. 2015. Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing. J Mol Diagn 17(4):456-61.
- Beck TF, Mullikin JC, Program NCS, Biesecker LG. 2016. Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clin Chem* 62(4):647-54.
- Berwick M, Satagopan JM, Ben-Porat L, Carlson A, Mah K, Henry R, Diotti R, Milton K, Pujara K, Landers T, Dev Batish S, Morales J, Schindler D, Hanenberg H, Hromas R, Levran O, Auerbach AD. 2007. Genetic heterogeneity among Fanconi anemia heterozygotes and risk of cancer. *Cancer Res* 67(19):9591-6.
- Bluteau D, Masliah-Planchon J, Clairmont C, Rousseau A, Ceccaldi R, Dubois d'Enghien C, Bluteau O, Cuccuini W, Gachet S, Peffault de Latour R, Leblanc T, Socie G, Baruchel A, Stoppa-Lyonnet D, D'Andrea AD, Soulier J. 2016. Biallelic inactivation of REV7 is associated with Fanconi anemia. J Clin Invest 126(9):3580-4.
- Bochkarev A, Pfuetzner RA, Edwards AM, Frappier L. 1997. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* 385(6612):176-81.
- Bodi K, Perera AG, Adams PS, Bintzler D, Dewar K, Grove DS, Kieleczawa J, Lyons RH, Neubert TA, Noll AC, Singh S, Steen R, Zianni M. 2013. Comparison of commercially available target enrichment methods for next-generation sequencing. J Biomol Tech 24(2):73-86.
- Bogliolo M, Schuster B, Stoepker C, Derkunt B, Su Y, Raams A, Trujillo JP, Minguillon J, Ramirez MJ, Pujol R, Casado JA, Banos R, Rio P, Knies K, Zuniga S, Benitez J, Bueren JA, Jaspers NG, Scharer OD, de Winter JP, Schindler D, Surralles J. 2013. Mutations in ERCC4, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *Am J Hum Genet* 92(5):800-6.
- Bogliolo M, Surralles J. 2015. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. *Curr Opin Genet Dev* 33:32-40.
- Bolderson E, Tomimatsu N, Richard DJ, Boucher D, Kumar R, Pandita TK, Burma S, Khanna KK. 2010. Phosphorylation of Exo1 modulates homologous recombination repair of DNA double-strand breaks. *Nucleic Acids Res* 38(6):1821-31.

- Boulton SJ, Jackson SP. 1996. Identification of a Saccharomyces cerevisiae Ku80 homologue: roles in DNA double strand break rejoining and in telomeric maintenance. *Nucleic Acids Res* 24(23):4639-48.
- Brett M, McPherson J, Zang ZJ, Lai A, Tan ES, Ng I, Ong LC, Cham B, Tan P, Rozen S, Tan EC. 2014. Massively parallel sequencing of patients with intellectual disability, congenital anomalies and/or autism spectrum disorders with a targeted gene panel. *PLoS One* 9(4):e93409.
- Bridge WL, Vandenberg CJ, Franklin RJ, Hiom K. 2005. The BRIP1 helicase functions independently of BRCA1 in the Fanconi anemia pathway for DNA crosslink repair. *Nat Genet* 37(9):953-7.
- Burma S, Chen BP, Murphy M, Kurimasa A, Chen DJ. 2001. ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *J Biol Chem* 276(45):42462-7.
- Byrd PJ, Stewart GS, Smith A, Eaton C, Taylor AJ, Guy C, Eringyte I, Fooks P, Last JI, Horsley R, Oliver AW, Janic D, Dokmanovic L, Stankovic T, Taylor AM. 2016. A Hypomorphic PALB2 Allele Gives Rise to an Unusual Form of FA-N Associated with Lymphoid Tumour Development. *PLoS Genet* 12(3):e1005945.
- Caldecott KW, McKeown CK, Tucker JD, Ljungquist S, Thompson LH. 1994. An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol Cell Biol* 14(1):68-76.
- Callen E, Casado JA, Tischkowitz MD, Bueren JA, Creus A, Marcos R, Dasi A, Estella JM, Munoz A, Ortega JJ, de Winter J, Joenje H, Schindler D, Hanenberg H, Hodgson SV, Mathew CG, Surralles J. 2005. A common founder mutation in FANCA underlies the world's highest prevalence of Fanconi anemia in Gypsy families from Spain. *Blood* 105(5):1946-9.
- Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA, Livingston DM. 2001. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* 105(1):149-60.
- Carter MS, Doskow J, Morris P, Li S, Nhim RP, Sandstedt S, Wilkinson MF. 1995. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. *J Biol Chem* 270(48):28995-9003.
- Castella M, Pujol R, Callen E, Trujillo JP, Casado JA, Gille H, Lach FP, Auerbach AD, Schindler D, Benitez J, Porto B, Ferro T, Munoz A, Sevilla J, Madero L, Cela E, Belendez C, de Heredia CD, Olive T, de Toledo JS, Badell I, Torrent M, Estella J, Dasi A, Rodriguez-Villa A, Gomez P, Barbot J, Tapia M, Molines A, Figuera A, Bueren JA, Surralles J. 2011. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood* 117(14):3759-69.
- Chandra S, Levran O, Jurickova I, Maas C, Kapur R, Schindler D, Henry R, Milton K, Batish SD, Cancelas JA, Hanenberg H, Auerbach AD, Williams DA. 2005. A rapid method for retrovirus-mediated identification of complementation groups in Fanconi anemia patients. *Mol Ther* 12(5):976-84.
- Chandrasekharappa SC, Lach FP, Kimble DC, Kamat A, Teer JK, Donovan FX, Flynn E, Sen SK, Thongthip S, Sanborn E, Smogorzewska A, Auerbach AD, Ostrander EA, Program NCS. 2013. Massively parallel sequencing, aCGH, and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. *Blood* 121(22):e138-48.
- Chaudhuri TK, Paul S. 2006. Protein-misfolding diseases and chaperone-based therapeutic approaches. *FEBS J* 273(7):1331-49.
- Chimpanzee Sequencing Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437(7055):69-87.
- Christensen DE, Brzovic PS, Klevit RE. 2007. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nat Struct Mol Biol* 14(10):941-8.
- Ciccia A, Ling C, Coulthard R, Yan Z, Xue Y, Meetei AR, Laghmani el H, Joenje H, McDonald N, de Winter JP, Wang W, West SC. 2007. Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Mol Cell* 25(3):331-43.
- Ciccia A, Elledge SJ. 2010. The DNA damage response: making it safe to play with knives. *Mol Cell* 40(2):179-204.
- Cleaver JE, Lam ET, Revet I. 2009. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nat Rev Genet* 10(11):756-68.
- Cole RS. 1973. Repair of DNA containing interstrand crosslinks in Escherichia coli: sequential excision and recombination. *Proc Natl Acad Sci U S A* 70(4):1064-8.

- Collins NB, Wilson JB, Bush T, Thomashevski A, Roberts KJ, Jones NJ, Kupfer GM. 2009. ATR-dependent phosphorylation of FANCA on serine 1449 after DNA damage is important for FA pathway function. *Blood* 113(10):2181-90.
- Collis SJ, Ciccia A, Deans AJ, Horejsi Z, Martin JS, Maslen SL, Skehel JM, Elledge SJ, West SC, Boulton SJ. 2008. FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. *Mol Cell* 32(3):313-24.
- Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, Zilversmit M, Cartwright R, Rouleau GA, Daly M, Stone EA, Hurles ME, Awadalla P, Genomes P. 2011. Variation in genome-wide mutation rates within and between human families. *Nat Genet* 43(7):712-4.
- Cummings CJ, Sun Y, Opal P, Antalffy B, Mestril R, Orr HT, Dillmann WH, Zoghbi HY. 2001. Overexpression of inducible HSP70 chaperone suppresses neuropathology and improves motor function in SCA1 mice. *Hum Mol Genet* 10(14):1511-8.
- de Vries Y, Lwiwski N, Levitus M, Kuyt B, Israels SJ, Arwert F, Zwaan M, Greenberg CR, Alter BP, Joenje H, Meijers-Heijboer H. 2012. A Dutch Fanconi Anemia FANCC Founder Mutation in Canadian Manitoba Mennonites. *Anemia* 2012:865170.
- de Winter JP, Waisfisz Q, Rooimans MA, van Berkel CG, Bosnoyan-Collins L, Alon N, Carreau M, Bender O, Demuth I, Schindler D, Pronk JC, Arwert F, Hoehn H, Digweed M, Buchwald M, Joenje H. 1998. The Fanconi anaemia group G gene FANCG is identical with XRCC9. Nat Genet 20(3):281-3.
- de Winter JP, Rooimans MA, van Der Weel L, van Berkel CG, Alon N, Bosnoyan-Collins L, de Groot J, Zhi Y, Waisfisz Q, Pronk JC, Arwert F, Mathew CG, Scheper RJ, Hoatlin ME, Buchwald M, Joenje H. 2000a. The Fanconi anaemia gene FANCF encodes a novel protein with homology to ROM. *Nat Genet* 24(1):15-6.
- de Winter JP, Leveille F, van Berkel CG, Rooimans MA, van Der Weel L, Steltenpool J, Demuth I, Morgan NV, Alon N, Bosnoyan-Collins L, Lightfoot J, Leegwater PA, Waisfisz Q, Komatsu K, Arwert F, Pronk JC, Mathew CG, Digweed M, Buchwald M, Joenje H. 2000b. Isolation of a cDNA representing the Fanconi anemia complementation group E gene. *Am J Hum Genet* 67(5):1306-8.
- de Winter JP, van der Weel L, de Groot J, Stone S, Waisfisz Q, Arwert F, Scheper RJ, Kruyt FA, Hoatlin ME, Joenje H. 2000c. The Fanconi anemia protein FANCF forms a nuclear complex with FANCA, FANCC and FANCG. *Hum Mol Genet* 9(18):2665-74.
- Deans AJ, West SC. 2009. FANCM connects the genome instability disorders Bloom's Syndrome and Fanconi Anemia. *Mol Cell* 36(6):943-53.
- Demple B, Herman T, Chen DS. 1991. Cloning and expression of APE, the cDNA encoding the major human apurinic endonuclease: definition of a family of DNA repair enzymes. *Proc Natl Acad Sci U S A* 88(24):11450-4.
- Demuth I, Włodarski M, Tipping AJ, Morgan NV, de Winter JP, Thiel M, Grasl S, Schindler D, D'Andrea AD, Altay C, Kayserili H, Zatterale A, Kunze J, Ebell W, Mathew CG, Joenje H, Sperling K, Digweed M. 2000. Spectrum of mutations in the Fanconi anaemia group G gene, FANCG/XRCC9. *Eur J Hum Genet* 8(11):861-8.
- Dodd RB, Allen MD, Brown SE, Sanderson CM, Duncan LM, Lehner PJ, Bycroft M, Read RJ. 2004. Solution structure of the Kaposi's sarcoma-associated herpesvirus K3 N-terminal domain reveals a Novel E2binding C4HC3-type RING domain. J Biol Chem 279(51):53840-7.
- Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, Bennett RL, Jones KL, Tokita MJ, Bennett JT, Kim JH, Rosenthal EA, Kim DS, National Heart L, Blood Institute Grand Opportunity Exome Sequencing P, Tabor HK, Bamshad MJ, Motulsky AG, Scott CR, Pritchard CC, Walsh T, Burke W, Raskind WH, Byers P, Hisama FM, Nickerson DA, Jarvik GP. 2013. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet* 93(4):631-40.
- Dorsman JC, Levitus M, Rockx D, Rooimans MA, Oostra AB, Haitjema A, Bakker ST, Steltenpool J, Schuler D, Mohan S, Schindler D, Arwert F, Pals G, Mathew CG, Waisfisz Q, de Winter JP, Joenje H. 2007. Identification of the Fanconi anemia complementation group I gene, FANCI. *Cell Oncol* 29(3):211-8.
- Drummond JT, Li GM, Longley MJ, Modrich P. 1995. Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science* 268(5219):1909-12.

- Ebersberger I, Metzler D, Schwarz C, Paabo S. 2002. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am J Hum Genet* 70(6):1490-7.
- Ellis J. 1987. Proteins as molecular chaperones. Nature 328(6129):378-9.
- Evans T, Rosenthal ET, Youngblom J, Distel D, Hunt T. 1983. Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. *Cell* 33(2):389-96.
- Fanconi G. 1927. Familiäre infantile perniziosaartige Anämie (perniziöses Blutbild und Konstitution). Jahrbuch für Kinderheilkunde und physische Erziehung. Wien. 117: p 257-280.
- Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Tippin Davis B, Baxter RM, Zeng W, Mroske C, Parra MC, Gandomi SK, Lu I, Li X, Lu H, Lu HM, Salvador D, Ruble D, Lao M, Fischbach S, Wen J, Lee S, Elliott A, Dunlop CL, Tang S. 2015. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med* 17(7):578-86.
- Fassihi H, Sethi M, Fawcett H, Wing J, Chandler N, Mohammed S, Craythorne E, Morley AM, Lim R, Turner S, Henshaw T, Garrood I, Giunti P, Hedderly T, Abiona A, Naik H, Harrop G, McGibbon D, Jaspers NG, Botta E, Nardo T, Stefanini M, Young AR, Sarkany RP, Lehmann AR. 2016. Deep phenotyping of 89 xeroderma pigmentosum patients reveals unexpected heterogeneity dependent on the precise molecular defect. *Proc Natl Acad Sci U S A* 113(9):E1236-45.
- Finley D, Ciechanover A, Varshavsky A. 1984. Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. *Cell* 37(1):43-55.
- Flynn EK, Kamat A, Lach FP, Donovan FX, Kimble DC, Narisu N, Sanborn E, Boulad F, Davies SM, Gillio AP, 3rd, Harris RE, MacMillan ML, Wagner JE, Smogorzewska A, Auerbach AD, Ostrander EA, Chandrasekharappa SC. 2014. Comprehensive analysis of pathogenic deletion variants in Fanconi anemia genes. *Hum Mutat* 35(11):1342-53.
- Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struewing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BA, Gayther SA, Zelada-Hedman M, et al. 1998. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* 62(3):676-89.
- Frosina G, Fortini P, Rossi O, Carrozzino F, Raspaglio G, Cox LS, Lane DP, Abbondandolo A, Dogliotti E. 1996. Two pathways for base excision repair in mammalian cells. *J Biol Chem* 271(16):9573-8.
- Futaki M, Yamashita T, Yagasaki H, Toda T, Yabe M, Kato S, Asano S, Nakahata T. 2000. The IVS4 + 4 A to T mutation of the fanconi anemia gene FANCC is not associated with a severe phenotype in Japanese patients. *Blood* 95(4):1493-8.
- Gan GN, Wittschieben JP, Wittschieben BO, Wood RD. 2008. DNA polymerase zeta (pol zeta) in higher eukaryotes. *Cell Res* 18(1):174-83.
- Garcia-Higuera I, Kuang Y, Naf D, Wasik J, D'Andrea AD. 1999. Fanconi anemia proteins FANCA, FANCC, and FANCG/XRCC9 interact in a functional nuclear complex. *Mol Cell Biol* 19(7):4866-73.
- Garcia-Higuera I, Kuang Y, Denham J, D'Andrea AD. 2000. The fanconi anemia proteins FANCA and FANCG stabilize each other and promote the nuclear accumulation of the Fanconi anemia complex. *Blood* 96(9):3224-30.
- Garcia-Higuera I, Taniguchi T, Ganesan S, Meyn MS, Timmers C, Hejna J, Grompe M, D'Andrea AD. 2001. Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* 7(2):249-62.
- Gari K, Decaillet C, Delannoy M, Wu L, Constantinou A. 2008. Remodeling of DNA replication structures by the branch point translocase FANCM. *Proc Natl Acad Sci U S A* 105(42):16107-12.
- Giampietro PF, Verlander PC, Davis JG, Auerbach AD. 1997. Diagnosis of Fanconi anemia in patients without congenital malformations: an international Fanconi Anemia Registry Study. *Am J Med Genet* 68(1):58-61.
- Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M, van Reeuwijk J, Kant SG, Roepman R, Knoers NV, Veltman JA, Brunner HG. 2010. Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am J Hum Genet* 87(3):418-23.
- Gille JJ, Floor K, Kerkhoven L, Ameziane N, Joenje H, de Winter JP. 2012. Diagnosis of Fanconi Anemia: Mutation Analysis by Multiplex Ligation-Dependent Probe Amplification and PCR-Based Sanger Sequencing. Anemia 2012:603253.
- Gilles A, Meglecz E, Pech N, Ferreira S, Malausa T, Martin JF. 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* 12:245.
- Godthelp BC, Artwert F, Joenje H, Zdzienicka MZ. 2002. Impaired DNA damage-induced nuclear Rad51 foci formation uniquely characterizes Fanconi anemia group D1. Oncogene 21(32):5002-5.
- Gonzalez-Murillo A, Lozano ML, Alvarez L, Jacome A, Almarza E, Navarro S, Segovia JC, Hanenberg H, Guenechea G, Bueren JA, Rio P. 2010. Development of lentiviral vectors with optimized transcriptional activity for the gene therapy of patients with Fanconi anemia. *Hum Gene Ther* 21(5):623-30.
- Gordon SM, Buchwald M. 2003. Fanconi anemia protein complex: mapping protein interactions in the yeast 2- and 3-hybrid systems. *Blood* 102(1):136-41.
- Gordon SM, Alon N, Buchwald M. 2005. FANCC, FANCE, and FANCD2 form a ternary complex essential to the integrity of the Fanconi anemia DNA damage response pathway. *J Biol Chem* 280(43):36118-25.
- Goryshin IY, Miller JA, Kil YV, Lanzov VA, Reznikoff WS. 1998. Tn5/IS50 target recognition. *Proc Natl Acad Sci U S A* 95(18):10716-21.
- Gottlieb TM, Jackson SP. 1993. The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* 72(1):131-42.
- Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG, American College of Medical G, Genomics. 2013. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 15(7):565-74.
- Gregg SQ, Robinson AR, Niedernhofer LJ. 2011. Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. DNA Repair (Amst) 10(7):781-91.
- Groisman R, Polanowska J, Kuraoka I, Sawada J, Saijo M, Drapkin R, Kisselev AF, Tanaka K, Nakatani Y. 2003. The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* 113(3):357-367.
- Gross M, Hanenberg H, Lobitz S, Friedl R, Herterich S, Dietrich R, Gruhn B, Schindler D, Hoehn H. 2002. Reverse mosaicism in Fanconi anemia: natural gene therapy via molecular self-correction. *Cytogenet Genome Res* 98(2-3):126-35.
- Handley PM, Mueckler M, Siegel NR, Ciechanover A, Schwartz AL. 1991. Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1. Proc Natl Acad Sci U S A 88(1):258-62.
- Hanenberg H, Hashino K, Konishi H, Hock RA, Kato I, Williams DA. 1997. Optimization of fibronectinassisted retroviral gene transfer into human CD34+ hematopoietic cells. *Hum Gene Ther* 8(18):2193-206.
- Hanenberg H, Batish SD, Pollok KE, Vieten L, Verlander PC, Leurs C, Cooper RJ, Gottsche K, Haneline L, Clapp DW, Lobitz S, Williams DA, Auerbach AD. 2002. Phenotypic correction of primary Fanconi anemia T cells with retroviral vectors as a diagnostic tool. *Exp Hematol* 30(5):410-20.
- Haring SJ, Mason AC, Binz SK, Wold MS. 2008. Cellular functions of human RPA1. Multiple roles of domains in replication, repair, and checkpoints. J Biol Chem 283(27):19095-111.
- Hartmann L, Neveling K, Borkens S, Schneider H, Freund M, Grassman E, Theiss S, Wawer A, Burdach S, Auerbach AD, Schindler D, Hanenberg H, Schaal H. 2010. Correct mRNA processing at a mutant TT splice donor in FANCC ameliorates the clinical phenotype in patients and is enhanced by delivery of suppressor U1 snRNAs. *Am J Hum Genet* 87(4):480-93.
- Harutyunyan A, Gisslinger B, Klampfl T, Berg T, Bagienski K, Gisslinger H, Kralovics R. 2011. Rare germline variants in regions of loss of heterozygosity may influence clinical course of hematological malignancies. *Leukemia* 25(11):1782-4.
- Hashimoto K, Wada K, Matsumoto K, Moriya M. 2015. Physical interaction between SLX4 (FANCP) and XPF (FANCQ) proteins and biological consequences of interaction-defective missense mutations. DNA Repair (Amst) 35:48-54.
- Hass CS, Lam K, Wold MS. 2012. Repair-specific functions of replication protein A. J Biol Chem 287(6):3908-18.
- Hejna JA, Timmers CD, Reifsteck C, Bruun DA, Lucas LW, Jakobs PM, Toth-Fejel S, Unsworth N, Clemens SL, Garcia DK, Naylor SL, Thayer MJ, Olson SB, Grompe M, Moses RE. 2000. Localization of the

Fanconi anemia complementation group D gene to a 200-kb region on chromosome 3p25.3. Am J Hum Genet 66(5):1540-51.

- Hicks S, Wheeler DA, Plon SE, Kimmel M. 2011. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat* 32(6):661-8.
- Hira A, Yoshida K, Sato K, Okuno Y, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Shimamoto A, Tahara H, Ito E, Kojima S, Kurumizaka H, Ogawa S, Takata M, Yabe H, Yabe M. 2015. Mutations in the gene encoding the E2 conjugating enzyme UBE2T cause Fanconi anemia. *Am J Hum Genet* 96(6):1001-7.
- Hodson C, Purkiss A, Miles JA, Walden H. 2014. Structure of the human FANCL RING-Ube2T complex reveals determinants of cognate E3-E2 selection. *Structure* 22(2):337-44.
- Houghtaling S, Timmers C, Noll M, Finegold MJ, Jones SN, Meyn MS, Grompe M. 2003. Epithelial cancer in Fanconi anemia complementation group D2 (Fancd2) knockout mice. *Genes Dev* 17(16):2021-35.
- Howlett NG, Taniguchi T, Olson S, Cox B, Waisfisz Q, De Die-Smulders C, Persky N, Grompe M, Joenje H, Pals G, Ikeda H, Fox EA, D'Andrea AD. 2002. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* 297(5581):606-9.
- Huang M, Kim JM, Shiotani B, Yang K, Zou L, D'Andrea AD. 2010. The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response. *Mol Cell* 39(2):259-68.
- Huang Y, Leung JW, Lowery M, Matsushita N, Wang Y, Shen X, Huong D, Takata M, Chen J, Li L. 2014. Modularized functions of the Fanconi anemia core complex. *Cell Rep* 7(6):1849-57.
- Hustedt N, Durocher D. 2016. The control of DNA repair by the cell cycle. Nat Cell Biol 19(1):1-9.
- Hyman ED. 1988. A new method of sequencing DNA. Anal Biochem 174(2):423-36.
- Imai K, Sarker AH, Akiyama K, Ikeda S, Yao M, Tsutsui K, Shohmori T, Seki S. 1998. Genomic structure and sequence of a human homologue (NTHL1/NTH1) of Escherichia coli endonuclease III with those of the adjacent parts of TSC2 and SLC9A3R2 genes. *Gene* 222(2):287-95.
- Ishiai M, Kitao H, Smogorzewska A, Tomida J, Kinomura A, Uchida E, Saberi A, Kinoshita E, Kinoshita E, Koike T, Tashiro S, Elledge SJ, Takata M. 2008. FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nat Struct Mol Biol* 15(11):1138-46.
- Joenje H, Levitus M, Waisfisz Q, D'Andrea A, Garcia-Higuera I, Pearson T, van Berkel CG, Rooimans MA, Morgan N, Mathew CG, Arwert F. 2000. Complementation analysis in Fanconi anemia: assignment of the reference FA-H patient to group A. *Am J Hum Genet* 67(3):759-62.
- Joenje H, Patel KJ. 2001. The emerging genetic and molecular basis of Fanconi anaemia. Nat Rev Genet 2(6):446-57.
- Joo W, Xu G, Persky NS, Smogorzewska A, Rudge DG, Buzovetsky O, Elledge SJ, Pavletich NP. 2011. Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway. *Science* 333(6040):312-6.
- Kalb R, Neveling K, Hoehn H, Schneider H, Linka Y, Batish SD, Hunt C, Berwick M, Callen E, Surralles J, Casado JA, Bueren J, Dasi A, Soulier J, Gluckman E, Zwaan CM, van Spaendonk R, Pals G, de Winter JP, Joenje H, Grompe M, Auerbach AD, Hanenberg H, Schindler D. 2007. Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. *Am J Hum Genet* 80(5):895-910.
- Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT. 2017. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med* 19(2):249-255.
- Kashiyama K, Nakazawa Y, Pilz DT, Guo C, Shimada M, Sasaki K, Fawcett H, Wing JF, Lewin SO, Carr L, Li TS, Yoshiura K, Utani A, Hirano A, Yamashita S, Greenblatt D, Nardo T, Stefanini M, McGibbon D, Sarkany R, Fassihi H, Takahashi Y, Nagayama Y, Mitsutake N, Lehmann AR, Ogi T. 2013. Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. *Am J Hum Genet* 92(5):807-19.
- Kent WJ. 2002. BLAT--the BLAST-like alignment tool. Genome Res 12(4):656-64.
- Kim H, Yang K, Dejsuphong D, D'Andrea AD. 2012. Regulation of Rev1 by the Fanconi anemia core complex. Nat Struct Mol Biol 19(2):164-70.

- Kim JM, Kee Y, Gurtan A, D'Andrea AD. 2008. Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. *Blood* 111(10):5215-22.
- Kim K, Biade S, Matsumoto Y. 1998. Involvement of flap endonuclease 1 in base excision DNA repair. J Biol Chem 273(15):8842-8.
- Kim Y, Lach FP, Desetty R, Hanenberg H, Auerbach AD, Smogorzewska A. 2011. Mutations of the SLX4 gene in Fanconi anemia. *Nat Genet* 43(2):142-6.
- Kim Y, Spitz GS, Veturi U, Lach FP, Auerbach AD, Smogorzewska A. 2013. Regulation of multiple DNA repair pathways by the Fanconi anemia protein SLX4. *Blood* 121(1):54-63.
- Klungland A, Lindahl T. 1997. Second pathway for completion of human DNA base excision-repair: Reconstitution with purified proteins and requirement for DNase IV (FEN1). *Embo Journal* 16(11):3341-3348.
- Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D. 2011. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One* 6(11):e28240.
- Knies K, Schuster B, Ameziane N, Rooimans M, Bettecken T, de Winter J, Schindler D. 2012. Genotyping of fanconi anemia patients by whole exome sequencing: advantages and challenges. *PLoS One* 7(12):e52648.
- Koc A, Pronk JC, Alikasifoglu M, Joenje H, Altay C. 1999. Variable pathogenicity of exon 43del (FAA) in four Fanconi anaemia patients within a consanguineous family. Br J Haematol 104(1):127-30.
- Kopic S, Eirich K, Schuster B, Hanenberg H, Varon-Mateeva R, Rittinger O, Schimpl G, Schindler D, Jones N. 2011. Hepatoblastoma in a 4-year-old girl with Fanconi anaemia. *Acta Paediatr* 100(5):780-3.
- Kubota Y, Nash RA, Klungland A, Schar P, Barnes DE, Lindahl T. 1996. Reconstitution of DNA base excision-repair with purified human proteins: Interaction between DNA polymerase beta and the XRCC1 protein. *Embo Journal* 15(23):6662-6670.
- Kuiper RP, Hoogerbrugge N. 2015. NTHL1 defines novel cancer syndrome. Oncotarget 6(33):34069-70.
- Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4(7):1073-81.
- Kutler DI, Singh B, Satagopan J, Batish SD, Berwick M, Giampietro PF, Hanenberg H, Auerbach AD. 2003. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood* 101(4):1249-56.
- Kutler DI, Patel KR, Auerbach AD, Kennedy J, Lach FP, Sanborn E, Cohen MA, Kuhel WI, Smogorzewska A. 2016. Natural history and management of Fanconi anemia patients with head and neck cancer: A 10-year follow-up. *Laryngoscope* 126(4):870-9.
- Lamarche BJ, Orazio NI, Weitzman MD. 2010. The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett* 584(17):3682-95.
- Landwehr R, Bogdanova NV, Antonenkova N, Meyer A, Bremer M, Park-Simon TW, Hillemanns P, Karstens JH, Schindler D, Dork T. 2011. Mutation analysis of the SLX4/FANCP gene in hereditary breast cancer. *Breast Cancer Res Treat* 130(3):1021-8.
- Levitus M, Rooimans MA, Steltenpool J, Cool NF, Oostra AB, Mathew CG, Hoatlin ME, Waisfisz Q, Arwert F, de Winter JP, Joenje H. 2004. Heterogeneity in Fanconi anemia: evidence for 2 new genetic subtypes. *Blood* 103(7):2498-503.
- Levitus M, Waisfisz Q, Godthelp BC, de Vries Y, Hussain S, Wiegant WW, Elghalbzouri-Maghrani E, Steltenpool J, Rooimans MA, Pals G, Arwert F, Mathew CG, Zdzienicka MZ, Hiom K, De Winter JP, Joenje H. 2005. The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. Nat Genet 37(9):934-5.
- Levran O, Erlich T, Magdalena N, Gregory JJ, Batish SD, Verlander PC, Auerbach AD. 1997. Sequence variation in the Fanconi anemia gene FAA. *Proc Natl Acad Sci U S A* 94(24):13051-6.
- Levran O, Diotti R, Pujara K, Batish SD, Hanenberg H, Auerbach AD. 2005a. Spectrum of sequence variations in the FANCA gene: an International Fanconi Anemia Registry (IFAR) study. *Hum Mutat* 25(2):142-9.
- Levran O, Attwooll C, Henry RT, Milton KL, Neveling K, Rio P, Batish SD, Kalb R, Velleuer E, Barral S, Ott J, Petrini J, Schindler D, Hanenberg H, Auerbach AD. 2005b. The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. *Nat Genet* 37(9):931-3.
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz

SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC. 2007. The diploid genome sequence of an individual human. *PLoS Biol* 5(10):e254.

- Li W, Bengtson MH, Ulbrich A, Matsuda A, Reddy VA, Orth A, Chanda SK, Batalov S, Joazeiro CA. 2008. Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. *PLoS One* 3(1):e1487.
- Liang CC, Li Z, Lopez-Martinez D, Nicholson WV, Venien-Bryan C, Cohn MA. 2016. The FANCD2-FANCI complex is recruited to DNA interstrand crosslinks before monoubiquitination of FANCD2. *Nat Commun* 7:12124.
- Lim ET, Wurtz P, Havulinna AS, Palta P, Tukiainen T, Rehnstrom K, Esko T, Magi R, Inouye M, Lappalainen T, Chan Y, Salem RM, Lek M, Flannick J, Sim X, Manning A, Ladenvall C, Bumpstead S, Hamalainen E, Aalto K, Maksimow M, Salmi M, Blankenberg S, Ardissino D, Shah S, Horne B, McPherson R, Hovingh GK, Reilly MP, Watkins H, Goel A, Farrall M, Girelli D, Reiner AP, Stitziel NO, Kathiresan S, Gabriel S, Barrett JC, Lehtimaki T, Laakso M, Groop L, Kaprio J, Perola M, McCarthy MI, Boehnke M, Altshuler DM, Lindgren CM, Hirschhorn JN, Metspalu A, Freimer NB, Zeller T, Jalkanen S, Koskinen S, Raitakari O, Durbin R, MacArthur DG, Salomaa V, Ripatti S, Daly MJ, Palotie A, Sequencing Initiative Suomi P. 2014. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet* 10(7):e1004494.
- Ling C, Ishiai M, Ali AM, Medhurst AL, Neveling K, Kalb R, Yan Z, Xue Y, Oostra AB, Auerbach AD, Hoatlin ME, Schindler D, Joenje H, de Winter JP, Takata M, Meetei AR, Wang W. 2007. FAAP100 is essential for activation of the Fanconi anemia-associated DNA damage response pathway. *EMBO J* 26(8):2104-14.
- Litman R, Peng M, Jin Z, Zhang F, Zhang J, Powell S, Andreassen PR, Cantor SB. 2005. BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCJ. *Cancer Cell* 8(3):255-65.
- Liu TX, Howlett NG, Deng M, Langenau DM, Hsu K, Rhodes J, Kanki JP, D'Andrea AD, Look AT. 2003. Knockdown of zebrafish Fancd2 causes developmental abnormalities via p53-dependent apoptosis. *Dev Cell* 5(6):903-14.
- Lo Ten Foe JR, Rooimans MA, Bosnoyan-Collins L, Alon N, Wijker M, Parker L, Lightfoot J, Carreau M, Callen DF, Savoia A, Cheng NC, van Berkel CG, Strunk MH, Gille JJ, Pals G, Kruyt FA, Pronk JC, Arwert F, Buchwald M, Joenje H. 1996. Expression cloning of a cDNA for the major Fanconi anaemia gene, FAA. *Nat Genet* 14(3):320-3.
- Lo Ten Foe JR, Kwee ML, Rooimans MA, Oostra AB, Veerman AJ, van Weel M, Pauli RM, Shahidi NT, Dokal I, Roberts I, Altay C, Gluckman E, Gibson RA, Mathew CG, Arwert F, Joenje H. 1997. Somatic mosaicism in Fanconi anemia: molecular basis and clinical significance. *Eur J Hum Genet* 5(3):137-48.
- Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30(5):434-9.
- Lopez-Bigas N, Audit B, Ouzounis C, Parra G, Guigo R. 2005. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett* 579(9):1900-3.
- MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, Albers CA, Zhang ZD, Conrad DF, Lunter G, Zheng H, Ayub Q, DePristo MA, Banks E, Hu M, Handsaker RE, Rosenfeld JA, Fromer M, Jin M, Mu XJ, Khurana E, Ye K, Kay M, Saunders GI, Suner MM, Hunt T, Barnes IH, Amid C, Carvalho-Silva DR, Bignell AH, Snow C, Yngvadottir B, Bumpstead S, Cooper DN, Xue Y, Romero IG, Genomes Project C, Wang J, Li Y, Gibbs RA, McCarroll SA, Dermitzakis ET, Pritchard JK, Barrett JC, Harrow J, Hurles ME, Gerstein MB, Tyler-Smith C. 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335(6070):823-8.
- Machida YJ, Machida Y, Chen Y, Gurtan AM, Kupfer GM, D'Andrea AD, Dutta A. 2006. UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation. *Mol Cell* 23(4):589-96.
- Mailand N, Bekker-Jensen S, Faustrup H, Melander F, Bartek J, Lukas C, Lukas J. 2007. RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. *Cell* 131(5):887-900.

- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376-80.
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18(9):1509-17.
- Masutani C, Sugasawa K, Yanagisawa J, Sonoyama T, Ui M, Enomoto T, Takio K, Tanaka K, van der Spek PJ, Bootsma D, et al. 1994. Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *EMBO J* 13(8):1831-43.
- Matsumoto Y, Kim K. 1995. Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science* 269(5224):699-702.
- McDowell DG, Burns NA, Parkes HC. 1998. Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. *Nucleic Acids Res* 26(14):3340-7.
- Medhurst AL, Laghmani el H, Steltenpool J, Ferrer M, Fontaine C, de Groot J, Rooimans MA, Scheper RJ, Meetei AR, Wang W, Joenje H, de Winter JP. 2006. Evidence for subcomplexes in the Fanconi anemia pathway. *Blood* 108(6):2072-80.
- Meetei AR, Sechi S, Wallisch M, Yang D, Young MK, Joenje H, Hoatlin ME, Wang W. 2003a. A multiprotein nuclear complex connects Fanconi anemia and Bloom syndrome. *Mol Cell Biol* 23(10):3417-26.
- Meetei AR, de Winter JP, Medhurst AL, Wallisch M, Waisfisz Q, van de Vrugt HJ, Oostra AB, Yan Z, Ling C, Bishop CE, Hoatlin ME, Joenje H, Wang W. 2003b. A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* 35(2):165-70.
- Meetei AR, Levitus M, Xue Y, Medhurst AL, Zwaan M, Ling C, Rooimans MA, Bier P, Hoatlin M, Pals G, de Winter JP, Wang W, Joenje H. 2004. X-linked inheritance of Fanconi anemia complementation group B. Nat Genet 36(11):1219-24.
- Meetei AR, Medhurst AL, Ling C, Xue Y, Singh TR, Bier P, Steltenpool J, Stone S, Dokal I, Mathew CG, Hoatlin M, Joenje H, de Winter JP, Wang W. 2005. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. Nat Genet 37(9):958-63.
- Mehta PA, Tolar J. 1993-2017. Fanconi anemia. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJ, Bird TD, Ledbetter N, Mefford HC, Smith RJ and others, editors. GeneReviews® [Internet]. Washington, Seattle.
- Mi J, Kupfer GM. 2005. The Fanconi anemia core complex associates with chromatin during S phase. *Blood* 105(2):759-66.
- Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, Jian M, Liu G, Greer D, Bhandari A, Wu W, Corominas R, Peoples A, Koren A, Gore A, Kang S, Lin GN, Estabillo J, Gadomski T, Singh B, Zhang K, Akshoomoff N, Corsello C, McCarroll S, Iakoucheva LM, Li Y, Wang J, Sebat J. 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151(7):1431-42.
- Miller SA, Dykes DD, Polesky HF. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16(3):1215.
- Moghrabi NN, Johnson MA, Yoshitomi MJ, Zhu X, Al-Dhalimy MJ, Olson SB, Grompe M, Richards CS. 2009. Validation of Fanconi anemia complementation Group A assignment using molecular analysis. *Genet Med* 11(3):183-92.
- Morgan NV, Tipping AJ, Joenje H, Mathew CG. 1999. High frequency of large intragenic deletions in the Fanconi anemia group A gene. *Am J Hum Genet* 65(5):1330-41.
- Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh T, McDonald H, Varhol R, Jones S, Marra M. 2008. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* 45(1):81-94.

- Mosedale G, Niedzwiedz W, Alpi A, Perrina F, Pereira-Leal JB, Johnson M, Langevin F, Pace P, Patel KJ. 2005. The vertebrate Hef ortholog is a component of the Fanconi anemia tumor-suppressor pathway. *Nat Struct Mol Biol* 12(9):763-71.
- Mouw KW, D'Andrea AD. 2014. Crosstalk between the nucleotide excision repair and Fanconi anemia/BRCA pathways. DNA Repair (Amst) 19:130-4.
- Mu W, Lu HM, Chen J, Li S, Elliott AM. 2016. Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. J Mol Diagn 18(6):923-932.
- Mullis KB, Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155:335-50.
- Murer-Orlando M, Llerena JC, Jr., Birjandi F, Gibson RA, Mathew CG. 1993. FACC gene mutations and early prenatal diagnosis of Fanconi's anaemia. *Lancet* 342(8872):686.
- Myers K, Davies SM, Harris RE, Spunt SL, Smolarek T, Zimmerman S, McMasters R, Wagner L, Mueller R, Auerbach AD, Mehta PA. 2012. The clinical phenotype of children with Fanconi anemia caused by biallelic FANCD1/BRCA2 mutations. *Pediatr Blood Cancer* 58(3):462-5.
- Neidhardt G, Hauke J, Ramser J, Gross E, Gehrig A, Muller CR, Kahlert AK, Hackmann K, Honisch E, Niederacher D, Heilmann-Heimbach S, Franke A, Lieb W, Thiele H, Altmuller J, Nurnberg P, Klaschik K, Ernst C, Ditsch N, Jessen F, Ramirez A, Wappenschmidt B, Engel C, Rhiem K, Meindl A, Schmutzler RK, Hahnen E. 2016. Association Between Loss-of-Function Mutations Within the FANCM Gene and Early-Onset Familial Breast Cancer. JAMA Oncol.
- Neveling K, Endt D, Hoehn H, Schindler D. 2009. Genotype-phenotype correlations in Fanconi anemia. *Mutat* Res 668(1-2):73-91.
- New HV, Cale CM, Tischkowitz M, Jones A, Telfer P, Veys P, D'Andrea A, Mathew CG, Hann I. 2005. Nijmegen breakage syndrome diagnosed as Fanconi anaemia. *Pediatr Blood Cancer* 44(5):494-9.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 42(1):30-5.
- Nicchia E, Greco C, De Rocco D, Pecile V, D'Eustacchio A, Cappelli E, Corti P, Marra N, Ramenghi U, Pillon M, Farruggia P, Dufour C, Pallavicini A, Torelli L, Savoia A. 2015. Identification of point mutations and large intragenic deletions in Fanconi anemia using next-generation sequencing technology. *Mol Genet Genomic Med* 3(6):500-12.
- Nick McElhinny SA, Snowden CM, McCarville J, Ramsden DA. 2000. Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol* 20(9):2996-3003.
- Niedernhofer LJ, Garinis GA, Raams A, Lalai AS, Robinson AR, Appeldoorn E, Odijk H, Oostendorp R, Ahmad A, van Leeuwen W, Theil AF, Vermeulen W, van der Horst GT, Meinecke P, Kleijer WJ, Vijg J, Jaspers NG, Hoeijmakers JH. 2006. A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature* 444(7122):1038-43.
- Nijman SM, Huang TT, Dirac AM, Brummelkamp TR, Kerkhoven RM, D'Andrea AD, Bernards R. 2005. The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Mol Cell* 17(3):331-9.
- Nishi R, Okuda Y, Watanabe E, Mori T, Iwai S, Masutani C, Sugasawa K, Hanaoka F. 2005. Centrin 2 stimulates nucleotide excision repair by interacting with xeroderma pigmentosum group C protein. *Mol Cell Biol* 25(13):5664-74.
- Ocampo MT, Chaung W, Marenstein DR, Chan MK, Altamirano A, Basu AK, Boorstein RJ, Cunningham RP, Teebor GW. 2002. Targeted deletion of mNth1 reveals a novel DNA repair enzyme activity. *Mol Cell Biol* 22(17):6111-21.
- Ochi T, Blackford AN, Coates J, Jhujh S, Mehmood S, Tamura N, Travers J, Wu Q, Draviam VM, Robinson CV, Blundell TL, Jackson SP. 2015. DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science* 347(6218):185-8.
- Oostra AB, Nieuwint AW, Joenje H, de Winter JP. 2012. Diagnosis of fanconi anemia: chromosomal breakage analysis. *Anemia* 2012:238731.
- Osborn M, Lonetree CL, Webber BR, Patel D, Dunmire S, McElroy AN, DeFeo AP, MacMillan ML, Wagner J, Balzar BR, Tolar J. 2016. CRISPR/Cas9 Targeted Gene Editing and Cellular Engineering in Fanconi Anemia. *Stem Cells Dev* 25(20):1591-603.

- Osborn MJ, Gabriel R, Webber BR, DeFeo AP, McElroy AN, Jarjour J, Starker CG, Wagner JE, Joung JK, Voytas DF, von Kalle C, Schmidt M, Blazar BR, Tolar J. 2015. Fanconi anemia gene editing by the CRISPR/Cas9 system. *Hum Gene Ther* 26(2):114-26.
- Pace P, Johnson M, Tan WM, Mosedale G, Sng C, Hoatlin M, de Winter J, Joenje H, Gergely F, Patel KJ. 2002. FANCE: the link between Fanconi anaemia complex assembly and activity. *EMBO J* 21(13):3414-23.
- Paez JG, Lin M, Beroukhim R, Lee JC, Zhao X, Richter DJ, Gabriel S, Herman P, Sasaki H, Altshuler D, Li C, Meyerson M, Sellers WR. 2004. Genome coverage and sequence fidelity of phi29 polymerasebased multiple strand displacement whole genome amplification. *Nucleic Acids Res* 32(9):e71.
- Palombo F, Iaccarino I, Nakajima E, Ikejima M, Shimada T, Jiricny J. 1996. hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Curr Biol* 6(9):1181-4.
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40(12):1413-5.
- Park JY, Virts EL, Jankowska A, Wiek C, Othman M, Chakraborty SC, Vance GH, Alkuraya FS, Hanenberg H, Andreassen PR. 2016. Complementation of hypersensitivity to DNA interstrand crosslinking agents demonstrates that XRCC2 is a Fanconi anaemia gene. J Med Genet 53(10):672-80.
- Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, Buck D, Ragoussis J, Ponting CP, Fischer MD. 2012. Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. *Genome Res* 22(1):125-33.
- Pelzer C, Kassner I, Matentzoglu K, Singh RK, Wollscheid HP, Scheffner M, Schmidtke G, Groettrup M. 2007. UBE1L2, a novel E1 enzyme specific for ubiquitin. J Biol Chem 282(32):23010-4.
- Peterlongo P, Catucci I, Colombo M, Caleca L, Mucaki E, Bogliolo M, Marin M, Damiola F, Bernard L, Pensotti V, Volorio S, Dall'Olio V, Meindl A, Bartram C, Sutter C, Surowy H, Sornin V, Dondon MG, Eon-Marchais S, Stoppa-Lyonnet D, Andrieu N, Sinilnikova OM, Genesis, Mitchell G, James PA, Thompson E, kConFab, Swe B, Marchetti M, Verzeroli C, Tartari C, Capone GL, Putignano AL, Genuardi M, Medici V, Marchi I, Federico M, Tognazzo S, Matricardi L, Agata S, Dolcetti R, Della Puppa L, Cini G, Gismondi V, Viassolo V, Perfumo C, Mencarelli MA, Baldassarri M, Peissel B, Roversi G, Silvestri V, Rizzolo P, Spina F, Vivanet C, Tibiletti MG, Caligo MA, Gambino G, Tommasi S, Pilato B, Tondini C, Corna C, Bonanni B, Barile M, Osorio A, Benitez J, Balestrino L, Ottini L, Manoukian S, Pierotti MA, Renieri A, Varesco L, Couch FJ, Wang X, Devilee P, Hilbers FS, van Asperen CJ, Viel A, Montagna M, Cortesi L, Diez O, Balmana J, Hauke J, Schmutzler RK, Papi L, Pujana MA, Lazaro C, Falanga A, Offit K, Vijai J, Campbell I, Burwinkel B, Kvist A, Ehrencrona H, Mazoyer S, Pizzamiglio S, Verderio P, Surralles J, Rogan PK, Radice P. 2015. FANCM c.5791C>T nonsense mutation (rs144567652) induces exon skipping, affects DNA repair activity and is a familial breast cancer risk factor. *Hum Mol Genet* 24(18):5345-55.
- Qiao F, Mi J, Wilson JB, Zhi G, Bucheimer NR, Jones NJ, Kupfer GM. 2004. Phosphorylation of fanconi anemia (FA) complementation group G protein, FANCG, at serine 7 is important for function of the FA pathway. J Biol Chem 279(44):46035-45.
- Rabbani B, Tekin M, Mahdieh N. 2014. The promise of whole-exome sequencing in medical genetics. J Hum Genet 59(1):5-15.
- Rackoff WR, Orazi A, Robinson CA, Cooper RJ, Alter BP, Freedman MH, Harris RE, Williams DA. 1996. Prolonged administration of granulocyte colony-stimulating factor (filgrastim) to patients with Fanconi anemia: a pilot study. *Blood* 88(5):1588-93.
- Rajendra E, Oestergaard VH, Langevin F, Wang M, Dornan GL, Patel KJ, Passmore LA. 2014. The genetic and biochemical basis of FANCD2 monoubiquitination. *Mol Cell* 54(5):858-69.
- Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E, Working Group of the American College of Medical G, Genomics Laboratory Quality Assurance C. 2013. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 15(9):733-47.
- Reid S, Schindler D, Hanenberg H, Barker K, Hanks S, Kalb R, Neveling K, Kelly P, Seal S, Freund M, Wurm M, Batish SD, Lach FP, Yetgin S, Neitzel H, Ariffin H, Tischkowitz M, Mathew CG, Auerbach AD, Rahman N. 2007. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat Genet* 39(2):162-4.

- Riballo E, Woodbine L, Stiff T, Walker SA, Goodarzi AA, Jeggo PA. 2009. XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Res* 37(2):482-92.
- Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, Lyon E, Ward BE, Molecular Subcommittee of the ALQAC. 2008. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med* 10(4):294-300.
- Rickman KA, Lach FP, Abhyankar A, Donovan FX, Sanborn EM, Kennedy JA, Sougnez C, Gabriel SB, Elemento O, Chandrasekharappa SC, Schindler D, Auerbach AD, Smogorzewska A. 2015. Deficiency of UBE2T, the E2 Ubiquitin Ligase Necessary for FANCD2 and FANCI Ubiquitination, Causes FA-T Subtype of Fanconi Anemia. *Cell Rep* 12(1):35-41.
- Rio P, Banos R, Lombardo A, Quintana-Bustamante O, Alvarez L, Garate Z, Genovese P, Almarza E, Valeri A, Diez B, Navarro S, Torres Y, Trujillo JP, Murillas R, Segovia JC, Samper E, Surralles J, Gregory PD, Holmes MC, Naldini L, Bueren JA. 2014. Targeted gene therapy and cell reprogramming in Fanconi anemia. *EMBO Mol Med* 6(6):835-48.
- Rivera B, Castellsague E, Bah I, van Kempen LC, Foulkes WD. 2015. Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. N Engl J Med 373(20):1985-6.
- Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 242(1):84-9.
- Ronaghi M, Uhlen M, Nyren P. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281(5375):363, 365.
- Ronaghi M. 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res 11(1):3-11.
- Rosenberg PS, Tamary H, Alter BP. 2011. How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel. *Am J Med Genet A* 155A(8):1877-83.
- Rosendorff J, Bernstein R, Macdougall L, Jenkins T. 1987. Fanconi anemia: another disease of unusually high prevalence in the Afrikaans population of South Africa. *Am J Med Genet* 27(4):793-7.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primerdirected enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239(4839):487-91.
- Samorodnitsky E, Datta J, Jewell BM, Hagopian R, Miya J, Wing MR, Damodaran S, Lippus JM, Reeser JW, Bhatt D, Timmers CD, Roychowdhury S. 2015. Comparison of custom capture for targeted nextgeneration DNA sequencing. J Mol Diagn 17(1):64-75.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci* USA 74(12):5463-7.
- Sato K, Toda K, Ishiai M, Takata M, Kurumizaka H. 2012. DNA robustly stimulates FANCD2 monoubiquitylation in the complex with FANCI. *Nucleic Acids* Res 40(10):4553-61.
- Sauna ZE, Kimchi-Sarfaty C. 2011. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* 12(10):683-91.
- Savino M, Ianzano L, Strippoli P, Ramenghi U, Arslanian A, Bagnara GP, Joenje H, Zelante L, Savoia A. 1997. Mutations of the Fanconi anemia group A gene (FAA) in Italian patients. *Am J Hum Genet* 61(6):1246-53.
- Savino M, Borriello A, D'Apolito M, Criscuolo M, Del Vecchio M, Bianco AM, Di Perna M, Calzone R, Nobili B, Zatterale A, Zelante L, Joenje H, Della Ragione F, Savoia A. 2003. Spectrum of FANCA mutations in Italian Fanconi anemia patients: identification of six novel alleles and phenotypic characterization of the S858R variant. *Hum Mutat* 22(4):338-9.
- Sawyer SL, Tian L, Kahkonen M, Schwartzentruber J, Kircher M, University of Washington Centre for Mendelian G, Consortium FC, Majewski J, Dyment DA, Innes AM, Boycott KM, Moreau LA, Moilanen JS, Greenberg RA. 2015. Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. *Cancer Discov* 5(2):135-42.
- Scaglione KM, Basrur V, Ashraf NS, Konen JR, Elenitoba-Johnson KS, Todi SV, Paulson HL. 2013. The ubiquitin-conjugating enzyme (E2) Ube2w ubiquitinates the N terminus of substrates. J Biol Chem 288(26):18784-8.

- Scheckenbach K, Morgan M, Filger-Brillinger J, Sandmann M, Strimling B, Scheurlen W, Schindler D, Gobel U, Hanenberg H. 2012. Treatment of the bone marrow failure in Fanconi anemia patients with danazol. *Blood Cells Mol Dis* 48(2):128-31.
- Schindler D, Kubbies M, Hoehn H, Schinzel A, Rabinovitch PS. 1987. Confirmation of Fanconi's anemia and detection of a chromosomal aberration (1Q12-32 triplication) via BrdU/Hoechst flow cytometry. Am J Pediatr Hematol Oncol 9(2):172-7.
- Schindler D, Endt D, Neveling K. 2011. Fanconi Anemia. In: Schwab M, editor. Encyclopedia of Cancer. Berlin Heidelberg, Germany: Springer-Verlag. p 1374-1379.
- Schlacher K, Wu H, Jasin M. 2012. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* 22(1):106-16.
- Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 30(12):e57.
- Schroeder TM, Anschutz F, Knopp A. 1964. Spontaneous chromosome aberrations in familial panmyelopathy. *Humangenetik* 1(2):194-6.
- Schuster B, Knies K, Stoepker C, Velleuer E, Friedl R, Gottwald-Muhlhauser B, de Winter JP, Schindler D. 2013. Whole exome sequencing reveals uncommon mutations in the recently identified Fanconi anemia gene SLX4/FANCP. *Hum Mutat* 34(1):93-6.
- Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deepsequencing age. *Nat Methods* 11(4):361-2.
- Schwertman P, Bekker-Jensen S, Mailand N. 2016. Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nat Rev Mol Cell Biol* 17(6):379-94.
- Seal S, Barfoot R, Jayatilake H, Smith P, Renwick A, Bascombe L, McGuffog L, Evans DG, Eccles D, Easton DF, Stratton MR, Rahman N, Breast Cancer Susceptibility C. 2003. Evaluation of Fanconi Anemia genes in familial breast cancer predisposition. *Cancer Res* 63(24):8596-9.
- Semlow DR, Zhang J, Budzowska M, Drohat AC, Walter JC. 2016. Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase. *Cell* 167(2):498-511 e14.
- Seyschab H, Sun Y, Friedl R, Schindler D, Hoehn H. 1993. G2 phase cell cycle disturbance as a manifestation of genetic cell damage. *Hum Genet* 92(1):61-8.
- Seyschab H, Friedl R, Sun Y, Schindler D, Hoehn H, Hentze S, Schroeder-Kurth T. 1995. Comparative evaluation of diepoxybutane sensitivity and cell cycle blockage in the diagnosis of Fanconi anemia. *Blood* 85(8):2233-7.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids* Res 29(1):308-11.
- Shimamura A, Alter BP. 2010. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood Rev* 24(3):101-22.
- Sijbers AM, de Laat WL, Ariza RR, Biggerstaff M, Wei YF, Moggs JG, Carter KC, Shell BK, Evans E, de Jong MC, Rademakers S, de Rooij J, Jaspers NG, Hoeijmakers JH, Wood RD. 1996. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* 86(5):811-22.
- Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, van den Berg MP, van Spaendonck-Zwarts KY, van Tintelen JP, Sijmons RH, Jongbloed JD, Sinke RJ. 2013. Targeted nextgeneration sequencing can replace Sanger sequencing in clinical diagnostics. *Hum Mutat* 34(7):1035-42.
- Sims AE, Spiteri E, Sims RJ, 3rd, Arita AG, Lach FP, Landers T, Wurm M, Freund M, Neveling K, Hanenberg H, Auerbach AD, Huang TT. 2007. FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. *Nat Struct Mol Biol* 14(6):564-7.
- Singh TR, Bakker ST, Agarwal S, Jansen M, Grassman E, Godthelp BC, Ali AM, Du CH, Rooimans MA, Fan Q, Wahengbam K, Steltenpool J, Andreassen PR, Williams DA, Joenje H, de Winter JP, Meetei AR. 2009. Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. *Blood* 114(1):174-80.

- Singh TR, Saro D, Ali AM, Zheng XF, Du CH, Killen MW, Sachpatzidis A, Wahengbam K, Pierce AJ, Xiong Y, Sung P, Meetei AR. 2010. MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. *Mol Cell* 37(6):879-86.
- Singh TR, Ali AM, Paramasivam M, Pradhan A, Wahengbam K, Seidman MM, Meetei AR. 2013. ATRdependent phosphorylation of FANCM at serine 1045 is essential for FANCM functions. *Cancer Res* 73(14):4300-10.
- Skibbens RV. 2004. Chl1p, a DNA helicase-like protein in budding yeast, functions in sister-chromatid cohesion. *Genetics* 166(1):33-42.
- Sleeth KM, Sorensen CS, Issaeva N, Dziegielewski J, Bartek J, Helleday T. 2007. RPA mediates recombination repair during replication stress and is displaced from DNA by checkpoint signalling in human cells. J Mol Biol 373(1):38-47.
- Smogorzewska A, Matsuoka S, Vinciguerra P, McDonald ER, 3rd, Hurov KE, Luo J, Ballif BA, Gygi SP, Hofmann K, D'Andrea AD, Elledge SJ. 2007. Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* 129(2):289-301.
- Soulier J, Leblanc T, Larghero J, Dastot H, Shimamura A, Guardiola P, Esperou H, Ferry C, Jubert C, Feugeas JP, Henri A, Toubert A, Socie G, Baruchel A, Sigaux F, D'Andrea AD, Gluckman E. 2005. Detection of somatic mosaicism and classification of Fanconi anemia patients by analysis of the FA/BRCA pathway. *Blood* 105(3):1329-36.
- Stewart GS, Maser RS, Stankovic T, Bressan DA, Kaplan MI, Jaspers NG, Raams A, Byrd PJ, Petrini JH, Taylor AM. 1999. The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. *Cell* 99(6):577-87.
- Stewart GS, Wang B, Bignell CR, Taylor AM, Elledge SJ. 2003. MDC1 is a mediator of the mammalian DNA damage checkpoint. *Nature* 421(6926):961-6.
- Stewart MD, Ritterhoff T, Klevit RE, Brzovic PS. 2016. E2 enzymes: more than just middle men. *Cell Res* 26(4):423-40.
- Stoepker C, Hain K, Schuster B, Hilhorst-Hofstee Y, Rooimans MA, Steltenpool J, Oostra AB, Eirich K, Korthof ET, Nieuwint AW, Jaspers NG, Bettecken T, Joenje H, Schindler D, Rouse J, de Winter JP. 2011. SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. Nat Genet 43(2):138-41.
- Strathdee CA, Gavish H, Shannon WR, Buchwald M. 1992a. Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* 358(6385):434.
- Strathdee CA, Duncan AM, Buchwald M. 1992b. Evidence for at least four Fanconi anaemia genes including FACC on chromosome 9. *Nat Genet* 1(3):196-8.
- Strom SP, Lee H, Das K, Vilain E, Nelson SF, Grody WW, Deignan JL. 2014. Assessing the necessity of confirmatory testing for exome-sequencing results in a clinical molecular diagnostic laboratory. *Genet Med* 16(7):510-5.
- Sugasawa K, Ng JM, Masutani C, Iwai S, van der Spek PJ, Eker AP, Hanaoka F, Bootsma D, Hoeijmakers JH. 1998. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell* 2(2):223-32.
- Sugasawa K, Okamoto T, Shimizu Y, Masutani C, Iwai S, Hanaoka F. 2001. A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev* 15(5):507-21.
- Sugasawa K, Okuda Y, Saijo M, Nishi R, Matsuda N, Chu G, Mori T, Iwai S, Tanaka K, Tanaka K, Hanaoka F. 2005. UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* 121(3):387-400.
- Swuec P, Renault L, Borg A, Shah F, Murphy VJ, van Twest S, Snijders AP, Deans AJ, Costa A. 2017. The FA Core Complex Contains a Homo-dimeric Catalytic Module for the Symmetric Mono-ubiquitination of FANCI-FANCD2. *Cell Rep* 18(3):611-623.
- Syed F, Grunenwald H, Caruccio N. 2009. Optimized library preparation method for next-generation sequencing. *Nature Methods* 6(10):I-II.
- Taniguchi T, Garcia-Higuera I, Xu B, Andreassen PR, Gregory RC, Kim ST, Lane WS, Kastan MB, D'Andrea AD. 2002. Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. *Cell* 109(4):459-72.

- Taniguchi T, D'Andrea AD. 2002. The Fanconi anemia protein, FANCE, promotes the nuclear accumulation of FANCC. *Blood* 100(7):2457-62.
- Thomashevski A, High AA, Drozd M, Shabanowitz J, Hunt DF, Grant PA, Kupfer GM. 2004. The Fanconi anemia core complex forms four complexes of different sizes in different subcellular compartments. *J Biol Chem* 279(25):26201-9.
- Timmers C, Taniguchi T, Hejna J, Reifsteck C, Lucas L, Bruun D, Thayer M, Cox B, Olson S, D'Andrea AD, Moses R, Grompe M. 2001. Positional cloning of a novel Fanconi anemia gene, FANCD2. *Mol Cell* 7(2):241-8.
- Tkac J, Xu G, Adhikary H, Young JT, Gallo D, Escribano-Diaz C, Krietsch J, Orthwein A, Munro M, Sol W, Al-Hakim A, Lin ZY, Jonkers J, Borst P, Brown GW, Gingras AC, Rottenberg S, Masson JY, Durocher D. 2016. HELB Is a Feedback Inhibitor of DNA End Resection. *Mol Cell* 61(3):405-18.
- Tomida J, Itaya A, Shigechi T, Unno J, Uchida E, Ikura M, Masuda Y, Matsuda S, Adachi J, Kobayashi M, Meetei AR, Maehara Y, Yamamoto K, Kamiya K, Matsuura A, Matsuda T, Ikura T, Ishiai M, Takata M. 2013. A novel interplay between the Fanconi anemia core complex and ATR-ATRIP kinase during DNA cross-link repair. *Nucleic Acids Res* 41(14):6930-41.
- Tubbs A, Nussenzweig A. 2017. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* 168(4):644-656.
- van der Lelij P, Oostra AB, Rooimans MA, Joenje H, de Winter JP. 2010. Diagnostic Overlap between Fanconi Anemia and the Cohesinopathies: Roberts Syndrome and Warsaw Breakage Syndrome. *Anemia* 2010:565268.
- van Twest S, Murphy VJ, Hodson C, Tan W, Swuec P, O'Rourke JJ, Heierhorst J, Crismani W, Deans AJ. 2017. Mechanism of Ubiquitination and Deubiquitination in the Fanconi Anemia Pathway. *Mol Cell* 65(2):247-259.
- Varon R, Vissinga C, Platzer M, Cerosaletti KM, Chrzanowska KH, Saar K, Beckmann G, Seemanova E, Cooper PR, Nowak NJ, Stumm M, Weemaes CM, Gatti RA, Wilson RK, Digweed M, Rosenthal A, Sperling K, Concannon P, Reis A. 1998. Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* 93(3):467-76.
- Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, Erven V, Neveling K, Endt D, Kesterton I, Autore F, Fraternali F, Freund M, Hartmann L, Grimwade D, Roberts RG, Schaal H, Mohammed S, Rahman N, Schindler D, Mathew CG. 2010. Mutation of the RAD51C gene in a Fanconi anemia-like disorder. *Nat Genet* 42(5):406-9.
- Veal CD, Freeman PJ, Jacobs K, Lancaster O, Jamain S, Leboyer M, Albanes D, Vaghela RR, Gut I, Chanock SJ, Brookes AJ. 2012. A mechanistic basis for amplification differences between samples and between genome regions. BMC Genomics 13:455.
- Verlander PC, Lin JD, Udono MU, Zhang Q, Gibson RA, Mathew CG, Auerbach AD. 1994. Mutation Analysis of the Fanconi-Anemia Gene Face. *American Journal of Human Genetics* 54(4):595-601.
- Verlander PC, Kaporis A, Liu Q, Zhang Q, Seligsohn U, Auerbach AD. 1995. Carrier frequency of the IVS4 + 4 A-->T mutation of the Fanconi anemia gene FAC in the Ashkenazi Jewish population. *Blood* 86(11):4034-8.
- Vetro A, Iascone M, Limongelli I, Ameziane N, Gana S, Della Mina E, Giussani U, Ciccone R, Forlino A, Pezzoli L, Rooimans MA, van Essen AJ, Messa J, Rizzuti T, Bianchi P, Dorsman J, de Winter JP, Lalatta F, Zuffardi O. 2015. Loss-of-Function FANCL Mutations Associate with Severe Fanconi Anemia Overlapping the VACTERL Association. *Hum Mutat* 36(5):562-8.
- Virts EL, Jankowska A, Mackay C, Glaas MF, Wiek C, Kelich SL, Lottmann N, Kennedy FM, Marchal C, Lehnert E, Scharf RE, Dufour C, Lanciotti M, Farruggia P, Santoro A, Savasan S, Scheckenbach K, Schipper J, Wagenmann M, Lewis T, Leffak M, Farlow JL, Foroud TM, Honisch E, Niederacher D, Chakraborty SC, Vance GH, Pruss D, Timms KM, Lanchbury JS, Alpi AF, Hanenberg H. 2015. AluY-mediated germline deletion, duplication and somatic stem cell reversion in UBE2T defines a new subtype of Fanconi anemia. *Hum Mol Genet* 24(18):5093-108.
- Volker M, Mone MJ, Karmakar P, van Hoffen A, Schul W, Vermeulen W, Hoeijmakers JH, van Driel R, van Zeeland AA, Mullenders LH. 2001. Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell* 8(1):213-24.

- Wagner JE, Tolar J, Levran O, Scholl T, Deffenbaugh A, Satagopan J, Ben-Porat L, Mah K, Batish SD, Kutler DI, MacMillan ML, Hanenberg H, Auerbach AD. 2004. Germline mutations in BRCA2: shared genetic susceptibility to breast cancer, early onset leukemia, and Fanconi anemia. *Blood* 103(8):3226-9.
- Waisfisz Q, Morgan NV, Savino M, de Winter JP, van Berkel CG, Hoatlin ME, Ianzano L, Gibson RA, Arwert F, Savoia A, Mathew CG, Pronk JC, Joenje H. 1999. Spontaneous functional correction of homozygous fanconi anaemia alleles reveals novel mechanistic basis for reverse mosaicism. Nat Genet 22(4):379-83.
- Waltes R, Kalb R, Gatei M, Kijas AW, Stumm M, Sobeck A, Wieland B, Varon R, Lerenthal Y, Lavin MF, Schindler D, Dork T. 2009. Human RAD50 deficiency in a Nijmegen breakage syndrome-like disorder. Am J Hum Genet 84(5):605-16.
- Walther AP, Gomes XV, Lao Y, Lee CG, Wold MS. 1999. Replication protein A interactions with DNA. 1. Functions of the DNA-binding and zinc-finger domains of the 70-kDa subunit. *Biochemistry* 38(13):3963-73.
- Wang AT, Kim T, Wagner JE, Conti BA, Lach FP, Huang AL, Molina H, Sanborn EM, Zierhut H, Cornes BK, Abhyankar A, Sougnez C, Gabriel SB, Auerbach AD, Kowalczykowski SC, Smogorzewska A. 2015. A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. *Mol Cell* 59(3):478-90.
- Wang H, Zhu X. 2014. De novo mutations discovered in 8 Mexican American families through whole genome sequencing. BMC Proc 8(Suppl 1):S24.
- Wang W, Malcolm BA. 2002. Two-stage polymerase chain reaction protocol allowing introduction of multiple mutations, deletions, and insertions, using QuikChange site-directed mutagenesis. *Methods Mol Biol* 182:37-43.
- Wang X, Kennedy RD, Ray K, Stuckert P, Ellenberger T, D'Andrea AD. 2007. Chk1-mediated phosphorylation of FANCE is required for the Fanconi anemia/BRCA pathway. *Mol Cell Biol* 27(8):3098-108.
- Ward JD, Barber LJ, Petalcorin MI, Yanowitz J, Boulton SJ. 2007. Replication blocking lesions present a unique substrate for homologous recombination. EMBO J 26(14):3384-96.
- Weren RD, Ligtenberg MJ, Kets CM, de Voer RM, Verwiel ET, Spruijt L, van Zelst-Stams WA, Jongmans MC, Gilissen C, Hehir-Kwa JY, Hoischen A, Shendure J, Boyle EA, Kamping EJ, Nagtegaal ID, Tops BB, Nagengast FM, Geurts van Kessel A, van Krieken JH, Kuiper RP, Hoogerbrugge N. 2015. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nat Genet* 47(6):668-71.
- Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452(7189):872-6.
- Whitney M, Thayer M, Reifsteck C, Olson S, Smith L, Jakobs PM, Leach R, Naylor S, Joenje H, Grompe M. 1995. Microcell mediated chromosome transfer maps the Fanconi anaemia group D gene to chromosome 3p. Nat Genet 11(3):341-3.
- Wijker M, Morgan NV, Herterich S, van Berkel CG, Tipping AJ, Gross HJ, Gille JJ, Pals G, Savino M, Altay C, Mohan S, Dokal I, Cavenagh J, Marsh J, van Weel M, Ortega JJ, Schuler D, Samochatova E, Karwacki M, Bekassy AN, Abecasis M, Ebell W, Kwee ML, de Ravel T, Mathew CG, et al. 1999. Heterogeneous spectrum of mutations in the Fanconi anaemia group A gene. *Eur J Hum Genet* 7(1):52-9.
- Wilson TE, Grawunder U, Lieber MR. 1997. Yeast DNA ligase IV mediates non-homologous DNA end joining. *Nature* 388(6641):495-8.
- Wong MW, Nordfors C, Mossman D, Pecenpetelovska G, Avery-Kiejda KA, Talseth-Palmer B, Bowden NA, Scott RJ. 2011. BRIP1, PALB2, and RAD51C mutation analysis reveals their relative importance as genetic susceptibility factors for breast cancer. *Breast Cancer Res Treat* 127(3):853-9.
- Wyatt DW, Feng W, Conlin MP, Yousefzadeh MJ, Roberts SA, Mieczkowski P, Wood RD, Gupta GP, Ramsden DA. 2016. Essential Roles for Polymerase theta-Mediated End Joining in the Repair of Chromosome Breaks. *Mol Cell* 63(4):662-73.

- Xia B, Dorsman JC, Ameziane N, de Vries Y, Rooimans MA, Sheng Q, Pals G, Errami A, Gluckman E, Llera J, Wang W, Livingston DM, Joenje H, de Winter JP. 2007. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. *Nat Genet* 39(2):159-61.
- Yamashita T, Barber DL, Zhu Y, Wu N, D'Andrea AD. 1994. The Fanconi anemia polypeptide FACC is localized to the cytoplasm. *Proc Natl Acad Sci U S A* 91(14):6712-6.
- Yan Z, Delannoy M, Ling C, Daee D, Osman F, Muniandy PA, Shen X, Oostra AB, Du H, Steltenpool J, Lin T, Schuster B, Decaillet C, Stasiak A, Stasiak AZ, Stone S, Hoatlin ME, Schindler D, Woodcock CL, Joenje H, Sen R, de Winter JP, Li L, Seidman MM, Whitby MC, Myung K, Constantinou A, Wang W. 2010. A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Mol Cell* 37(6):865-78.
- Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13:134.
- You Z, Bailis JM. 2010. DNA damage and decisions: CtIP coordinates DNA repair and cell cycle checkpoints. *Trends Cell Biol* 20(7):402-9.
- Youssoufian H. 1994. Localization of Fanconi anemia C protein to the cytoplasm of mammalian cells. *Proc* Natl Acad Sci U S A 91(17):7975-9.
- Zakrzewski S, Sperling K. 1980. Genetic heterogeneity of Fanconi's anemia demonstrated by somatic cell hybrids. *Hum Genet* 56(1):81-4.
- Zecca M, Strocchio L, Pagliara D, Comoli P, Bertaina A, Giorgiani G, Perotti C, Corbella F, Brescia L, Locatelli F. 2014. HLA-haploidentical T cell-depleted allogeneic hematopoietic stem cell transplantation in children with Fanconi anemia. *Biol Blood Marrow Transplant* 20(4):571-6.
- Zhang F, Ma J, Wu J, Ye L, Cai H, Xia B, Yu X. 2009. PALB2 links BRCA1 and BRCA2 in the DNA-damage response. *Curr Biol* 19(6):524-9.
- Zhang M, Windheim M, Roe SM, Peggie M, Cohen P, Prodromou C, Pearl LH. 2005. Chaperoned ubiquitylation--crystal structures of the CHIP U box E3 ubiquitin ligase and a CHIP-Ubc13-Uev1a complex. *Mol Cell* 20(4):525-38.
- Zhang Y, Zhou X, Zhao L, Li C, Zhu H, Xu L, Shan L, Liao X, Guo Z, Huang P. 2011. UBE2W interacts with FANCL and regulates the monoubiquitination of Fanconi anemia protein FANCD2. *Mol Cells* 31(2):113-22.
- Zhang Z, Carriero N, Zheng D, Karro J, Harrison PM, Gerstein M. 2006. PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics* 22(12):1437-9.

## Abkürzungsverzeichnis

aa	Amino acid
ACMG	American College of Medical Genetics and Genomics
alt-EJ	Alternatives end joining
AML	Akute myeloische Leukämie
AP	Apurin/ Apyrimidin
AT	Ataxia telangiectasia
ATLD	AT-like disease
ATP	Adenosintriphosphat
BER	Basenexzisionsreparatur
bp	Basenpaar
CCD	Charge-coupled device
cDNA	Complementary DNA
CDK	Cyclin-dependent kinase
CGH	Comparative genomic hybridization
CHX	Cycloheximid
CNV	Copy number variation
Co-IP	Co-Immunpräzipitation
CPD	Cyclobutan-Pyrimidin-Dimere
CS	Cockayne-Syndrom
DAPI	4',6-Diamidin-2-phenylindol
DBD	DNA-Bindedomäne
ddH2O	Demineralisiertes Wasser
ddNTP	2'-3'-Didesoxyribonukleosid-5'-Triphosphat
DEB	Diepoxybutan
D-Loop	Displacement loop
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxid
dNTP	2'-Desoxyribonukleosid-5'-Triphosphat
DSB	Doppelstrangbruch
dsDNA	Doppelsträngige DNA
EBV	Epstein-Barr-Virus
EDTA	Ethylendiamintetraacetat

EdU	5-Ethynyl-2'-desoxyuridin
emPCR	Emulsions-PCR
ESE	Exonic splicing enhancer
EU	5-Ethynyl-uridin
ExAC	Exome Aggregation Consortium
FA	Fanconi-Anämie
FAAP	Fanconi-Anämie-assoziiertes Protein
FCS	Fetal calf serum
FU	Fluorescence unit
g	Erdbeschleunigung; Gramm
Gb	Gigabasen
G-CSF	Granulocyte-colony stimulating factor
gDNA	Genomische DNA
GG-NER	Globale Genom-NER
GvHD	Graft-versus-Host-Disease
h	Stunde
HeBS	HEPES Buffered Saline
HEPES	Hydroxyethylpiperazin-Ethansulfonsäure-Puffer
HJ	Holliday Junction
HNPCC	Hereditary non-polyposis colon cancer
HNSCCs	Head and neck squamous cell carcinomas
HR	Homologe Rekombination
HRP	Horseradish peroxidase
HSC	Hematopoetic stem cell
HSZT	Hämatopoetische Stammzelltransplantation
HU	Hydroxyurea
ICL	Interstrand crosslink
IFAR	Internationales Fanconi-Anämie-Register
IMDM	Iscove's Modified Dulbecco's Medium
IVS	Intervening sequence
kb	Kilobasen
kDa	Kilodalton
LCL	Lymphoblastoid cell line
LM-PCR	Ligation-mediated PCR
MAF	Minor allele frequency
Mb	Megabasen

MDS	Myelodisplastisches Syndrom
MID	Multiplex Identifier
min	Minuten
MLPA	Multiplex Ligation-dependent Probe Amplification
MMC	Mitomycin C
MMEJ	Microhomology-mediated end-joining
MMR	Mismatch-Reparatur
NBS	Nijmegen-Breakage-Syndrom
NER	Nukleotidexzisionsreparatur
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NMD	Nonsense-mediated mRNA decay
PBS	Phosphate-Buffered Saline
PCR	Polymerase chain reaction
PKcs	Protein kinase catalytic subunit
РТР	Pikotiterplatte
RFU	Relative fluorescence unit
ROI	Region of interest
ROS	Reaktive Sauerstoffspezies
RRS	Recovery of RNA synthesis
RT	Raumtemperatur
SDS	Natriumdodecylsulfat
SDSA	Synthesis-dependent strand-annealing
sec	Sekunden
SNP	Single nucleotide polymorphism
SOLiD	Sequencing by Oligonucleotide ligation and detection
SPRI	Solid phase reverse immobilization
SSA	Single strand DNA annealing
ssDNA	Single stranded DNA
TC-NER	Transkriptions-gekoppelte NER
TLS	Transläsionssynthese
UDS	Unscheduled DNA synthesis
VUS	Variante unbekannter Signifikanz
WES	Whole exome sequencing
WGS	Whole genome sequencing
ХР	Xeroderma pigemtosum

## Anhang

### Primersequenzen

## **Mutagenese-Primer**

Bezeichnung	Sequenz (5'→3')
XPF_c.1765C_T_F_1	CGCAGAGCTAACCTITGTTTGGCAGCTTGAAATTTACAG
XPF_c.1765C_T_R_1	CTGTAAATTTCAAGCTGCCAAACAAAGGTTAGCTCTGCG
XPF_c.1765C_T_F_2	CAGAGCTAACCTTTGTTTGGCAGCTTGAAATTTAC
XPF_c.1765C_T_R_2	GTAAATTTCAAGCTGCCAAACAAAGGTTAGCTCTG

## InFusion-Primer

Bezeichnung	Sequenz (5'→3')
pLXV_InF2_F	TATTTCCGGTGAATTATGGAGTCAGGGCAGCCG
pLXV_InF2_R	GCGGGATCCGCGGCCTCACTTTTTCCCTTTTCCTTT

#### Primer zur Bestätigung von Mutationen durch Sanger-Sequenzierung

Im Folgenden sind die Primer aufgeführt, welche zur Bestätigung der Mutationen verwendet wurden, die zuvor per NGS detektiert wurden. Zudem sind auch alle Primer enthalten, die zur Mutationsanalyse von FA-I- sowie FA-D2-Patienten während der Routinediagnostik verwendet wurden. Die gelisteten *FANCD2*-Primer sind dabei eine Ergänzung der bei [Kalb et al., 2007] angegebenen Primersequenzen. gDNA- sowie cDNA-Primerpaare sind jeweils separat aufgelistet.

## gDNA-Primer

<b>E</b>		Forward Reve		Revers
EX011	Bezeichnung	Sequenz (5'→3')	Bezeichnung	Sequenz (5'→3')
2	FANCA_Ex2A_for	CCCGTGTGTGAATTGTGCTGT	FANCA_Ex2A_rev	GCCGCCTCGGGTGTTTTCTTAGGAAA
9	FANCA_Ex9_for	TTCTCTTGTGTGATGCAGGT	FANCA_Ex10_rev	TCCAGAGGCGGAACAGGATA
12	FANCA_Ex12_for	CCCACAACTTTTTGATCTCTGAC	FANCA_Ex13_rev	CCTCTCAGAGCAGCAGTCAGTG
14	FANCA_Ex14_F	CTTGATCACCCAGCATAACTC	FANCA_Ex14_F	GTTGCTCACTCACATGACAGA
15	FANCA_Ex15_for	AGGAGGCCGACTACAGC	FANCA_Ex15_rev	CTTGGGGAGGCCAAGGCAGTC
16-17	FANCA_Ex16_for	CAGCACTGTGGATGTTGGAAG	FANCA_Ex17_rev	GGGAAGGCTGAAAAACTCAACTC
24-26	FANCA_Ex24_for	CGATGTCTGTGGCGGGAGCG	FANCA_Ex26_rev	GACAGATAAAATTCTGGAAGG
25	FANCA_Ex25_for	TAGGACGCAGCAGAAGTTGG	FANCA_Ex25_rev	AACGAGAATGAGGGTGGCAG
27	FANCA_Ex27_for	ATGCTCAGGCCATCCAGTTC	FANCA_Ex27_rev	TCCGAAAGCTGCGTAAACCT
28	FANCA_Ex28_for	GTGTGGGCTGTTGATGGTCTG	FANCA_Ex28_rev	CTGTTCTTGCCCGAGGAGCAC
29	FANCA_Ex29_for	ACACAGGGACATGGAGGACT	FANCA_Ex29_rev	GGCAATGTGGAAAACCGTGT
33	FANCA_Ex33_for	CTCAGAGTGCAGGCTTTGGA	FANCA_Ex33_rev	AGATCACCCTGCAACTCACG
34-35	FANCA_Ex34_for	GCATGTTGGTCTCGCCTCTCTT	FANCA_Ex35_rev	TTTCCCTGAGATGGTAACACC
36	FANCA_Ex36_for	GTCATGGCTGGGGCAGCGGAG	FANCA_Ex36_rev	TCTTTTCTCCCTGCTCACACGA
37	FANCA_Ex37_for	GGTGACAGGTGGGAATAAGGAC	FANCA_Ex37_rev	CTTGCTCCAAGCCACATATTTG
38	FANCA_Ex38_for	GACCGCAGCAGTCTTCTCTT	FANCA_Ex38_rev	CTTCCGCAAACACAAGGAGC
39-40	FANCA_Ex39_40_for	CTGAGGGCTGTACCCTCCTA	FANCA_Ex39_40_rev	CAGAGGCACAGACAACCCTT
41-43	FANCA_Ex41_43_for	AAGGGTTGTCTGTGCCTCTG	FANCA_Ex41_43_rev	CCCCACTAAAGCAGTCGAGG
43	FANCA_FA-13_Ex43_for	ACCTGTGAGTITGCCTGTIT	FANCA_FA-13_Ex43_rev	TATTTCCAGCAGCTGTGAG
3	FANCB_Ex3B_for	TGGAATGACTAGCAAACAAGC	FANCB_Ex3B_rev	TAAGGACCCTTAGGCCATCC
3	FANCB_Ex3C_for	AGAACTGGAATTAACCTCCCTTACATTG	FANCB_Ex3C_rev	ACATAAGTCACCACACTGCTGTAAGCA
2	FANCC_Ex2_for	AATTAGCGTGTGCCTGTGGA	FANCC_Ex2_rev	CCGATTCTGGGTGGCATTCT
5	FANCC_Ex5_for	TCATAGAGACCACCCCCATGT	FANCC_Ex5_rev	GCCATAAGTCTGCCCAAGGT
6	FANCC_Ex6_for	CTGATGTAATCCTGTTTGCAG	FANCC_Ex6_rev	CCTCTCATAACCAAACTGATAC
7	FANCC_Ex7_for	GTCCTTAATTATGCATGGCTC	FANCC_Ex7_rev	CAACACACCACAGCCTTCTAAG
15	FANCC_Ex15_for	GCGCTTGTTTGTGGGCCTTTA	FANCC_Ex15_rev	ACAGCTCATTCTCACAGCCC
2	FANCD2_Ex2_for	TCCTTCAGCAACAGCGAAGT	FANCD2_Ex2_rev	TGGAGACTGAGGTCCAGAGG
3	FANCD2_Ex3_for	AGCCACTATTGCTGAAGACAGT	FANCD2_Ex3_rev	TGTCTCAAGATGGATGGCCC

4-5	FANCD2_Ex4_for	TCATCAGGCAAGAAACTTGGGT	FANCD2_Ex5_rev	TGGCTCTTTGCAAGGGAGAG
6	FANCD2_Ex6_for	TGCAAAGAGCCATCTGCTCA	FANCD2_Ex6_rev	GCCGCCAGATGTATTAGGCT
7	FANCD2_Ex7_for	TCCAGGAGCTCTAGTTGGAA	FANCD2_Ex7_rev	CCAACCAACTGGCTACCCTC
9	FANCD2_Ex8_for	TGTGTCTAGTGCAGTGCCGA	FANCD2_Ex9_rev	GCCAGACTTCACGCTCATAAC
10	FANCD2_Ex10_for	GCTCCTGTAGCACCACCTAC	FANCD2_Ex10_rev	ACTCCCAAGGCAATGACTGA
11-12	FANCD2_Ex11_for	AGTGGGAAGATGGAGTAAGAGA	FANCD2_Ex12_rev	GATCGGAGTTCCTGCCATGT
13-14	FANCD2_Ex13_for	CGGTTACAGGCTTCCCAAGT	FANCD2_Ex14_rev	AGAAGAGTGACTGGAGGGGG
15-16	FANCD2_Ex15_for	TGGCAACAAATACACGCTGG	FANCD2_Ex16_rev	AGACCCAGGTCAGAGTTCCT
17	FANCD2_Ex17_for	GGAGCAAAATCAGGAGGGCT	FANCD2_Ex17_rev	CTTGGGCACCAGTTTGCTTC
19	FANCD2_Ex19_for	CTTCTCCATTTCGGGCCAGG	FANCD2_Ex19_rev	GGCTAGGATCACATTCCAGCA
20	FANCD2_Ex20_for	TCCTCTAATTATTGCTAATTTGCCC	FANCD2_Ex20_rev	AGCACTITCTGGTGAACTGACA
21-25	FANCD2_Super21_25b_fo r	TGTCAGTTCACCAGAAAGTGCT	FANCD2_Super21_25b_re v	TGGGAGGGAAACTACCACCA
2	FANCE_Ex2b_for	CAGGAAACCACTGTTGCTGC	FANCE_Ex2b_rev	TAACTGCCATGACAGGCTGG
2	FANCE_Ex2c_for	TTGTGCTCCCTTCAGCCAAA	FANCE_Ex2c_rev	CTTTGGAGCTGTCTCTGGCA
1	FANCF_Ex1_1_for	AGGCGTATCATTTCGCGGAT	FANCF_Ex1_1_rev	CGCCTGGGTCTTCATCAGAG
1	FANCF_Ex1_2_for	ACCTGGTGCAGCAACTCTTT	FANCF_Ex1_2_rev	ACTITGGACACACGAAGGCA
4	FANCG_Ex4_for	GTGTGGGGAGATGGAGGATGAGG	FANCG_Ex4_rev	GGAGGAGGAAGGAAGGAGGAGGAGACC
7	FANCG_Ex7_for	TAAGGGTAGCTGAGGCCCGTGTCC	FANCG_Ex7_rev	TCAAGTCACCCCATCACAAGCACC
10	FANCG_Ex10A_for	AGAGGTTGGGGATGGTGGCTCATG	FANCG_Ex10A_rev	AGCCGGGACATCTTGGGTAGCAGA
13	FANCG_Ex13_for	CCCCGCTTCCATATGTGAGTGTA	FANCG_Ex13_rev	CTCTAGGACACCAACTGCCCCTCA
14	FANCG_Ex14_for	TGAACACTGAAAGCGGGGGGTATGG	FANCG_Ex14_rev	CACAGGCCTACCACCAATCTCACC
10	FANCI_Ex10_for	TGCTCATCGTAGCACCTGTC	FANCI_Ex10_rev	GGCCATAGCTGAAGGTCACA
19	FANCI_Ex19_for	AGACTAGGACATTTTGGGTGGG	FANCI_Ex19_rev	TGGCAGACAGGTAAGTGGTG
20	FANCI_Ex20_for	AAGAGCACAGGATGTGGGTG	FANCI_Ex20_rev	AGCAGGTCCCCTAGGTTCAT
23-24	FANCI_Ex23_for	GATTGCTGTGACCTGGGAGT	FANCI_Ex24_rev	CCTTCATCCTCTGTTCCCAGT
27	FANCI_Ex27_for	TGGTGCCTTGTCTAAGCCTG	FANCI_Ex27_rev	TTCTTGCCCTGTATGGCTGC
28	FANCI_Ex28_for	CTGGGGGGCATGGGAAAGAAA	FANCI_Ex28_rev	TTTTCCAAATGCCACTGGAGG
31-32	FANCI_Ex31+32_for	GTGTCCTGAGGCCTGTAACC	FANCI_Ex31+32_rev	ACCATGACCTGCCAAGATGG
37	FANCI_Ex37_for	GGCAAACTGAAGCAGCACAA	FANCI_Ex37_rev	TGACTGCTCCCAGATAGGCT
38	FANCI_Ex38_for	AGTGGGAAAGTGGGGAAAGC	FANCI_Ex38_rev	TCGGTGTGGACTACAGGACA
1	FANCL_Ex1_for	AGGATCTTCCCGCCAAGG	FANCL_Ex1_rev	CCTAGCCCGTCACAGACTTC
4	FANCL_Ex4_for	TTCACTTGGGCCTACATTITT	FANCL_Ex4_rev	AAAGAAAACATCAAATGACTGAGA

12	FANCL_Ex12_for	CCCTCCCAAATATGTACAACG	FANCL_Ex12_rev	CAGGAATACTTCCTATGTTGTGTTAG
14	FANCL_Ex14_for	TTCAGAAGTGTTTTCCAAACTGT	FANCL_Ex14_rev	TGAAGATGATACCAAAATTCCTTT
7	FANCM_Ex7_for	TTCCTGGAGAAAAGGAAACCA	FANCM_Ex7_rev	ACCACTAACAATTACCAACATCTG
11	FANCM_Ex11_for	TGGATCGGGGTTTAATTTTATG	FANCM_Ex11_rev	ATTTGCCACATGAATAATCTCG
4	PALB2_Ex4a_for	CAGTGACCTTACTACTCACAGCC	PALB2_Ex4a_rev	CTGCCATCAAGAGTGTCACTGG
4	PALB2_Ex4b_for	GGTAGCAGTGAACTTACTACTCACG	PALB2_Ex4b_rev	AGGAAGTGCCAGGCAAATAGTAATTGT
7	PALB2_Ex7B_for	AAGCTCTTTCTTTTCACCTGC	PALB2_Ex7B_rev	TTGCATGGTCATAGCTCCCAA
8	PALB2_Ex8_for	CATTAGGTAAATGCTCAGTAAGCAC	PALB2_Ex8_rev	CAGTCAGGGATGTCATTCTATTG
5	SLX4_Ex5_for	GACCCACATTTGCTCCAATC	SLX4_Ex5_rev	AGGGCTCTTTTTCCCTCCA
8	SLX4_Ex8_for	ATGTGATGGCTTCTGCAGTG	SLX4_Ex8_rev	AGAGGATTCACTCGCTGTGG
10	SLX4_Ex10_for	GGGTCACTCAGAGGTTGAGG	SLX4_Ex10_rev	GCAGGAAGTGAGGGAGAGTG
4	XPF_Ex4_for	CATAGCTGCTGAAACTCTAGAAAATTGTTGAAA	XPF_Ex4_rev	AGTCAGAGTGATGCTTATATGCCAATCCAC
5	XPF_Ex5_for	TACACAGGAAATAATCCTTTTGAAAGTATG	XPF_Ex5_rev	CAGTATAACATATAGTTGAATATAGCACTT
8	XPF_Ex8b_for	CCCCAAAACAAAGAACGGGC	XPF_Ex8b_rev	AGCAGCACGTAACGAGAACA
11	XPF_Ex11b_for	CTTGTTTCAGGAGATCTCCAGCAATG	XPF_Ex11b_rev	TATGATGTCTGGCAAGGAGCCGCT
1-7	UBE2T_1_7F_Virts	GCTTCTTTCCCGGTGGATTA	UBE2T_1_7R_Virts	CCCAGACACACATTCAGGATAAA
4	UBE2T_Ex4_for	CATCCAGACAAATCCTTCCAGCAG	UBE2T_Ex5_rev	TGGGTTCTGACATGAGCAGCTGAA
1-7	UBE2T_474Fwd_Rickman	GCGTTGCTGCGTTGTGAGG	UBE2T_479Rev_Rickman	TTAACTAAGATGAACCAGGACAAG
3	NTHL1_Ex3_for	GAGGCTGATGCCTCAAGTGT	NTHL1_Ex3_rev	TTGACCCTCACTTCCTGCAC
3	NTHL1_Ex3b_for	ACAAACCAGGGTGTGTCGAG	NTHL1_Ex3b_rev	ACCCATCTGAGAAACTGCGG
11	RPA1_Ex11b_for	TCAGTGCTTGCCGTTCATCA	RPA1_Ex11b_rev	TAGGACAAGATGCCCCCTCC
12	RPA1_Ex12_for	TACGCGGGGGAAATGCTCTTT	RPA1_Ex12_rev	ATTGCCTCCTCCGACTCTCT

## cDNA-Primer

Forward		Revers	
Bezeichnung	Sequenz (5'→3')	Bezeichnung	Sequenz (5'→3')
FANCA_c.788_for	CGCAGGTCACGGTTGATGTA	FANCA_c.1942_rev	CCTCAGCAGAGTTGGGTTCTGC
FANCA_c.1974_for	AGAGCTGAGAGCCTCCATGA	FANCA_c.2490_rev	CAAGGAATCCCTCGTCCTACAG
FANCA_c.3464_for	ATACTGGCCAAGTGCCAGAC	FANCAc.3919_rev	GAAAGAGTGCCAGCCAGGAT
FANCA_c.751 for	TCAGATCTGAGAAGAACTGTG	FANCA_c.962_rev	CTGAAGAACCTCTTCAGAGGA
FANCA_c.1728_for	CAAAGTCCCTGACTCCCGTGT	FANCA_c.2490_rev	CAAGGAATCCCTCGTCCTACAG
FANCA_c.3277_for	AGCAGCTTCCAGGCAGAACAG	FANCA_c.3602_rev	TGTAGCTTCTGCAGTTCCCG

FANCA_c.1017_for	TGTTCAGATGCAGAGAGAGTGG	FANCA_c.1724_rev	CTGAATATGCTGGCCTCCATGAC
FANCA_c.3627_for	TTTCCTCTCCCCTGAGGCTG	FANCA_c.4142_rev	CCAGCTGGAGGTGAAACTGT
FANCA_c.3823_for	TCAAATAGCACCACAGACCT	FANCA_c.4153_rev	TGCCCTGACCCTTGAGCTC
FANCA_c.3989_for	TGCTGCCTITCGCTITTTAC	FANCA_c.4363_rev	AGAGATGAGGCTCCTGGGAC
FANCC_c.391_for	GTTGCTCTTTTCACTCAAGGTCT	FANCC_c.734_rev	CGAAGCCAGAGGCAGACTAC
FANCD2_c46_for	CGACGGCTTCTCGGAAGTAA	FANCD2_c.576_rev	GAGGTCCTTGCCATCCACAA
FANCD2_c.44_for	GCCTGACAGAAGATGCCTCC	FANCD2_c.621_rev	CAGGTTCTCTGGAGCAATACTG
FANCD2_c.240_for	TCAGACCCTGAGGAGACACC	FANCD2_c.760_rev	GTCGGAGGCTTGAAAGGACA
FANCD2_c.700_for	CTACTGATAGAGAATACTTCACTCACTG	FANCD2_c.1183_rev	TCTGAGTATTGGTGCTATAGATGATG
FANCD2_c.1134_for	GGTGTTTGACCTGGTGATGC	FANCD2_c.1806_rev	CAGGTTGGCTCTCTCTTGGG
FANCD2_c.1680_for	AAAGCAGCTCTCTAGCACCG	FANCD2_c.2516_rev	GGAGGGACATAGTCTGGGGT
FANCD2_c.2394_for	TGCCTTCTGCCAGGAAACAT	FANCD2_c.3072_rev	ACACATTGGGGTCAGCAGTT
FANCD2_c.2624_for	ATGGCAGCAAGACATCCTCC	FANCD2_c.3230_rev	CCACTCCAAGCAAAAAGCCC
FANCD2_c.2886_for	CCCTGAGCTGCTTTTCTTGC	FANCD2_c.3515_rev	TTATCCCCACTTGGCCACAC
FANCD2_c.3268_for	GCCCTCCATGTCCTTAGTAGC	FANCD2_c.3674_rev	GTAGGGAATGTGGAGGAAGATG
FANCD2_c.3375_for	TCCCAGTTTCCAGTGTGCTC	FANCD2_c.4076_rev	AGAGGCACATGTTGGGTGAG
FANCD2_c.3571_for	GAGCACACAGAGAGCATTCTGAAG	FANCD2_c.4211_rev	GAATITTGGGACTTAATCTCTTCACCCTG
FANCD2_c.3919_for	AAGCAATGTATGCCGCTCCT	FANCD2_c.*+84_rev	AGGCAAACAGCGGATTTCAA
FANCI_c.618_for	TCTTCAAGAAATACCACCTTTGG	FANCI_c.1014_rev	AAAGCTCTTTACAACCGAAGTCT
FANCI_c.1616_for	GGTTTTTGCTGCTCCTGAAG	FANCI_c.2100_rev	TTCGTAGAATGCCTCTTCCTC
FANCI_c.1985_for	AACCACTGGATTATCTGCTGTG	FANCI_c.2454_rev	GAAAAGAGCAGTGAGAAGACTGG
FANCI_c.2190_for	TCAGAGCACCAGTATTGGCA	FANCI_c.2953_rev	GCAGGAGGGCTTCTTTGCTA
FANCI_c.2358_for	TGAAAAAGCGGGTAAAGCCA	FANCI_c.2730_rev	CTCCAAGCACAGCAGTGAGA
FANCI_c.2404_for	GATAGTCTTTTGTCCATGAAATTTGT	FANCI_c.2982_rev	CTTGGACAAACTGGTAAGAACC
FANCI_c.2711_for	TCTCACTGCTGTGCTTGGAG	FANCI_c.3286_rev	GAACCTTCTCGGCCTGACTC
FANCI_c.3320_for	AGGGACAAGTGAGCCAAGAA	FANCI_c.3769_rev	TGGCAAAGATGAGGTTAGGG
FANCL_c.1_for	ATGGCGGTGACGGAAGCGAGCCTG	FANCL_c.580_rev	CTGACTATAAATGCTTATTAAGGAGC
XPF_c.578_for	GGCCAAGGTTCCATGTAGCA	XPF_c.1212_rev	TGGACCACCAAGAGCTTCAC
UBE2T_c64_for	TGCATCCCAGGCAGCTCTTA	UBE2T_c.569_rev	TCTATGCCTACTAGCTGACTGG
UBE2T_c.118_for	GGTGGAGCCAACACCCTTA	UBE2T_c.523_rev	GTACTCTGGAGTCACCAGCC
UBE2T_c136_for	AGTCAGAGGTCGCGCAGGCGCTG	UBE2T_c.*63_rev	GGTAGGCAACTTAGATCACCTTGGCA

## Vektorkarten



XVII



## Genliste der SeqCap EZ-Anreicherungspanels

Allen vier *SeqCap EZ*-Anreicherungspanel war ein Set von 47 Genen gemeinsam. Zudem enthielten OID38614 und OID41218 weitere acht, OID42419 sowie OID43014 zwölf zusätzliche Gene. Diese Gene sind, mit Angabe des Haupttranskriptes, in der folgenden Tabelle zusammengefasst. Die zusätzlichen acht Gene der ersten beiden Panels sind hellgrau, die zwölf Gene der letzten beiden Panels dunkelgrau hinterlegt.

Gen	Alias	Bezeichnung	Transkript (RefSeq)
ATR	FCTCS, FRP1, MEC1, SCKL, SCKL1	Ataxia telangiectasia and Rad3 relatedserine/threonine kinase	NM_001184
ATRIP	-	ATR interacting protein	NM_130384
BRCA1	BRCAI, BRCC1, BROVCA1, FANCS, IRIS, PNCA4, PPP1R53, PSCP, RNF53	Breast cancer 1, early onset	NM_007294
BRCA2	BRCC2, BROVCA2, FACD, FAD, FAD1, FANCD, FANCD1, GLM3, PNCA2, XRCC11	Breast cancer 2, early onset	NM_000059
BRIP1	FANCJ, BACH1, OF	BRCA1 interacting protein C-terminal helicase 1	NM_032043
CENPS	APITD1, CENP-S, FAAP16, MHF1	Centromere protein S	NM_199294
CENPX	CENP-X, D9, FAAP10, MHF2, STRA13	Centromere protein X	NM_001271006
CTBP1	BARS	C-terminal binding protein 1	NM_001328
CTBP2	-	C-terminal binding protein 2	NM_022802
DCLRE1B	SNM1B, Apollo, SNMIB	DNA cross-link repair 1B	NM_022836
EME1	SLX2A, MMS4L	Essential meiotic endonuclease 1 homolog 1 (S. pombe)	NM_152463
ERCC1	RAD10, COFS4, UV20	ERCC excision repair 1, endonuclease non-catalytic subunit	NM_202001
ERCC4	RAD1, FANCQ, XPF, XFEPS, ERCC11	ERCC excision repair 4, endonuclease catalytic subunit	NM_005236
EWSR1	EWS, EWS-FL11, bK984G1.4	EWS RNA binding protein 1	NM_013986
FAAP100	c17orf70	Fanconi anemia core complex associated protein 100	NM_025161
FAAP20	c1orf86, FP7162	Fanconi anemia core complex associated protein 20	NM_182533
FAAP24	c19orf40	Fanconi anemia core complex associated protein 20	NM_152266
FANCA	FA, FA-H, FA1, FAA, FACA, FAH, FANCH	Fanconi anemia complementation group A	NM_000135
FANCB	FA2, FAAP90, FAAP95, FAB, FACB	Fanconi anemia complementation group B	NM_001018113
FANCC	FA3, FAC, FACC	Fanconi anemia complementation group C	NM_000136
FANCD2	FA-D2, FA4, FACD, FAD, FAD2, FANCD	Fanconi anemia complementation group D2	NM_001018115
FANCE	FACE, FAE	Fanconi anemia complementation group E	NM_021922
FANCF	FAF	Fanconi anemia complementation group F	NM_022725
FANCG	FAG, XRCC9	Fanconi anemia complementation group G	NM_004629
FANCI	KIAA1794	Fanconi anemia complementation group I	NM_001113378
FANCL	FAAP43, PHF9, POG	Fanconi anemia complementation group L	NM_018062
FANCM	FAAP250, KIAA1596	Fanconi anemia complementation group M	NM_020937
HELQ	Helicase, POLQ-lik	HEL308	NM_133636
HES1	HES-1, HHL, HRY, bHLHb39	Hes family bHLH transcription factor 1	NM_005524

MMS22L	c6orf167, dJ39B17.2	MMS22 like, DNA repair protei	NM_198468
MORF4L1	Eaf3, FWP006, HsT17725, MEAF3, MORFRG15, MRG15, S863-6	Mortality factor 4 like 1	NM_206839
MUS81	SLX3	MUS81 structure-specific endonuclease subunit	NM_025128
PALB2	FANCN, PNCA3	Partner and localizer of BRCA2	NM_024675
POLH	RAD30, RAD30A, XP-V, XPV	DNA polymerase eta	NM_006502
RAD51	BRCC5, FANCR, HRAD51, HsRad51, HsT16930, MRMV2A, RECA	RAD51 recombinase	NM_002875
RAD51B	R51H2, RAD51L1, REC2	RAD51 paralog B	NM_133509
RAD51C	BROVCA3, FANCO, R51H3, RAD51L2	RAD51 paralog C	NM_058216
RAD51D	BROVCA4, R51H3, RAD51L3, TRAD	RAD51 paralog D	NM_001142571
REV1	AIBP80L	REV1, DNA directed polymerase	NM_016316
RF₩D3	RNF201	Ring finger and WD repeat domain 3	NM_018124
RNF8	bRNF8	Ring finger protein 8	NM_003958
SFPQ	POMP100, PPP1R140, PSF	Splicing factor proline and glutamine rich	NM_005066
SLX4	BTBD12, FANCP, MUS312	SLX4 structure-specific endonuclease subunit	NM_032444
TONSL	IKBR, NFKBIL2	Tonsoku-like, DNA repair protein	NM_013432
TP53BP1	53BP1, TDRD30, TP53, p202, p53BP1	Tumor protein p53 binding protein 1	NM_001141980
XRCC2	FANCU	X-ray repair cross complementing 2	NM_005431
XRCC3	СММ6	X-ray repair cross complementing 3	NM_001100118
CDKN1A	CAP20, CDKN1, CIP1, MDA-6, P21, SDI1, WAF1, p21CIP1	Cyclin dependent kinase inhibitor 1A	NM_001291549
EMSY	c11orf30, GL002	EMSY, BRCA2 interacting transcriptional repressor	NM_020193
FOXO3	AF6q21, FKHRL1, FKHRL1P2, FOXO2A, FOXO3	Forkhead box O3	NM_001455
KMT2C	HALR, MLL3, KIAA1506	Lysine methyltransferase 2C	NM_170606
RFFL	CARP-2, CARP2, FRING, RIFIFYLIN, RNF189, RNF34L	Ring finger and FYVE like domain containing E3 ubiquitin protein ligase	NM_001017368
RIF1	-	Replication timing regulatory factor 1	NM_018151
SIPA1L1	E6TP1	Signal induced proliferation associated 1 like 1	NM_015556
SLX4IP	c20orf94, bA204H22.1, bA254M13.1, dJ1099D15.3	SLX4 interacting protein	NM_001009608
DKK1	DKK-1, SK	Dickkopf WNT signaling pathway inhibitor 1	NM_012242
EEPD1	HSPC107	Endonuclease/exonuclease/phosphatase family domain containing 1	NM_030636
GEN1	Gen	GEN1, Holliday junction 5' flap endonuclease	NM_182625
POLN	POLAP	DNA polymerase nu	NM_181808

POLQ	PRO0327	DNA polymerase theta	NM_199420
RAD50	NBSLD2, hRad50	RAD50 double strand break repair protein	NM_005732
REV3L	POLZ, REV3	REV3 like, DNA directed polymerase zeta catalytic subunit	NM_002912
SLX1B	GIYD2	SLX1 homolog B, structure-specific endonuclease subunit	NM_024044
SPRTN	c1orf124, DVC1, PRO4323, spartan	SprT-like N-terminal domain	NM_032018
UBE2T	FANCT, HSPC150, PIG50	Ubiquitin conjugating enzyme E2 T	NM_014176
USP1	UBP	Ubiquitin specific peptidase 1	NM_003368
WDR48	P80, SPG60, UAF1	WD repeat domain 48	NM_020839

## Genliste des TruSight Cancer-Anreicherungspanels

Im Folgenden sind die mit Hilfe des TruSight Cancer-Panels angereicherten Gene aufgeführt. Darunter sind die rs-Nummern der zur gezielten Anreicherung enthaltenen SNPs gelistet.

AIP	CDKN1C	ERCC5	FH	NBN	RB1	SUFU
ALK	CDKN2A	EXT1	FLCN	NF1	RECQL4	TMEM127
APC	CEBPA	EXT2	GATA2	NF2	RET	TP53
ATM	CEP57	EZH2	GPC3	NSD1	RHBDF2	TSC1
BAP1	CHEK2	FANCA	HNF1A	PALB2	RUNX1	TSC2
BLM	CYLD	FANCB	HRAS	PHOX2B	SBDS	VHL
BMPR1A	DDB2	FANCC	KIT	PMS1	SDHAF2	WRN
BRCA1	DICER1	FANCD2	MAX	PMS2	SDHB	WT1
BRCA2	DIS3L2	FANCE	MEN1	PRF1	SDHC	XPA
BRIP1	EGFR	FANCF	MET	PRKAR1A	SDHD	XPC
BUB1B	EPCAM	FANCG	MLH1	PTCH1	SLX4	
CDC73	ERCC2	FANCI	MSH2	PTEN	SMAD4	
CDH1	ERCC3	FANCL	MSH6	RAD51C	SMARCB1	
CDK4	ERCC4	FANCM	MUTYH	RAD51D	STK11	

rs138213197	rs11655512	rs17181170	rs2736100	rs4474514	rs6691170	rs790356
rs149617956	rs116909374	rs17272796	rs28421666	rs4487645	rs671	rs7931342
rs10052657	rs11940551	rs17401966	rs2853676	rs4488809	rs6734275	rs798766
rs10058728	rs11978267	rs17483466	rs2858870	rs45430	rs674313	rs801114
rs10069690	rs1210110	rs17505102	rs2860580	rs4624820	rs6763931	rs8030672
rs10088218	rs1219648	rs17674580	rs2894207	rs4635969	rs6774494	rs8034191
rs1011970	rs12210050	rs1801516	rs2900333	rs4712653	rs685449	rs8042374
rs1014971	rs1229984	rs1805007	rs29232	rs4765623	rs6903608	rs807624
rs1016343	rs12413624	rs1859962	rs2981575	rs4767364	rs6939340	rs8102137
rs10187424	rs12500426	rs1945213	rs2981579	rs4775302	rs6983267	rs8102476
rs1027643	rs12621278	rs2014300	rs2981582	rs4775699	rs7014346	rs8170
rs10411210	rs12653946	rs2019960	rs3112612	rs4779584	rs7023329	rs865686
rs10484561	rs130067	rs2046210	rs3117582	rs4784227	rs7040024	rs872071
rs10484761	rs13016963	rs204999	rs3123078	rs4785763	rs704010	rs889312
rs10505477	rs1321311	rs20541	rs3129055	rs4793172	rs7089424	rs891835
rs1051730	rs1327301	rs2072590	rs31489	rs4795519	rs710521	rs902774
rs1052501	rs13281615	rs2074356	rs339331	rs4809324	rs7105934	rs910873
rs10795668	rs13361707	rs210138	rs36600	rs4905366	rs7127900	rs9268542
rs10821936	rs13387042	rs2121875	rs372883	rs4924410	rs7130881	rs9275572
rs10822013	rs13393577	rs2131877	rs3755132	rs4925386	rs7176508	rs9284813
rs10875943	rs13397985	rs2157719	rs3765524	rs4939827	rs718314	rs9293511
rs10896449	rs1364054	rs2180341	rs3768716	rs4973768	rs7210100	rs9352613
rs10934853	rs1393350	rs2242652	rs3781264	rs4975616	rs721048	rs9364554
rs10936599	rs1412829	rs224278	rs3782181	rs4977756	rs7238033	rs9430161
rs10936632	rs1432295	rs2248462	rs3790844	rs498872	rs7335046	rs944289
rs10937405	rs1447295	rs2255280	rs3802842	rs505922	rs735665	rs9485372
rs10974944	rs1456315	rs2274223	rs3803662	rs5759167	rs738722	rs9510787

rs1465618	rs2283873	rs3814113	rs5768709	rs7412746	rs9543325
rs1494961	rs2284063	rs3817198	rs5919432	rs7501939	rs9572094
rs1495741	rs2292884	rs3824999	rs5934683	rs7538876	rs9573163
rs1512268	rs2294008	rs391525	rs5945572	rs753955	rs9600079
rs1528601	rs2395185	rs401681	rs5945619	rs755383	rs961253
rs1547374	rs242076	rs4132601	rs5955543	rs7555566	rs9642880
rs1562430	rs2439302	rs4242382	rs6010620	rs757978	rs965513
rs1572072	rs2456449	rs4242384	rs614367	rs7579899	rs966423
rs161792	rs2517713	rs4295627	rs6435862	rs7584330	rs971074
rs16892766	rs258322	rs4324798	rs6457327	rs7584993	rs9841504
rs16901979	rs2596542	rs4415084	rs6465657	rs7679673	rs9929218
rs16902094	rs2647012	rs4430796	rs651164	rs7758229	rs995030
rs17021918	rs2660753	rs4444235	rs6603251	rs7808249	
rs17119461	rs2735839	rs445114	rs6687758	rs7837688	
	rs1465618 rs1494961 rs1495741 rs1512268 rs1528601 rs1547374 rs1562430 rs1572072 rs161792 rs16892766 rs16901979 rs16902094 rs17021918 rs17119461	rs1465618rs2283873rs1494961rs2284063rs1495741rs2292884rs1512268rs2294008rs1528601rs2395185rs1547374rs242076rs1562430rs2439302rs1572072rs2456449rs161792rs2517713rs16892766rs258322rs16901979rs2596542rs16902094rs2647012rs17021918rs2660753rs17119461rs2735839	rs1465618rs2283873rs3814113rs1494961rs2284063rs3817198rs1495741rs2292884rs3824999rs1512268rs2294008rs391525rs1528601rs2395185rs401681rs1547374rs242076rs4132601rs1562430rs2439302rs4242382rs1572072rs2456449rs4242384rs161792rs2517713rs4295627rs16892766rs258322rs4324798rs16901979rs2596542rs4415084rs16902094rs2647012rs4430796rs17021918rs2660753rs445114	rs1465618rs2283873rs3814113rs5768709rs1494961rs2284063rs3817198rs5919432rs1495741rs2292884rs3824999rs5934683rs1512268rs2294008rs391525rs5945572rs1528601rs2395185rs401681rs5945619rs1547374rs242076rs4132601rs595543rs1562430rs2439302rs4242382rs6010620rs1572072rs2456449rs422076rs424384rs161792rs2517713rs4295627rs6435862rs16892766rs258322rs4324798rs6457327rs16901979rs2596542rs4430796rs651164rs17021918rs260753rs444235rs603251rs17119461rs2735839rs445114rs6687758	rs1465618rs2283873rs3814113rs5768709rs7412746rs1494961rs2284063rs3817198rs5919432rs7501939rs1495741rs2292884rs3824999rs5934683rs7538876rs1512268rs2294008rs391525rs5945572rs753955rs1528601rs2395185rs401681rs5945619rs755383rs1547374rs242076rs4132601rs595543rs755566rs1562430rs2439302rs4242382rs610620rs757978rs1572072rs2456449rs42295627rs6435862rs7584330rs16792rs2517713rs4295627rs6435862rs7584330rs16892766rs258322rs415084rs645657rs7679673rs16901979rs2647012rs4430796rs651164rs7758229rs17021918rs2660753rs445114rs6687758rs7837688

## Eigene Publikationen, die im Rahmen der Dissertation entstanden sind

Rost I, Endt D, Gehrig A, Neveling K, Gille JJ, Leblanc T, Stoppa-Lyonnet D, Soulier J, Smogorzewska A, Schindler D. 2017. A Master Switch of the FA/BRCA Pathway: FANCD2 – Review and Mutation Update. *Hum Mutat.* Submitted

Rost I, Punekar M, Telford N, Stivaros SM, Chandler K, Minnis M, Castleton A, Higham C, Hopewell L, Evans G, Raams A, Theil A, Meyer S, Schindler D. 2017. Fanconi anemia with sun-sensitivity caused by a Xeroderma pigmentosum associated missense mutation in XPF - case report of the third FA-Q patient. *BMC Med Genet.* In revision.

Stivaros SM, Punekar M, Chandler K, Rost I, Schindler D, Meyer S. 2014. Pollicization of the index finger in Fanconi anaemia: appearances and functionality 40 years after the intervention. *Br J Haematol* 166(6):807.

## Weitere Publikationen

Gavvovidis I, Rost I, Trimborn M, Kaiser FJ, Purps J, Wiek C, Hanenberg H, Neitzel H, Schindler D. 2012. A novel MCPH1 isoform complements the defective chromosome condensation of human MCPH1-deficient cells. *PLoS One* 7(8):e40387.

## Beiträge zu Fachkonferenzen

# Reanalysis of FANCD2 pseudogenes and customized regional enrichment strategies improve FANCD2 mutation identification

I. Rost, D. Schindler

Posterpräsentation, Jahrestagung der Gesellschaft für Humangenetik (GfH), Lübeck, Deutschland, 2016 Posterpräsentation, Jahrestagung des Fanconi Anemia Research Fund (FARF), Toronto, Kanada, 2015

## Unusual characteristics in Fanconi anemia subtype FA-Q

I. Rost, K. Chantler, M. Punekar, M. Meenakshi, D. Schindler, S. Meyer Posterpräsentation, Jahrestagung der European Society of Human Genetics (ESHG), Glasgow, UK, 2015

#### A 48-year-old Fanconi anemia patient of unusual subtype

I. Rost, K. Chantler, M. Punekar, M. Meenakshi, D. Schindler, S. Meyer Vortrag, Jahrestagung des Fanconi Anemia Research Fund (FARF), Bethesda, USA, 2014

XXIV

# Experiences in using target enrichment and next-generation sequencing in Fanconi anemia diagnostics

I. Rost, D. Schindler

Posterpräsentation, Jahrestagung der Gesellschaft für Humangenetik (GfH), Essen, Deutschland, 2014

### Target enrichment and next-generation sequencing in Fanconi anemia diagnostics

I. Rost, D. Schindler

Posterpräsentation, Jahrestagung der Gesellschaft für Humangenetik (GfH), Dresden, Deutschland, 2013

# Cellular localization of MCPH1 isoforms and effects of MCPH1 mutations on G2/M checkpoint release

I. Rost, T. Staab, D. Schindler

Posterpräsentation, Jahrestagung der European Society of Human Genetics (ESHG), Nürnberg, Deutschland, 2012